

SPRINGER
REFERENCE

D. Ganten
K. Ruckpaul
Editors

VOLUME 1
A-L

Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine

 Springer

**Encyclopedic Reference of Genomics and Proteomics
in Molecular Medicine**

DETLEV GANTEN · KLAUS RUCKPAUL (Eds.)

Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine

**Volume 1
A-L**

With 297 Figures and 112 Tables

 Springer

Professor Dr. Detlev Ganten

Charité
Campus Charité Mitte
Schumannstr. 20/21
10117 Berlin
ganten@charite.de

Professor Em. Dr. Klaus Ruckpaul

Max Delbrück Center for Molecular Medicine (MDC) Berlin-Buch
Robert-Rössle-Straße 10
13125 Berlin
Germany
brandenburg@mdc-berlin.de

ISBN-10 3-540-44244-8 Springer Berlin Heidelberg New York
ISBN-13 978-3-540-44244-8 Springer Berlin Heidelberg New York

This publication is available also as

Electronic publication under ISBN-10: 3-540-29623-9/ISBN-13: 978-3-540-29623-2 and
Electronic and print bundle under ISBN-10: 3-540-30954-3/ISBN-13: 978-3-540-30954-3

Library of Congress Control Number: 2001012345

Bibliographic information published by *Die Deutsche Bibliothek*.

Die Deutsche Bibliothek lists this publication in the Deutsche Nationalbibliografie;
detailed bibliographic data is available in the Internet at: <http://dnb.ddb.de>

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in other ways, and storage in data banks. Duplication of this publication or parts thereof is only permitted under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer-Verlag. Violations are liable for prosecution under the German Copyright Law.

Springer is part of Springer Science+Business Media

springeronline.com

© Springer-Verlag Berlin Heidelberg New York 2006
Printed in Germany

The use of registered names, trademarks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

Product liability: The publishers cannot guarantee the accuracy of any information about the application of operative techniques and medications contained in this book. In every individual case the user must check such information by consulting the relevant literature.

Development Editor: Dr. Michaela Bilic, Dr. Natasja Sherif, Heidelberg
Production Editor: Frank Krabbes, Heidelberg
Cover design: Frido Steinen-Broo, Girona
Printed on acid free paper

SPIN: 10893219 2109fk - 5 4 3 2 1 0

Preface

The development of molecular medicine is closely linked with the rapidly growing knowledge in molecular biology, molecular genetics and genome research. Research findings in these fields have led to a shift in therapeutic targets. Whilst until recently only gene products such as enzymes and cell receptors represented targets of diagnostic processes, today the information carriers DNA and RNA themselves are evolving into target molecules, or medicinal drugs. Alongside researching into interactive processes on a cellular level – drawn from the biochemical advances made in the past century – mutation and disorders of mechanisms of gene expression are becoming the object of clinical research and of application in diagnosis and therapy. Hence, genomics and proteomics are developing into new fields of research, and their results are substantially influencing the future orientation of the fundamentals for diagnosis and therapy.

These new research fields including molecular biology have generated a huge amount of new terms, abbreviations (acronyms) which need explanation. Therefore these encyclopedic references are aimed at making all those acquainted with selected terms most frequently used in genomics and proteomics in molecular medicine who are not directly involved in those termini by their research.

The theoretical background of molecular medicine is based on the assumption that one or more molecular and/or genetic causes (e.g. changes caused by point mutation, deletions, shift in the reading frame etc.) underlie the outbreak of all diseases. Inversely, however, this does not necessarily imply that every mutated gene leads to a disorder in the sense of genetic determinism. There are forms of mutation that result in a transformed gene product without triggering any disease. For an illness to take shape many other factors besides a changed gene play a role, such as the endogenous disposition (e.g. inherent damage caused by previous illnesses) and exogenous factors (of environmental nature, strain through incompatibility of medication, drinking and smoking etc.). Genomics and proteomics are evolving into key technologies of advances in molecular medicine.

The transfer of molecular biological knowledge into therapeutic treatment is still at the initial stage. So far, results in gene therapy have fallen short of high-flying expectations. Instead, diagnostic methods based on molecular biology have found their way into medical practice.

Molecular medicine is scientifically based both on classical medicine, which is characterized by a phenotypical description of symptoms, and on the specific genotypical characterization of methods employed in molecular biology and genetic engineering. For the first time, this enabled a systematic analysis of the molecular causes of illnesses in a precise and rapid manner hitherto unattained. Molecular biological methods such as high-throughput techniques based on biochips are already being applied in various medical fields, and have significantly improved diagnosis, for instance in the prediction of risk estimation in certain illnesses, or with regard to sensitivity and specificity of treatment. This widens diagnostic scope in areas such as hereditary diseases of monogenetic origin, which until now could only be described phenomenologically, and enables preventive treatment and causal therapy of such diseases in the future. Evidently, even the molecular causes of uninfluenceable pleiotropic diseases are becoming increasingly tangible, moving their therapeutic treatment within reach.

This has a decisive impact on medical work. Nevertheless, molecular medicine, too, will retain its basic character which is precise clinical observation and integral medical care. Without thorough medical examination and detailed phenotypical description neither phenotype-genotype association of any significance can be derived, nor can the possibilities of detailed gene standardisation be fully exploited. The conventional physical examination and the analysis of the gene profile remain of equal importance in the domains of both research as well as medical care.

The spectacular publication on the human genome comprising 3.2 billion base sequences, which appeared simultaneously in the journals “Nature” (International non-commercial project, Head: Francis Collins) and “Science” (Genetic engineering company Celera Genomics, Head: Craig Venter) on the 50th anniversary of the double helix discovery by James Watson and Francis Crick in April 2003, marked a new era in biological basic research and in the knowledge of the components of human life. The successful decoding of the sequence raises the question about its function in the organism.

Recent sequencing of the human genome (published in 2004; Nature 431, p.915, p.927, p.933) revealed that (with an accuracy of 99.999%) 2.85 billions of base pairs out of a total of 3.08 billions of base pairs have been determined; thus 99% of the genome which contains the genes have been identified. Consequently, the number of genes had to be corrected to 20,000-25,000 from the estimated 100,000 genes in the 90's, and 30,000-40,000 in 2001. The genetic information for proteins comprises only 1.5% of the genome. 98.5% of the genome may be assigned to the so-called junk-DNA with so far largely unknown functions. Such data are erroneous. As parts of the gene expression products, introns, for example, contain regulative sequences for gene expression and sequences that are responsible for correct splicing. These sequences must be allocated to a specific gene as essential constituents and can therefore affect the assessment significantly.

Decoding the entire sequence of the genome is but the first step of a far more complex task: Understanding the functional significance of the sequences. So the determination of the sequences logically entails their functional deciphering as the following step. This field of research is known as "Functional genomics". It deals with the allocation of parts of the entire sequence to defined gene structures. This also includes the attribution of intron sequences to functions within the regulation process of gene expression that are only partly understood so far, as well as the control of all subsequent steps up to the final protein synthesis.

Alternative splicing is a major cause of transcriptome diversification. A single primary transcript yields different mature RNAs which lead to the production of proteins with various functions. This can be performed by alternative promoters. For 23,245 gene loci in the human genome over 43,000 transcripts are known. The alternative transcripts range from 2 to 40 (for details see Functional genomics).

Comparison with analysed genomes of other organisms are extremely useful in the process of allocating sequence parts to whole gene sequences. In the meantime a number of gene sequences have been fully deciphered, as for example those of prokaryote micro-organisms such as *Escherichia coli* and *Helicobacter pylori*, and eukaryote microorganisms such as baker's yeast *Saccharomyces cerevisiae* and of polycellular organisms like the worm *Caenorhabditis elegans*, or those of the fruit fly *Drosophila melanogaster* and of different vertebrates: e.g. man, mouse, rat and the green puffer-fish *Tetraodon nigroviridis*. Also the sequence of the bovine genome is available in a rough sketch with the exact decoding to be expected in 2005. These comparative studies revealed only 1183 species specific genes in the human genome. In all, about 200 genomes of various species have been sequenced and published.

The human chromosomes 21, 22, 14, 7, 6, and Y have been fully analysed and the results were published in 2003. Recently, an academic research team decoded the base sequence of the second smallest human chromosome 22 and made it accessible to the researching public. Chromosome 22 has 33 million base pairs with 545 genes. Defects on this chromosome are likely to be the cause of diseases such as schizophrenia, leukemia, immune deficiencies, bone cancer and brain tumours. Similarly, a German-Japanese research team succeeded in completely decoding the smallest carrier of genetic information in humans (chromosome 21) with 225 genes, some of which play a role in diseases like Alzheimer, ALS (amyotrope lateral sclerosis), myoclonic epilepsy, innate deafness, and Down's syndrome.

Two further chromosomes whose anomalies cause diseases are chromosomes 5 and 6. Chromosome 6 represents the largest human chromosome with 166.880 million base pairs. It has 1557 functional genes for, inter alia, the major histocompatibility complex (MHC), hereditary haemochromatosis with multi-organ dysfunction, juvenile-onset form of Parkinson's disease, and gene abnormalities are implicated as a contributory cause of schizophrenia, epilepsy, cancer and heart disease. Chromosome 5 has 177.7 million base pairs with 923 protein coding genes including for instance those for protocadherine and the interleukine gene families. In some regions deletion can generate disorders including spinal muscular dystrophy.

After enormous 12 years lasting efforts, recently (2005), the sequence of the human X-chromosome (the determining chromosome for women: XX = female, XY = male) was completely deciphered. It comprises 1098 native genes as compared with only 78 genes of the male Y-chromosome. It is of interest that about 10% of the X-chromosome play an important role in man and do not have any function in women. The X-chromosome contains only 4% of all human genes but is linked with every 10th hereditary monogenetic disease.

Recently, American scientists succeeded in introducing a human gene associated with the generation of Parkinson's disease into the genome of *Drosophila*, which produced impaired balance and other typical symptoms of nervous disease similar to those of people suffering from Parkinson's. This example demonstrates the immense significance of model organisms within the framework of functional genomics in the quest of grasping the human genome in its entirety.

So far a functional role could be allocated to specific gene products of about 5,000 genes, accounting for about 20% of the estimated total of 20,000 to 25,000 genes and about 3% of the total stock of human DNA. Thus a large number of genes still needs to be similarly allocated, and their molecular structures to be determined. This is the basic content of proteomics which in analogy to genomics has two complimentary objectives: the functional analysis and localization of gene products and the comprehension of their molecular structure.

As in the quest of the genome sequence which united scientists of six different nations to cooperate in a joint project, once again in 2001 a team of international researchers founded the Human Proteome Organisation (HUPO) to investigate into the significance of the enzymes coded by the genes. Proteome stands for the entire protein in a cell. In view of the magnitude of the task it seems quite plausible that the main project was split into 5 individual sub-projects: Human Plasma Proteome Project (HPPP), Sweden, USA; Human Liver Proteome Project (HLPP), Canada, China, France; Proteomics Standard Initiative (PSI), all countries; Human Brain Proteome Project (HBPP), Germany; International Mouse and Rat Proteome Project (MRPP) Canada, Germany.

All projects aim at decoding the functional network of proteins in the human organism, at the characterization and localization of proteins in normal and diseased humans and those of model organisms, and at disseminating knowledge and respective technologies so as to find clues for the treatment of diseases.

The second objective of proteomics is to determine the structure of gene products, a field that is primarily a part of basic research. Any overview of the latest developments in molecular medicine in this encyclopedia would be incomplete if it were restricted to clinical findings alone. The fundamental principles of molecular medicine are essentially based on the knowledge of cellular and molecular biological processes which are introduced into clinical practice through application in diagnosis. Moreover, the application of the whole range of genetic engineering tools resulted in radical changes in biotechnology and medical drug research. The application of genetic engineering in the pharmaceutical industry has led to a notable rationalization of production processes. In some cases the application of the processes made production of certain substances available that had so far been inaccessible for medical treatment. This marked the onset of the therapeutic use of medical drugs which could previously only be produced in chemical synthetic processes, rendering them unsuitable for large scale production.

The forthcoming gain in knowledge in functional genomics and proteomics with regard to very complex processes of growth, cell division and differentiation, and their respective mechanisms of regulation (such as first, second and third messenger, transcription factors and the corresponding cis-elements on the promoters) also pave the way for new strategies in the development of medical drugs.

Transforming these scientific results into useful medical drugs requires the knowledge of the structure of molecules, which is necessary for understanding the functional interaction of nucleic acids, proteins and ligands on a molecular level. This is made possible by ascertaining the molecular structures through X-ray radiation, synchrotrone radiation, nuclear magnetic resonance by super-conducting high performance magnets (up to 900 MHz) or other spectroscopic methods. All these processes complement each other in terms of applicability and significance of the evidence yielded. Another possibility of structure determination is the application of theoretical methods for predictive structure assessment of potential drugs. In this way pharmacological effects can be estimated. These techniques utilize highly sophisticated electronic simulation methods. When linking methods of the combination theory with the knowledge of the topography of bonding points, structure based drug design becomes feasible.

To date (i.e. in 2003) biotechnology and genetical engineering have brought forth approximately 10,800 products world-wide. Of these about 15% have been commercialized, a further 15-20% are in the licensing phase or are on the verge of coming on the market, 30% are in the clinical test phase (Phase III) and the remaining 40% are in the clinical study phase (Phase I and II). The decoding of the human genome and its functional analysis through genomics and proteomics will further enhance this process as a multitude of further molecule structures is

unravelling, thus considerably widening the range of rational drug development. It was the comprehension of biotechnological methods that made the production of certain therapeutic drugs possible as for example in the case of human insulin, erythropoietin, coagulation factor VIII, and interferon, resulting in their widespread therapeutic application.

Until now, however, only a few drugs have been developed with the help of structure based drug design. Among these are an inhibitor of HIV-1 protease 1 and an inhibitor substance of neuroaminidase of the influenza virus 2. These findings form the fundamental principles of molecular medicine. They enable rationally developed drugs to influence new target structures such as gene regulators (transcription factors), or the gene itself to become the object of therapeutic manipulation, for instance by functioning as a substance-producing drug.

We would like to point out that although substantial efforts were made to compose factually correct and well understandable presentations, there may be places where a definition is incomplete or a phrase in an essay is flawed. All contributors to this encyclopedia will be extremely happy to receive corrections or revised passages for inclusion in future editions of the “Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine”.

This encyclopedia endeavours to accompany current developments and convey the present level in knowledge on molecular causes of illnesses from a practice-oriented point of view. Acknowledged experts from various specialized fields such as human genetics, molecular biology, cell biology, biochemistry, physics and other bioscience disciplines explain the most important terms, complementing information by topical surveys, numerous figures and tables, and keywords. It is to be hoped that this compendium may contribute to understanding the advances in molecular medicine and may find many interested readers.

Berlin, 2005

DETLEV GANTEN
KLAUS RUCKPAUL

Editors-in-Chief

PROF. DR. DETLEV GANTEN
Charité
University Medicine Berlin
Berlin, Germany

PROF. EM. DR. KLAUS RUCKPAUL
Max Delbrück Center for Molecular Medicine (MDC)
Berlin-Buch
Berlin, Germany

Field Editors

PROF. DR. WALTER BIRCHMEIER
Cancer Research Program, Division of Signal Transduction,
Invasion, and Metastasis of Epithelial Cell
Max Delbrück Center for Molecular Medicine (MDC)
Berlin-Buch
Berlin, Germany

PROF. DR. JÖRG T. EPPLER
Department of Neurogenetics and Behavioural Genetics
Institute of Human Genetics
Ruhr University Bochum
Bochum, Germany

PROF. DR. KLAUS GENSER
Department of Proteomics and Molecular Mechanisms of
Neurodegenerative Diseases
Max Delbrück Center for Molecular Medicine (MDC)
Berlin-Buch
Berlin, Germany

DR. MANFRED GOSSEN
Cancer Research Program, Division of Control of DNA
Replication
Max Delbrück Center for Molecular Medicine (MDC)
Berlin-Buch
Berlin, Germany

DR. BIRGIT KERSTEN
Department of Proteomics and Molecular Mechanisms of
Neurodegenerative Diseases
Max Delbrück Center for Molecular Medicine (MDC)
Berlin-Buch
Berlin, Germany

PROF. DR. HANS LEHRACH
Department of Vertebrate Genomics
Max Planck Institute for Molecular Genetics
Berlin, Germany

PROF. DR. HARTMUT OSCHKINAT
Department of NMR-Supported Structural Biology
Forschungsinstitut für Molekulare Pharmakologie (FMP)
Berlin-Buch
Berlin, Germany

PROF. DR. PATRIZIA RUIZ
Division of Molecular Genetics
Center for Cardiovascular Research
Charité University Medicine
Berlin, Germany

DR. PETER SCHMIEDER
Department of NMR-Supported Structural Biology
Forschungsinstitut für Molekulare Pharmakologie (FMP)
Berlin-Buch
Berlin, Germany

PROF. DR. ERICH WANKER
Department of Proteomics and Molecular Mechanisms of
Neurodegenerative Diseases
Max Delbrück Center for Molecular Medicine (MDC)
Berlin-Buch
Berlin, Germany

Editorial Assistant

DR. CHRISTIANE NOLTE
Max Delbrück Center for Molecular Medicine (MDC)
Berlin-Buch
Berlin, Germany

List of Contributors

SALIM ABDELILAH-SEYFRIED
Max Delbrück Center (MDC) for Molecular Medicine
Berlin
Germany
salim@mdc-berlin.de

TILL ACKER
Edinger Institute, Neuropathology
Johann-Wolfgang Goethe University
Frankfurt
Germany
till.acker@med.uni-frankfurt.de

DAVID J. ADAMS
Wellcome Trust Genome Campus Hinxton
The Sanger Institute
Cambridge
UK

UELI AEBI
M.E. Müller Institute for Structural Biology, Biozentrum
University of Basel
Basel
Switzerland
ueli.aebi@unibas.ch

SAIMA AIJAZ
Division of Cell Biology
Institute of Ophthalmology
University College London
London
UK

LEILA N. ALBERS
Department of Chemistry
University of Minnesota
Minneapolis, MN
USA

REGINA ALBERT
Molecular Genome Analysis
DKFZ
Heidelberg
Germany

THOMAS ALTMANN
Institute of Biochemistry and Biology
University Potsdam
Potsdam-Golm
Germany
taltmann@rz.uni-potsdam.de

PETER ANGEL
Deutsches Krebsforschungszentrum, Division
of Signal Transduction and Growth Control
Heidelberg
Germany
p.angel@dkfz-heidelberg.de
p.angel@dkfz.de

MARTIN HRABÉ DE ANGELIS
GSF Research Center for Environment and Health
Neuherberg
Germany
hrabe@gsf.de

ADAM ANTEBI
Baylor College of Medicine
Huffington Center on Aging
Houston, TX
USA
aantebi@bcm.tmc.edu

STYLIANOS E. ANTONARAKIS
University Hospital of Geneva
Geneva
Switzerland
stylianos.antonarakis@medecine.unige.ch

RAJESH ARASADA
Adolf-Butenandt-Institute/Cell Biology
Ludwig-Maximilians-University
Munich
Germany

HAYLEY ARCHER
Institute of Medical Genetics
University Hospital of Wales, Heath Park
Cardiff
UK
archerhl@cardiff.ac.uk

HANS-HENNING ARNOLD
Cell and Molecular Biology
Technical University of Braunschweig
Braunschweig
Germany
h.arnold@tu-bs.de

KHUSRU ASADULLAH
Corporate Research Business Area Dermatology
Schering AG
Berlin
Germany
khusru.asadullah@schering.de

MANFRED AUER
Laboratory of Sensory Neuroscience
The Rockefeller University, New York
NY
USA
auerm@mail.rockefeller.edu

ANDREAS G. BADER
The Scripps Research Institute
La Jolla, CA
USA

MARIA S. BALDA
Division of Cell Biology
Institute of Ophthalmology
University College London
London
UK
m.balda@ucl.ac.uk

MELISA J. BAPTISTA
National Institute on Aging and National Institute of
Neurological Diseases and Stroke
National Institutes of Health
Bethesda, MD
USA
baptista@mail.nih.gov

RALF BARGOU
Department of Hematology, Oncology
and Tumor Immunology
Robert Rössle Clinic at the HELIOS
Clinical Center Berlin-Buch
Charité
Berlin
Germany

OLIVER BARTSCH
Institute of Human Genetics
Mainz University School of Medicine
Mainz
Germany
bartsch@humgen.klinik.uni-mainz.de

AMANDA E. BASS
Departments of Pathology, Dermatology, and the
Robert H. Lurie Cancer Center
Northwestern University Feinberg School of Medicine
Chicago, IL
USA
bass@northwestern.edu

INGO BECHMANN
Experimental Neuroimmunology, Institute of Anatomy,
Cell and Neurobiology
University Hospital Charité
Berlin
Germany
ingo.bechmann@charite.de

STEPHANIE BECHTEL
Department of Molecular Genome Analysis
German Cancer Research Center
Heidelberg
Germany
s.bechtel@dkfz.de

JOHANNES BECKERS
GSF Research Center for Environment and Health
Neuherberg
Germany

ROLAND BECKMANN
Institute for Biochemistry
Charité University Medicine Berlin
Germany
roland.beckmann@charite.de

MARK T. BEDFORD
The University of Texas M.D. Anderson Cancer Center
Science Park – Research Division
Smithville, MO
USA
mtbedford@mdanderson.org

DAVID J. BEEBE
Biomedical Engineering
University of Wisconsin
Madison, WI
USA
djbeebe@wisc.edu

CHRISTIAN BEHL
Department of Pathobiochemistry
Johannes Gutenberg University
Mainz
Germany
cbehl@uni-mainz.de

DIETRICH BEHNE
Department of Molecular Trace Element Research
in the Life Sciences
Hahn Meitner Institute Berlin
Berlin
Germany
behne@hmi.de

JÜRGEN BEHRENS
Nikolaus-Fiebiger-Center for Molecular Medicine
University Erlangen-Nürnberg
Erlangen
Germany
jbehrens@molmed.uni-erlangen.de

SAVERIO BELLUSCI
Developmental Biology Program, Department
of Surgery
USC Keck School of Medicine and the Saban
Research Institute of Childrens Hospital Los Angeles
Los Angeles, CA
USA
sbellusci@chla.usc.edu

OLIVIER BENSUADE
Régulation de l'Expression Génétique
Ecole Normale Supérieure (ENS)
Paris
France
bensaude@wotan.ens.fr

LEE BERGMAN
Department of Molecular Genome Analysis
German Cancer Research Center
Heidelberg
Germany

ERNEST BEUTLER
The Scripps Research Institute
Department of Molecular and Experimental Medicine
La Jolla, CA
USA
beutler@scripps.edu

MARTIN BEUTLER
Randall Division of Cell and Molecular Biophysics
King's College
London
UK
martin.beutler@kcl.ac.uk

GIUSEPPE BIAMONTI
Istituto di Genetica Molecolare, CNR
Pavia
Italy
biamonti@igm.cnr.it

ETHAN BIER
UCSD, and The Burnham Institute
La Jolla, CA
USA

WALTER BIRCHMEIER
Max-Delbrueck-Center for Molecular Medicine
Berlin
Germany
wbirch@mdc-berlin.de

MARIEL BIRNBAUMER
Receptor Regulation Section, Laboratory of Signal
Transduction
National Institute of Environmental Health Sciences
Research Triangle Park, NC
USA
birnbau2@niehs.nih.gov

CAROLINE BODEN
Max Delbrück Centre for Molecular Medicine
Berlin
Germany

ROLF BODMER
UCSD, and The Burnham Institute
La Jolla, CA
USA

JUDITH M. BOER
Center for Human and Clinical Genetics
Leiden University Medical Center
Leiden
The Netherlands
j.m.boer@lumc.nl

KENNETH R. BOHELER
Kenneth R. Boheler, Laboratory of Cardiovascular Science
NIA/NIH
Baltimore, MD
USA
bohelerk@grc.nia.nih.gov

GUILLAUME BOSSIS
University of Göttingen
Göttingen
Germany

DONALD W. BOWDEN
Center for Human Genomics
Wake Forest University School of Medicine
Winston-Salem, NC
USA
dbowden@wfubmc.edu

ALAN BRADLEY
Wellcome Trust Genome Campus Hinxton
The Sanger Institute
Cambridge
UK
abradley@sanger.ac.uk

OLE BRANDT
Functional Genome Analysis
Deutsches Krebsforschungszentrum
Heidelberg
Germany

SUSANNE BRAUN
Department of Biology
Institute of Cell Biology
ETH Zurich, Zurich
Switzerland

THOMAS BRAUN
M.E. Müller Institute for Structural Biology, Biocenter
University of Basel
Basel
Switzerland

ALVIS BRAZMA
European Bioinformatics Institute
European Molecular Biology Laboratory
Outstation-Hinxton
Cambridge
UK
brazma@ebi.ac.uk

MICHAEL BREITENBACH
Department of Cell Biology
University of Salzburg
Salzburg
Austria
michael.breitenbach@sbg.ac.at

KENNETH J. BRESLAUER
Department of Chemistry and Chemical Biology, Rutgers
State University of New Jersey
Piscataway, NJ
USA
kjbdna@rci.rutgers.edu

WILLIAM R. BRINKLEY
Department of Molecular and Cellular Biology
Baylor College of Medicine
Houston, TX
USA
brinkley@bcm.tmc.edu

STEFAN BRITSCH
Max Delbrück Center for Molecular Medicine (MDC)
Berlin-Buch
Berlin
Germany
sbritsch@mdc-berlin.de

M. JULIA BROSNAN
BHF Glasgow Cardiovascular Research Centre
University of Glasgow
Scotland
UK
mjb8n@clinmed.gla.ac.uk

SHIRLEY R. BRUCE
Department of Immunology
The University of Texas
MD Anderson Cancer Center
Houston, TX
USA

CHRISTIAN BUCHHOLZ
Paul-Ehrlich-Institute
Division of Medical Biotechnology
Langen
Germany

MALTE BUCHHOLZ
Department of Internal Medicine I
University of Ulm
Ulm
Germany
malte.buchholz@medizin.uni-ulm.de

MARGARET E. BUCKINGHAM
Department of Developmental Biology
CNRS, Pasteur Institute
Paris
France
margab@pasteur.fr

HEIKE BUDDE
MOLISA GmbH
Mageburg
Germany

JEAN-MARIE BUERSTEDDE
Institute of Molecular Radiobiology, GSF
Neuherberg
Germany
buersted@gsf.de

ALEXANDER BÜRKLE
Molecular Toxicology, Department of Biology
University Konstanz
Konstanz
Germany
alexander.buerkle@uni-konstanz.de

KONRAD BÜSSOW
Max Planck Institute for Molecular Genetics
Berlin
Germany
buessow@molgen.mpg.de

ANNA MARIA CALELLA
European Molecular Biology Laboratory
Mouse Biology Unit
Monterotondo
Italy

M. CRISTINA CARDOSO
Max Delbrueck Center for Molecular Medicine
Berlin
Germany
cardoso@mdc-berlin.de

MICAELA CASERTA
Istituto di Biologia e Patologia Molecolari
CNR
Rome
Italy
micaela.caserta@uniroma1.it

GIUSEPPE CASTALDO
CEINGE – Biotecnologie Avanzate, scrl and Dipartimento
di Biochimica e Biotecnologie Mediche
University of Napoli “Federico II”
Napoli
Italy

Facoltà di Scienze Matematiche, Fisiche e Naturali
Univeristy of Molise
Isernia
Italy

ANDREW C.B. CATO
Forschungszentrum Karlsruhe
Institute of Toxicology and Genetics
Karlsruhe
Germany
andrew.cato@itg.fzk.de

DORA CAVALLO-MEDVED
Department of Pharmacology
Wayne State University and Barbara Ann Karmanos
Cancer Institute
Detroit, MI
USA

FLORENCE CHAINIAUX
University of Namur
Namur
Belgium

SANG YUN CHO
Yonsei Proteome Research Center
Biomedical Proteome Research Center
Seoul
Korea

ALINE CHRÉTIEN
University of Namur
Namur
Belgium

MARKUS CHRISTMANN
Department of Toxicology
University of Mainz
Mainz
Germany

RICHARD I. CHRISTOPHERSON
School of Molecular and Microbial Biosciences
University of Sydney
Sydney, NSW
Australia
ric@mmb.usyd.edu.au

GEORGE P. CHROUSOS
National Institute of Child Health and
Human Development
Pediatric and Reproductive Endocrinology Branch
Bethesda, MD
USA
chrousog@mail.nih.gov

ELENA CIBRIAN-UHALTE
Max Delbrück Center
Berlin
Germany

SVEN CICHON
Department of Medical Genetics
University of Antwerp
Belgium

Institute of Human Genetics
University of Bonn
Germany
sven.cichon@uni-bonn.de

KLAUS CICHUTEK
Paul-Ehrlich-Institute
Division of Medical Biotechnology
Langen
Germany
cickl@pei.de

ANGUS J. CLARKE
Institute of Medical Genetics
University Hospital of Wales, Heath Park
Cardiff
UK
clarkeaj@cardiff.ac.uk

JAMES E. CLEAVER
Auerback Melanoma Laboratory, UCSF Cancer Center
University of California
San Francisco, CA
USA
jcleaver@cc.ucsf.edu

JOHN M. COFFIN
Department of Molecular Biology and Microbiology
Tufts University
Boston, MA
USA
john.coffin@tufts.edu

PIERRE COLAS
Aptanomics
Lyon
France
pierre.colas@aptanomics.com

ALASTAIR COMPSTON
University of Cambridge Neurology Unit
Cambridge
UK
alastair.compston@medschl.cam.ac.uk

KATJA CONRATH
Cellular and Molecular Immunology, Vlaams
Inter-University Institute of Biotechnology
Free University of Brussels
Brussels
Belgium

MARK R. COOKSON
National Institute on Aging and National Institute of
Neurological Diseases and Stroke
National Institutes of Health
Bethesda, MD
USA

KEITH D. COON
The Translational Genomics Research Institute
Phoenix, AZ
USA

DAVID N. COOPER
Institute of Medical Genetics
University of Wales
College of Medicine
Cardiff
UK
cooperdn@cardiff.ac.uk

RICHARD D. CUMMINGS
University of Oklahoma Health Services Center
Department of Biochemistry and Molecular Biology
Oklahoma City
Oklahoma, OK
USA
richard-cummings@ouhsc.edu

KERSTIN DANKER
Institute for Biochemistry und Molecular
Biology, Campus Benjamin Franklin
Charité-University Medicine Berlin
Berlin
Germany
kerstin.danker@charite.de

PHILIPPE DAUBAS
Department of Developmental Biology
CNRS, Pasteur Institute
Paris
France

EDOUARD DELAIVE
University of Namur
Namur
Belgium

SAMUEL DEUTSCH
Department of Genetic Medicine and Development
University of Geneva Medical School
Geneva
Switzerland
samuel.deutsch@medecine.unige.ch

JEAN-FRANÇOIS DIERICK
Biovallee
Gosselies
Belgium
jf.dierick@biovallee.be

MARTIN DIGWEED
Institute of Human Genetics
Charité University Medicine Berlin
Berlin
Germany
martin.digweed@charite.de

MARK D. DISTEFANO
Department of Chemistry
University of Minnesota
Minneapolis, MN
USA
distefan@chem.umn.edu

OLIVER DISTLER
Center of Experimental Rheumatology
University Hospital Zurich
Zurich
Switzerland
oliver.distler@usz.ch

GUNNAR DITTMAR
Max Delbrück Center for Molecular Medicine
Berlin
Germany

ANNA F. DOMINICZAK
BHF Glasgow Cardiovascular Research Centre
University of Glasgow
Scotland
UK
anna.dominiczak@clinmed.gla.ac.uk

ANJA DRÖGE
Max Delbrück Center for Molecular Medicine (MDC)
Berlin
Germany
a.droege@mdc-berlin.de

RAINER DUDEN
School of Biological Sciences
Royal Holloway University of London
Egham, Surrey
UK
rainer.duden@rhul.ac.uk

TRAVIS DUNCKLEY
The Translational Genomics Research Institute
Phoenix, AZ
USA

ELAINE DZIERZAK
Department of Cell Biology and Genetics
Erasmus University Medical Center
Rotterdam
The Netherlands
e.dzierzak@erasmusmc.nl

CLAUDIA EDER
Institute of Physiology
Medical Faculty Charité
Berlin
Germany
claudia.eder@charite.de

ANN EHRENHOFER-MURRAY
Otto-Warburg-Laboratories
Max-Planck-Institute of Molecular Genetics
Berlin
Germany
ehrenhof@molgen.mpg.de

MICHEL EICHELBAUM
Dr Margarete Fischer-Bosch Institute of Clinical
Pharmacology
Stuttgart
Germany
michel.eichelbaum@ikp-stuttgart.de

PETER EKBLOM
Department of Cell and Molecular Biology
Lund University
Lund
Sweden
peter.ekblom@medkem.lu.se

JAMES ELLIS
Developmental Biology Program
Hospital for Sick Children
Toronto, ON
Canada
jellis@sickkids.ca

NATHAN A. ELLIS
Department of Medicine
University of Chicago
Chicago, IL
USA
nellis@medicine.bsd.uchicago.edu

ANDREAS ENGEL
M.E. Müller Institute for Structural Biology, Biocenter
University of Basel
Basel
Switzerland
andreas.engel@unibas.ch

MARTIN ENGELHARD
Max Planck Institute of Molecular Physiology
Dortmund
Germany
martin.engelhard@mpi-dortmund.mpg.de

RICHARD A. ENGH
Max Planck Institute of Biochemistry
Martinsried
Germany
engh@biochem.mpg.de

MARKUS ENZELBERGER
MorphoSys AG
Martinsried
Germany
enzelberger@morphosys.com

JÖRG T. EPPLER
Department of Human Genetics
Ruhr University Bochum
Germany
joerg.t.epplen@rub.de

ELISABETH ERHARDTSEN
Novo Nordisk A/S
Bagsværd
Denmark
eer@novonordisk.com

W. HOWARD EVANS
Medical Biochemistry and Immunology
Cardiff University School of Medicine
Cardiff, Wales
UK
wmbwhe@cf.ac.uk

TIMOTHY M. EVANS
The Institute for Molecular Bioscience
The University of Queensland
St Lucia, QLD
Australia

BIRTHE FAHRENKROG
M.E. Müller Institute for Structural Biology, Biozentrum
University of Basel
Basel
Switzerland
birthe.fahrenkrog@unibas.ch

ARTURO FALASCHI
International Centre for Genetic Engineering and
Biotechnology
Trieste
Italy
faldaschi@icgeb.org

PHILIP FARABAUGH
Department of Biological Sciences
University of Maryland Baltimore County
Baltimore, MD
USA
farabaugh@umbc.edu

DIANE FATKIN
Sr Bernice Research Program in Inherited Heart Diseases
Molecular Cardiology Unit
Victor Chang Cardiac Research Institute
Darlinghurst NSW
Australia
d.fatkin@victorchang.unsw.edu.au

MONICA FEDELE
Istituto di Endocrinologia ed Oncologia Sperimentale
University "Federico II"
Naples
Italy

MARC FELDMANN
The Kennedy Institute of Rheumatology Division
Faculty of Medicine
Imperial College London
London
UK
m.feldmann@imperial.ac.uk

DIRK FELDMEYER
Institute of Medicine
Research Centre Jülich
Jülich
Germany
d.feldmeyer@fz-juelich.de

OLIVER FIEHN
UC Davis Genome Center
Davis, CA
USA
ofiehn@ucdavis.edu

ULRICH FINCKH
Laboratory Medicine Dortmund, Dr. Eberhard & Partner
Dortmund
Germany
finckh@labmed.de

VÉRONIQUE FLAMENT
University of Namur
Namur
Belgium

BERNHARD FLECKENSTEIN
Institute of Clinical and Molecular Virology
Erlangen
Germany

BERND K. FLEISCHMANN
Institute of Physiology I
University of Bonn
Bonn
Germany

LEOPOLD FLOHÉ
MOLISA GmbH
Mageburg
Germany
l.flohe@t-online.de

EGBERT FLORY
Paul-Ehrlich-Institute
Division of Medical Biotechnology
Langen
Germany

CLEMENS FRANZ

Ludwig Institute for Cancer Research, Royal Free and
University College School of Medicine, Department of
Biochemistry and Molecular Biology
University College London
London
UK

CHRISTIAN FREUND

Freie Universität Berlin and Institute for Molecular
Pharmacology
Berlin
Germany
freund@fmp-berlin.de

JEAN-NOËL FREUND

INSERM U682
Strasbourg
France

MATI FRIDKIN

Departments of Biological Chemistry and
Organic Chemistry
Weizmann Institute of Science
Rehovot
Israel
mati.fridkin@weizmann.ac.il

EVELYNE FRIEDERICH

Laboratory of Molecular Biology, Genetic Analysis
and Modelling
Public Research Center for Health
Luxembourg
evelyne.friederich@crp-sante.healthnet.lu

MATTHIAS G. FRIEDRICH

Department of Cardiac Sciences
University of Calgary
Calgary, AB
Canada
matthias.friedrich@ucalgary.ca

GERHARD FRITZ

Division of Applied Toxicology, Institute of Toxicology
University of Mainz
Mainz
Germany

CORNELIUS FRÖMME

Faculty of Medicine
Georg August University
Göttingen
Germany
c.froemmel@med.uni-goettingen.de

MASATOSHI FUJITA

Virology Division
National Cancer Center Research Institute
Chuohku, Tokyo
Japan
mafujita@gan2.res.ncc.go.jp

YASUYUKI FUJITA

MRC Laboratory of Molecular Cell Biology and
Cell Biology Unit, and Department of Biology
University College
London
UK
y.fujita@ucl.ac.uk

YASUHIRO FURUICHI

GeneCare Research Institute
Kamakura
Japan
furuichi@genecare.co.jp

ALFREDO FUSCO

Istituto di Endocrinologia ed Oncologia Sperimentale
University "Federico II"
Naples
Italy
afusco@napoli.com

GIULIO GABBANI

Department of Pathology
CMU-University of Geneva
Geneva
Switzerland
giulio.gabbiani@medecine.unige.ch

UDO S. GAJPL

Institute of Clinical Immunology and Rheumatology,
Department of Internal Medicine III
Friedrich-Alexander University of
Erlangen-Nuernberg
Erlangen
Germany

NIELS H. GEHRING

University Clinic Heidelberg
Children's Hospital
Heidelberg
Germany
niels.gehring@med.uni-heidelberg.de

ROBERT GERLAI

Psychology Department
University of Toronto at Mississauga
Mississauga, ON
Canada
robert_gerlai@yahoo.com

PHILIP F. GIAMPIETRO

Department of Medical Genetic Services
Marshfield Clinic
Marshfield, WI
USA
giampietro.philip@marshfieldclinic.org

MARK H. GINSBERG

Department of Medicine, Division of Rheumatology
University of California
San Diego, La Jolla, CA
USA
mhginsberg@ucsd.edu

ANTON GLIEDER
Research Centre Applied Biocatalysis, Institute of
Molecular Biotechnology
Graz University of Technology
Austria
glieder@glieder.com

JÖRN GLÖKLER
Department of Vertebrate Genomics
Max Planck Institute of Molecular Genetics
Berlin
Germany

RiNA GmbH
Berlin
Germany
gloekler@ma-network.com

JOHAN GOBOM
Max Planck Institute for Molecular Genetics
Berlin
Germany
gobom@molgen.mpg.de

WOLFGANG GOEDECKE
Institute for Genetics
University of Essen
Essen
Germany
wolfgang.goedecke@uni-essen.de

LAWRENCE E. GOLDFINGER
Department of Medicine, Division of Rheumatology
University of California
San Diego, La Jolla, CA
USA

SARAH J. GOODFELLOW
Institute of Biomedical and Life Sciences
University of Glasgow
Glasgow
UK

STEVEN GOODRICK
Division of Cell Biology, Institute of Ophthalmology
University College London
London
UK

ERYNN S. GORDON
Children's National Medical Center
Washington, DC
USA

MICHAEL GOTTHARDT
MDC
Berlin
Germany
gotthardt@mdc-berlin.de

RALPH GRASSMANN
Institute of Clinical and Molecular Virology
Erlangen
Germany
grassmann@viro.med.uni-erlangen.de

KATHLEEN J. GREEN
Departments of Pathology, Dermatology, and the Robert H.
Lurie Cancer Center
Northwestern University Feinberg School of Medicine
Chicago, IL
USA
kgreen@northwestern.edu

THOMAS M. GRESS
Department of Internal Medicine I
University of Ulm
Ulm
Germany
thomas.gress@medizin.uni-ulm.de

ULRICH GROSS
Veterinary Clinics Dr. Schäfer
Bad Nauheim
Germany
ulrich.gross@aranea.de

KATJA S. GROSSMANN
Max Delbrück Center for Molecular Medicine
Berlin
Germany

MATTHEW W. GROW
Biochemistry and Molecular Biology,
Indiana University School of Medicine, Biotechnology
Research and Training Center
Indianapolis, IN
USA

ALEXANDER GRUNDT
University Clinic Heidelberg
Inner Medicine III
Heidelberg
Germany

THOMAS GUDERMANN
Phillips-University Marburg
Marburg
Germany
gudermann@staff.uni-marburg.de

VILMUNDUR GUDNASON
Icelandic Heart Association Heart Preventive Clinic and
Research Institute
Kopavogur
Iceland
v.gudnason@hjarta.is

KATRINA GWINN-HARDY
National Institute on Aging and National Institute of
Neurological Diseases and Stroke
National Institutes of Health
Bethesda, MD
USA

VOLKER HAGEN
Research Institute for Molecular Pharmacology
Berlin
Germany
hagen@fmp-berlin.de

WILLIAM C. HAHN
Department of Medical Oncology
Dana-Farber Cancer Institute
Boston, MA
USA
william_hahn@dfci.harvard.edu

HIROSHI HANDA
Tokyo Institute of Technology
Graduate School of Bioscience and Biotechnology,
Midori-ku
Yokohama
Japan
hhanda@bio.titech.ac.jp

CHRISTINE HARTMANN
Research Institute for Molecular Pathology
Vienna
Austria
hartmann@imp.univie.ac.at

FRANZ HARTNER
Research Centre Applied Biocatalysis, Institute of
Molecular Biotechnology
Graz University of Technology
Austria

VOLKER HAUCKE
Institute for Chemistry-Biochemistry
Free University
Berlin
Germany
vhaucke@chemie.fu-berlin.de

THOMAS HÄUPL
Department of Rheumatology, Charité
Humboldt-University
Berlin
Germany
thomas.haeupl@charite.de

JIAN-QING HE
University of British Columbia McDonald Research
Laboratories / iCAPTURE Center
Vancouver, BC
Canada
jhe@mrl.ubc.ca

MARGARETE M.S. HECK
Wellcome Trust Centre for Cell Biology
University of Edinburgh
Edinburgh
UK
margarete.heck@ed.ac.uk

ULF HEDIN
Department of Surgical Sciences
Karolinska Hospital
ulf.hedin@kirurgi.ki.se

UDO HEINEMANN
Max Delbrück Center for Molecular Medicine
Berlin
Germany
heinemann@mdc-berlin.de

RAINER HEINTZMANN
Randall Division of Cell and Molecular Biophysics
King's College
London
UK
rainer.heintzmann@kcl.ac.uk

CARL-HENRIK HELDIN
Ludwig Institute for Cancer Research
Uppsala
Sweden
c-h.heldin@licr.uu.se

FRITJOF HELMCHEN
Brain Research Institute
University of Zurich
Switzerland
helmchen@hifo.unizh.ch

VOLKHARD HELMS
Center for Bioinformatics
Saarland University
Saarbrücken
Germany
volkhard.helms@bioinformatik.uni-saarland.de

GEOFFREY N. HENDY
Calcium Research Laboratory,
Royal Victoria Hospital
and Departments of Medicine, Physiology and
Human Genetics
McGill University
Montreal, QC
Canada
geoffrey.hendy@mcgill.ca

LUDGER HENGST
Biocenter-Insbruck Medical University
Innsbruck
Austria
ludger.hengst@uibk.ac.at

STEFFEN HENNIG
RZPD - Deutsches Ressourcenzentrum
für Genomforschung
Berlin
Germany
hennig@rzpd.de

ANDREAS HERRMANN
Institute of Biology, and Center of Biophysics and
Bioinformatics
Humboldt University Berlin
Berlin
Germany
andreas.herrmann@rz.hu-berlin.de

BERNHARD G. HERRMANN
Department of Developmental Genetics
Max Planck Institute for Molecular Genetics
Berlin
Germany
herrmann@molgen.mpg.de

MARTIN HERRMANN
Institute of Clinical Immunology and Rheumatology,
Department of Internal Medicine III
Friedrich-Alexander University of Erlangen-Nuernberg
Erlangen
Germany

JÜRGEN HESCHELER
Institute of Neurophysiology
University of Cologne
Cologne
Germany
j.hescheler@uni-koeln.de

MASAHIKO HIBI
Center for Developmental Biology (CDB) RIKEN
Laboratory for Vertebrate Axis Formation
Hyogo
Japan
hibi@cdb.riken.go.jp

MARC HILD
Novartis Institutes for BioMedical Research
Cambridge, MA
USA
marc.hild@pharma.novartis

HANS-JÜRGEN HINZ
Institute for Physical Chemistry
WWU Münster
Münster
Germany
hinz@uni-muenster.de

MICHIO HIRANO
Department of Neurology
Columbia University College
of Physicians and Surgeons
New York, NY
USA
mh29@columbia.edu

ERIC P. HOFFMAN
Children's National Medical Center
Washington, DC
USA
ehoffman@cnmcresearch.org

SABINE HOFFJAN
Department of Human Genetics
Ruhr-University of Bochum
Bochum
Germany
sabine.hoffjan@ruhr-uni-bochum.de

KLAUS PETER HOFMANN
Institute of Medical Physics and Biophysics, Charité, and
Center of Biophysics
Humboldt University
Berlin
Germany
kph@charite.de

CATHERINE HOGAN
MRC Laboratory of Molecular Cell Biology and
Cell Biology Unit, and Department of Biology
University College
London
UK

JÖRG D. HOHEISEL
Functional Genome Analysis
Deutsches Krebsforschungszentrum
Heidelberg
Germany
j.hoheisel@dkfz.de

MATTHIAS HOHL
Department of Medical Biochemistry and
Molecular Biology
University of Saarland Medical Center
Homburg
Germany

MARTIN HOLČÍK
Apoptosis Research Center
Children's Hospital of Eastern Ontario Research Institute
Ottawa, ON
Canada
martin@mgcheo.med.uottawa.ca

HELMUT HOLTSMANN
Medizinische Hochschule Hannover
Hannover
Germany
holtmann.helmut@mh-hannover.de

BERNHARD HORSTHEMKE
Institute for Human Genetics
University Clinic Essen
Essen
Germany
b.horsthemke@uni-essen.de

RÜDIGER HORSTKORTE
Institute for Biochemistry and Molecular Biology
Charité
Berlin
Germany
ruediger.horstkorte@charite.de

RICHARD HORUK
Department of Immunology
Berlex Biosciences
Richmond, CA
USA
horuk@pacbell.net

LARS C. HUBER
Center of Experimental Rheumatology
University Hospital Zurich
Zurich
Switzerland

NORBERT HÜBNER
Max Delbrück Center for Molecular Medicine (MDC)
Berlin
Germany
nhuebner@mdc-berlin.de

ULRICH HÜBSCHER
Institute of Veterinary Biochemistry and Molecular Biology
University of Zürich
Zürich
Switzerland
hubscher@vetbio.unizh.ch

JOERG HUELSKEN
Epithelial Stem Cell Biology
ISREC
Epalinges
Switzerland
joerg.huelsken@isrec.unil.ch

TAKUYA IMAMURA
Ecole Pratique des Hautes Etudes
Evry
France

BEAT A. IMHOF
Department of Pathology and Immunology and
Department of Cellular Physiology and Metabolism
University of Geneva
Geneva
Switzerland
beat.imhof@medecine.unige.ch

AKIHIRO INAZU
Molecular Biochemistry and Molecular Biology
Laboratory, Graduate School of Medical Sciences
Kanazawa University
Kanazawa
Japan
inazua@mhs.mp.kanazawa-u.ac.jp

ZOLTÁN IVICS
Max Delbrück Center for Molecular Medicine
Berlin
Germany
zivics@mdc-berlin.de

ZSUZSANNA IZSVÁK
Max Delbrück Center for Molecular Medicine
Berlin
Germany

ANETTE JACOB
Functional Genome Analysis
Deutsches Krebsforschungszentrum
Heidelberg
Germany

TIMO JAEGER
MOLISA GmbH
Magdeburg
Germany

WOLFGANG JAHNKE
Novartis Institutes for BioMedical Research, Novartis
Pharma AG
Discovery Technologies
Basel
Switzerland
wolfgang.jahnke@novartis.com

MICHAEL JANITZ
Max Planck Institute for Molecular Genetics
Department Vertebrate Genomics
Berlin
Germany
janitz@molgen.mpg.de

DIETER E. JENNE
Dieter E. Jenne, Max Planck Institute of Neurobiology
Martinsried
Germany
djenne@neuro.mpg.de

HENRICK K. JENSEN
Department of Cardiology
Aarhus University Hospital, Skejby Hospital
Aarhus
Denmark
hkjensen@dadlnet.dk

MATTHEW C. JONES
The Wellcome Trust Sanger Institute, Wellcome Trust
Genome Campus
Hinxton
Cambridgeshire
UK
mcj@sanger.ac.uk

PETER R. JUNGBLUT
Max Planck Institute for Infectious Diseases, Protein
Analysis
Berlin
Germany
jungblut@mpiib-berlin.mpg.de

JERZY JURKA
Genetic Information Research Institute
Mountain View, CA
USA
jurka@girinst.org

KLAUS H. KAESTNER
Department of Cell and Developmental Biology
University of Pennsylvania
Philadelphia, PA
USA
kaestner@mail.med.upenn.edu

BERND KAINA
Division of Applied Toxicology, Institute of Toxicology
University of Mainz
Mainz
Germany
kaina@uni-mainz.de

BERND KAINA
Department of Toxicology
University of Mainz
Mainz
Germany

CHAYA KALCHEIM
Department of Anatomy and Cell Biology
Hebrew University-Hadassah Medical School
Jerusalem
Israel
kalcheim@nn-shum.cc.huji.ac.il

JOACHIM R. KALDEN
Institute of Clinical Immunology and Rheumatology,
Department of Internal Medicine III
Friedrich-Alexander University of Erlangen-Nuernberg
Erlangen
Germany

WOLFGANG E. KAMINSKI
Institute for Clinical Chemistry and Laboratory Medicine
University Regensburg
Regensburg
Germany
wolfgang.kaminski@klinik.uni-regensburg.de

VLADIMIR V. KAPITONOV
Genetic Information Research Institute
Mountain View, CA
USA

MARIA KARTALOU
Division of Biological Engineering and Center for
Environmental Health Sciences
Massachusetts Institute of Technology
Cambridge, MA
USA

TACIANA KASCIUKOVIC
Division of Gene Regulation and Expression, Wellcome
Trust Biocentre
University of Dundee
UK

RONI KASHER
Departments of Biological Chemistry and Organic Chemistry
Weizmann Institute of Science
Rehovot
Israel
roni.kasher@weizmann.ac.il

EPHRAIM KATCHALSKI-KATZIR
Departments of Biological Chemistry and Organic Chemistry
Weizmann Institute of Science
Rehovot
Israel
ephraim.katzir@weizmann.ac.il

THOMAS C. KAUFMANN
M.E. Müller Institute for Structural Biology, Biocenter
University of Basel
Basel
Switzerland

AJAMETE KAYKAS
Department of Pharmacology, Howard Hughes
Medical Institute
University of Washington School of Medicine
Seattle, WA
USA

MICHÈLE KEDINGER
INSERM U682
Strasbourg
France
michele.kedinger@inserm.u-strasbg.fr

GERD KEMPERMANN
Gerd Kempermann, Max-Delbrück-Center for Molecular
Medicine (MDC) Berlin-Buch
Berlin
Germany
gerd.kempermann@mdc-berlin.de

BIRGIT KERSTEN
Department of Neuroproteomics
Max Delbrück Center for Molecular
Medicine Berlin-Buch
Berlin
Germany
b.kersten@mdc-berlin.de

TOMOSHIGE KINO
National Institute of Child Health and
Human Development
Pediatric and Reproductive Endocrinology Branch
Bethesda, MD
USA
kinot@mail.nih.gov

FRANK KIRCHHOFF
Department of Neurogenetics
Max Planck Institute of Experimental Medicine
Göttingen
Germany
kirchhoff@em.mpg.de

FRANK KIRCHHOFF
Max Planck Institute of Experimental Medicine
Göttingen
Germany

GERHARD KLEBE
Institute of Pharmaceutical Chemistry
Philipps-University
Marburg
Germany
klebe@mail.uni-marburg.de

CHRISTOPH A. KLEIN
Institute of Immunology
Ludwig Maximilian University
Munich
Germany
christoph.klein@med.uni-muenchen.de

THOMAS KLEIN
Institute for Genetics
University of Cologne
Cologne
Germany
th.klein@uni-koeln.de

MICHAEL W. KLYMKOWSKY
Molecular, Cellular and Developmental Biology
University of Colorado
Boulder, CO
USA
klym@spot.colorado.edu

KARL-WILHELM KOCH
Division of Biochemistry, Institute of Biology and
Environmental Sciences
Carl von Ossietzky University
Oldenburg
Germany
karl.w.koch@uni-oldenburg.de

ZOLTÁN KONTHUR
Department of Vertebrate Genomics
Max Planck Institute of Molecular Genetics
Berlin
Germany
konthur@molgen.mpg.de

JOACHIM KOPKA
Max Planck Institute of Molecular Plant Physiology
Golm
Germany
kopka@mpimp-golm.mpg.de

PETER KOPP
Division of Endocrinology, Metabolism & Molecular
Medicine, Feinberg School of Medicine
Northwestern University
Chicago, IL
USA
p-kopp@northwestern.edu

ULRIKE KORF
Molecular Genome Analysis, DKFZ
Heidelberg
Germany
u.korf@dkfz.de

BERNHARD KORN
RZPD
Deutsches Ressourcenzentrum für Genomforschung GmbH
Heidelberg
Germany
korn@rzpd.de

MICHAEL KRACHT
Medizinische Hochschule Hannover
Hannover
Germany
kracht.michael@mh-hannover.de

ACHIM KRAMER
Institute for Medical Immunology, Laboratory of
Chronobiology
Charité University Medicine
Berlin
Germany
achim.kramer@charite.de

DANIEL KRAPPMANN
Max Delbrück Center for Molecular Medicine
Berlin
Germany
scheidereit@mdc-berlin.de

EBERHARD KRAUSE
Institute for Molecular Pharmacology
Berlin
Germany
ekrause@fmp-berlin.de

MICHAEL KRAUSS
Institute for Chemistry-Biochemistry
Free University
Berlin
Germany

MICHAEL KRAWCZAK
Institute for Medical Informatics und Statistics
Christian-Albrechts-University Kiel
Kiel
Germany
krawczak@medinfo.uni-kiel.de

HANS A. KRETZSCHMAR
Institute for Neuropathology
Ludwig Maximilian University
Munich
Germany
hans.kretzschmar@inp.med.uni-muenchen.de

ULRICH KUBITSCHKE
Institute of Physical and Theoretical Chemistry
Bonn
Germany
u.kubitschke@uni-bonn.de

RONALD KÜHNE
Research Institute for Molecular Pharmacology
Berlin
Germany
kuehne@fmp-berlin.de

RYOKO KURIYAMA
Department of Genetics, Cell Biology, and Development
University of Minnesota, MN
USA
ryoko@lenti.med.umn.edu

DAVID J. KWIATKOWSKI
Brigham & Woman's Hospital
Boston, MA
USA
dk@rics.bwh.harvard.edu

MIN-SEOK KWON
Yonsei Proteome Research Center
Biomedical Proteome Research Center
Seoul
Korea

Department of Biochemistry
Yonsei University
Seoul
Korea

ANTONIOS KYRIAKOPOULOS
Department of Molecular Trace Element Research
in the Life Sciences
Hahn Meitner Institute Berlin
Berlin
Germany

PATRICIA A. LABOSKY
Department of Cell and Developmental Biology
University of Pennsylvania
Philadelphia, PA
USA
plabosky@mail.med.upenn.edu

APARNA LAKKARAJU
Margaret M. Dyson Vision Research Institute and
Departments of Cell Biology and Ophthalmology
Weill Medical College of Cornell University
New York, NY
USA

SEAN LAL
Muscle Research Unit, Institute for Biomedical Research
University of Sydney
Sydney, NSW
Australia
sean@anatomy.usyd.edu.au

VICTOR S. LAMZIN
European Molecular Biology Laboratory (EMBL),
Hamburg c/o DESY
Hamburg
Germany
victor@embl-hamburg.de

E. BIRGITTE LANE
Division of Cell & Developmental Biology
University of Dundee School of Life Sciences
MSI/WTB Complex
Dundee
UK
e.b.lane@dundee.ac.uk

ULRICH LANGENBECK
Institute of Human Genetics
University Hospital
Frankfurt/Main
Germany
u.langenbeck@em.uni-frankfurt.de

STIJW P DE LANGHE
Developmental Biology Program, Department of Surgery
USC Keck School of Medicine and the Saban Research
Institute of Childrens Hospital Los Angeles
Los Angeles, CA
USA

CLAUDIA LANGLAIS
RZPD
Deutsches Ressourcenzentrum für
Genomforschung GmbH
Heidelberg
Germany
langlais@rzpd.de

NILS-GÖRAN LARSSON
Karolinska Institute, Department of Laboratory Medicine,
NOVUM
Karolinska University Hospital-Huddinge
Stockholm
Sweden
nils-goran.larsson@labmed.ki.se

PETER LAUN
Department of Cell Biology
University of Salzburg
Austria

JEANNIE T. LEE
Howard Hughes Medical Institute,
Department of Molecular Biology,
Massachusetts General Hospital,
Department of Genetics
Harvard Medical School
Boston, MA
USA
lee@molbio.mgh.harvard.edu

JEAN LEHMANN
Theoretical Biophysics, Department of Physics
Royal Institute of Technology
Stockholm
Sweden
jean@theophys.kth.se

HANS LEHRACH
Max Planck Institute for Molecular Genetics
Berlin
Germany
lehrach@molgen.mpg.de

TROND P. LEREN
Medical Genetics Laboratory, Department of Medical
Genetics
Rikshospitalet University Hospital
Oslo
Norway
trond.leren@rikshospitalet.no

ANCY LEROY
Laboratory of Experimental Cancerology
Ghent University Hospital
Ghent
Belgium
ancy.leroy@ugent.be

JEFFREY M. LEVSKY
Anatomy & Structural Biology
Albert Einstein College of Medicine
Bronx, New York, NY
USA

The Wellcome Trust Sanger Institute, Wellcome Trust
Genome Campus
Hinxton
Cambridgeshire, UK
kem@sanger.ac.uk

CHIN LI
Institute of Biomedical Sciences
Academia Sinica
Taipei
Taiwan

MICHAEL LIETZ
Department of Medical Biochemistry and
Molecular Biology
University of Saarland Medical Center
Homburg
Germany

CHOONG-CHIN LIEW
Harvard Medical School
Boston, MA
USA
cliew@rics.bwh.harvard.edu

DIETER LINK
Department Robotics & Screening
Xantos Biomedicine AG
Munich
Germany
d.link@xantos.de

SANDY H.M. LITJENS
Division of Cell Biology
The Netherlands Cancer Institute
Amsterdam
The Netherlands
s.litjens@nki.nl

EVELINE S. LITSCHER
Brookdale Department of Molecular
Cell & Developmental Biology
Mount Sinai School of Medicine
New York, NY
USA
eveline.litscher@mssm.edu

ESTHER LIZANO
Institute for Biochemistry
University of Leipzig
Leipzig
Germany
elizano@uni-leipzig.de

FELIX LOOSLI
Developmental Biology Programme
European Molecular Biology Laboratory, EMBL
Heidelberg
Germany
felix.loosli@embl.de

DANIEL LOUVARD
Subcellular structure and cellular dynamics, UMR 144
CNRS/IC
Institute Curie – Research Division
Paris
France
daniel.louvard@curie.fr

JOACHIM LÜBKE
Institute of Medicine
Research Centre Jülich
Jülich
Germany
j.luebke@fz-juelich.de

WOLF-DIETER LUDWIG
Department of Hematology, Oncology and
Tumor Immunology
Robert Rössle Clinic at the HELIOS Clinical
Center Berlin-Buch
Charité, Berlin
Germany
ludwig@rrk.charite-buch.de

FRIEDRICH C. LUFT
Franz Volhard Clinic, HELIOS Clinical Center Berlin
Max Delbrück Center for Molecular Medicine, Medical
Faculty of the Charité
Berlin
Germany
luft@fvk-berlin.de

RODNEY LUI
Muscle Research Unit, Institute for Biomedical Research
University of Sydney
Sydney, NSW
Australia
rod@anatomy.usyd.edu.au

ALBERTO LUINI
Department of Cell Biology and Oncology
Laboratory of Molecular Neurobiology
S. Maria Imbaro (Chieti)
Italy
luini@negrisud.it

JAMES R. LUPSKI
Department of Molecular and Human Genetics
Baylor College of Medicine
Houston, TX
USA
jlupski@bmc.tmc.edu

BARBARA LUSTIG
Clinic for Abdominal, Endocrine and Thoracic Surgery
Clinical Center Nürnberg
Nürnberg
Germany

HIROSHI MABUCHI
Molecular Biochemistry and Molecular Biology Laboratory,
Graduate School of Medical Sciences
Kanazawa University
Kanazawa
Japan

ARNAUD A. MAILLEUX
Developmental Biology Program, Department of Surgery
USC Keck School of Medicine and the Saban Research
Institute of Childrens Hospital Los Angeles
Los Angeles, CA
USA

MARY L. MARAZITA
Craniofacial and Dental Genetics, School of Dental Medicine
University of Pittsburgh, Pittsburgh
PA
USA
marazita@sdmgenetics.pitt.edu

MARC MAREEL
Laboratory of Experimental Cancerology
Ghent University Hospital
Ghent
Belgium

FLORIAN MARKOWETZ
Max Planck Institute for Molecular Genetics
Berlin Center for Genome Based Bioinformatics
Berlin
Germany

KENKICHI MASUTOMI
Department of Medical Oncology
Dana-Farber Cancer Institute
Boston, MA
USA

KARL MATTER
Division of Cell Biology, Institute of Ophthalmology
University College London
London
UK

MATTHIAS P. MAYER
Center for Molecular Biology Heidelberg (ZMBH)
Heidelberg
Germany
m.mayer@zmbh.uni-heidelberg.de

NAYEF A. MAZLOUM
Weill College of Medicine
Cornell University
New York, NY
USA

KIRSTEN E. MCLAY
The Wellcome Trust Sanger Institute
Wellcome Trust Genome Campus
Hinxton
Cambridgeshire, UK
kem@sanger.ac.uk

KARYN MÉGY
Sygen International plc, Department of Pathology
University of Cambridge
Cambridge
UK
km369@cam.ac.uk

FRAUKE MELCHIOR
University of Göttingen
Göttingen
Germany
f.melchior@medizin.uni-goettingen.de

DIETER K. MEYER
Institute for Experimental and Clinical Pharmacology and
Toxicology
Albert Ludwig University Freiburg
Freiburg
Germany
dieter.meyer@pharmakol.uni-freiburg.de

BURKHARD MICHEEL
Institute of Biochemistry and Biology
Department of Biotechnology
University of Potsdam
Germany
bmicheel@rz.uni-potsdam.de

ERIC MILOT
Centre de Recherche Guy-Bernier
Hopital Maissoneuve-Rosemont
Montreal, QC
Canada
ericmilot.hmr @ssss.gouv.qc.ca

LILIANA MINICHELLO
European Molecular Biology Laboratory
Mouse Biology Unit
Monterotondo
Italy
minichiello@embl-monterotondo.it

ALEXANDER A. MIRONOV
Department of Cell Biology and Oncology
Laboratory of Molecular Neurobiology
S. Maria Imbaro (Chieti)
Italy

CHANDRA MOHAN
Simmons Arthritis Research Center and Department of
Internal Medicine/Rheumatology
The University of Texas Southwestern Medical
Center at Dallas
Dallas, TX
USA
chandra.mohan@utsouthwestern.edu

ROLAND MOLL
Institute of Pathology
Philipp University
Marburg
Germany
mollr@med.uni-marburg.de

ALESSANDRA MONTECUCCO
Istituto di Genetica Molecolare
CNR, Pavia
Italy
montecucco@igm.cnr.it

RANDALL T. MOON
Department of Pharmacology, Howard Hughes
Medical Institute
University of Washington School of Medicine
Seattle, WA
USA
rtmoon@u.washington.edu

KEVIN L. MOORE
Oklahoma Medical Research Foundation
Oklahoma City, OK
USA
kevin-moore@omrf.ouhsc.edu

JAISREE MOORTHY
Biomedical Engineering
University of Wisconsin
Madison, WI
USA

PETRA MOOSMAYER
Molecular Genome Analysis
DKFZ
Heidelberg
Germany

PATRICE MORAND
EMBL
Grenoble Outstation and University Hospital
Grenoble
France

INGO MORANO
Max Delbrück Center for Molecular Medicine, Department
of Molecular Muscle Physiology, Berlin and, University
Medicine Berlin (Charité)
Johannes-Müller Institute for Physiology
Berlin
Germany
imorano@mdc-berlin.de

MARIO MÖRL
Institute for Biochemistry
University of Leipzig
Leipzig
Germany
moerl@uni-leipzig.de

ARISTIDIS MOUSTAKAS
Ludwig Institute for Cancer Research
Uppsala
Sweden

MARTINA MUCKENTHALER
Department of Hematology, Oncology and Immunology
University of Heidelberg
Heidelberg
Germany
martina.muckenthaler@med.uni-heidelberg.de

CHRISTOPH W. MÜLLER
EMBL
Grenoble Outstation and University Hospital
Grenoble
France
mueller@embl-grenoble.fr

EVA-CHRISTINA MÜLLER
Max Delbrück Center for Molecular Medicine
Berlin
Germany
emu@mdc-berlin.de

JÜRGEN J. MÜLLER
Max Delbrück Center for Molecular Medicine
Berlin
Germany
jjm@mdc-berlin.de

THOMAS MÜLLER
Department of Medical Genetics
Max Delbrück Center for Molecular Medicine
Berlin
Germany
thomu@mdc-berlin.de

JOCHEN MUELLER-DIECKMANN
EMBL Hamburg Outstation
c/o DESY
Hamburg
Germany
jochenmd@embl-hamburg.de

STEFAN MUNDLOS
Institute for Medical Genetics
Charité-University Medicine Berlin
Berlin
Germany
stefan.mundlos@charite.de

THOMAS MURRAY-RUST
The Oxford Centre for Molecular Sciences and
The Chemistry Research Laboratory
Oxford
UK
thomas.murray-rust@chem.ox.ac.uk

SERGE MUYLDERMANS
Cellular and Molecular Immunology, Vlaams
Inter-University Institute of Biotechnology
Free University of Brussels
Brussels
Belgium
svmuyld@vub.ac.be

OWEN W. NADEAU
Department of Biochemistry and Molecular Biology
Kansas University Medical Center
Kansas City, KS
USA
onadeau@kumc.edu

AKIRA NAGAFUCHI
Department of Cellular Interactions
Institute of Molecular Embryology and Genetics, Kumamoto
University
Kumamoto
Japan
naga-san@kaiju.medic.kumamoto-u.ac.jp

FRANÇOIS LE NAOUR
INSERM U602
Hospital Paul Brousse
Villejuif
France
flenaour@vjf.inserm.fr

MAHESH NARAYAN
Baker Laboratory of Chemistry and Chemical Biology
Cornell University
Ithaca, NY
USA

V. NARRY KIM
School of Biological Sciences and Institute
of Molecular Biology and Genetics
Seoul National University
Seoul
Korea
narrykim@snu.ac.kr

HEINZ-PETER NASHEUER
Department of Biochemistry
National University of Ireland
Galway
Ireland
h.nasheuer@nuigalway.ie

KLAUS-ARMIN NAVE
Department of Neurogenetics
Max Planck Institute of Experimental Medicine
Goettingen
Germany
nave@em.mpg.de

LORENZ NEIDHARDT
Department of Developmental Genetics
Max Planck Institute for Molecular Genetics
Berlin
Germany
neidhardt@molgen.mpg.de

EVA NELIS
University of Antwerpen, Molecular Genetics Laboratory,
Peripheral Neuropathy Group
Antwerpen
Belgium
eva.nelis@ua.ac.be

ALAIN NEPVEU
Departments of Biochemistry, Medicine, and Oncology
McGill University
Montreal, QC
Canada
alain.nepveu@mcgill.ca

WALTER NEUPERT
Department of Physiological Chemistry
Ludwig-Maximilians-University Munich
Munich
Germany
neupert@med.uni-muenchen.de

ALISTAIR E.T. NEWALL
Human Cytogenetics Laboratory
Cancer Research UK
London Research Institute
London
UK

ERICH A. NIGG
Max-Planck-Institute for Biochemistry
Martinsried
Germany
nigg@biochem.mpg.de

MICHAEL NILGES
Unité de Bio-Informatique Structurale
Institute Pasteur
Paris
France
nilges@pasteur.fr

JOHN L. NITISS
Department of Molecular Pharmacology
St. Jude Children's Research Hospital
Memphis, TN
USA
john.nitiss@stjude.org

NADINE NORTON
Department of Psychological Medicine, Wales School of
Medicine
Cardiff University
Cardiff
UK
nortonn@cardiff.ac.uk

MARKUS M. NÖTHEN
Department of Medical Genetics
University of Antwerp
Belgium

Department of Genomics
Live and Brain Center
University of Bonn
Germany

Institute of Human Genetics
University of Bonn
Germany
markus.noethen@uni-bonn.de

FILIPP OESTERHELT
Institute for Physical Chemistry II, AG Nanoanalytics
Heinrich-Heine University
Düsseldorf
Germany
filipp.oesterhelt@uni-duesseldorf.de

SHUJI OGINO

Department of Pathology, Brigham and Women's Hospital;
Department of Medical Oncology
Dana-Farber Cancer Institute; and
Harvard Medical School
Boston, MA
USA
sogino@partners.org

MARIA JOSÉ OLIVEIRA

Laboratory of Experimental Cancerology
Ghent University Hospital
Ghent
Belgium

IPATIMUP, Institute of Molecular Pathology and
Immunology
University of Porto
Porto
Portugal

CLAUDIA ORELIO

Department of Cell Biology and Genetics
Erasmus University Medical Center
Rotterdam
The Netherlands
c.orelio@erasmusmc.nl

MICHAEL J. OWEN

Department of Psychological Medicine
Wales School of Medicine
Cardiff University
Cardiff
UK
owenmj@cardiff.ac.uk

YOUNG-KI PAIK

Yonsei Proteome Research Center
Biomedical Proteome Research Center
Seoul
Korea

Department of Biochemistry
Yonsei University
Seoul
Korea
paiky@yonsei.ac.kr

BENOÎT PALANCADE

Régulation de l'Expression Génétique
Ecole Normale Supérieure (ENS)
Paris
France

ANDRAS PALDI

Ecole Pratique des Hautes Etudes
Evry
France
paldi@genethon.fr

IGNACIO PALMERO

Institute of Biomedical Research
CSIC-UAM
Madrid
Spain
ipalmero@iib.uam.es

ANJA PANNEK

Max-Delbrück-Center for Molecular Medicine
Berlin
Germany

ELS PARDON

Cellular and Molecular Immunology, Vlaams
Inter-University Institute of Biotechnology
Free University of Brussels
Brussels
Belgium

RENATO PARO

ZMBH
University of Heidelberg
Heidelberg
Germany
paro@zmbh.uni-heidelberg.de

EBERHARD PASSARGE

Institute of Human Genetics
University Clinic Essen
Essen
Germany
eberhard.passarge@uni-essen.de

ALBERT PASTINK

Department of Toxicogenetics
LUMC
Leiden
The Netherlands
a.pastink@lumc.nl

ADAM PAVLICEK

Genetic Information Research Institute
Mountain View, CA
USA

MIREILLE PELLIS

Cellular and Molecular Immunology, Vlaams
Inter-University Institute of Biotechnology
Free University of Brussels
Brussels
Belgium

PÄIVI PELTOMÄKI

Department of Medical Genetics
University of Helsinki
Helsinki
Finland
paivi.peltomaki@helsinki.fi

MARCO PETRELLA

McGill University AIDS Centre, Lady Davis Institute for
Medical Research, Jewish General Hospital
Montreal, QC
Canada

CHRISTOPH PLASS

The Ohio State University
Columbus, OH
USA
plass-1@medctr.osu.edu

KARL H. PLATE
Neurological Institute, Clinical Center of the
Johann Wolfgang Goethe University
Frankfurt
Germany
karl-heinz.plate@kgu.de

G. ERIC PLUM
Department of Chemistry and Chemical Biology,
Rutgers State University of New Jersey
Piscataway, NJ
USA

IBET, Inc.,
Columbus, OH
USA

ROMAN S. POLISHCHUK
Department of Cell Biology and Oncology
Laboratory of Molecular Neurobiology
S. Maria Imbaro (Chieti)
Italy

THOMAS POLLMÄCHER
Center of Mental Health
Clinical Center Ingolstadt
Ingolstadt
Germany
thomas.pollmaecher@klinikum-ingolstadt.de

THOMAS POMORSKI
Institute of Biology, and Center of Biophysics
and Bioinformatics
Humboldt University Berlin
Berlin
Germany
thomas.pomorski@rz.hu-berlin.de

HELMUT POSPIECH
Department of Biochemistry
National University of Ireland
Galway
Ireland

Biocenter Oulu and Department of Biochemistry
University of Oulu
Oulu
Finland

THOMAS PREISS
Victor Chang Cardiac Research Institute
Molecular Genetics Program
Darlinghurst/Sydney, NSW
Australia
t.preiss@victorchang.unsw.edu.au

GABRIELE PROETZEL
The Jackson Laboratory
Maine, ME
USA
g.proetzel@scil.com

KATY RADDATZ
MDC
Berlin
Germany

SALLY RADOVICK
Section of Pediatric Endocrinology
The University of Chicago
Chicago, IL
USA
sradovici@peds.bsd.uchicago.edu

TIM W. RATZLAFF
Department of Neurology
University of Pennsylvania School of Medicine
Philadelphia, PA
USA

BERND RAUTENSTRAUß
Friedrich-Alexander-University, Institute of Human Genetics
Erlangen
Germany
bernd.rautenstrauss@humgenet.uni-erlangen.de

GURU REDDY
CIPHERGEN Biosystems, Inc.
Fremont, CA 94555
USA

WESTLEY H. REEVES
Division of Rheumatology & Clinical Immunology, Center
for Autoimmune Diseases
University of Florida, Gainesville
FL
USA
whreeves@ufl.edu

ANDREAS S. REICHERT
Department of Physiological Chemistry
Ludwig-Maximilians-University Munich
Munich
Germany
andreas.reichert@med.uni-muenchen.de

EVAN REID
Department of Medical Genetics
University of Cambridge
Cambridge
UK
ereid@hgmp.mrc.ac.uk

JÖRG REIMANN
Institute for Medical Microbiology and Immunology
University of Ulm
Ulm
Germany

T. REINISCH
Institute for Physical Chemistry
WWU Münster
Münster
Germany

CRISTOBAL G. DOS REMEDIOS
Muscle Research Unit, Institute for Biomedical Research
University of Sydney
Sydney, NSW
Australia
crisdos@anatomy.usyd.edu.au

HERVÉ-W. RÉMIGY
M.E. Müller Institute for Structural Biology, Biocenter
University of Basel
Basel
Switzerland

RAINER RENKAWITZ
Institute for Genetics
Justus-Liebig-University
Giessen
Germany
rainer.renkawitz@gen.bio.uni-giessen.de

ALEXANDRE REYMOND
Center for Integrative Genomics
University of Lausanne
Lausanne
Switzerland
alexandre.reymond@medecine.unige.ch

GABRIELE RICHARD
Department of Dermatology and Cutaneous Biology, and
Jefferson Institute of Molecular Medicine
Thomas Jefferson University
Philadelphia, PA
USA

ISABELLE RICHARD
Généthon CNRS
Evry
France
richard@genethon.fr

ANNE RIDLEY
Ludwig Institute for Cancer Research, Royal Free and
University College School of Medicine, Department of
Biochemistry and Molecular Biology
University College London
London
UK
anne@ludwig.ucl.ac.uk

MARCELLA RIETSCHEL
Division of Genetic Epidemiology in Psychiatry
Central Institute of Mental Health
Mannheim
Germany
rietschel@zi-mannheim.de

LEONIE RINGROSE
Centre for Molecular Biology (ZMBH)
University of Heidelberg
Germany
l.ringrose@zmbh.uni-heidelberg.de

SILVANO RIVA
Istituto di Genetica Molecolare, CNR
Pavia
Italy
riva@igm.cnr.it

ELIZABETH A. ROCHOWICZ
Section of Pediatric Endocrinology
The University of Chicago
Chicago, IL
USA

MARINA V. RODNINA
Institutes of Molecular Biology and
Physical Biochemistry
University of Witten/Herdecke
Germany
rodnina@uni-wh.de

ENRIQUE RODRIGUEZ-BOULAN
Margaret M. Dyson Vision Research Institute and
Departments of Cell Biology and Ophthalmology
Weill Medical College of Cornell University
New York, NY
USA
boulan@med.cornell.edu

STEPHANE ROMBAUTS
Department of Plant Systems Biology
University of Ghent
Ghent
Belgium

JÖRG RÖSGEN
Institute of Human Biological Chemistry and Genetics
University of Texas Medical Branch
Galveston, TX
USA

ROSSELLA ROSSI
Istituto di Genetica Molecolare
CNR, Pavia
Italy

WOLFGANG ROTTBAUER
University Clinic Heidelberg
Inner Medicine III
Heidelberg
Germany
wolfgang.rottbauer@med.uni-heidelberg.de

PIERRE ROUZÉ
Department of Plant Systems Biology
University of Ghent
Ghent
Belgium
pierre.rouze@psb.ugent.be

ANANDA L. ROY
Departments of Pathology and Biochemistry, Programs in
Immunology and Genetics
Tufts University School of Medicine
Boston, MA
USA
ananda.roy@tufts.edu

JOY ROY
Department of Surgery
St Gorans Hospital
Stockholm
Sweden
ulf.hedin@kirurgi.ki.se

RICHARD RUBIN
CIPHERGEN Biosystems, Inc.
Fremont, CA 94555
USA

THOMAS VON RÜDEN
MorphoSys AG
Martinsried
Germany
vonRueden@morphosys.com

PATRICIA RUIZ
Center for Cardiovascular Research
Charité University Medicine Berlin and
Max Planck Institute for Molecular Genetics
Berlin
Germany
ruiz@molgen.mpg.de

LAURA J. RUSH
The Ohio State University
Columbus, OH
USA
rush.61@osu.edu

DIRK SAERENS
Cellular and Molecular Immunology, Vlaams
Inter-University Institute of Biotechnology
Free University of Brussels
Brussels
Belgium

FREDERIC SALA
Developmental Biology Program, Department of Surgery
USC Keck School of Medicine and the Saban Research
Institute of Childrens Hospital Los Angeles
Los Angeles, CA
USA

FRANCESCO SALVATORE
CEINGE – Biotecnologie Avanzate, scrl and Dipartimento
di Biochimica e Biotecnologie Mediche
University of Napoli “Federico II”
Napoli
Italy
salvator@unina.it

LEONA D. SAMSON
Division of Biological Engineering and Center for
Environmental Health Sciences
Massachusetts Institute of Technology
MA
USA
lsamson@mit.edu

ANDREW SANDFORD
University of British Columbia McDonald Research
Laboratories / iCAPTURE Center
Vancouver, BC
Canada
asandford@mrl.ubc.ca

ROZANNE M. SANDRI-GOLDIN
Department of Microbiology and Molecular Genetics
University of California
Irvine, CA
USA
rmsandri@uci.edu

MASAYUKI SANO
Gene Function Research Center
National Institute of Advanced Industrial Science and
Technology (AIST)
Tsukuba Science City
Japan

HANS MARTIN SASS
Kennedy Institute of Ethics
Georgetown University
Washington DC
USA
sassh@georgetown.edu

BRIAN SAUER
Stowers Institute for Medical Research, Kansas City
MO
USA
bls@stowers-institute.org

SASCHA SAUER
Max-Planck-Institute for Molecular Genetics
Berlin
Germany
sauer@molgen.mpg.de

CAROLINE O.S. SAVAGE
Renal Immunobiology, The Medical School
University of Birmingham
Birmingham
UK

ELLADA SAVVIDOU
Wellcome Trust Centre for Cell Biology
University of Edinburgh
Edinburgh
UK
elladasav@hotmail.com

UTE SCHAEFER
Max Delbrück Center for Molecular Medicine
Berlin
Germany
uschaep@mde-berlin.de

REINHOLD SCHÄFER
Laboratories of Molecular Tumour Pathology and of
Functional Genomics, Charité
University Medicine Berlin
Berlin
Germany
reinhold.schaefer@charite.de

WALTER SCHAFFNER
Institute of Molecular Biology
Zurich University
Zurich
Switzerland
walter.schaffner@molbio.unizh.ch

LEONARD C. SCHALKWYK
Social, Genetic and Developmental Psychiatry
Institute of Psychiatry
London
UK
spjgls@iop.kcl.ac.uk

CLAUS SCHEIDEREIT
Max Delbrück Center for Molecular Medicine
Berlin
Germany
scheidereit@mdc-berlin.de

HAROLD A. SCHERAGA
Baker Laboratory of Chemistry and Chemical Biology
Cornell University
Ithaca, NY
USA
has5@cornell.edu

CORDELIA SCHIENE-FISCHER
Max Planck Research Unit for Enzymology of Protein
Folding
Halle
Germany
schiene@enzyme-halle.mpg.de

REINHOLD SCHIRMBECK
Institute for Medical Microbiology and Immunology
University of Ulm
Ulm
Germany
reinhold.schirmbeck@medizin.uni-ulm.de

NIKOLAUS L. SCHLAICH
Institute Bio III (Plant Physiology)
RWTH Aachen
Aachen
Germany
schlaich@bio3.rwth-aachen.de

MARTIN SCHLEEF
Institute for Medical Microbiology and Immunology
University of Ulm
Ulm
Germany

MICHAEL SCHLEICHER
Adolf-Butenandt-Institute / Cell Biology
Ludwig-Maximilians-University
Munich
Germany
schleicher@lrz.uni-muenchen.de

KARIN SCHLEINKOFER
Department of Bioinformatics, Biocenter
University of Würzburg
Würzburg
Germany

MANFRED SCHLIWA
Institute for Cell Biology
University of Munich
Munich
Germany
schliwa@med.uni-muenchen.de

CHRISTINE SCHMÄL
Division of Genetic Epidemiology in Psychiatry
Central Institute of Mental Health
Mannheim
Germany
schmael@as200.zi-mannheim.de

MICHAEL F.G. SCHMIDT
Department of Immunology and Molecular Biology,
Faculty of Veterinary Medicine
Free University
Berlin
Germany
schmidt.mfg@vetmed.fu-berlin.de

ANDREA SCHMIDT
European Molecular Biology Laboratory (EMBL),
Hamburg c/o DESY
Hamburg
Germany
andrea@embl-hamburg.de

JÖRG SCHMIDTKE
Institute of Human Genetics
Hannover Medical School
Hannover
Germany
Schmidtke.Joerg@mh-hannover.de

PETER SCHMIEDER
FMP, Institute for Molecular Pharmacology
Berlin
Germany
schmieder@fmp-berlin.de

GERD SCHMITZ
Institute for Clinical Chemistry and Laboratory Medicine
University Regensburg
Regensburg
Germany
gerd.schmitz@klinik.uni-regensburg.de

CHRISTOPHER SCHOFIELD
The Oxford Centre for Molecular Sciences and
The Chemistry Research Laboratory
Oxford
UK
christopher.schofield@chem.ox.ac.uk

MARINA SCHORPP-KISTNER
Deutsches Krebsforschungszentrum, Division of Signal
Transduction and Growth Control
Heidelberg
Germany
marina.schorpp@dkfz.de

ANJA SCHRAMME
Laboratories of Molecular Tumour Pathology and of
Functional Genomics, Charité
University Medicine Berlin
Berlin
Germany

STEFAN SCHREIBER
Institute for Clinical Molecular Biology and Department of
General Internal Medicine
Christian-Albrechts-University
Kiel
Germany
s.schreiber@mucosa.de

PETER SCHUCK
Division of Bioengineering and Physical Science, OD
National Institutes of Health
Bethesda, MD
USA
pschuck@helix.nih.gov

JÖRG SCHULTZ
Department of Bioinformatics, Biocenter
Würzburg University
Würzburg
Germany
joerg.schultz@biozentrum.uni-wuerzburg.de

INGO SCHUPP
Molecular Genome Analysis
DKFZ
Heidelberg
Germany

MATTHIAS SCHWAB
Dr. Margarete Fischer-Bosch Institute of Clinical
Pharmacology
Stuttgart
Germany

MANFRED SCHWAB
DKFZ
Heidelberg
Germany
m.schwab@dkfz.de

ANNEMARIE SCHWAN
Human Genetics
Ruhr-University Bochum
Bochum
Germany

MATTHIAS SCHWEIZER
Paul-Ehrlich-Institute
Division of Medical Biotechnology
Langen
Germany

PETRA SCHWILLE
Institute for Biophysics/BioTec
Dresden University of Technology
Dresden
Germany
petra.schwille@biotec.tu-dresden.de

EVA SEEMANOVA
Institute of Biology and Medical Genetics
2nd Medical School, Charles University
Prague
Czech Republic
eva.seemanova@lfmotol.cuni.cz

AARON J. SHATKIN
Center for Advanced Biotechnology and Medicine
Piscataway, NY
USA
shatkin@cabm.rutgers.edu

LIUBOV SHATKINA
Forschungszentrum Karlsruhe
Institute of Toxicology and Genetics
Karlsruhe
Germany

DENISE SHEER
Human Cytogenetics Laboratory
Cancer Research UK London Research Institute
London
UK
denise.sheer@cancer.org.uk

AHMED SHERIFF
Institute of Clinical Immunology and Rheumatology,
Department of Internal Medicine III
Friedrich-Alexander University of Erlangen-Nuernberg
Erlangen
Germany
sheriff@genethor.com

IGOR SHEVELEV
Institute of Veterinary Biochemistry and Molecular Biology
University of Zürich
Zürich
Switzerland

SEBASTIAN M. SHIMELD
Department of Zoology
University of Oxford
Oxford
UK
sebastian.shimeld@zoo.ox.ac.uk

HANS-UWE SIMON
Department of Pharmacology
University of Bern
Bern
Switzerland
hus@pki.unibe.ch

JEREMY C. SIMPSON
Cell Biology and Biophysics Programme, European
Molecular Biology Laboratory (EMBL)
Heidelberg
Germany
simpson@embl-heidelberg.de

ROBERT H. SINGER
Anatomy & Structural Biology
Albert Einstein College of Medicine
Bronx, New York, NY
USA
rhsinger@aecom.yu.edu

KARL SKRINER
Department of Rheumatology and Clinical Immunology
Charité University Medicine Berlin
Berlin
Germany
karl.skriner@charite.de

SCOTT SLATTERY
Department of Molecular and Cellular Biology
Baylor College of Medicine
Houston, Texas
USA

BONNIE F. SLOANE
Department of Pharmacology
Wayne State University and Barbara Ann Karmanos
Cancer Institute, Detroit
MI
USA
bsloane@med.wayne.edu

ALAN J. SLUSARENKO
Institute Bio III (Plant Physiology)
RWTH Aachen
Aachen
Germany
alan.slusarenko@bio3.rwth-aachen.de

ERIK LEE SNAPP
Department of Anatomy and Structural Biology
Albert Einstein College of Medicine
Bronx, NY
USA
esnapp@aecom.yu.edu

DIAN SOEWARTO
GSF Research Center for Environment and Health
Neuherberg
Germany
hrabe@gsf.de

THOMAS SOMMER
Max Delbrück Center for Molecular Medicine
Berlin
Germany
sommer@mdc-berlin.de

RALF J. SOMMER
Max Planck Institute for Developmental
Biology, Department of Evolutionary Biology
Tübingen
Germany
ralf.sommer@tuebingen.mpg.de

ARNOUD SONNENBERG
Division of Cell Biology
The Netherlands Cancer Institute
Amsterdam
The Netherlands
a.sonnenberg@nki.nl

LENE SØRENSEN
Karolinska Institute, Department of Laboratory Medicine,
NOVUM
Karolinska University Hospital-Huddinge
Stockholm
Sweden
lene.sorensen@labmed.ki.se

NATALIA SOSHNIKOVA
Max-Delbrueck-Center for Molecular Medicine
Berlin
Germany
nsoschni@mdc-berlin.de

CLAUDIO SOTO
Department of Neurology
University of Texas Medical Branch
TX
USA
clsoto@utmb.edu

CHRISTOPH A. SOTRIFFER
Department of Pharmaceutical Chemistry
Philipps-University
Marburg
Germany
sottriffer@staff.uni-marburg.de

SERHIY SOUCHELNITSKYI
Ludwig Institute for Cancer Research
Uppsala
Sweden

RAINER SPANG
Max Planck Institute for Molecular Genetics
Berlin Center for Genome Based Bioinformatics
Berlin
Germany
spang@molgen.mpg.de

JAGAN SRINIVASAN
Max Planck Institute for Developmental
Biology, Department of Evolutionary Biology
Tübingen
Germany

KATRIN STADE
Max-Delbrück-Center for Molecular Medicine
Berlin
Germany
kstade@mdc-berlin.de

IGOR STAGLJAR
Institute of Veterinary Biochemistry and
Molecular Biology
University of Zurich
Zurich
Switzerland
stagljar@vetbio.unizh.ch

STEFAN STAMM
Institute of Biochemistry
Erlangen
Germany
stefan@stamms-lab.net

ULRIKE STEIN
Max-Delbrück-Center for Molecular Medicine
Berlin
Germany
ustein@mdc-berlin.de

PETER STEINBACH
Department of Human Genetics
University of Ulm
Ulm
Germany
peter.steinbach@medizin.uni-ulm.de

ALEXANDER STEPANOV
Department of Molecular Pharmacology
St. Jude Children's Research Hospital
Memphis, TN
USA

DIETRICH A. STEPHAN
The Translational Genomics Research Institute
Phoenix, AZ
USA
dstephan@tgen.org

WOLFRAM STERRY
Department for Dermatology and
Allergology, Medical School Charité
Humboldt University Berlin
Berlin
Germany

RANDY STRICH
Fox Chase Cancer Center
Philadelphia, PA
USA
r_strich@fccc.edu

MILTON T. STUBBS II
Institute for Biotechnology
Martin Luther University Halle-Wittenberg
Halle (Saale)
Germany
stubbs@biochemtech.uni-halle.de

BRYAN K. SUN
Howard Hughes Medical Institute,
Department of Molecular Biology, Massachusetts General
Hospital, Department
of Genetics
Harvard Medical School
Boston, MA
USA
bryansun@post.harvard.edu

ZHEN-YU J. SUN
Department of Biological Chemistry and Molecular
Pharmacology
Harvard Medical School
Boston, MA
USA
zhen-yu_sun@hms.harvard.edu

MICHAEL SUTTERS
Division of Renal Medicine
Johns Hopkins University School of Medicine
Baltimore, MD
USA
msutters@jhmi.edu

JUHANI SYVÄOJA
Department of Biochemistry
National University of Ireland
Galway
Ireland

Department of Biology
University of Joensuu
Joensuu
Finland

KAZUNARI TAIRA
Department of Chemistry and Biotechnology
School of Engineering
University of Tokyo
Tokyo
Japan
taira@chembio.t.u-tokyo.ac.jp

HIROYUKI TAKEDA
Department of Biological Sciences
Graduate School of Science
University of Tokyo
Tokyo
Japan
htakeda@biol.s.u-tokyo.ac.jp

MASATOSHI TAKEICHI
RIKEN Center for Developmental Biology
Chuo-ku, Kobe
JAPAN
takeichi@cdb.riken.jp

JAN F. TALTS
Animal and Veterinary Basic Sciences
The Royal Veterinary and Agricultural University
Frederiksberg
Denmark
jft@kvl.dk

HENGLI TANG
Department of Biological Science
Florida State University
Tallahassee, FL
USA
tang@bio.fsu.edu

WOAN-YUH TARN
Institute of Biomedical Sciences
Academia Sinica
Taipei
Taiwan
wtarn@ibms.sinica.edu.tw

J. PAUL TAYLOR
Department of Neurology
University of Pennsylvania School of Medicine
Philadelphia, AZ
USA
jpt@mail.med.upenn.edu

PETER C. TAYLOR
The Kennedy Institute of Rheumatology Division, Faculty
of Medicine
Imperial College London
London
UK
peter.c.taylor@imperial.ac.uk

CHARLES TEASE
Department of Biological Sciences
University of Warwick
Coventry
UK
ctease@btinternet.com

HEINRICH TERLAU
Institute of Experimental and Clinical
Pharmacology and Toxicology
University Clinic of Schleswig-Holstein, Campus Lübeck
Lübeck
Germany
heinrich.terlau@uni-luebeck.de

KAY TERPE
Proteins and Nucleic Acids
IBA GmbH
Göttingen
Germany
terpe@iba-go.com

GERALD THIEL
Department of Medical Biochemistry and Molecular Biology
University of Saarland Medical Center
Homburg
Germany
gerald.thiel@uniklinik-saarland.de

KATSUSHI TOKUNAGA
University of Tokyo, Graduate School of Medicine,
Department Of Human Genetics
Tokyo
Japan
tokunaga@m.u-tokyo.ac.jp

IAN TOMLINSON
London Research Institute
Cancer Research UK
London
UK
ian.tomlinson@cancer.org.uk

NIELS TOMMERUP
Department of Medical Biochemistry and Genetics, Wilhelm
Johannsen Center for Functional Genome Research
University of Copenhagen
Copenhagen
Denmark
tommerup@imbg.ku.dk

OLIVIER TOUSSAINT
University of Namur
Namur
Belgium
olivier.toussaint@fundp.ac.be

MARY TRUSCOTT
Departments of Biochemistry, Medicine, and Oncology
McGill University
Montreal, QC
Canada

JACK UETRECHT
Pharmacy and Medicine
University of Toronto
Toronto, ON
Canada
jack.uetrecht@utoronto.ca

PETER UETZ
Institute for Genetics
Forschungszentrum Karlsruhe
Karlsruhe
Germany
peter.uetz@itg.fzk.de

JOUNI UITTO
Department of Dermatology and Cutaneous Biology, and
Jefferson Institute of Molecular Medicine
Thomas Jefferson University, Philadelphia
Pennsylvania, MA
USA
jouni.uitto@jefferson.edu

MEENA UPADHYAYA
Institute of Medical Genetics
University of Wales College of Medicine
Cardiff
UK
upadhyaya@cardiff.ac.uk

SEPPO J. VAINIO
Biocenter Oulu and Department of Biochemistry, Linnanmaa,
Faculties of Science and Medicine
University of Oulu
Oulu
Finland
seppo.vainio@oulu.fi

Y. VAN DE PEER
Department of Plant Systems Biology
University of Ghent
Ghent
Belgium

J. VAN DEN HOFF
PET Center
Institute of Bioinorganic and Radiopharmaceutical
Chemistry, Research Center Rossendorf
Dresden
Germany
j.van_den_hoff@fz-rossendorf.de

DOMINIQUE VANHECKE
Max Planck Institute for Molecular Genetics, Department
Vertebrate Genomics
Berlin
Germany
vanhecke@molgen.mpg.de

MICHAEL VEIT
Department of Immunology and Molecular Biology, Faculty
of Veterinary Medicine
Free University
Berlin
Germany
mveit@zedat.fu-berlin.de

LOREDANA VERDONE
Dipartimento di Genetica e Biologia Molecolare
University La Sapienza
Rome
Italy

CÉCILE VINCKE
Cellular and Molecular Immunology, Vlaams
Inter-University Institute of Biotechnology
Free University of Brussels
Brussels
Belgium

MARTIN VINGRON
Max Planck Institute of Molecular Genetics
Berlin
Germany
vingron@molgen.mpg.de

PETER K. VOGT
The Scripps Research Institute
La Jolla, CA
USA
pkvogt@scripps.edu

JENS VÖLKER
Department of Chemistry and Chemical Biology, Rutgers
State University of New Jersey
Piscataway, NJ
USA

IBET, Inc.
Columbus, OH
USA

RUDOLF VOLKMER-ENGERT
Molecular Libraries, Institute for Medical Immunology
Charite – University Clinical Center Berlin
Berlin
rve@charite.de

REINHARD E. VOLL
Institute of Clinical Immunology and Rheumatology,
Department of Internal Medicine III
Friedrich-Alexander University of Erlangen-Nuernberg
Erlangen
Germany

CAROLINA S. VOLLERT
Institute for Genetics
Forschungszentrum Karlsruhe
Karlsruhe
Germany

ANNEMIEKE DE VRIES
National Institute of Public Health and the Environment
Laboratory of Toxicology, Pathology and Genetics
Bilthoven
The Netherlands
annemieke.de.vries@rivm.nl

STEPHEN S. WACHTEL
Reproductive Genetics
University of Tennessee
Memphis, TN
USA
swachtel@utmem.edu

TADASHI WADA
Tokyo Institute of Technology
Graduate School of Bioscience and Biotechnology,
Midori-ku
Yokohama
Japan

REBECCA C. WADE
Molecular and Cellular Modeling Group
EML Research gGmbH
Heidelberg
Germany
rebecca.wade@eml-r.villa-bosch.de

SIBYLLE WAGNER
GSF Research Center for Environment and Health
Neuherberg
Germany

GERHARD WAGNER
Department of Biological Chemistry and Molecular
Pharmacology
Harvard Medical School
Boston, MA
USA
gerhard_wagner@hms.harvard.edu

ELMAR WAHLE
Institute for Biochemistry
Martin-Luther-University Halle-Wittenberg
Halle
Germany
ewahle@biochemtech.uni-halle.de

MARK A. WAINBERG
McGill University AIDS Centre, Lady Davis Institute for
Medical Research, Jewish General Hospital
Montreal, QC
Canada
mark.wainberg@mcgill.ca

HARALD WAJANT
Department of Molecular Internal Medicine
Medical Clinic and Polyclinic II
Würzburg
Germany
harald.wajant@mail.uni-wuerzburg.de

OLIVER WALISKO
Max Delbrück Center for Molecular Medicine
Berlin
Germany

RONALD J.A. WANDERS
Pediatrics and Clinical Chemistry
University of Amsterdam, Academic Medical Centre
Amsterdam
The Netherlands
r.j.wanders@amc.uva.nl

JAYSON WANG
Human Cytogenetics Laboratory
Cancer Research UK London Research Institute
London
UK

TING WANG
Molecular and Cellular Modeling Group
EML Research gGmbH
Heidelberg
Germany

ERICH E. WANKER
Department of Neuroproteomics
Max Delbrück Center for Molecular Medicine Berlin-Buch
Berlin
Germany
ewanker@mdc-berlin.de

DERICK G. WANSINK
Department of Cell Biology, Nijmegen Center for Molecular
Life Sciences
Radboud University Nijmegen Medical Center
Nijmegen
The Netherlands
r.wansink@ncmls.ru.nl

PAUL M. WASSARMAN
Brookdale Department of Molecular
Cell & Developmental Biology
Mount Sinai School of Medicine
New York, NY
USA
paul.wassarman@mssm.edu

WOLFRAM WECKWERTH
Max-Planck-Institute of Molecular Plant Physiology
Golm/Potsdam
Germany

BERHARD WEHRLE-HALLER
Department of Pathology and Immunology and
Department of Cellular Physiology and Metabolism
University of Geneva
Geneva
Switzerland

SCOT WEINBERGER
CIPHERGEN Biosystems, Inc.
Fremont, CA 94555
USA
sweinberger@ciphergen.com

ROLAND WEIS
Research Centre Applied Biocatalysis, Institute of
Molecular Biotechnology
Graz University of Technology
Austria

MANFRED S. WEISS
EMBL Hamburg Outstation
c/o DESY
Hamburg
Germany
msweiss@embl-hamburg.de

ERVIN WELKER
Baker Laboratory of Chemistry and Chemical Biology
Cornell University
Ithaca, NY
USA

ANDREAS WERNER
Institute of Cell and Molecular Biosciences
University of Newcastle
Newcastle upon Tyne
UK
andreas.werner@ncl.ac.uk

HAUKE WERNER
Department of Neurogenetics
Max Planck Institute of Experimental Medicine
Goettingen
Germany
hauke@em.mpg.de

SABINE WERNER
Department of Biology
Institute of Cell Biology
ETH Zurich, Zurich
Switzerland
sabine.werner@cell.biol.ethz.ch

LOUISE VAN DER WEYDEN
Wellcome Trust Genome Campus Hinxton
The Sanger Institute
Cambridge
UK

ROBERT J. WHITE
Institute of Biomedical and Life Sciences
University of Glasgow
Glasgow
UK
rwhite@udcf.gla.ac.uk

CAROL WICKING
The Institute for Molecular Bioscience
The University of Queensland
St Lucia, QLD
Australia
c.wicking@imb.uq.edu.au

STEFAN WIECZOREK
Human Genetics
Ruhr-University
Bochum
Germany
stefan.wieczorek@rub.de

STEFAN WIEMANN
Department of Molecular Genome Analysis
German Cancer Research Center
Heidelberg
Germany
s.wiemann@dkfz.de

BÉ WIERINGA
Department of Cell Biology, Nijmegen
Center for Molecular Life Sciences
Radboud University Nijmegen Medical Center
Nijmegen
The Netherlands
b.wieringa@ncmls.ru.nl

MICHAEL WIESE
Rheinische Friedrich-Wilhelms University
Bonn
Germany
mwiese@uni-bonn.de

BURKHARD WIESNER
Research Institute for Molecular Pharmacology
Berlin
Germany
wiesner@fmp-berlin.de

SUSAN WIJNHOFEN
National Institute of Public Health and the Environment
Laboratory of Toxicology, Pathology and
Genetics
Bilthoven
The Netherlands
susan.wijnhoven@rivm.nl

MICHAEL V. WILES
The Jackson Laboratory
Maine, ME
USA
mvwiles@jax.org

MILES F. WILKINSON
Department of Immunology
The University of Texas, MD Anderson Cancer Center
Houston, TX
USA
mwilkins@mdanderson.org

JULIE M. WILLIAMS
Renal Immunobiology
The Medical School, University of Birmingham
Birmingham
UK
j.m.williams.med@bham.ac.uk

MATTHIAS WILMANN
EMBL Hamburg Outstation
c/o DESY
Hamburg
Germany
wilmanns@embl-hamburg.de

WOLFGANG WINTERMEYER
Institutes of Molecular Biology and Physical Biochemistry
University of Witten/Herdecke
Germany
winterme@uni-wh.de

ERHARD WINTERSBERGER
Institute of Medical Biochemistry, Division of Molecular
Biology
Medical University of Vienna
Vienna
Austria
erhard.wintersberger@univie.ac.at

JOACHIM WITTBRODT
Developmental Biology Programme, European Molecular
Biology Laboratory
EMBL
Heidelberg
Germany
jochen.wittbrodt@embl.de

STEFAN WÖFL
Institut für Pharmazie und Molekulare Biotechnologie,
Abteilung Biologie, Arbeitsgruppe Bioanalytik und
Molekularbiologie
Ruprecht-Karls-Universität Heidelberg
Heidelberg
wolfl@uni-hd.de

CHUN XIE
Simmons Arthritis Research Center and Department of
Internal Medicine/Rheumatology
The University of Texas Southwestern Medical
Center at Dallas
Dallas, TX
USA

JU-HUA XU

Department of Chemistry
University of Minnesota
Minneapolis, MN
USA

YUKI YAMAGUCHI

Tokyo Institute of Technology
Graduate School of Bioscience and Biotechnology
Midori-ku, Yokohama
Japan

UWE ZANGEMEISTER-WITTKE

Department of Oncology
University of Zurich
Zurich
Switzerland
uwe.zangemeister@USZ.ch

MALGORZATA Z. ZDZIENICKA

Department of Molecular Cell Genetics, The L. Rydygier
Collegium Medicum
Nicolaus Copernicus University
Bydgoszcz
Poland

ULRIKE ZIEBOLD

Max Delbrück Centre for Molecular Medicine
Berlin
Germany
uziebold@mdc-berlin.de

WALTER ZIEGLGÄNSBERGER

Max Planck Institute of Psychiatry
Munich
Germany
zieglgaensberger@mpipsykl.mpg.de

HEIKE ZIMDAHL

Max Delbrück Center for Molecular Medicine (MDC)
Berlin
Germany

JOOST C.B.M. ZOMERDIJK

Division of Gene Regulation and Expression, Wellcome
Trust Biocentre
University of Dundee
UK
j.zomerdiijk@dundee.ac.uk

EBERHART ZRENNER

University Eye Hospital
Eberhard Karl University Tübingen
Germany
ezrenner@uni-tuebingen.de

AAA ATPase

Definition

AAA ATPases form a large and extremely heterogeneous family of enzymes belonging to the AAA superfamily containing a ring-shaped NTPase domain. These enzymes are involved in a wide variety of cellular processes, hence the name (ATPases Associated with various cellular Activities).

► [Molecular Motors](#)

AAV

► [Adeno-Associated Virus](#)

Abasic Site

► [Apurinic or Apyrimidinic \(AP\) Site](#)
 ► [DNA Polymerases](#)

ABCB1

► [Multidrug Resistance Gene 1 \(MDR1\) alias ABCB1](#)

ABC Transporters

► [ATP-Binding Cassette \(ABC\) Transporters](#)

Absorption Coefficient

Definition

The absorption coefficient (a.c.) is defined as the fraction of light absorbed per unit distance in an absorbing medium. The unit is cm^{-1} . The a.c. of a solution is a measure for the general ability of the dissolved molecules to absorb photons and act as potential dyes.

Absorption Edge

Definition

The absorption of X-rays by an atom varies with the wavelength. When an x-ray photon is absorbed by an atom, it enters an excited state for an extremely short period of time. When the atom de-excites it will emit an electron from near the core of the atom. This will only happen if the energy of the x-ray is higher than the energy level that is binding that electron to the core of the atom. When the x-ray has an energy equal to that of the binding energy of an electron, there is a sharp rise in the absorption, referred to as the absorption edge and this indicates that a photo-electron has been released. An X-ray photon of lower energy may be absorbed by an inner-shell electron (e.g. from the K-shell) which is elevated to a free place in a higher electron shell. If an electron returns to the K-shell immediately, a photon will be re-emitted at the same energy (strong coupling to absorption-edge energy), but the scattered photon is phase-shifted compared to a normally scattered photon, i.e. the scattering factor (atomic form factor) of the atom acquires an imaginary component. Thus, the f'' scattering coefficient becomes non-zero, and the scattering factor is no longer a scalar quantity but a complex number. The imaginary scattering component f'' is proportional to the directly measurable absorption and fluorescence. The real scattering component f' is related to f'' via the Kramers-Kronig relationship.

► [MAD Phasing](#)

Absorptive Intestinal Cells

Definition

Absorptive intestinal cells are epithelial cells lining the lumen of the gut. They exhibit an increased ► [apical surface](#), the ► [brush border](#), which is specialised in the up-take of material including sugars, lipids, immunoglobulins, ions and water.

► [Microvilli](#)

ACAMP

Apoptotic cell-associated molecular pattern.

► [Inflammatory Response](#)

Acamprosate

Definition

The anti-craving compound acamprosate acts as a weak glutamate (NMDA)-receptor antagonist and modulates the number of specific glutamate (NMDA)-receptor subunits. Acamprosate is the Calcium-salt of N acetyl-homotaurinate. It interacts with NMDA-receptor mediated glutamatergic transmission in various brain regions.

► [Addiction](#), [Molecular Biology](#)

Acantholysis

Definition

Acantholysis is the loss of adhesion between epidermal keratinocytes.

► [Desmosomes](#)

Acanthosis

Definition

Acanthosis is the thickening of the epidermis, a typical finding in histological examination resulting from enhanced keratinocyte proliferation.

► [Psoriasis](#), [Molecular Basis](#)

Acceptor

Definition

Molecule involved in Förster resonance transfer that receives energy from a different excited molecule (the donor).

► [FRET](#)

Accessible Surface Area

Definition

Accessible surface area describes the extent to which protein atoms can form contacts with water. It is defined as the area of a sphere of radius R, on each point of which the center of a solvent molecule can be placed in contact with an atom, without penetrating any other atoms of the molecule. The radius is given by the sum of the ► [van der Waals](#) radius of the atom and the chosen radius of the solvent molecule.

► [Protein-Protein Interaction](#)

► [Two Hybrid System](#)

ACE

► [Angiotensin Converting Enzyme](#)

ACE Inhibitor

Definition

► [Angiotensin Converting Enzyme](#) inhibitor is a medication that is used to treat heart failure and helps to remodel the diseased heart.

► [Cardiac Signaling: Cellular, Molecular and Clinical Aspects](#)

► [Diabetes Mellitus](#), [Genetics](#)

Acetabular Protrusion

Definition

Acetabular protrusion is the intrapelvic displacement of the medial wall of the acetabulum (the large cup-shaped

cavity in the hip in which the femoral head articulates) into the pelvic cavity, and may occur as a manifestation of severe rheumatoid arthritis or a number of other conditions.

► [Marfan Syndrome](#)

Acetal Linkage

Definition

Acetal linkage denotes a linkage formed between organic compounds derived from a hemiacetal by reaction with an alcohol. If the hemiacetal is a sugar, the acetal is a glycoside.

► [Glycosylation of Proteins](#)

Acetyl Choline Esterase

Definition

Acetyl choline esterase (AChE) is the major physiological degrading enzyme for the neurotransmitter acetyl choline.

► [Alzheimer's Disease](#)

AChE

► [Acetyl Choline Esterase](#)

Achondrogenesis

Definition

Achondrogenesis describes a lethal autosomal-recessively inherited chondrodysplasia caused by mutation(s) in the collagen Type II gene.

► [Bone Disease and Skeletal Disorders, Genetics](#)

Achondroplasia

Definition

Achondroplasia refers to the most frequent form of chondrodysplasia with short limbs, large head, dominant inheritance and mutations in the fibroblast growth factor receptor 3 (FGFR3) gene.

► [Bone Disease and Skeletal Disorders, Genetics](#)

Acousto-Optical Filters

Definition

Acousto-optical filters refer to electro-optical devices that allow the modulation of the light intensity as well as wavelength. Light penetrating the optical crystal is deflected depending on its wavelength and the wavelength of the ultrasonic field applied to the crystal.

► [Fluorescence Microscopy: Spectral Imaging vs. Filter-Based Imaging](#)

Acousto-Optical Modulator

Definition

An acousto-optical modulator is an optical device for regulating the power of a laser beam in response to an incoming dc-voltage. It includes a refractive crystal and an electronic driver. The driver generates an ultrasound wave, which is transferred onto the crystal by means of a piezo element. The sound wave propagating through the crystal creates a regular pattern of dense and less dense material, which represents a diffractive grating. The amplitude of the sound wave determines the strength of the density differences proportional to the dc signal, which is fed into the driver. A light beam passing through the crystal is diffracted at the density grating, with the intensity of the diffracted beam proportional to the amplitude of the density wave, and hence the incoming dc-signal.

► [Fluorescence Microscopy: Single Particle Tracking](#)

Acousto-Optical Tunable Filter

Definition

An acousto-optical tunable filter (AOTF) is an extended version of an acousto-optical modulator

(AOM). Like the ►AOM, the AOTF comprises of a refractive crystal and an electronic driver. The AOTF, however, simultaneously operates on several co-linear laser beams with different wavelengths in response to incoming dc-voltage signals. It is usually employed in combination with lasers run in a multi-line operation, emitting light at different wavelengths simultaneously. By means of the AOTF, one or more of the lines can be selected, and their power can be defined.

►Fluorescence Microscopy: Single Particle Tracking

Acquired Immunity

►Adaptive Immunity

Acquisition Time

Definition

Within all methods that work via the detection of a signal, the time necessary to record the signal is called acquisition time. In the context of multidimensional NMR spectroscopy, the acquisition time is the time interval during which the free induction decay (FID) is recorded. The length of the acquisition time is determined by the width of the spectral window and the number of data points that are recorded, and ranges from several milliseconds to seconds. It occurs in all NMR pulse sequences regardless of the dimensionality of the experiment.

►Multidimensional NMR Spectroscopy

Acrocentric Chromosome

Definition

Acrocentric chromosome defines a chromosome on which the centromere is located closer to one end, so that the two arms flanking the centromere are of unequal length. Human chromosomes 13, 14, 15, 21 and 22 are acrocentric.

►Centromeres

►Chromosome 21 Disorders

►RNA Polymerase I

Acrosome

Definition

Acrosome is a large, membrane-bound, lysosome-like vesicle that is located in the anterior region of the sperm head, just beneath the plasma membrane and above the nucleus.

►Acrosome Reaction

►Mammalian Fertilization

Acrosome Reaction

Definition

Acrosome reaction refers to a form of cellular exocytosis that is induced by binding of sperm to the zona pellucida of the egg during ►fertilization. It involves multiple fusions of sperm plasma membrane with the outer acrosomal membrane, resulting in exposure of the inner acrosomal membrane and acrosomal enzymes.

►Mammalian Fertilization

ACTH

►Adrenocorticotropin

Actin

Definition

Actin is a major component of the cytoskeleton and is present in all eukaryotic cells. Its most characteristic feature is the reversible polymerisation of globular actin monomers (G-actin) into filaments (►F-actin). Filaments can be associated into stem fibers that are anchored in focal contacts. Monomeric actin has a molecular mass of about 42,000 Daltons. The actin cytoskeleton is required in many cellular processes like cell migration, axonal growth, phagocytosis, cytoplasmic streaming and organelle transport.

►Actin Cytoskeleton

►Cell polarity

- ▶ Desmosomes
- ▶ Microvilli
- ▶ Focal Complexes/Focal Contacts

Actin Cytoskeleton

RAJESH ARASADA, MICHAEL SCHLEICHER
 Adolf-Butenandt-Institute / Cell Biology,
 Ludwig-Maximilians-University, Munich, Germany
 schleicher@lrz.uni-muenchen.de

Definition

The ▶actin ▶cytoskeleton is a network of actin and actin binding proteins that together with the other components of the cytoskeleton such as microtubules and intermediate filaments is responsible for essential cellular processes like cell migration, axonal growth, ▶phagocytosis, cytoplasmic streaming and organelle transport (1, 2). It was first isolated as a part of the actomyosin complex from muscle cells. Actin was subsequently identified in all eukaryotic cells. Its most characteristic feature is the reversible polymerization of globular actin monomers (G-actin) into filaments (F-actin) that build either a relatively permanent structure as in sarcomeres of fully differentiated muscle cells or, in nearly all other cell types, a continuously restructured filamentous meshwork. Given its role in many essential cellular processes, the amino acid sequence of actin is highly conserved throughout evolution. Deletion of the single actin gene in yeast is lethal. In many other organisms more than one gene codes for this protein. In humans for example there are at least six isoforms of actin that differ only slightly in their protein sequence and cell type specific expression.

Characteristics

It is very helpful for understanding actin functions in the living cell if one considers major features of the protein in the test tube. Actin has a molecular mass of about 42,000 Daltons and can be stored in its monomeric form. As soon as simple salts like KCl or $MgCl_2$ are added, the monomers form filaments and turn a sol into a gel with extremely high viscosity in a matter of minutes.

The polymerization of G-actin into F-actin occurs in three phases (Fig. 1):

• Lag Phase

The polymerization of actin requires trimeric intermediates ('actin nuclei') which are highly unstable

and therefore delay rapid elongation. It is obvious that a cellular factor with a nuclei-stabilizing activity drastically shifts the G:F-actin ratio towards a high viscosity of the cytoplasm.

• Elongation

The elongation of ▶actin filaments is very rapid and continues until it reaches an equilibrium at which nearly all of the actin molecules are part of filaments. A cellular factor which blocks this elongation by, e.g. capping filament ends, will delay elongation and thus cause higher fluidity of the cytoplasm.

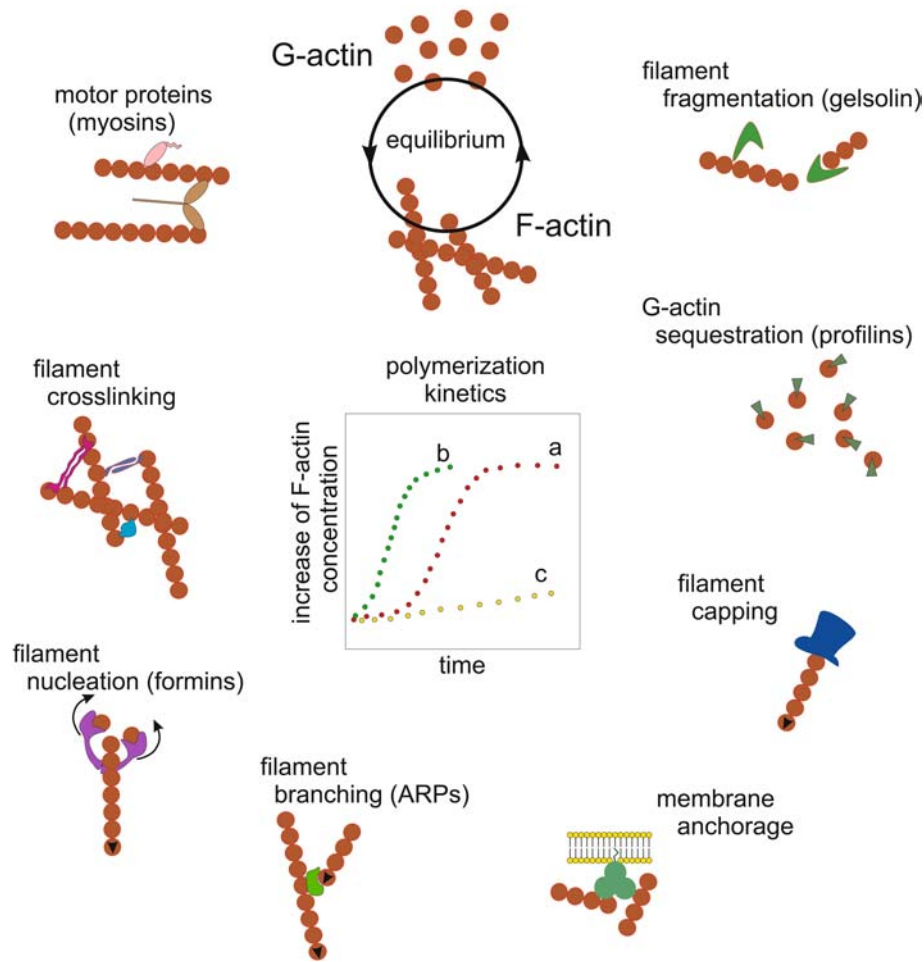
• Steady State

At the steady state a continuous exchange of actin monomers occurs at both ends of filaments but the net polymerization of actin is zero. This constant concentration of G-actin that remains in equilibrium with filament ends is termed the 'critical concentration'. Under these conditions a cell would be able to finely tune the cytosolic viscosity. If a cellular factor binds to actin monomers and thus lowers the critical concentration of free monomers at steady state, then depolymerization of filaments occurs and continues until the critical concentration is reached again. This decreases the number and length of actin filaments and leads to a less viscous cytosol.

Actin filaments are polar with the two ends being different from each other. The polarity of the filaments can be observed by decorating the filaments with myosin. Coating with myosins gives the filament an appearance of an arrowhead hence the names 'pointed' and 'barbed' ends. The growth of the filaments differs at the two ends with actin monomers being added at the barbed end about $10\times$ faster than at the pointed end. The critical concentrations also differ for the ends, with barbed ends having a critical concentration of about $0.1\ \mu M$ in contrast to the pointed ends of approximately $0.6\ \mu M$ G-actin.

The ability of actin to polymerize under conditions normally used to generate crystals made it difficult to obtain actin's atomic structure. However, crystals could be obtained for actin in complex with actin binding proteins such as DNase I, ▶profilin and ▶gelsolin, or when it was covalently modified with small molecules that inhibit ▶actin polymerization and do not change the overall conformation of actin. The actin monomer consists of two major lobes that are connected by a hinge region; they form the ATP binding cleft, and harbor sites for the binding of metal ions.

ATP-actin has a higher affinity for the barbed end than ADP-actin. However, ATP hydrolysis is not required for polymerization. The very end of an actin filament at the fast growing side consists of a small stack of ATP-actin units, followed by a stretch of hydrolyzed ADP/ P_i -actin complexes. After the release of P_i , actin

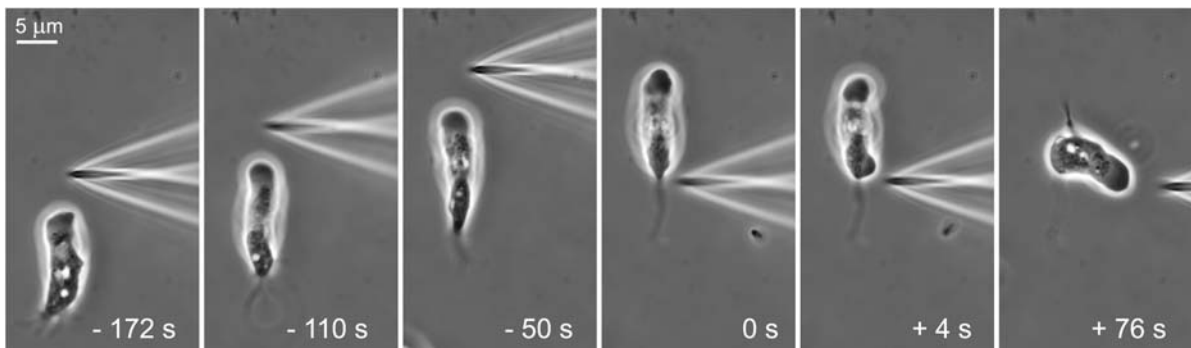


Actin Cytoskeleton. Figure 1 Kinetics of actin polymerization and function of actin-binding proteins. The graph in the center shows three typical kinetics of actin polymerization *in vitro*. Actin alone (curve a) starts to polymerize with some delay because the formation of trimeric actin seeds is highly unfavourable. After this lag-phase, filaments elongate rapidly until almost all actin monomers are part of filaments, only a typical critical concentration of G-actin is at equilibrium with filaments. A nucleating protein favors filament formation and shortens the lag time (curve b). Capping proteins inhibit the addition of monomers at filament ends and thus delay polymerization (curve c). Numerous classes of actin-binding proteins finely tune those kinetics or change the strength of existing filamentous networks. This leads to altered viscoelasticities in the cytoplasm and consequently to major changes in cell migration, cell morphology, tissue formation and other essential cellular reactions.

filaments consist almost entirely of ADP-actin, which leads to remarkably decreased stability if certain cellular proteins, e.g. actin-depolymerization factor/cofilin, attack the filament. Thus ATP serves as a timer for the depolymerization and filament turnover.

As compared to the ordered organization of actin- and myosin-filaments between the Z-disks in a sarcomere, the dynamic rearrangements of the actin cytoskeleton in a migrating cell are very complex and integrate numerous intra- and extracellular parameters (1). A superb model for chemotactic movement is the migration of *Dictyostelium* amoebae (3). Like migrating leukocytes, *Dictyostelium* amoebae are, at

a certain developmental stage, polar and show directed movement in a gradient of chemoattractant (Fig. 2). The moving front extends the membrane towards the higher concentration of chemoattractant and is characterized by an extreme polymerization of actin. The filament meshwork is so dense that even small vesicles are excluded and the leading pseudopodium appears hyaline. At the rear the cellular reactions are completely different because only contraction and depolymerization of F-actin guarantee cellular integrity and directed movement. It is amazing how quickly this pattern can be changed if the extracellular gradient is readjusted. In Fig. 2a a chemotactically active *Dictyostelium* cell



A

Actin Cytoskeleton. Figure 2 Cytoskeletal dynamics during cell migration. A *Dictyostelium* amoeba migrates towards a microcapillary which is filled with the chemoattractant cAMP. The moving front of the polarized cell is filled with a dense F-actin meshwork. At time 0 sec the capillary was relocated to the rear of the cell, and already after 4 sec the cell starts to reorganize and to create a new moving front. This quick reaction suggests extremely efficient signal transduction and cytoskeletal rearrangement.

moves towards a microcapillary that is filled with cAMP as chemoattractant. The gradient is sensed by cAMP receptors in the plasma membrane that transfer the signal *via* G-proteins to intracellular effectors. After replacing the capillary at the rear end of the cell, it takes only 2–4 s before the cell transforms the rear into a new front. This is a drastic example of the dynamics in the actin cytoskeleton, which involve not only the polymerization patterns of actin but also signal transduction and other cytoskeletal proteins.

Molecular Interactions

The actin cytoskeleton interacts constantly with various other proteins that either regulate the polymerization, stability and length of the actin filaments or perform specialized functions. The interaction of actin and myosin in the skeletal muscles during muscle contraction is well characterized. In non-muscle cells actin and myosin II interaction plays a very important role in the cleavage furrow. Studies from *Dictyostelium* cells show that myosin I is localized at the moving front of the cell whereas myosin II is at the rear end. In erythrocytes actin interacts with spectrin and ankyrin, two membrane associated proteins. Tropomyosin interacts with the actin filaments in the muscle cells providing mechanical strength to the filaments.

The presence of actin in the nucleus was under debate for a long time. Failure to find any phalloidin stained actin in the nucleus argued against its presence and actin obtained from nuclear extracts was seen as a cytoplasmic contamination. *In vivo* crosslinking using cell permeable chemical crosslinkers recently showed close association of actin with the nuclear matrix. Using ▶**leptomycin B**, a cell permeable nuclear export inhibitor, nuclear actin filaments could even be stained with phalloidin. A large number of actin binding proteins like cofilin, profilin,

protein 4.1 and c-Abl were also found in the nucleus. Protein 4.1 was immunoprecipitated along with the components of the RNA splicing machinery suggesting a role of this protein in RNA splicing. ▶**Actin related proteins** were reported to exist as subunits of larger protein complexes that are involved in chromatin remodeling. All these recent findings suggest that studies on the function and regulation of actin in the nucleus will speed up tremendously, opening a completely new field of actin's role during replication and transcription.

Regulatory Mechanisms

The average eukaryotic cell contains a large number of actin-binding proteins of different function, location and concentration; only a few can be described here. These proteins themselves are regulated, in most cases by phosphorylation, changes in free Ca^{2+} or interaction with lipids (4, 5). The physiological concentrations of Mg^{2+} or K^{+} are such that in the absence of such proteins all actin would be polymerized and the cells immobile due to their high cytoplasmic viscosity.

- Only the presence of G-actin sequestering proteins like profilin or thymosin $\beta 4$ keeps the G-actin concentration at about 50% of total actin and thus maintains the cell's ability to react quickly to external signals.
- Another important class of actin-binding proteins is the already mentioned capping proteins. They usually cover the fast growing end of an actin filament, inhibit elongation and keep cytoplasmic viscosity low. The complexity of cytoskeletal dynamics becomes obvious if one considers 'uncapping' of a locked filament upon a signal. In this case the cap is suddenly removed and the large number of now free filament ends triggers elongation at high

speed due to bypassing the lag-phase (see above). Uncapping of filaments would be a reasonable mechanism to create a new moving front in a ►[chemotaxis](#) gradient in a few seconds.

- c) A very well known example of a third class of actin-binding proteins is gelsolin, whose name describes the transformation of an actin gel into a less viscous sol. Gelsolin severs the filaments into short fragments and leads to a weakening of sub-membranous actin cortices. In higher vertebrates including man, gelsolin is also present at high concentrations in the bloodstream, most probably to remove F-actin aggregates that originate from broken cells and might be harmful in capillaries.
- d) In addition to the above mentioned actin binding proteins that regulate its polymerization, numerous other proteins like fimbrin, α -actinin, filamin and cortexillin regulate the 3-dimensional architecture of the meshwork by bundling or crosslinking the filaments.
- e) Actin related proteins (Arps) 2 and 3 were identified as part of a complex of 7 different proteins that functions as a nucleator of filament formation and initiates branching of filaments. The family of ►[WASp/SCAR](#) proteins regulates the activity of the Arp2/3 complex, probably by inducing a conformational change.

Cytoskeleton and Diseases

In view of the importance of the cytoskeleton for major cellular functions it is not surprising that many diseases can be traced back to aberrant cytoskeletal components. Only a small number can be mentioned in this context. Actin is an essential gene and mutations in actin usually lead to severe pathological alterations. The familial hypertrophic cardiomyopathy (FHC) mutation was mapped to the actin domain that might bind myosin. The mutation prevents adequate contraction leading to hypertrophy. In idiopathic dilated cardiomyopathy (IDC) the mutation was mapped to the immobilized region of actin that binds to other thin filament binding proteins like troponin I and α -tropomyosin that are involved in force transmission.

Mutations in actin binding proteins also lead to severely altered phenotypes. Profilin I minus mutants of mice are defective in cytokinesis and die at an early stage of embryogenesis. Profilin II null mice show behavioral impairments resembling a hyperdopaminergic phenotype. The mutant mice are hyperactive and show reduced anxiety. This is similar to the ADHS (attention deficient hyperactive syndrome) in humans. Changes in myosin genes are often the basis for disturbed dynamics of the actin cytoskeleton. Among others the ►[shaker/Usher syndrome](#) is of special interest because a myosin VII mutation results in aberrant cochlear hair cells and consequently deafness.

Recently, mutations have been identified in the filamin A gene that lead to otopalatodigital (OPD)-spectrum disorders. These missense mutations were mapped in the highly conserved regions of the actin-binding domain (ABD) and in the rod repeats which recruit multiple proteins. Mutations in the same regions were also reported in the ABDs of other cytoskeletal proteins like dystrophin (Duchenne and Becker muscular dystrophy), β -spectrin (hereditary spherocytosis) and α -actinin-4 (focal segmental glomerulosclerosis). Bacterial pathogens exploit the cytoskeleton either to gain entry into the cell or to propagate. *Listeria monocytogenes* recruits the actin machinery for its intracellular motility and to infect neighboring cells (6). Toxins produced by certain bacteria disorganize the cytoskeleton and change the permeability of the cell barrier leading to severe gastrointestinal diseases. Toxins produced by the members of the *Clostridium* family influence the cytoskeleton directly by ADP-ribosylation of G-actin thus inhibiting its polymerization at the barbed end, or indirectly by modifying regulatory small GTPases by ADP-ribosylation, glucosylation or deamidation (7).

References

1. Pollard TD, Borisy GG (2003) Cellular motility driven by assembly and disassembly of actin filaments. *Cell* 112:453–65
2. Engqvist-Goldstein AE, Drubin DG (2003) Actin assembly and endocytosis: from yeast to mammals. *Annu Rev Cell Dev Biol* 19:287–332
3. Noegel AA, Schleicher M (2000) The actin cytoskeleton of Dictyostelium: a story told by mutants. *J Cell Sci* 113:759–66
4. Kreis T, Vale R (1999) Guidebook to the Cytoskeletal and Motor Proteins. A Sambrook & Tooze Publication at Oxford University Press
5. dos Remedios CG, Chhabra D, Kekic M et al (2003) Actin binding proteins: regulation of cytoskeletal microfilaments. *Physiol Rev* 83:433–73
6. Carlier MF, Le Clainche C, Wiesner S et al (2003) Actin-based motility: from molecules to movement. *Bioessays* 25:336–45
7. Barbieri JT, Riese MJ, Aktories K (2002) Bacterial toxins that modify the actin cytoskeleton. *Annu Rev Cell Dev Biol* 18:315–44

Actin Filaments

Definition

Actin filaments are helical protein filaments that are formed by polymerization of globular actin (G-actin) molecules. Actin filaments (F-actin) are a major constituent of the eukaryotic cytoskeleton and part of the contractile apparatus of skeletal muscle.

Actin can exist as a monomer, but in the presence of simple salts like KCl or MgCl₂, it polymerises in a head to tail fashion forming filaments. Actin filaments are polar, with the two ends being different from each other. The polarity of the filaments can be observed by decorating the filaments with myosin. Coating with ►myosins gives the filament an appearance of an arrow head, hence the names ‘pointed’ (slow growing) and ‘barbed’ (fast growing) ends.

►Actin Cytoskeleton

►Rho, Rac, Cdc42

Actin Polymerisation

Definition

Polymerisation of actin initially requires the formation of dimeric and trimeric actin nuclei, which are highly unstable. Once formed, the elongation is rapid and continues until the equilibrium between F- and G-actin is attained. At this stage a distinct concentration remains monomeric. For ATP-actin, this ‘critical concentration’ is about 0.1 mM at the barbed end. Actin polymerisation works best with ATP-actin, but ATP hydrolysis is not required for elongation. Hydrolysis of ATP is delayed, thus the very end of a filament is composed of about 5–10 ATP-actin units. The vast majority of an actin filament consists of ADP-actin.

►Actin Cytoskeleton

Actin Related Proteins

Definition

Actin related proteins (ARPs) are ubiquitous among eukaryotes and share significant sequence and structural similarities with conventional actins. ARPs are generally larger molecules than the conventional actins, and are involved in various cellular processes like actin nucleation, transport of vesicles and control of gene expression.

►Actin Cytoskeleton

Action Potential

Definition

Action potential is the electric signal generated in a neuron by the activation of voltage-gated Na⁺ channels

in an all-or-none manner, and terminated by the activation of K⁺ channels. The rate of action potentials is the neuronal signalling code. Apart from neurons, certain types of cardiac muscle cells are able to generate action potentials.

►Heart

►Neurons

Action Potential Backpropagation

Definition

Dendrites, in particular apical dendrites pyramidal neurones in the cortex, may sustain the propagation of Na⁺ action potentials generated at the axon initial segment. These ‘backpropagated’ action potentials play an important role in the initiation of synaptic plasticity.

►Neurons

Activation (of a Molecule)

Definition

In order to participate in biochemical pathways, some molecules must gain chemical energy so that the reaction(s) in which the molecule in question is involved is thermodynamically favoured. The activation is realised by an enzyme generally through the hydrolysis of ATP, which is thus covalently bound to the molecule in the form of AMP.

►Genetic Code

Activators of G-Protein Signaling

Definition

Activators of G-protein signaling (AGS) are structurally diverse proteins, which engage G-protein-dependent signaling pathways in the absence of a classical receptor.

►G-Proteins

Active Site

Definition

Active site designates a region on the surface of an enzyme to which substrate molecules bind. In molecular motors, the active site is the binding site for ATP.

► [Molecular Motors](#)

Activin

Definition

Activin is a member of the ► [TGF- \$\beta\$](#) superfamily, which plays a critical role in early embryogenesis as a mesoderm inducer, as an inhibitor of erythropoiesis and an inducer of ► [follicle stimulating hormone](#) (FSH) in the pituitary gland.

► [Receptor Serine/Threonine Kinase](#)

Activin Receptor-Like Kinase

► [ALK](#)

Actomyosin

Definition

Actomyosin is the protein complex in muscle fibers, composed of actin and myosin filaments. ATP-driven movement of myosin filaments along actin filaments drives contractility in stress fibres and muscle.

► [Rho](#), [Rac](#), [Cdc42](#)

ActR-IB/-IC

Definition

ActR-IB/-IC describes an activin Type IB receptor also known as ALK-4, and an activin Type IC receptor also known as ALK-7.

► [Receptor Serine/Threonine Kinase](#)

ActR-IIA/-IIB

Definition

ActR-IIA/-IIB are activin Type IIA and IIB receptors.

► [Receptor Serine/Threonine Kinase](#)

Acute Intermittent Porphyria

ULRICH GROSS

Veterinary Clinics Dr. Schäfer, Bad Nauheim, Germany
ulrich.gross@arana.de

Definition

► [Porphyrias](#) are diseases of metabolism, which develop mostly from genetic disturbances of heme biosynthesis. Eight enzymes are involved in the biosynthesis of heme. With the exception of the first enzyme, δ -aminolevulinic acid synthase, every step of this cascade correlates with a porphyria. The enzymatic diminutions of the activities are not fully expressed. A complete loss of heme biosynthesis would not be consistent with life. However, in rare cases homozygous or ► [compound-heterozygous](#) genotypes may occur with nearly complete enzyme deficiencies.

Typical for porphyrias is the overproduction of heme precursors in liver and bone marrow. Thus, hepatic and ► [erythropoietic porphyrias](#) are distinguished. This distinction depends on the localization of the enzymatic deficiency. Porphyrias are divided into acute and non-acute types from the clinical aspect. The non-acute forms are often called chronic porphyrias.

Four types of acute hepatic porphyrias are known, acute intermittent porphyria, ► [hereditary coproporphyria](#), ► [variegate porphyria](#) and ► [Doss porphyria](#). Acute porphyrias show abdominal, neurological and cardiovascular symptoms. Non-acute, chronic hepatic porphyrias such as ► [porphyria cutanea tarda](#) or the erythropoietic porphyrias show skin symptoms. Porphyria cutanea tarda is the commonest porphyria in Europe, followed by acute intermittent porphyria and ► [erythropoietic protoporphyria](#).

The molecular genetics of porphyrias are heterogeneous. In acute porphyrias a large number of mutations have been described, especially in acute intermittent porphyrias. Nearly every family carries its own mutation. The genes for the different enzymes are localized on different chromosomes.

Characteristics

In acute porphyria patients the following triggering factors were noted. Drugs and alcohol were factors in 54% of all patients, followed in frequency by fasting and alcohol, ►[premenstrual manifestation](#), estrogens, anticonvulsant treatment, several infections and stress. In this context two issues in clinical porphyrinology demand investigation. These are the pathogenesis of neuropsychiatric symptoms and the wide individual variability in biochemical and clinical expression of porphyria including unexplained individual differences in porphyrinogenic response with a similar enzyme deficiency (1). Acute porphyria has been considered to be a permanent system overload of oxidative stress, with long-term effects on hepatic and renal tissue and with instances of a periodic overload of free radicals giving rise to acute neurological involvement. Oxidative stress triggered by δ -aminolevulinic acid in the brain accompanied by iron metabolism alterations and GABAergic receptor damage may be implicated in the neuropsychiatric manifestations of the acute hepatic porphyrias. The precipitating factors may be either of endogenous or exogenous origin. They can be divided into five classes (1).

1. Inducers of non-specific δ -aminolevulinic acid synthase in the liver. An overproduction of δ -aminolevulinic acid affects the partially deficient and the secondary controlling functions of porphobilinogen deaminase and causes it to become rate limiting.
2. Endocrine factors include cyclical premenstrual exacerbations. Estrogens and gestagens are capable of inducing porphyria. On the other hand, the beneficial effect of low dose, gestagen-accented hormonal oral contraceptives has been observed and such doses recommended as a therapeutic measure (5).
3. Calorie intake. Fasting, often in combination with alcohol intake is a very important ►[precipitating factor](#). This agrees with the benefit of early consideration of the “glucose effect” and nutrition management.
4. Certain drugs and chemicals foreign to the human body have the potential to induce cytochrome P450. Depletion of the regulatory heme leads to an induction of hepatic δ -aminolevulinic acid synthase and the initiation of the porphyria disease process resulting in an exacerbation of acute hepatic porphyria. A compilation of safe and unsafe drugs in acute porphyrias is given by Doss et al. (2).
5. Stress in all its variable forms can lead to a manifestation of acute hepatic porphyria. Stress may result from strong emotions, alcohol excess,

surgery, intercurrent illness or social problems (e.g. stress of relatives, examinations etc.).

It should be noted that in some porphyria patients, not just one but several factors, acting cumulatively, are the cause of the manifestation of symptomatic acute porphyria.

Cellular and Molecular Regulation

Molecular Genetics of Acute Intermittent Porphyria

Acute intermittent porphyria is an autosomal dominant disorder resulting from a partial deficiency of porphobilinogen deaminase activity. In acute intermittent porphyria a lot of different mutations have been identified. More than half of these mutations were single-base substitutions and a quarter of them consist of small insertions or deletions (3). Founder effects of single mutations from selected geographic areas have been reported (4).

The porphobilinogen deaminase gene shows that two different ►[isoenzymes](#) may be coded by a single gene. Such a single gene with two different ►[promoters](#) produces two different mRNAs, which differ in their 5' ends.

With respect to porphobilinogen deaminase activity in erythrocytes, three different groups can be distinguished. In Germany 87% of the observed gene carriers with acute intermittent porphyria show a lowered porphobilinogen deaminase activity, 8% develop an activity within an intermediate range and 5% have a normal porphobilinogen deaminase activity in erythrocytes (6).

The variant of acute intermittent porphyria in which patients have normal erythroid activity, but half-normal activity of the housekeeping isozyme which is also representative in the liver, results from three different splicing defects located in the intron 1 donor splice site. The mutations entailed in the activation of a cryptic splice site 67 bp downstream in intron 1 lead to a ►[frameshift](#) and a premature stop codon in exon 4 (9). The steady state mRNA levels of the porphobilinogen deaminase gene were analyzed in patients with acute intermittent porphyria. The levels of the mutant transcripts show large variations independent of the class of cross-reacting immunological material, the location of the mutation or the clinical phenotype.

In rare cases conventional biochemical techniques have limited ability to discover asymptomatic gene carriers, as porphobilinogen deaminase activity can be normal. Furthermore in some gene carriers, affected children or under therapy with heme arginate, the urinary excretion of heme precursors may be borderline. In these instances, direct mutation testing is the definitive method for the diagnosis of acute intermittent porphyria (7).

An association between acute intermittent porphyria and hepatocellular carcinoma was reported. However, the porphobilinogen deaminase gene does not function as a tumor suppressor gene. Intrinsic aberrations in acute intermittent porphyria, including a reduced heme synthesis and endogenous oxidative damage to DNA, may incite carcinogenic mutations in the genome of the liver cells.

Molecular Dysregulation and the Disorder-Disease Process

Acute hepatic porphyrias (Table 1) are considered to be molecular regulatory diseases in contrast to non-acute, chronic hepatic porphyria, appearing clinically as porphyria cutanea tarda, which is always associated with liver injury. Porphyrins do not accumulate in the liver in acute porphyrias, whereas in chronic hepatic porphyrias they do. Thus, porphyria cutanea tarda is a porphyrin-accumulation disease – comparable with congenital erythropoietic porphyria – whereas acute hepatic porphyrias are heme-pathway-dysregulation

diseases, which are characterized by the induction of δ -aminolevulinic acid synthase in the liver.

In the disorder-disease progression of acute intermittent porphyria, four phases may be differentiated.

1. The enzyme deficiency alone or the genetic phase,
2. a compensated latent phase with a slight to moderate increase in metabolite excretion, but without clinical symptoms,
3. followed by the decompensated latent phase with a remarkable increase in porphyrin precursor and porphyrin excretion,
4. and the phase of the acute syndrome.

A relationship has been established between the extent of heme precursor excretion, and acute symptoms and the probability of clinical precipitation. As drugs can be disease-inducing factors, acute hepatic porphyrias are classed as pharmacogenetic diseases.

The enzyme deficiency (Table 1) itself is not the factor causing illness in acute hepatic porphyrias. Hereditary enzyme deficiencies lead to a destabilization of the

Acute Intermittent Porphyrria. Table 1 Hereditary enzyme defects in the heme biosynthetic pathway as genetic causes of the porphyrias

Enzymes	Metabolites	Porphyrias
	Glycine + Succinyl-CoA	
δ -Aminolevulinic Acid Synthase	↓	
	δ -Aminolevulinic Acid	
δ -Aminolevulinic Acid Dehydratase	↓	Doss-Porphyria*
	Porphobilinogen	
Porphobilinogen Deaminase	↓	Acute Intermittent Porphyria*
	Hydroxymethylbilane	
Uroporphyrinogen III Synthase	↓	Congenital Erythropoietic Porphyria
	Uroporphyrinogen	
Uroporphyrinogen Decarboxylase	↓	Porphyria Cutanea Tarda
	Coproporphyrinogen	
Coproporphyrinogen Oxidase	↓	Hereditary Coproporphyria*
	Protoporphyrinogen	
Protoporphyrinogen Oxidase	↓	Variegate Porphyria*
	Protoporphyrin	
Ferrochelatase	↓	Erythropoietic Protoporphyria
	Heme	

*Acute hepatic porphyrias

heme control on hepatic δ -aminolevulinic acid synthase. δ -Aminolevulinic acid synthase is regulated by endogenous heme and glucose and can be repressed by exogenous glucose and heme. A decrease in regulatory liver heme, caused by drugs for example, is followed by an induction of the enzyme. This mechanism is considered to be a counter-regulatory modulation compensating for the specific enzyme deficiency. By these means, porphyrinogen synthesis should be maximized, providing sufficient hepatic heme biosynthesis. Furthermore, induction of δ -aminolevulinic acid synthase will be stimulated by high porphyrinogen substrate loss from hepatocytes. When δ -aminolevulinic acid synthase is induced, δ -aminolevulinic acid and porphobilinogen will be overproduced, but their metabolic conversion into porphyrinogens is limited by the secondary controlling sequence of porphobilinogen deaminase. This limiting function of hepatic porphobilinogen deaminase explains the simultaneous increase in porphyrin precursors (δ -aminolevulinic acid and porphobilinogen) in all three types of dominant acute porphyrias in the acute clinical syndrome. Addition of protoporphyrinogen IX and coproporphyrinogen III to human lymphoblasts or to purified human erythrocyte porphobilinogen deaminase lowered this enzyme activity. This may be an explanation for the increased δ -aminolevulinic acid and porphobilinogen during acute attacks.

Clinical Relevance

Clinical findings include gastro-intestinal, neurological and cardio-vascular symptoms. Colicky abdominal pain, associated with vomiting and constipation or diarrhea is very common and in most cases an initial sign of an acute crisis. ► **Tachycardia** is observed in about one third of all patients. Tachycardia is a very sensitive sign of the clinical activity of the porphyria disease process. Neurological symptoms, which follow the abdominal manifestation, are highly diverse. Peripheral neuropathy leads to paresis in arms and legs. In cases with a severe course, the cranial nerves are involved, leading to a bulbar paralysis and respiratory impairment. In the acute porphyria syndrome, seizures can occur, especially in patients with hyponatremia due to the Schwartz-Bartter syndrome (inappropriate antidiuretic hormone release). Early regulatory treatment of the acute symptomatic expression of acute intermittent porphyria is the treatment of choice to avoid a complicated course. There is no general rule for the appearance and disappearance of clinical symptoms. The course of the acute hepatic porphyria syndrome is highly diverse in the previously undiagnosed and known porphyria patients.

Diagnosis

The clinical manifestation of acute intermittent porphyria is due to the induction of hepatic δ -aminolevulinic acid synthase. The manifestation occurs mostly in the third decade of life. The course of the disease follows latent and acute phases. The latent phase lacks symptoms more or less. However, the excretion of heme precursors is enhanced. δ -aminolevulinic acid synthase is induced by various drugs, alcohol, fasting and stress as well as hormones. Thus the synthesis of porphyrin precursors and porphyrins becomes stimulated. The excessive amounts of these metabolites lead to an abdominal, neurological and cardiovascular syndrome. The results of clinical examination and the anamnesis lead to systematic studies of the urinary and fecal heme precursors. In the acute phase of acute intermittent porphyria, the urinary excretion of δ -aminolevulinic acid, porphobilinogen, uroporphyrin and coproporphyrin is excessively enhanced, whereas the fecal excretion of porphyrins is normal.

The secondary coproporphyrinurias, which are characterized by a mild to moderate coproporphyrinuria without uroporphyrinuria, δ -aminolevulinic aciduria, or porphobilinogenuria are the most frequent disturbances of heme biosynthesis. They arise as a result of the variable and complex effects that many disorders and conditions have on normal porphyrin biosynthesis. These secondary asymptomatic coproporphyrinurias do not develop clinical symptoms by themselves. Therefore, secondary porphyrinurias cannot explain clinical symptoms, even acute porphyria-like symptoms. Thus, porphyrinuria does not necessarily mean porphyria. Treatment of these subclinical disturbances of porphyrin metabolism is not indicated (with the exception of subclinical lead poisoning); instead, the underlying associated disease has to be diagnosed and treated. From the clinical point of view the over-interpretation of secondary coproporphyrinuria as hepatic porphyria is the most frequent reason for errors in diagnosis and therapy in this field.

Therapy

Treatment of acute porphyrias includes the omission of precipitating factors, symptomatic measures according to drug lists and regulatory treatment by glucose or heme (1) (the procedure for the therapy of the acute clinical porphyria crisis is listed below). Both glucose and heme effects on the porphyrin pathway have a proven experimental basis in studies both *in vitro* and *in vivo* (1). In all four types of acute porphyrias, the “glucose effect” by itself causes a significant decrease of between 50 and 90% in urinary δ -aminolevulinic acid, porphobilinogen and porphyrin excretion. This

biochemical response is associated with the clinical response when intravenous glucose is administered in the ascending phase of the acute porphyria manifestation.

Prerequisite

1. Confirmation of diagnosis by urinary δ -aminolevulinic acid and prothobilinogen excess
2. Elimination of all precipitating factors

Regulatory Treatment

3. High dose glucose (400 g per day per os or i.v.)
4. Heme arginate (3 mg/kg body weight per day i.v. for 4 days)

Symptomatic Treatment

5. Pain: opiates
6. Hypertension, tachycardia: beta-blockers (e.g. propranolol)
7. Nausea, vomiting: chlorpromazine or other phenothiazines
8. Psychosis: neuroleptics or other antipsychotic drugs
9. Epilepsy: diazepam, check for hyponatremia
10. Hyponatremia: saline infusion, fluid restriction if signs of inappropriate secretion of antidiuretic hormone
11. Motor neuropathy: physiotherapy

Instability of hematin was one of the possible reasons for the variable clinical experience that has dampened initial enthusiasm. Heme arginate, however, is a stable and safe heme preparation. Its clinical importance has been mentioned in an investigation (8).

Apart from glucose and heme, another therapeutic approach should be mentioned. It concerns the application of long-acting synthetic luteinizing-hormone-releasing hormone-agonistic analogues in those porphyric women who have clinical manifestations of acute porphyria in relation to their menses. Application of luteinizing-hormone-releasing hormone-agonistic analogues prevents ovulation and the resultant progesterone production to prevent cyclical attacks of porphyria.

It was shown that three females with repeated premenstrual clinical expressions of an acute hepatic porphyria (acute intermittent porphyria and hereditary coproporphyria) could be treated successfully with a hormonal oral contraceptive or other exogenous hormones to stabilize the latent, subclinical phase of the disease (5). Luteinizing-hormone-releasing hormone-agonistic analogues cause menopause. Thus, luteinizing-hormone-releasing hormone-agonistic analogues should be used to treat only those patients in whom the application of hormonal oral contraceptives does not improve the metabolic status of porphyrin

parameters and the clinical symptoms. We recommend a restriction of the application of luteinizing-hormone-releasing hormone-agonistic analogues to patients not responding to hormonal oral contraceptives. Finally, research was undertaken into certain synthetic heme analogues in which the central iron is replaced by another element. Tin-protoporphyrin, a potent inhibitor of heme degradation by inhibition of heme oxygenase, suppresses chemically induced hepatic porphyria; unlike heme derivatives, tin protoporphyrin is neither degraded nor does it induce the activity of heme oxygenase. These circumstances may have potential clinical implications for the future. However, photosensitivity is a side effect of treatment with tin protoporphyrin.

Acknowledgements

This article contains results of work supported by the German Research Association (grant GR 1363/2-3).

References

1. Anderson KE, Sassa S, Bishop DF et al (2001) Disorders of heme biosynthesis: X-linked sideroblastic anemia and the porphyrias. In: Scriver CR, Beaudet AL, Sly WS et al (eds) *The metabolic bases of inherited disease*. 8th edn. McGraw-Hill, New York
2. Doss MO, Stölzel U, Doss M (2003) Arzneistoffe bei akuten hepatischen Porphyrien und Empfehlungen zur Anästhesie. In: Rote Liste. Editio Cantor Verlag, Aulendorf/Württ., pp 493–494
3. Grandchamp B (1998) Acute intermittent porphyria. In: Schmid R (ed) *The porphyrias*. Semin Liver Dis 18:17–24
4. Greene-Davis ST, Neumann PE, Mann OE et al (1997) Detection of a R173W mutation in the prothobilinogen deaminase gene in the Nova Scotian “foreign protestant” population with acute intermittent porphyria: a founder effect. Clin Biochem 30:607–612
5. Groß U, Honcamp M, Daume E et al (1995) Hormonal oral contraceptives, urinary porphyrin excretion and porphyrias. Horm metab Res 27:379–383
6. Groß U, Honcamp M, Doss MO (1996) Heterogeneity of AIP: a subtype with normal PBG deaminase activity in erythrocytes in Germany. Eur J Clin Chem Clin Biochem 34:613–618
7. Groß U, Puy H, Doss M et al (1999) New mutations of the hydroxymethylbilane synthase gene in German patients with acute intermittent porphyria. Mol Cell Probes 14:443–447
8. Mustajoki P, Nordmann Y (1993) Early administration of heme arginate for acute porphyric attacks. Arch Intern Med 153:2004–2008
9. Puy H, Groß U, Deybach JC et al (1998) Exon 1 donor splice site mutations in the prothobilinogen deaminase gene in the non-erythroid variant form of acute intermittent porphyria. Hum Genet 103:570–575

Acute Lymphoblastic Leukaemia

Definition

Acute lymphoblastic leukaemia is a clonal malignant disorder of lymphoid progenitor or precursor cells. Blockade of lymphoid cell differentiation can occur at distinct stages of B- or T-cell development, including primitive cells with multilineage potential, and leads to persistent proliferation, defective cell death, and accumulation of leukaemic lymphoblasts in tissues, causing organomegaly and bone marrow failure.

► [Leukemia](#)

Acute Myeloid Leukaemia

Definition

Acute myeloid leukaemia is a clonal malignant disorder of haematopoietic tissue in the bone marrow that is characterized by: (1) the proliferation and accumulation of abnormal (leukaemic) blast cells, principally in the bone marrow; and (2) the suppression of normal stem cell expression resulting in impaired production of normal blood cells. Most cases of adult AML probably arise in a primitive, multipotent stem cell, whereas childhood AML might derive from somewhat more differentiated, lineage-restricted stem or progenitor cells that are active in early development.

► [Leukemia](#)

Acute Phase Proteins

Definition

Acute phase proteins are a group of proteins whose concentration in the blood plasma changes (minimally with 25%) during (certain) inflammatory disorders/reactions. Acute phase proteins are predominantly produced in the liver, and include, for example, C-reactive protein (CRP) and serum amyloid A (SAA).

► [Interleukins and Their Receptors, a Role in Regulating Hematopoietic Stem Cells](#)

Acute Phase Reaction

Definition

The acute phase response (APR) of inflammation is a well-known clinical phenomenon, characterized by an increased white blood cell count (leukocytosis), fever, alterations in the metabolism of many organs, and changes in the plasma concentrations of various so-called acute phase proteins like ► [C-reactive protein \(CRP\)](#) (in humans).

► [Inflammatory Response](#)

Acute Promyelocytic Leukaemia

Definition

Acute promyelocytic leukaemia is a form of acute myeloid leukaemia, and is characterized by an increase of enzyme-containing granula in pathological promyelocytes, which may lead to severe bleeding due to hyperfibrinolysis. Complications comprise of cerebral bleedings, therefore early therapy with antibiotics is indicated.

Acyltransferase

► [Transacylation](#)

AD

► [Alzheimer's Disease](#)

ADAM

Definition

ADAM refers to "a disintegrin and metalloproteinase", a family of membrane-anchored metalloproteases.

ADAM33 is the first gene identified by positional cloning that is involved with a predisposition to asthma bronchiale.

► [Atopy Genetics](#)

Adaptive Immunity

Definition

Adaptive immunity collectively comprises of immune responses mediated by T and/or B lymphocytes that are antigen-specific, and results in either the production of antibodies (by B lymphocytes) or in cellular activation (of T lymphocytes). Adaptive immune responses generally lead to long-lasting immunological memory, in which re-exposure to the same antigen results in a rapid recall response.

► [Autoimmune Diseases](#)
► [Inflammatory Response](#)

Adaptor Complexes

Definition

Adaptors are heterotetrameric protein complexes (AP-1 - AP-4) that serve in the sorting of proteins in several different intracellular transport steps, through direct binding to distinct di-leucine or tyrosine-based trafficking motifs present in the cytoplasmic tails of cargo proteins. Proteins with adaptor-like functions that do not possess the classical AP structure have also recently been identified, e.g. GGAs.

► [Vesicular Traffic](#)

Adaptor Protein Complexes

Definition

Four different adaptor protein complexes with homologous structural organization have been identified (AP1, AP2, AP3, AP4). They are hetero-tetramers composed of two large, one medium and one small subunit. AP1 and AP2 bind to cargo proteins present at the *trans*-Golgi network (TGN) and the plasma membrane, respectively. They also recruit clathrin, thereby promoting the formation of clathrin-coated vesicles. AP3 and AP4 have been less well characterized. Both are believed to be recruited in an ARF-dependent manner to distinct regions of the TGN. AP3 may also associate with a

subpopulation of endosomes. Although AP3 has been shown to associate with clathrin *in vitro*, it is not yet clear if it functions as a clathrin adaptor *in vivo*. AP4 is most likely part of a non-clathrin coat.

► [Protein and Membrane Transport in Eukaryotic Cells](#)

Adaptor Proteins

Definition

Adaptor proteins contain several protein-protein interaction domains that serve to assemble a multiprotein complex. They contain an SH2 domain that binds to phosphotyrosine residues of activated tyrosine kinase receptors, and an SH3 domain that binds to proline rich regions of effector proteins

► [Neurotrophic Factors](#)
► [Tight Junctions](#)
► [Tyrosine Kinases](#)

Addiction, Molecular Biology

WALTER ZIEGLGÄNSBERGER

Max Planck Institute of Psychiatry, Munich, Germany
zieglgaensberger@mpipsykl.mpg.de

Synonym

Drug abuse

Definition

Addiction is a sociological term that refers to a chronic relapsing disease that is characterized by ► [compulsive drug seeking](#) and use. Following periodic or chronic intoxication the addict loses control of drug-taking behavior despite being aware of the adverse health, social or legal consequences. Drugs of abuse change the way the brain works. Under the influence of addictive drugs such as ► [cocaine](#) and ► [amphetamines](#) (psychostimulants), ► [heroin](#) (opiate) or ► [alcohol](#) and ► [nicotine](#) the brain changes its structure and function. The various molecular targets responsible for the habit-forming action of drugs of abuse in humans and in experimental animals are similar and are presently examined by various techniques.

Characteristics

The recent advances in addiction medicine illustrate the analytical power of modern neurosciences in a field

previously accessible only to methods of systems biology. The introduction of the concepts of neuronal plasticity mediated, e.g. *via* the activation of L-glutamate-, ►GABA-, opioid- and dopamine-receptors or transport mechanisms for the re-uptake of transmitters after synaptic release has already led to important therapeutic consequences. Drug use pervades all life activities and becomes central to the life of a user, who will continuously lose social compatibility (e.g. loss of social contacts, job etc.). At present it is not possible to predict an individual's predisposition to develop an addiction.

Addictive Drugs

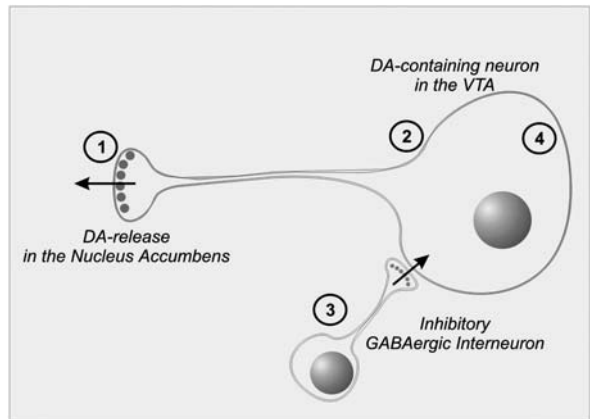
All psychoactive drugs that are abused are ►positive reinforcers and can act as a discriminative stimulus. Positive reinforcement is rewarding. Typical natural reinforcers are e.g. food, water and sex. Drugs of abuse usurp the brain's rewarding system. With most drugs of abuse only brief learning about the significance of the effects is necessary; the brain will soon start to respond differently to environmental stimuli. Formerly neutral stimuli such as a bottle or certain instruments or smells gain saliency through a learned association. These stimuli will become putative triggers for relapse and further compulsive drug taking. Drugs are good tutors; their lessons rarely get forgotten. This teaching of the brain to take drugs will finally result in a loss of control of drug-taking behavior (conditioning). It should be noted that brains cannot discriminate between illicit and legal drugs, the latter being optimally supported by advertising. Legal drugs are the more deleterious to society.

Cellular and Molecular Regulation

The Brain's ►Reward System

Internal states and memories of previous psychoactive reinforcing events are the main gateways to drug use. These processes involve learning and memory in numerous neuronal structures of the central nervous system.

One of the major targets of all drugs of abuse is a brain circuit called the reward system. The activation of this neuronal assembly reinforces behaviors. The key players are ►dopamine (DA)-containing neurons that have their cell bodies in the ventral tegmental area (VTA) and send their axons to the nucleus accumbens, the frontal cortex and parts of the striatum. When this neuronal circuit is activated dopamine is released into these structures. DA release probably induces the reward-associated feelings of pleasure in humans. There are various ways by which DA concentrations in the vicinity of neurons can be influenced (Fig. 1). Actions that increase DA release tend to be repeated. The sensitization process that follows early drug consumption is characterized by a pronounced increase in dopamine levels. There is evidence that the release of DA increases after repeated drug use and is reduced



Addiction, Molecular Biology. Figure 1 Drugs of abuse increase extracellular dopamine concentrations in midbrain structures, such as the nucleus accumbens. Cocaine (1) blocks the presynaptic dopamine transporter and the serotonin transporter and increases the extracellular dopamine by blocking reuptake. Nicotine (2) activates DA-containing neurons directly. Opioid receptor (3) agonists (μ or δ) increase the release of dopamine in the nucleus accumbens by reducing the activity of inhibitory GABAergic interneurons. Like opioid-induced dopamine release alcohol also (4) disinhibits neurons in the VTA. Transcription factors are up- or down-regulated—the excitability of the neurons is altered.

during withdrawal from the drug. It has to be noted that the responses of midbrain dopaminergic neurons represent learning signals. Rather than signaling pleasure, DA highlights significant stimuli and their motivational significance, suggesting that drug addiction might be a DA-dependent associative learning disorder. Activity-dependent gene regulation is involved in drug addiction. Many kinds of drugs of abuse can increase transcription factors in neurons of reward-related structures.

The increase of extracellular DA in the nucleus accumbens, striatum and the frontal cortex, which is of pivotal importance for the initiation and the maintenance of drug-seeking behavior, is controlled by endogenous opioids. Most principal components of the brain reward system also receive glutamatergic input from heterogeneous structures such as the medial prefrontal cortex and are influenced by local GABAergic interneuronal activity. Inhibitory GABAergic interneurons control the neuronal discharge activity of DA-containing neurons in the VTA. Drugs such as alcohol or opioids inhibit these inhibitory interneurons. A reduction in their activation leads to disinhibition of DA-containing neurons and the subsequent release of DA (Fig. 1). The reward system is linked to numerous other neuronal circuits in the brain. Drugs of abuse influence the brain as a whole.

Tolerance, Dependence and Withdrawal

The brain adapts to the continued presence of the drug of abuse. ► **Tolerance** to a drug of abuse is usually characterized by the need to take increasing amounts of the drug to elicit the same effect as this dose evoked previously. Tolerance is a more general phenomenon and does not only develop with drugs of abuse. There is evidence that in the case of some receptors, G proteins uncouple from the receptor inducing desensitization.

► **Dependence** describes the physical and/or psychological reliance on drugs. The repeated or prolonged presence of an addictive drug induces alterations in neuronal signaling cascades. After this adaptation the brain and the body will require the presence of the drug to function properly (psychological and physical dependence). Chronic administration of opioids and alcohol leads to physical dependence. This is in contrast to the effects following chronic administration of psychostimulants, which are highly addictive but hardly express physical dependence.

After cessation of drug use, the addict usually undergoes ► **withdrawal**, which is often associated with very aversive physical symptoms. For the addict it becomes now even more difficult to choose to take the drug or not, since the reduction of these aversive symptoms becomes a rewarding signal. In general, most drug-taking behavior is not cognitively perceived and controlled. Environmental situations can become conditioned to withdrawal and neutral stimuli can become motivationally significant signals that evoke ► **craving** and relapse. After chronic consumption of opiates the application of the antagonist induces dramatic withdrawal symptoms. Most importantly, after the physical withdrawal symptoms have disappeared and the drug addict is detoxified he/she is still an addict. The neuroplastic changes, which took place during the formation of the memory of addiction, are still functioning. To help an addict to finally quit drug taking and to permanently kick the habit, detoxification will be useless if it is not combined with psychotherapy and extensive long-term case management. Only the combination of these procedures will help the addict to regain the skills for successful living. It should be noted that detoxification processes are uncomfortable and no trivial interventions even under hospital supervision.

A relapse to drug taking behavior can occur after many years of abstinence. Stress can induce the reinstatement of drug-seeking behavior and self-administration.

Chronic alcohol abuse can induce chronic disorders in learning and memory (Wernicke's syndrome or so-called Korsakow's psychosis), which are characterized by confabulation as a key symptom and delirium characterized by neuron hyperactivity, tremor, insomnia and anxiety. Following chronic administration the brain gets hyperexcitable and the susceptibility to

seizures is massively increased. Neural hyperactivity can destroy neurons by overexcitation (excitotoxicity).

Craving

Craving is the uncontrollable desire for a drug of abuse. Drug-taking usually induces feelings of pleasure, alters emotional responsiveness and perception and impairs judgment and performance. Drug-induced learning and robust long-term memory consolidation in reward-related structures take place. Previously neutral stimuli acquire the ability to trigger craving. This hunger for drugs may then become an important gateway to relapse. However, there is evidence that relapse occurs without any obvious craving and that even strong craving signs are not necessarily followed by relapse. Potentially independent pathways may induce alcohol craving and relapse by negative motivational states including conditioned withdrawal and stress. These systems involve the glutamatergic systems and the corticotropin ► **CRH system**. Cues, which are associated with prior alcohol intake but are not followed by actual drug consumption, may induce conditioned withdrawal. Withdrawal symptoms are alterations in blood pressure, pulse rate and body temperature, restlessness, diarrhea, weight loss, anxiety, and depressive mood.

The ► **limbic system**, a heterogeneous nuclear system that is involved in the generation of feelings, emotions and motivations is intimately involved in learning and memory consolidation related to drug effects. Following repeated drug use these brain structures lay down and store memories in various circuits. Functional imaging techniques in humans demonstrate that craving for drugs of abuse involves areas predicted from animal experiments.

Craving as a multi-dimensional phenomenon is most readily measured through language-based descriptions following e.g. the presentation of cognitive stimuli or by inducing certain mood states. It is difficult to measure craving in laboratory animals and in each model only aspects of craving might be described. Most animal models measure the behavioral responses rather than internal states and are therefore better models for relapse rather than craving *per se*. However, some of the available animal models can serve as powerful tools for designing human craving studies.

Clinical Relevance

Drug Substitution and Anti-Craving Medication

To prevent withdrawal and craving that often provoke relapse, drug addicts are treated with either substitution or anti-craving compounds. Both substitution and anti-craving drugs prevent relapses and through this effect will help to regain control over behavior. Deconditioning is one of the strategies in the treatment of addicts. This process can either neutralize the

learned cues or can help to relearn other cues, provided no relapse occurs which might then reinstate the previous behavior. A stimulus is said to reinstate drug-seeking behavior if it causes renewed responding. Various conditions can reinstate drug taking, drug priming, i.e. the injection of a small dose of the drug, stress or conditioned stimuli. Acquisition and storage of memories is one of the basic principles of nervous systems. In the absence of reinforcement the behavioral response will gradually diminish, to be finally extinct. It is well established that the activation of the brain's ►glutamate circuitry contributes to the learning of addictive behavior. Drugs that block glutamatergic activity may help the addicts to kick their habit.

Anti-craving Substances

After several months of voluntary alcohol consumption the drug taking behavior following deprivation (withdrawal) is characterized by increased alcohol intake and preference. This behavior of animals is interpreted as craving.

In this model for craving, several months of alcohol availability are followed by a period of alcohol deprivation (i.e., a withdrawal phase). When alcohol is subsequently made available the animals increase their alcohol consumption and preference for alcohol. These animals clearly demonstrate a preference of alcohol over water and exhibit changes in their alcohol intake pattern. This alcohol deprivation effect leads to consumption of highly concentrated alcohol solutions, even at inappropriate times during the inactive light phase when drinking activity is usually low. These data show that there is a high motivation to drink alcohol following a period of deprivation. Animals will continue to work to obtain alcohol significantly longer than they would before the alcohol deprivation.

Anti-craving substances have been registered for relapse prophylaxis in weaned alcoholics in various European countries (acamprosate) and the United States (naltrexone). Acamprosate and the opiate antagonist naltrexone most probably reduce ethanol abuse through different neuronal mechanisms. ►Acamprosate, the Ca-salt of N-acetyl-homotaurinate interacts with NMDA-receptor mediated glutamatergic transmission in various brain regions. Naltrexone most probably interferes primarily with the mesolimbic/mesotelencephalic dopaminergic brain-reinforcement systems. This structure, more recently attributed to the extended amygdala, involves the shell of the nucleus accumbens, the bed nucleus of the stria terminalis and the central nucleus of the ►amygdala.

Ethanol (alcohol) can be seen as an ►NMDA receptor antagonist at concentrations reached in the brains of alcohol abusers. There is some *in vitro* evidence that the effects of ethanol may be related to selective actions

at NR2B subunit containing ionotropic glutamate receptors. We found in a recent study that the enhanced and delayed stress-induced alcohol drinking in mice lacking functional CRH1 receptors is associated with an up-regulation of the NR2B subunit (3).

The anti-craving compound acamprosate acts as a weak glutamate (NMDA)-receptor antagonist and modulates the number of specific glutamate (NMDA)-receptor subunits.

Taken together these data indicate that NMDA receptor activation is probably involved in plastic changes following chronic ethanol-induced receptor adaptation and that the inhibitory effects of acamprosate on neuronal excitability are more important for its therapeutic effects. Acamprosate produced similar increases in glutamate receptor expression *in vivo* to those seen following acute treatment with the use-dependent open-channel blocker MK-801 and memantine indicating that acamprosate may produce changes in the CNS which are similar to those seen following NMDA receptor antagonists and that these changes may, in turn, underlie the effects of both kinds of drugs in the treatment of alcohol abuse.

Moreover, it is important to note that acamprosate lacked both reinforcing properties and discriminative stimulus properties suggesting that it has little or no abuse potential in its own regard. Not all neurobiological phenotypes can be treated sufficiently for relapse prevention.

References

1. Berridge KC (2003) Pleasures of the brain. *Brain Cogn* 52:106–128
2. Littleton J, Zieglgänsberger W (2002) Pharmacological Mechanisms of Naltrexone and Acamprosate in the prevention of relapse in alcohol dependence. *Am J Addict* 12:S3–S11
3. Sillaber I, Rammes G, Zimmermann S et al (2002) Enhanced and delayed stress-induced alcohol drinking in mice lacking functional CRH1 receptors. *Nature* 206:931–933
4. Spanagel R, Zieglgänsberger W (1997) Anti-craving compounds for ethanol: new pharmacological tools to study addictive processes. *Trends Pharmacol Sci* 18:54–59

Addison's Disease

Definition

Addison's disease is the synonym for chronic adrenocortical insufficiency caused by idiopathic atrophy or destruction of both adrenal glands.

►Hypoparathyroidism

Adeno-Associated Virus

Definition

Adeno-associated virus (AAV) is a small virus that requires adenovirus for growth. This virus does not cause any known human disease and induces little immunogenicity. Recombinant AAV is used as a gene therapy vector.

► [Limb Girdle Muscular Dystrophies](#)

binds to and downregulates β -catenin, a key effector of the Wnt signalling pathway; APC binds to axin and conductin also involved in the Wnt pathway; APC interacts with microtubules and microtubule-binding proteins, regulating the chromosomal segregation; and, APC interaction with the Rac-specific factor, Asef, negatively regulates cell adhesion and favours cell migration.

► [Gut Epithelium](#)

► [Hereditary Nonpolyposis Colorectal Cancer](#)

► [Wnt/Beta-Catenin Signaling Pathway](#)

► [Colorectal Cancer](#)

Adenoma

Definition

Adenoma, in general, refers to a usually well-circumscribed benign epithelial neoplasm in which the tumor cells form gland-like structures.

► [Hereditary Non-Polyposis Colon Carcinoma](#)

► [Hypoparathyroidism](#)

Adenylyl Cyclase

Definition

Adenylyl cyclase is an enzyme that produces cAMP from ATP; it is controlled by ► [heterotrimeric G-proteins](#) in mammalian cells.

► [Diabetes Insipidus, a Water Homeostasis Disease](#)

Adenomas

Definition

Adenomas are precancerous lesions characterized by columnar epithelium with atypical nuclei, dysplasias and increased mitotic activity. There is a high risk that colorectal adenomas develop into carcinomas. The risk is correlated with adenoma size and histology.

► [Peutz-Jeghers Syndrome](#)

ADH

Definition

ADH refers to antidiuretic hormone or arginine vasopressin (AVP).

► [Diabetes Insipidus, a Water Homeostasis Disease](#)

Adenomatous Polyposis Coli

Definition

Tumor suppressor gene encoding the multifunctional adenomatous polyposis coli (APC) protein that binds Axin, conductin, β -catenin, PP2A, and microtubules, and is required for the degradation of β -catenin. Mutation of the APC gene is linked to ► [Familial adenomatous polyposis](#).

The majority of the colorectal tumors are caused by APC mutations resulting in truncated proteins. Functional domains of APC are involved in various functions: APC

ADHD

► [Attention Deficit Hyperactivity Disorder](#)

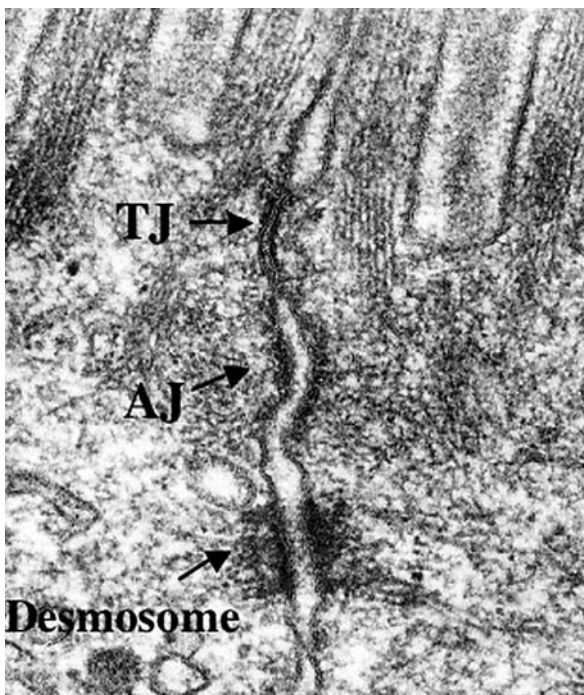
Adherens Junction

YASUYUKI FUJITA, CATHERINE HOGAN
MRC Laboratory of Molecular Cell Biology and Cell Biology Unit, and Department of Biology, University College, London, UK
y.fujita@ucl.ac.uk

Definition

Adherens junctions are large multiprotein complexes that join the actin cytoskeleton to the plasma membrane to form adhesive contacts between cells, or between cells and extracellular matrix. Adherens junctions occur in different cell types, and either form plaque-containing linear (zonula adhaerens) or punctate (punctum adhaerens) intracellular junctions. The junctions are composed of transmembrane cell-cell adhesion proteins of the cadherin family (cadherins and catenins), and of plaque-forming proteins such as plakoglobin and β -catenin, and are attached to actin microfilaments.

In vertebrates, cells adhere to their neighbouring cells through three types of junction, ►tight junctions (TJ), adherens junctions (AJ) and ►desmosomes, which together constitute the intracellular junctional complex. TJ act as a diffusion barrier that regulates epithelial or endothelial permeability and restricts the apical-basolateral diffusion of membrane components. AJ and desmosomes are adhesive junctions that anchor to actin and intermediate filaments, respectively. TJ generally localise to the apical side of the lateral membrane of cells and AJ localise just basal to TJ (Fig. 1). Desmosomes are generally localised basal to AJ. The basic structure and components of AJ are conserved



Adherens Junction. Figure 1 Electron micrograph of cell-cell contact (epithelium of small intestine). TJ: tight junction, AJ: adherens junction. (The image was kindly gifted by Dr. Sh. Tsukita, Kyoto University).

during evolution and they are found from lower organisms (e.g. *Drosophila*) to mammals. In contrast, *Drosophila* does not have TJ or desmosomes, but has septate junctions that are structurally comparable to TJ. AJ were originally identified by EM analyses as electron dense plaques of closely opposed membranes between epithelial cells, anchored to a belt of cortical actin. This structure has been well characterised in epithelial cells, however, AJ are also present in other types of cells such as endothelial cells and neuronal cells where they form synapses. AJ are fundamental to tissue development, organisation and function. AJ play a major role in the establishment of tight cell-cell adhesions. The adhesiveness of AJ is conferred by two families of single transmembrane proteins, cadherins and nectins. The extracellular domains of these proteins form homophilic interactions and the cytoplasmic domains are connected to actin filaments *via* binding proteins, thus linking the cell adhesion complex to the actin cytoskeleton.

Characteristics Components

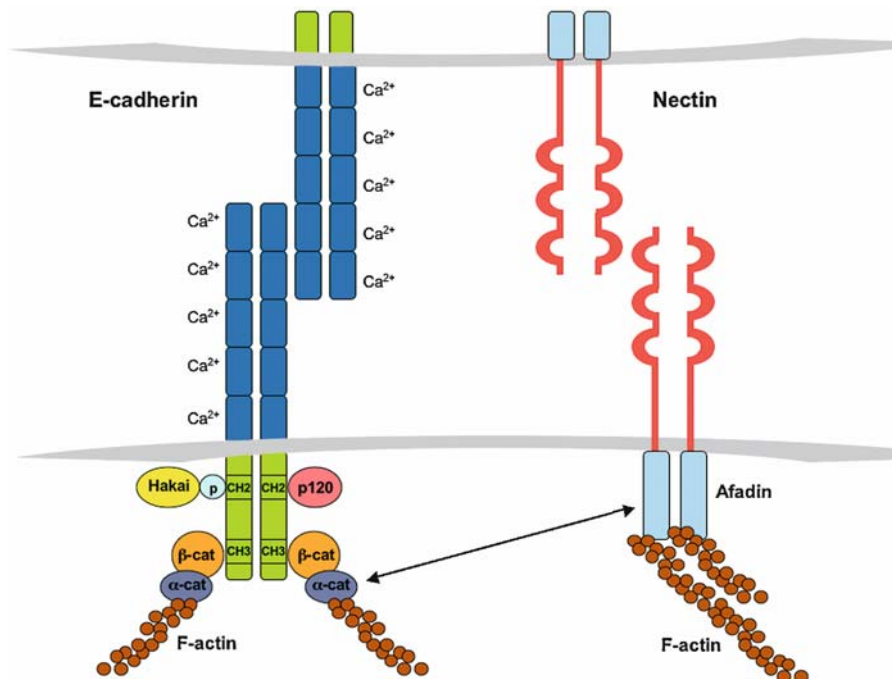
As described above, there are two families of membrane proteins that are involved in the formation of adherens junctions (Fig. 2).

Cadherin System

►Cadherins are a family of calcium-dependent cell-cell adhesion molecules that exist in both vertebrates and invertebrates. The cadherin superfamily contains >80 members and can be sub-divided into several groups, classical cadherins, desmosomal cadherins, protocadherins and atypical cadherin-like domain containing proteins. Among them, classical cadherins are found at AJ. Many classical cadherins have been identified and each of them shows a characteristic tissue distribution. For example, E-cadherin, the prototype and the best-characterised classical cadherin, is primarily expressed in epithelial cells. N-cadherin is expressed in neuronal and fibroblastic cells, while VE-cadherin is in endothelial cells.

During development, cadherin expression is highly regulated, as the pattern of expression of cadherin determines tissue and organ morphogenesis (for a review, see 5). The important role of cadherin in development was confirmed by the phenotype of E-cadherin- or N-cadherin-KO-mice that died at an early stage of embryogenesis. Cadherin expression is also important for maintaining tissue integrity since loss of cadherin adhesive activity is highly correlated with many invasive carcinomas.

Cadherin is synthesised as an inactive precursor molecule that must be proteolytically cleaved to generate a functional, mature protein. This processing



Adherens Junction. Figure 2 Molecules in adherens junctions.

involves cleavage of a propeptide fragment at the N-terminal end of an immature protein, mediated by a family of calcium-dependent endoproteases that recognise a specific recognition sequence. Expression studies have shown that the precursor forms of cadherin do not have adhesive activity.

The extracellular domain of classical cadherins is composed of five structurally homologous domains, EC1-EC5. Each of the five extracellular domains contains Ca^{2+} binding motifs. The current model of cadherin-mediated cell-cell adhesion has been derived from crystallographic analyses of the EC1 and EC2 domains of E-cadherin and N-cadherin and of the entire extracellular domain of *Xenopus* C-cadherin. In this model, cadherin molecules form a *cis* dimer between monomers aligned in a parallel orientation. Binding of calcium ions induces *trans* dimerization between the amino terminal domains of *cis*-dimers in an anti-parallel orientation, which eventually leads to a 'zipper'-like structure between cadherin-expressing cells. Single amino acid mutations that affect the formation of *cis*-dimers abolish cadherin adhesive activity, indicating that the formation of the *cis*-dimer is essential for the formation of a functional *trans*-dimer. The extracellular domain of cadherins only forms homophilic ligations, i.e. a cell expressing a certain cadherin will only interact with another cell expressing the same type of cadherin. This idea was supported by the following experiment; when cells

expressing E-cadherin or N-cadherin were mixed, E-cadherin-expressing cells were segregated from N-cadherin-expressing cells, leading to the formation of two separate cell aggregates.

The cytoplasmic domain of cadherin includes two domains, the CH2 domain (located in the proximal region) and the CH3 domain (located in the distal region). CH denotes cadherin homology and these domains are found in most of the classical cadherins. The CH2 and CH3 domains of cadherins interact with p120^{ctn} and β -catenin, respectively. The juxtamembrane domain of E-cadherin interacts with Hakai. (For details on cadherin-binding proteins, see below.) Although the extracellular domain of cadherin induces cell-cell adhesion in the presence of Ca^{2+} , interaction between the cytoplasmic domain of cadherin and the underlying actin cytoskeleton is also required for construction of tight and compact cell-cell adhesions. β -catenin binds to α -catenin, which in turn associates with actin filaments. The anchoring of cadherin-catenin complex to the actin cytoskeleton promotes the clustering and stabilisation of AJ proteins to form the initial cluster or punctum. Additional adjacent complexes assemble to form the 'zipper'-like structure, which seals the membrane between neighbouring cells. In polarised epithelial cells, this assembly of a 'zipper'-like structure occurs at the apical side of the lateral membrane. However, it remains unclear how this process of AJ formation is regulated.

Nectin System

Nectin is a Ca^{2+} -independent immunoglobulin-like cell-cell adhesion molecule that comprises a family of at least four members, nectin -1, -2, -3 and -4, all of which, except nectin-4, have two or three splice variants (for a review, see 4). Nectin-1 and -2 were originally identified as poliovirus-receptor-related proteins and were named PPR-1 and -2, respectively. However, neither PPR-1 nor PPR-2 has been shown to act as a poliovirus receptor itself. Nectin is conserved from frog to human. Nectin-1, -2 and -3 are ubiquitously expressed in a variety of cells including epithelial cells, fibroblasts and neurons, while nectin-2 and -3 are also expressed in cells that lack cadherins such as monocytes and B cells.

Nectin is a single transmembrane protein of which the extracellular region comprises three immunoglobulin-like domains. Each nectin family member forms a homo-*cis*-dimer in a Ca^{2+} -independent manner, before forming an intercellular homo-*trans*-dimer. The formation of *cis*-dimer is essential for that of a *trans*-dimer. In addition, nectin-3 can form hetero-*trans*-dimers with nectin-1 and -2, while nectin-4 can also form hetero-*trans*-dimers with nectin-1. Nectin is associated with the actin cytoskeleton through binding at the cytoplasmic tail to the F-actin binding protein afadin. Afadin has two splice variants, l-afadin and s-afadin. S-afadin, the smaller splice variant, does not contain the F-actin binding domain. L-afadin is ubiquitously expressed, whereas s-afadin is mainly expressed in neural tissue. Human s-afadin is the protein product of *AF-6*, a gene associated with acute myeloid leukemias. Mice deficient in afadin show severe developmental defects at stages of gastrulation, particularly in structures derived from the ectoderm and the mesoderm. In the ectoderm and embryoid bodies of afadin-deficient mice, the organisation of AJs is highly impaired, indicating the important role of afadin in the formation of AJ.

Several lines of evidence show that the nectin-afadin and E-cadherin-catenin systems are physically and functionally associated. Both systems function cooperatively to induce the formation of AJ, where the nectin-afadin system induces the recruitment of E-cadherin to AJ. However, the molecular mechanism by which nectin-afadin and E-cadherin-catenin systems associate has not been fully elucidated.

Cadherin-mediated Signalling Pathways

Several data suggest that cadherin functions not only as “glue” but also as a signalling centre to transmit several signalling pathways. For example, inhibition of cadherin-based cell-cell adhesions, by a dominant-negative cadherin protein or a blocking antibody,

induces up-regulation of cell proliferation and apoptosis, and modifies cell differentiation. Other experiments have shown that E-cadherin^{-/-} ES cells have no capability to form organized structures. Interestingly, this phenotype can be rescued by constitutive expression of E-cadherin, which results in the formation of epithelial structures. In contrast, rescue by constitutive expression of N-cadherin leads to formation of neuroepithelium and cartilage. These data indicate that cadherins can mediate common and/or specific signalling pathways. Downstream signals from cadherins is still not very clear, but recent reports have revealed several molecular mechanisms (for a review, see 6).

Signalling Through ► Rho GTPases: Rac, Cdc42 and Rho

There is accumulating evidence showing both inside-out and outside-in signals between cadherin and Rho GTPases. Rac and Cdc42 positively regulate cadherin adhesive activity, while Rho negatively regulates cadherins. In addition, cadherin-mediated cell-cell contacts activate Rac and Cdc42, depending on the experimental systems. The modulation of Rho GTPases may regulate the actin cytoskeleton, closely connected to cadherin-based adhesions, during establishment and disruption of cell-cell contacts.

Signalling Through Cadherin-binding Proteins

There are several cadherin-binding proteins identified. Among them, β -catenin and p120^{ctn} (ctn: catenin) have a signalling role in the nucleus. Interaction with cadherin may regulate their localization and nuclear function (For details, see below).

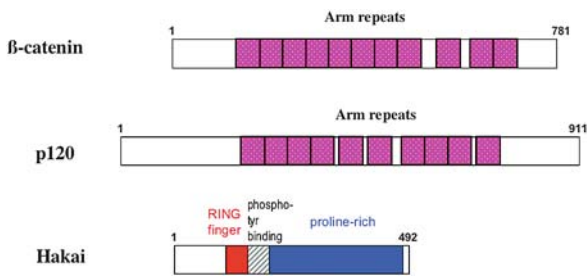
Signalling Role of the Cleaved Cytoplasmic Domain of Cadherin

The cytoplasmic domain of cadherin is cleaved by several ►proteases, such as presenilin, caspase and calpain. A recent report showed that the cytoplasmic domain of N-cadherin cleaved by presenilin has a signalling role (3). It binds to the transcriptional factor CBP and promotes its proteasomal degradation, inhibiting CRE-dependent transactivation. The functional role of the cleaved form of other cadherins remains unexplored.

Molecular Interactions

Cadherin Complex

As described above, the cytoplasmic domain of cadherin interacts with β -catenin, p120^{ctn} and Hakai. The domain structure of each cadherin-binding protein is shown in Fig. 3.



Adherens Junction. Figure 3 Domain structures of cadherin-binding proteins.

β-catenin

In the central region, **β-catenin** has eleven arm-repeats, characteristic 40–44 amino acids repetitive domains, which are responsible for the interaction with cadherins. It is still not very clear how cadherin-β-catenin interaction is regulated. However, crystal structure analyses suggest that the phosphorylation of serine residues in the CH3 domain of cadherin enhances the binding affinity for β-catenin. The N-terminal region of β-catenin interacts with α-catenin that binds to the actin filament. Through this interaction β-catenin mediates the interconnection between cadherin and the actin cytoskeleton, and this connection gives cadherin-based cell-cell adhesions strength.

In addition, β-catenin plays a central role in Wnt/Wingless signalling pathways (**Wnt/Beta-Catenin Signaling Pathway**). β-Catenin in the cytosol is constitutively phosphorylated by GSK-3β and the phosphorylated β-catenin is ubiquitinated and degraded in **proteasomes**. Upon activation of Wnt/Wingless pathways, GSK-3β-induced phosphorylation is blocked and the amount of β-catenin increases. The stabilized β-catenin interacts with a transcriptional factor Lef/Tcf and is transported into the nucleus. In the nucleus, the β-catenin-Lef/Tcf complex regulates the expression of several proteins affecting cell proliferation and differentiation. As the binding of β-catenin to cadherin and to Lef/Tcf is mutually exclusive, cadherin could regulate Wnt signalling by sequestering β-catenin. Indeed, the inhibitory role of cadherins in Wnt signalling pathways has been suggested in several experimental systems.

p120^{ctn}

p120^{ctn} also has ten arm repeats, which are responsible for the interaction with cadherins. p120^{ctn} has two functional roles in the establishment and maintenance of cadherin-based cell-cell adhesions. Firstly, p120^{ctn} regulates the cis-dimer formation of cadherins. As described above, cadherin forms not only trans-dimers (*via* the extracellular domain), but also cis-dimers on the cell surface. p120^{ctn} activates or inhibits the cis-dimerization depending on experimental conditions.

Secondly, p120^{ctn} inhibits endocytosis of cadherin, thus stabilizing cadherin-based cell-cell adhesions, though the molecular mechanism of this inhibitory role still remains to be resolved.

Several data suggest the involvement of p120^{ctn} in the regulation of cytoskeleton. p120^{ctn} can regulate the activity of Rho GTPase family members. Over-expression of p120^{ctn} in mammalian cells inhibits Rho activity and stimulates Cdc42 and Rac activity, leading to production of long neurite-like protrusions. In addition, p120^{ctn} interacts with kinesin and through this interaction p120^{ctn} is involved in microtubule-based vesicle transport.

Since p120^{ctn} and β-catenin share a similar domain structure (arm-repeats), p120^{ctn} has been postulated to have a functional role in the nucleus similar to β-catenin. Indeed, nuclear localization of p120^{ctn} has been reported under several experimental conditions. However, p120^{ctn} does not play a role in the Wnt signalling pathway and the functional role of p120^{ctn} in the nucleus is still largely unknown.

Hakai

Hakai has RING-finger, SH2 and polyproline domains (Fig. 3). *Via* the **SH2 domain**, it interacts with the juxtamembrane domain of E-cadherin in a tyrosine phosphorylation-dependent manner (2). *Via* the **RING finger domain**, it functions as an **E3 ubiquitin-ligase** for the E-cadherin complex and is involved in destruction of E-cadherin-based cell-cell contacts (For more details, see below). Hakai and p120^{ctn} compete for interaction with E-cadherin. Hakai interacts with E-cadherin, but not with other classical cadherins such as N- or VE-cadherin.

Nectin Complex

The C-terminus of the cytoplasmic domain of nectin contains a consensus amino acid sequence (Glu/Ala-X-Tyr-Val) that binds **PDZ domains**. The PDZ domain of afadin interacts with this domain. Afadin has an actin-binding site at the C-terminus, thus connecting nectin complexes to the actin cytoskeleton. Afadin also interacts with the small GTPase Rap and a deubiquitinating enzyme Fam, however the functional significance of these interactions remains unclear.

Regulatory Mechanisms

Cadherin-based cell-cell adhesion is not always static, but is quite dynamically modulated during several processes. During epithelial-mesenchymal transition (EMT) in embryonic development (e.g. gastrulation), E-cadherin protein expressed on the plasma membrane is endocytosed, and furthermore, the transcription of the *E-cadherin* gene is suppressed. In malignant cancer cells, the expression of cadherins is frequently

down-regulated (for a review, see 1). Thus, the expression level or subcellular distribution of cadherin is tightly regulated at transcriptional and posttranslational levels. Among the cadherin family members, the regulation of E-cadherin has been most intensively studied.

Transcriptional Regulation of E-Cadherin

Expression of the *E-cadherin* gene is regulated by several elements located in the 5' proximal sequences of its promoter. Among them, an E-pal element containing two adjacent E-boxes plays a vital role in the regulation of E-cadherin expression. Several proteins, such as Snail, Slug, E47 and SIP-1, have been identified as transcriptional repressors of E-cadherin acting through interaction with the E-pal element. Over-expression of these repressors in epithelial cells leads to a loss of E-cadherin expression and triggers transformation with fibroblastic and invasive phenotypes. Indeed, up-regulation of these molecules has been found in several malignant cells. Interestingly, some of these E-cadherin repressors were previously characterized as crucial regulators during the epithelial-mesenchymal transition of embryonic development. Taken together, these *E-cadherin* repressors are involved in down-regulation of E-cadherin during tumour progression and embryonic development.

Post-translational Regulation of E-Cadherin

The half-life of E-cadherin in epithelial cells is 8–12 h, and the stability or subcellular localization of the E-cadherin protein is regulated by several mechanisms.

Endocytosis

E-cadherin is endocytosed during several processes such as epithelial-mesenchymal transition and cell division. The molecular mechanism for E-cadherin endocytosis is still unclear, but the activation of tyrosine kinases including receptor tyrosine kinases (EGFR, HGFR, etc.) and src induces endocytosis of E-cadherin, leading to disruption of cell-cell adhesions. The RING-finger type E3-►ubiquitin ligase, Hakai, is involved in this process (2). Activation of ►tyrosine kinases induces tyrosine phosphorylation of the cytoplasmic domain of E-cadherin with which Hakai interacts *via* its SH2 domain. Upon interaction, Hakai mediates ubiquitination of the E-cadherin complex, leading to endocytosis and lysosomal degradation of the E-cadherin complex.

In addition to tyrosine kinase activation, there are several stimuli that cause endocytosis of E-cadherin, such as low calcium treatment and activation of cPKC.

It remains to be clarified whether the Hakai-ubiquitin system is also involved in these processes.

Proteolysis by Several Proteases

Cadherins can be cleaved by several proteases. The extracellular domain is cleaved by ►metalloproteases, resulting in secretion of the cleaved extracellular domain into an extracellular space. The secreted extracellular domain would act as a dominant negative protein to block the homophilic interaction of intact cadherins, resulting in disruption of cell-cell contacts. The cytoplasmic domain is also cleaved by presenilin (γ -secretase), calpain and ►caspase. The physiological significance of these cleavages is still not very clear. However, a recent report suggests that the presenilin-mediated cleavage liberates the cytoplasmic domain of cadherin from the membrane fraction. This free cytoplasmic domain could in turn play a role in signal transduction pathways (see above).

►Breast Cancer

►Rho, Rac, Cdc42

►Tight Junctions

►Wnt/Beta-Catenin Signaling Pathway

References

1. Birchmeier W, Behrens J (1994) Cadherin expression in carcinomas: role in the formation of cell junctions and the prevention of invasiveness. *Biochim Biophys Acta* 1198:11–26
2. Fujita Y, Krause G, Scheffner M et al (2002) Hakai, a c-Cbl-like protein, ubiquitinates and induces endocytosis of the E-cadherin complex. *Nat Cell Biol* 4:222–31
3. Marambaud P, Wen PH, Dutt A et al (2003) A CBP binding transcriptional repressor produced by the PS1/epsilon-cleavage of N-cadherin is inhibited by PS1 FAD mutations. *Cell* 114:635–45
4. Takai Y, Nakanishi H (2003) Nectin and afadin: novel organizers of intercellular junctions. *J Cell Sci* 116:17–27
5. Takeichi M (1995) Morphogenetic roles of classic cadherins. *Curr Opin Cell Biol* 7:619–27
6. Wheelock MJ, Johnson KR (2003) Cadherin-mediated cellular signaling. *Curr Opin Cell Biol* 15:509–14

Adhesion Molecules

KERSTIN DANKER

Institute for Biochemistry und Molecular Biology,
Campus Benjamin Franklin, Charité-University
Medicine Berlin, Berlin, Germany
kerstin.danker@charite.de

Definition

Cell adhesion is critical for the genesis and maintenance of both three-dimensional structure and normal function in tissues. The biochemical entities mediating cell adhesion are multiprotein complexes comprising three broad classes of macromolecules, the adhesion receptors, the ►extracellular matrix molecules and the adhesion plaque proteins.

Cell adhesion receptors are transmembrane cell surface proteins that mediate cellular interactions with neighboring cells or with the extracellular matrix. These interactions, called cell-cell or cell-matrix adhesion, profoundly influence a variety of signaling events including those involved in mitogenesis, survival and differentiation.

Signal transduction cascades induced by adhesion receptors are regulated by the cytoplasmic sequences of the receptors and by cis-interactions with other transmembrane receptors, thereby forming organized structures or scaffolds that permit the efficient flow of information in signaling pathways.

Characteristics

Adhesion Receptor Families

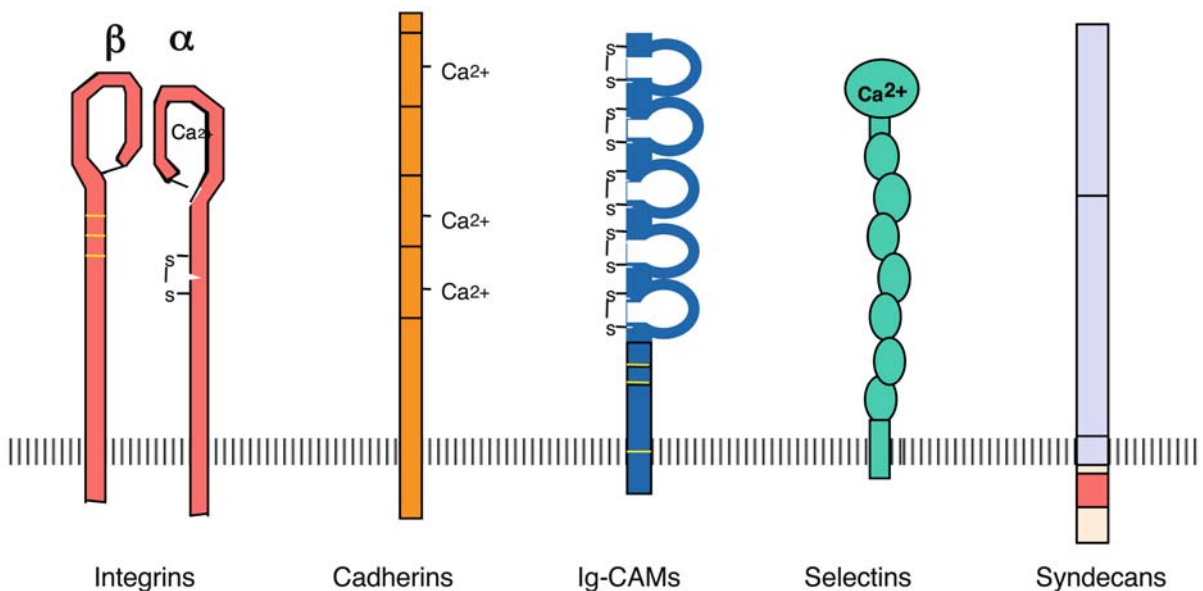
Receptors that mediate adhesion can be divided into at least five different families (Fig. 1) whose key characteristics are described in the following.

Integrins

►Integrins are heterodimeric transmembrane receptors consisting of one α and one β subunit each subunit having a large extracellular domain, a single

membrane-spanning region and, in most cases, a short cytoplasmic sequence (Fig. 1). The integrin receptor family of vertebrates includes at least 18 distinct α subunits and 8 β subunits, which can associate to form more than 20 distinct integrins. The pairing of the two subunits determines ligand specificity. Integrin ligands are extracellular matrix proteins, counter-receptors on other cells, soluble plasma proteins or microorganisms (1). The strength of ligand binding is modulated by divalent cations, by receptor clustering and by the association of integrins with accessory molecules (2). The name integrin was originally coined to denote the function of these proteins in linking the extracellular matrix (ECM) with the cytoskeleton. However, these receptors can regulate many other aspects of cell behavior. Signaling enzymes, adaptor and cytoskeletal proteins regulated by integrin engagement control physiological processes through interactions with the cytoskeleton and with other receptors. The cytoplasmic domains of the integrin subunits are key nexuses of interactions between the extracellular environment and intracellular structures and signaling cascades. Upon activation by ligand binding, integrins cluster and recruit cytoplasmic proteins, forming structures called focal adhesions. ►Focal adhesions can be considered as integrin-specific signal transduction entities, which serve as starting points for F-actin stress fibers and signaling cascades.

Both the α and the β subunit cytoplasmic sequence make important contributions to various aspects of overall integrin function, including cytoskeletal reorganization, cell motility, signal transduction and modulation of



Adhesion Molecules. Figure 1 Schematic presentation of the structures of the five families of adhesion receptors (modified from Hynes and Landers, 1992).

integrin affinity for the ligand. A large number of cytoskeletal, adaptor and signaling proteins can therefore interact with integrin cytoplasmic domains (2).

Cadherins

Cadherins are a superfamily of adhesion molecules that mediate Ca^{2+} -dependent homotypic cell-cell adhesion. This superfamily contains: 1. classical cadherins, which form the major class of cell-cell adhesive junctions; 2. desmosomal cadherins (desmocollins and desmogleins); 3. protocadherins; 4. some other cadherin-related molecules. Members of the cadherin family consist of a large extracellular domain, a single membrane-spanning region and a cytoplasmic sequence. The extracellular domain consists of repeats of an approximately 100 amino acid cadherin-specific module. Members of the classic cadherin subfamily, such as N-, P-, R-, B- and E-cadherins contain five such modules. The conformation of the cadherin molecule is stable only in the presence of Ca^{2+} , whose binding with the extracellular portion of the polypeptide chain is prerequisite for cadherin-mediated cell-cell adhesion. The N-terminal module is essential for homophilic interaction between two cadherin molecules. Classic cadherins localize in specialized sites of the membrane, such as the **▶zonula adherens**, where they can establish linkages with the F-actin cytoskeleton. Cadherin binding to the cytoskeleton is mediated by a group of intracellular proteins known as **▶catenins** (3). The β -catenin is directly attached to the cytoplasmic domain of cadherin *via* its central region, which contains so-called armadillo repeats (4). Other cytoplasmic proteins directly associated with cadherins are tyrosine phosphatases and the protein p120^{cas}.

In many studies it has been shown that, in addition to cell adhesion, cadherins also have roles in signal transduction processes. These include changes in cytoplasmic Ca^{2+} and activation of **▶G-proteins** and tyrosine kinases, as well as ligand-independent activation of the EGF receptor. Cadherins are connected to the Wnt/wingless signal pathway *via* β -catenin (**▶Wnt/Beta-Catenin Signaling Pathway**).

Many studies have correlated changes in cadherin expression or function with the onset of morphoregulatory processes such as the establishment of tissue boundaries, metastasis, tissue rearrangement, cell migration and cell differentiation. Misregulated cadherin expression or function disrupts embryonic morphogenesis and can alter the characteristics of differentiated cells. The capacity of tumor cells for uncontrolled growth, migration, invasion into surrounding tissues and metastasis is often associated with disruption of cadherin-dependent cell-cell junctions. E-cadherin expression is usually decreased in undifferentiated "aggressive" carcinomas with a high invasive potential.

Immunoglobulin Superfamily Cell Adhesion Molecules (IgCAMs)

Cell adhesion molecules of the **▶immunoglobulin superfamily** (IgCAMs) comprise a large, heterogeneous group of adhesive receptors mediating Ca^{2+} -independent homotypic and heterotypic cell-cell adhesion. Members of this family are defined by the presence of at least one **▶Ig domain**. IgCAMs possess a large amino terminal extracellular part containing the Ig domains, a single transmembrane helical segment and a cytoplasmic tail (5). Adhesion receptors of the Ig superfamily have important roles in development, the maintenance of adult tissue, including neurite fasciculation and neuronal plasticity, and the regulation of the immune response. The best-known example of Ig adhesion receptors is the neuronal cell adhesion molecule (NCAM), which contains five Ig domains in its extracellular portion and functions as a homotypic, calcium-independent cell-cell adhesion receptor. The members of the intercellular cell adhesion molecule subfamily (ICAM1-3) bind heterotypically to integrins of the $\beta 2$ -subfamily, thereby regulating the binding of lymphocytes to endothelial cells or antigen presenting cells.

Relatively little is known about the interactions of IgCAMs with cytoplasmic proteins. The receptor L1 is coupled to the **▶cytoskeleton** *via* binding to ankyrin. Members of the intercellular cell adhesion molecule (ICAM) subfamily interact **▶PIP2**-dependently with the protein **▶ezrin**, mediating a linkage to the cytoskeleton. Many IgCAMs contain **▶ITIM** motifs that serve as docking sites for **▶SH2 domains** of certain phosphatases. PECAM-1, a homotypic IgCAM receptor present on endothelial cells, platelets and some types of leukocytes and involved in the extravasation of leukocytes, interacts with SHP-1 and SHP-2 tyrosine phosphatases and the inositol phosphatase SHIP.

Like other classes of cell adhesion receptors, IgCAMs are able to induce signaling cascades. For NCAM a cooperation/interaction with the FGF receptor has been suggested, in which the phospholipase C is activated, leading to the production of diacylglycerol and eicosanoids, activation of calcium channels and an increase in intracellular calcium. One isoform of NCAM associates with the cytoplasmic tyrosine kinases pp59^{fyn} and pp125^{FAK}. Furthermore, NCAM fulfils its role in neurite outgrowth by signaling *via* the Erk/MAP kinase pathway. Cooperative interactions with other adhesion receptors, the integrin $\alpha\beta 3$ and cadherins, have been described for PECAM-1.

Another interesting subfamily of IgCAMs, the CEA (carcinoembryonic antigen) subfamily, is altered during epithelial tumor progression. The prototypic carcinoembryonic antigen, a glycosylated protein, which contains seven Ig domains and a GPI anchor, is

expressed in a controlled manner during development. In many tumors its expression is abnormally increased.

Syndecans

The syndecans are a four-member family (syndecan-1/syndecan, syndecan-2/fibroglycan, syndecan-3/N-syndecan, syndecan-4/ryudocan (amphyglycan)), of cell surface **▶heparan sulfate proteoglycans** (HSPGs) expressed on all adherent cells.

By way of their HS chains, syndecans can bind a wide variety of soluble and insoluble extracellular ligands including morphogens, growth factors and chemokines as well as ECM components and cell adhesion molecules. Syndecans have diverse functions ranging from participation in cell-cell adhesion, regulation of the signaling of heparan sulfate binding growth factors and organization of cell-matrix adhesion and signaling. The syndecan family members are type I integral membrane proteins with homologous transmembrane and cytoplasmic domains. The combined transmembrane/cytoplasmic domains contain four well-conserved tyrosine residues, which might have important roles in biological function.

The syndecans may function with several types of receptors. They are expressed at cell-cell adhesion sites, e.g. syndecan-1 on epithelial cells and syndecan-2 in neuronal synapses (6). Here, they are expressed with the PDZ protein CASK and the cytoskeletal protein 4.1 and β -catenin linked to cadherins. All three of these cytoplasmic proteins have nuclear functions and CASK binding to syndecans has been shown recently to alter its nuclear targeting. This suggests that coregulation of cadherins and syndecans may be important in the nucleus.

Selectins

Selectins are a small family of lectin-like adhesion receptors containing three members, L-, E- and P-selectins, expressed on leukocytes, endothelial cells and platelets (7). These adhesion receptors mediate heterotypic leukocyte adherence to endothelial cells and platelets during normal immune system cellular trafficking or during inflammation. For these adhesion processes, selectins are required for the very first cell-cell contact and the leucocyte rolling on the endothelium. Subsequently, selectins are involved in the activation of β 2-integrins that are responsible for firm adhesion of leukocytes to the vessel wall.

Selectins contain an amino terminal domain homologous to that of calcium-dependent animal lectins, followed by an epidermal growth factor (EGF)-type domain, two to nine complement regulatory protein repeats, a transmembrane helical segment and a short cytoplasmic sequence. The conserved lectin domain

mediates Ca^{2+} -dependent cell-cell adhesion by binding to sialylated glycans. Physiological selectin ligands include sialyl-Lewis^x saccharides in the context of macromolecular scaffolds. For example, PLGL-1, a ligand of P-selectin, is a mucin-like transmembrane glycoprotein expressed on leukocytes and lymphoid cells.

Selectin expression and function are tightly regulated in time, so that leukocytes can stick to the vessel wall only when it is necessary. Mediators and cytokines are also involved in this regulation of expression.

Recent studies indicate that selectins can also function as signaling receptors, but compared with the other adhesion receptor families relatively little is known about signaling by selectins.

▶Cell Polarity

References

1. Hynes RO (1992) Integrins: versatility, modulation, and signaling in cell adhesion. *Cell* 69:11–25
2. Liu S, Calderwood DA, Ginsberg MH (2000) Integrin cytoplasmic domain-binding proteins *J Cell Sci* 113:3563–3571
3. Kemler R (1993) From cadherins to catenins: cytoplasmic protein interactions and regulation of cell adhesion. *Trends Genet* 9:317–321
4. Hülsmen J, Behrens J, Birchmeier W (1994) Tumor-suppressor gene products in cell contacts: the cadherin-APC-armadillo connection. *Curr Opin Cell Biol* 6:711–716
5. Aplin AE, Howe A, Juliano RL (1999) Cell adhesion molecules, signal transduction and growth. *Curr Opin Cell Biol* 11:737–744
6. Rapraeger AC (2000) Syndecan-regulated receptor signaling. *J Cell Biol* 149:995–997
7. Lasky LA (1995) Selectin-carbohydrate interactions and the initiation of the inflammatory response. *Annu Rev Biochem* 64:113–139

Adipogenesis

Definition

Adipogenesis describes the process of formation of fat tissue (adipose tissue). It follows a multistep process, beginning with the clonal expansion of mesenchymal cells. These cells are capable of forming muscle, bone or fat cells (adipocytes). Some of these mesenchymal cells differentiate into preadipocytes and proliferate at the site of adipogenesis. In a second differentiation step, preadipocytes begin to fill with lipid. Initially, the lipid accumulates inside the cell in small droplets and later on the droplets fuse into one large droplet. By this

lipid accumulation, mesenchymal cells with an average diameter of 10–20 μm can reach diameters of up to 100–200 μm , thereby increasing the cell volume as much as thousand times. Molecular regulation of adipogenesis occurs via a complex transcription factor cascade. Transcription factors such as C/EBP α , C/EBP β , C/EBP δ , PPAR α/δ , and ADD1/SREBP1c have been shown, under various circumstances, to influence the extent of adipogenesis. Among the earliest events following hormonal stimulation is the induction of Krox20, which in turn, either directly or indirectly through a homeobox factor, activates the expression of C/EBP β and possibly C/EBP δ . C/EBP β and C/EBP δ then activate the expression of KLF (Kruppel-like factor) 5 that, in concert with C/EBP β and δ , activates the PPAR γ 2 gene and triggers conversion of preadipocytes to adipocytes. PPAR γ then activates the expression of C/EBP α that feeds back and regulates the [PPAR \$\gamma\$](#) gene. PPAR γ is the terminal factor that controls adipogenesis and the ultimate expression of genes expressed in mature adipocytes, including those encoding fatty-acid binding protein, lipoprotein receptor, lipoprotein lipase, glycerol kinase, and phosphoenolpyruvate carboxykinase. Endogenous negative regulators of adipocyte differentiation are Pref-1 and Wnt-10b.

Adipose tissue participates in the regulation of energy homeostasis as an important endocrine organ that secretes a number of biologically active factors, termed “adipokines”, which include [tumor necrosis factor \(TNF\)- \$\alpha\$](#) , [leptin](#), resistin, and [adiponectin](#). There is growing evidence that adipocyte secretory products are important determinants of insulin resistance, through either a traditional (circulating) hormonal effect or local effects on the adipocyte.

[►Obesity](#)

[►Wnt/Beta-Catenin Signaling Pathway](#)

Adiponectin

Adiponectin is an insulin-sensitizing, anti-inflammatory protein exclusively produced and released by adipocytes (fat cells). It is encoded by the APM1 gene which maps to chromosome 3q27. By regulating the metabolism of glucose and lipids adiponectin has central effects on energy homeostasis. Several studies have examined plasma adiponectin levels in humans and have found decreased levels in obese and diabetic subjects and significant inverse associations with some measure of insulin resistance. High adiponectin levels are associated with a reduced risk of heart attack.

Adjuvant

Definition

An adjuvant is a substance that enhances or modulates the immune response to an antigen.

[►DNA-based Vaccination](#)

ADMET or ADME/Tox

Definition

ADMET or ADME/Tox means Absorption, Distribution, Metabolism, Excretion and Toxicity of a drug or active compound.

[►Molecular Docking](#)

Adoptee's Family Studies

Definition

Adoptee's (family) studies compare rates of disorder in biological and adoptive relatives of affected adoptees, to those in biological and adoptive relatives of control adoptees.

[►Schizophrenia Genetics](#)

ADPKD

Definition

ADPKD (Autosomal Dominant Polycystic Kidney Disease) is a dominantly inherited disease of humans, characterized by the development of multitudes of renal cysts and progressive loss of renal function.

[►Polycystic Kidney Disease Autosomal Dominant](#)

Adrenal Insufficiency

Definition

Adrenal insufficiency refers to hypoadrenalism, and is caused by adrenal gland dysfunction with decreased or

lack of adrenal hormone production (cortisol, aldosterone and sex steroids: testosterone and estradiol). It is termed primary, secondary or tertiary with regards to where the cause of the adrenal hormone deficiency stems (primary – adrenal gland itself; secondary – pituitary gland; ACTH deficiency or tertiary-CRH (Cortico-releasing hormone) deficiency).

► [Hypothalamic and Pituitary Diseases Genetics](#)

Adrenocorticotropin

Definition

Adrenocorticotropin (ACTH) is a hormone that is produced by the anterior pituitary gland and stimulates hormone production from the adrenal gland (cortisol, ► [aldosterone](#) and sex steroids ► [testosterone](#) and estradiol).

► [Hypothalamic and Pituitary Diseases Genetics](#)

Adrenoleucodystrophy

Definition

Adrenoleucodystrophy (ALD) describes a group of disorders resulting from mutations or deletions in genes for structural components of myelin. It is a peroxisomal lipid storage disorder with different inherited forms (autosomal-recessive, X-chromosomal). The disorder is characterized by neurological symptoms such as psychomotor retardation (deterioration of vision and hearing disturbances), epileptic attacks and dementia in the final state.

► [Genetic Predisposition to Multiple Sclerosis](#)

► [Childhood Cerebral ALD](#)

Adrenomyeloneuropathy

Definition

Adrenomyeloneuropathy (AMN) is a neurologic disorder characterised by a predominant axonal sensory-motor polyneuropathy with gradually developing

paraparesis, disturbed vibration sense in the legs and sphincter dysfunction.

► [Peroxisomal Disorders](#)

Adult Neurogenesis

Definition

Adult neurogenesis is the development of new neurons from ► [precursor cells](#) in the adult brain.

► [Neural Stem Cells](#)

Adult Onset Polycystic Kidney Disease

► [Polycystic Kidney Disease, Autosomal Dominant](#)

Adult Stem Cells

Definition

Adult stem cells are multipotent and can differentiate into several types of tissues. They are synonymic to precursor or progenitor cells. It is generally assumed that these cells are already genetically committed to a certain cell type. Adult stem cells occur in various tissues of the adult body, and have the physiologic function to be able to regenerate some tissues which obey a strong degradation. For example, the hemopoietic stem cells regenerate blood and immunologic cells since they are quickly degraded.

► [Stem Cells: an Overview](#)

AER

Definition

The apical ectodermal ridge (AER) is a critical signaling center at the distal end of the developing limb bud that directs the outgrowth and patterning of limb mesoderm. Several bone morphogenetic proteins (► [BMPs](#)) are expressed in the AER.

► [Bone Disease and Skeletal Disorders, Genetics](#)

Affinity

Definition

Affinity is a measure of binding strength between two compounds, for example, in antigen-antibody reactions or interaction of a ligand and its receptor. An antibody that combines loosely with antigens and dissociates readily is said to possess low affinity.

► [Camel as a Model for Functional Genomics](#)

► [Monoclonal Antibodies](#)

Affinity Chromatography

Definition

Affinity chromatography designates a form of chromatography that is based on specific binding properties between the molecule to be isolated, and the molecule being fixed on the chromatographic column. A typical example is antigen-based purification of specific antibodies from cell culture media or patient sera.

► [Affinity Chromatography and *In Vitro* Binding \(Beads\)](#)

► [Protein Interaction Analysis: Chemical Cross-Linking](#)

► [Protein Tags](#)

► [Proteomics in Human-Pathogen Interactions](#)

Affinity Chromatography and *In Vitro* Binding (Beads)

KONRAD BÜSSOW
Max Planck Institute for Molecular Genetics,
Berlin, Germany
buessow@molgen.mpg.de

Synonyms

Affinity separation; Interaction chromatography

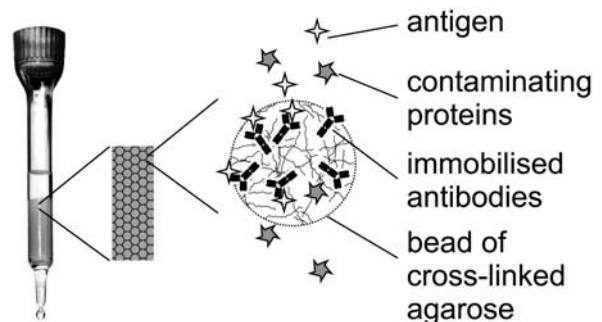
Definition

► [Affinity chromatography](#) exploits the specific binding of a biomolecule, e.g. a protein or nucleic acid, to a

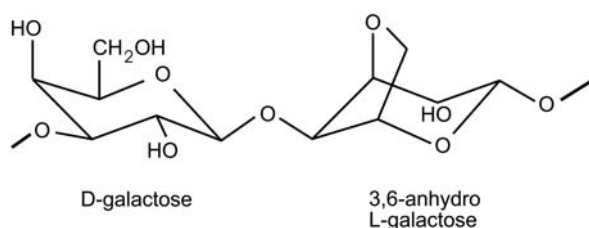
ligand that is immobilised on a solid support. It is a powerful means of preparative protein isolation, but also a valuable tool for scientific discovery. The combination of affinity chromatography approaches and ► [mass spectrometry](#) is an efficient approach for identification and characterisation of protein complexes.

Characteristics

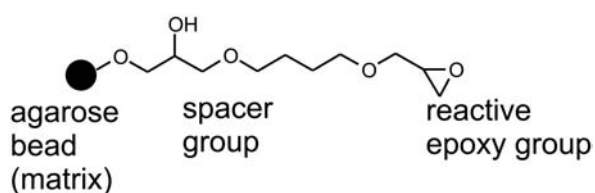
Affinity chromatography is a technique for isolation of biomolecules by virtue of their specific binding properties (1). The technique involves pairs of biomolecules, which specifically interact with and bind to each other, such as antibody/antigen, receptor/ligand and enzyme/substrate pairs or complementary nucleic acids (Fig. 1). Affinity chromatography is a common protein separation technique and is also used for analytical purposes and studies of protein interactions. As a prerequisite, one of the interacting biomolecules, the ligand, has to be immobilised on an inert support. The standard supports for affinity chromatography are natural polymers – ► [agarose](#) (Fig. 2), dextran and cellulose – which form gels with large diameter, non-rigid particles. More recently, small rigid particles of silica or synthetic polymers, suitable for high performance liquid chromatography (► [HPLC](#)), have been adapted to affinity chromatography. These modern chromatography supports withstand the high pressures and flow rates of HPLC. Traditional affinity columns have to be operated at a low flow rate because of slow mass transport by diffusion into the pores of the large chromatographic support particles (‘► [beads](#)’). The particles of modern supports are either very small or have large pores which reach through the entire



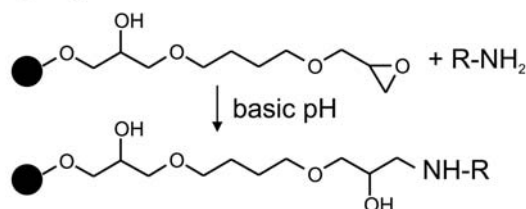
Affinity Chromatography and *In Vitro* Binding (Beads). Figure 1 Affinity chromatography applications. The photo shows a simple affinity chromatography column operated by gravity flow. More sophisticated columns are connected to pumps and UV-detectors with tubing. The schematic shows the isolation of an antigenic protein on an affinity column of immobilised antibodies.



Affinity Chromatography and *In Vitro* Binding (Beads). **Figure 2** Agarose. Structure of the agarose polymer. Agarose is found derivatised in natural sources. For chromatography, agarose beads are stabilised by chemical cross-linking.



Coupling reaction:



Affinity Chromatography and *In Vitro* Binding (Beads). **Figure 3** Immobilisation via flexible linkers. A typical preparation of an affinity resin with a flexible linker/spacer is shown. The linker carries a reactive group for immobilisation of ligands. The ligand (R) is covalently attached to the linker *via* its amino group. The affinity resin is then ready for purification of biomolecules that bind the ligand.

particle, allowing the liquid phase to flow through. Binding equilibrium is therefore established fast and overall run times are minimised.

Usually small molecules are immobilised, but antibodies, proteins and nucleic acids are also used as ligands. Spacer groups/flexible linkers attach the ligand to the solid support, allowing for sterically unimpeded access of the solutes (Fig. 3). When a protein mixture is applied to an affinity chromatography column, only proteins with specific affinity to the immobilised ligand are retained, while other proteins are washed away. Captured proteins are then released ('eluted') by a change of pH or ionic strength or with an excess of the soluble form of the immobilised ligand.

►Affinity Tags

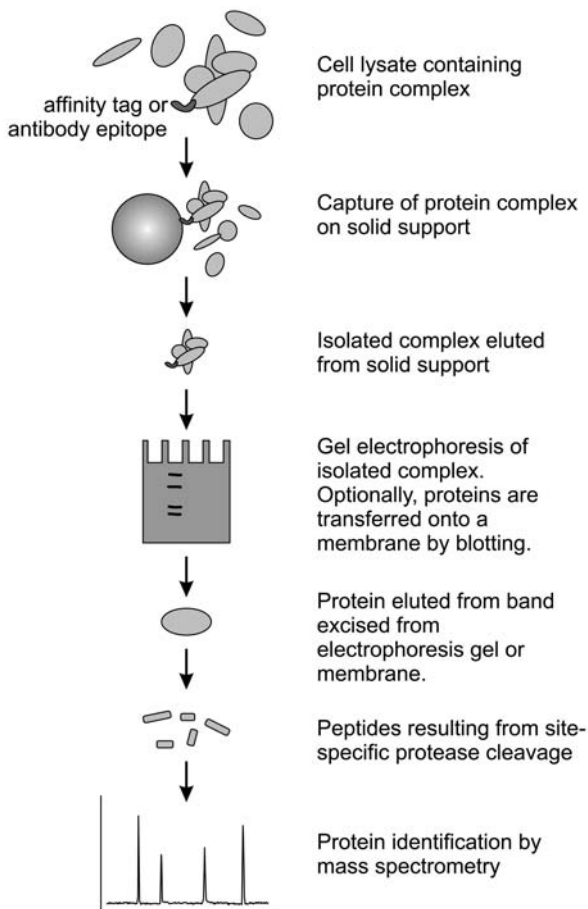
Affinity tags are proteins or peptides that are known to work well in an affinity chromatography system. An affinity tag is fused to a protein of interest by genetic engineering to obtain a fusion protein that is easily purified by a well established affinity chromatography system. Examples for affinity tags and corresponding affinity chromatography systems are glutathione-S-transferase and glutathione Sepharose (glutathione cross-linked to Sepharose), maltose binding protein and amylose, calmodulin binding peptide (CBP) and calmodulin beads, and the FLAG[®] epitope peptide tag and chromatography material carrying the anti-FLAG[®] antibody. The ►His-tag is a peptide of six or more histidines that binds to immobilised nickel or cobalt ions.

Identification of ►Protein Complexes

Oligomeric protein complexes are ideally studied under physiological, native expression levels and conditions (2). Over-expression of the subunits of a protein complex might lead to artificial results. Protein complexes have traditionally been isolated by ►immunoprecipitation or '►GST pull-down experiments'. Immunoprecipitation is a powerful approach for identification of a protein's binding partners. It requires a specific antibody of high affinity, which captures its antigen together with any bound proteins from a cell or tissue extract. Protein A of *Staphylococcus aureus* cross-links and precipitates antibody-protein complexes which are then isolated by centrifugation. Alternatively, Protein A bound to agarose beads is used. Captured proteins are separated and visualised by gel electrophoresis and identified by mass spectrometry (Fig. 4).

The GST pull-down approach uses a recombinant protein as a bait that is fused to an affinity tag to capture ('pull down') binding partners out of a cell lysate. The cell lysate is applied to the immobilised bait protein or, alternatively, bait protein and lysate are mixed in solution and complexes captured afterwards. Glutathione-S-transferase (GST) has most often been used as the affinity tag, but other tag systems have also been used successfully. Pull-down experiments are readily performed and do not require specific antibodies. However the high concentration of the recombinant bait protein might lead to non-physiological protein interactions. A potential binding partner of the recombinant bait protein might not be available for binding, if it is already sequestered in a complex with endogenous binding partners.

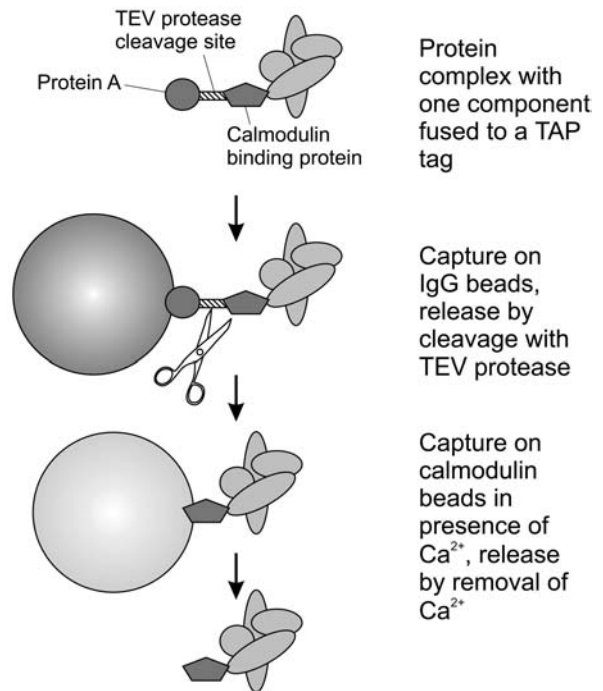
To circumvent the limitations of the GST pull-down and the requirement of immunoprecipitation for specific antibodies, novel approaches have been developed. The ►TAP tag (tandem affinity purification) is a highly



Affinity Chromatography and *In Vitro* Binding (Beads). Figure 4 Isolation of protein complexes and protein identification by mass spectrometry.

specific ‘double’ affinity tag that combines protein A of *Staphylococcus aureus*, which binds ►immunoglobulin G antibodies (IgG), and a calmodulin binding peptide (CBP). Both of these tag systems allow the efficient capture of proteins present in low concentrations. The combination of the two tags results in a highly specific isolation of protein complexes of low cellular abundance.

The TAP tag is typically used to isolate protein complexes from yeast cells but also works in mammalian cells. By genetic engineering of a cell’s genome, the open reading frame coding for a protein complex subunit is fused to the TAP tag coding sequence. The resulting fusion protein is expressed in native amounts from its original promoter and is incorporated into a protein complex. The complex is then isolated by two subsequent steps of affinity chromatography (Fig. 5) using immobilised IgG antibodies and calmodulin. First the protein complex



Affinity Chromatography and *In Vitro* Binding (Beads). Figure 5 Isolation of protein complexes with the TAP tag.

is captured on immobilised IgG. Since the interaction of protein A with immobilised IgG is not readily reversible, the TAP tag contains a cleavage site for the highly specific protease of tobacco etch virus (TEV). Treatment with TEV protease cleaves the TAP tag between protein A and CBP and releases the protein complex (Fig. 5). The released protein complex is still contaminated with unrelated proteins at this stage and a second affinity purification is required. The complex, still carrying the CBP tag, is captured on calmodulin beads. CBP binding to calmodulin is calcium-dependent. The immobilised protein complex is eluted with a calcium-chelating reagent, separated by electrophoresis and identified by mass spectrometry. The TAP tag is usually fused to the C-terminus of a protein complex subunit, but an N-terminal fusion can also be chosen if the tag interferes with protein function or complex assembly.

Affinity tags and mass spectrometry were used for genome-wide screens for protein-protein-interactions in the yeast *Saccharomyces cerevisiae*. Yeast genes were systematically genetically engineered to fuse their gene products with affinity tags. The TAP tag and the FLAG[®] epitope tag, the epitope of the anti-FLAG[®] antibody, were used. Thousands of new protein-protein interactions were discovered by this approach (3, 4).

Clinical Relevance

Affinity chromatography plays an important role in scientific discovery, clinical research and diagnostics and, more recently, in the preparation of therapeutic proteins (►[biopharmaceuticals](#)).

Clinical Diagnostics

In clinical research and diagnostics, affinity chromatography is used to isolate and quantitate various classes of biomolecules (5). When compared to ►[ELISA](#), affinity chromatography has the advantage of a direct linear correlation of signal strength and biomolecule concentration. When combined with modern HPLC equipment, affinity chromatography is a reliable and very fast method of high precision. It permits the assaying of a sample for more than one component – the flow through of a column containing ligand A can be directly forwarded onto a second column with a different ligand B and so forth. HPLC runs are fast, but are performed sequentially in contrast to ELISAs, which are usually performed in microtiter plates on hundreds of samples in parallel. Therefore, affinity chromatography is less ideal for diagnostic applications that require measurement of large numbers of samples. A classical clinical application of affinity chromatography is boronate affinity chromatography, which is used for capture and quantification of glycosylated proteins. The concentration of glycosylated serum proteins, namely ►[glycohemoglobin](#), is an important parameter in the therapy of ►[diabetes](#). Glycosylated haemoglobin isolated by boronate affinity chromatography is specifically monitored and quantified by absorption measurement at 410 nm.

Lectins are used in affinity chromatography to study ►[glycoproteins](#) and to separate ►[isoenzymes](#) and other closely related, but differently glycosylated proteins. Examples include alkaline phosphatase isoforms separated by wheat germ agglutinin affinity chromatography and lipoproteins separated by immobilized concavalin A.

Probably the most common application of affinity chromatography is the isolation and analysis of immunoglobulins from antisera, performed with immobilized protein A from *Staphylococcus aureus* or protein G of *Streptococcus*. Different immunoglobulin classes and subclasses can be differentiated by their affinity for protein A or protein G. Generally class G immunoglobulins are bound most strongly, but other classes can also be isolated.

Purification of Biopharmaceuticals

The introduction of proteins for clinical therapy, so-called biopharmaceuticals, ►[biologics](#) or ►[biologicals](#), requires safe, reliable and cost-effective methods for

protein purification (6). Proteins used in therapy have to be purified by robust methods that are compatible with harsh cleaning and sterilization procedures. Immobilized proteins or antibodies are not ideal since their stability is limited and they require careful quality control. They might contain contaminants from the biological source they originate from, such as pyrogens or viruses. Synthetic ligands, including peptides, are well suited for purification of biopharmaceuticals. Synthetic ►[biomimetic ligands](#) are designed on the basis of known ligand-protein interactions. The rational *de novo* design of such ligands makes use of three-dimensional structures of protein-ligand complexes. The most important class of biopharmaceuticals are immunoglobulins. Synthetic peptides and chemical compounds that mimic the contact region of protein A bound to immunoglobulin G have been developed. These compounds were designed according to the three-dimensional structure of a protein A-immunoglobulin complex. Useful synthetic ligands were obtained that were further refined by screening compound libraries created by combinatorial chemistry on the basis of the original ligands. This screening resulted in ligands that are used in affinity chromatography to isolate immunoglobulin G from plasma in excellent purity and yield. Synthetic ligands have also been successfully developed for insulin and ligands mimicking specific peptide substrates of proteases such as kallikrein or elastase have been designed.

Scientific Discovery

Affinity chromatography is not only a convenient means of protein purification, but also a very important tool of scientific discovery. It was and is being used to isolate and identify novel enzymes, antibodies and antigens, hormones and receptors. The isolation of the ►[insulin receptor](#), which was at that time unknown, using immobilised insulin, is an impressive example of the power of this technique.

More recently, affinity chromatography has been used in combination with mass spectrometry to identify large sets of novel protein-protein interactions.

References

1. Kline T (1993) Handbook of affinity chromatography, 1st edn. Marcel Dekker, New York
2. Drewes G, Bouwmeester T (2003) Global approaches to protein-protein interactions. *Curr Opin Cell Biol* 15:199–205
3. Ho Y, Gruhler A, Heilbut A et al (2002) Systematic identification of protein complexes in *Saccharomyces cerevisiae* by mass spectrometry. *Nature* 415:180–183
4. Gavin AC, Bosche M, Krause R et al (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415:141–147

5. Hage DS (1999) Affinity chromatography: A review of clinical applications. Clin Chem 45:593–615
6. Affinity chromatography issue (2001) J Biochem Biophys Methods 49:1–743

Affinity Constant

Definition

Affinity constant refers to a parameter that measures the strength of interactions between molecules. The affinity constant K_A is related to the free energy of binding $\Delta G = -RT \ln (K_A)$, and determines the equilibrium population of different molecular states through the mass action law.

► [Analytical Ultracentrifugation](#)

Affinity Separation

► [Affinity Chromatography and In Vitro Binding \(Beads\)](#)

Affinity Tag

Definition

Affinity tag refers to a protein or peptide that works well in an affinity chromatography system, and is fused to recombinantly expressed proteins by genetic engineering. Affinity tag fusion proteins are easily purified by standardised affinity chromatography systems.

► [Affinity Chromatography and In Vitro Binding \(Beads\)](#)

► [Protein Tags](#)

Agarose

Definition

Agarose is a linear polysaccharide extracted from sea weed (red algae) consisting of alternating D-galactose

and 3,6-anhydro-L-galactose and carrying non-uniform modifications, e.g. sulfate esterification. Agarose, at concentrations between 1% and 3%, is, for example, used for gel Electrophoresis of DNA.

► [Affinity Chromatography and In Vitro Binding \(Beads\)](#)

Aggrecan

Definition

Aggrecan is a large and highly complex macromolecule of the cartilage extracellular matrix. An average aggrecan monomer has a molecular mass of 3 Mda, with up to 90% of its mass contributed by unbranched sulphated glycosaminoglycan chains. Numerous aggrecan molecules connect via link proteins to a long linear hyaluronan chain. The negative charge of the sulphated glycosaminoglycans establishes a high degree of hydration needed to absorb compressive forces.

► [Extracellular Matrix](#)

Aggregation (Chimera)

Definition

Aggregation (► [chimera](#)) refers to a method to make chimeric animals by placing a small group of embryonic stem (ES) cells in close proximity to an eight cell embryo. The two groups of cells aggregate and may form a single embryo derived from the two cell sources (chimera).

► [Large-Scale Homologous Recombination Approaches in Mice](#)

► [Stem Cells: an Overview](#)

► [Transgenic and Knockout Animals](#)

Aggregophore

Definition

The aggregophore is an intracellular vesicle where the aquaporin2 channel accumulates after synthesis.

► [Diabetes Insipidus, a Water Homeostasis Disease](#)

Aging

Definition

Aging is defined as the age-associated disturbance of those physiological functions that are necessary for the survival and reproduction of an organism.

- ▶ *C. Elegans* as a Model Organism for Functional Genomics
- ▶ Molecular Aging Research

Agonist

Definition

An agonist is a drug or ligand of a target receptor protein that produces an intrinsic signal by binding to its receptor.

- ▶ Cytokine Receptors
- ▶ Structure-Based Drug Design

Agouti

Definition

Agouti denotes a secreted protein that leads to the production of agouti coat colour in a Tyr⁺ background mouse (a mouse expressing tyrosinase).

- ▶ Mouse Genomics

AGS

- ▶ Activators of G Protein Signaling

AGT

- ▶ Alkyltransferases

AHO

- ▶ Albright's Hereditary Osteodystrophy

AIDS

Definition

AIDS stands for acquired immune deficiency syndrome. It is caused by the human immunodeficiency (retro)virus type 1 (▶ HIV-1).

- ▶ Retroviruses

AJ

- ▶ Adherens Junction

Akt

Definition

Akt refers to a serine threonine kinase [also called protein kinase B (PKB)] which functions as an intracellular signaling molecule in the ▶ PI3K pathway.

- ▶ Breast Cancer

Alagille Syndrome

Definition

Alagille syndrome is a rare hereditary disorder (1:50,000 newborns) caused by functional loss of one allele of the JAG1 gene (▶ Jag1 and 2), in 99% of the cases by molecular mutation and ~1% by chromosome 20p12 deletion. Characteristic signs include arteriohepatic dysplasia, cholestasis, and facial and vertebral anomalies. Infants may die from hepatic and cardiac

malformations, and older patients from complications of cholestasis and hypercholesterolemia.

- ▶ Microdeletion Syndromes
- ▶ Notch Pathway

Albright's Hereditary Osteodystrophy

Definition

Albright's hereditary osteodystrophy (AHO) is a complex clinical syndrome that is characterized by short stature, obesity, skeletal abnormalities, mental retardation, and often subcutaneous ossification. It frequently concurs with resistance to parathyroid hormones and to other hormones acting via G_s -coupled receptors. Loss-of-function mutations in $G\alpha_s$ are the molecular cause of the disease.

- ▶ G-Proteins

Alcohol Abuse, Consequences

Definition

Chronic alcohol abuse can induce chronic disorders in learning and memory (Wernicke's syndrome or so-called Korsakow's psychosis), which are characterized by confabulation as a key symptom and delirium characterized by neuron hyperactivity, tremor, insomnia and anxiety. Following chronic administration, the brain gets hyperexcitable and the susceptibility to seizures is massively increased. Neural hyperactivity can destroy neurons by overexcitation. Alcohol can be seen as an N-Methyl-D-Aspartate (NMDA) receptor antagonist at concentrations reached in the brains of alcohol abusers. There is some *in vitro* evidence that the effects of ethanol may be related to selective actions at NMDA receptor 2B subunit containing ionotropic glutamate receptors.

- ▶ Addiction, Molecular Biology

ALD

- ▶ Adrenoleukodystrophy

Aldosterone

Definition

Aldosterone is a small hydrophobic molecule and belongs to the class of steroid hormones. Aldosterone is the major mineralocorticoid in the body. Its receptor belongs to the superfamily of steroid hormone receptors that are located intracellularly, and upon binding translocate into the cell nucleus, where they regulate the transcription of those genes that contain the appropriate responsive regulatory element. In the intestine, and in particular in the distal tubules of the kidney, aldosterone increases the expression of proteins that modulate the activity of the sodium-potassium ATPase and the amiloridsensitive sodium channel. As a consequence, sodium-reabsorption and secondarily potassium excretion increase.

- ▶ Mendelian Forms of Human Hypertension and Mechanisms of Disease

ALFPs

- ▶ Amplified Fragment Length Polymorphic Markers

ALK

Definition

Activin receptor-like kinase (ALK-1 to -7) is an alternative and common nomenclature for the I RSKs type (Receptor Serine Kinase).

- ▶ Receptor Serine/Threonine Kinases
- ▶ ActR-IB/-IC

Alkaline Phosphatase

Definition

Alkaline phosphatase is an enzyme that hydrolyzes phosphate esters in the pH range of 9 to 11. It can be used as a reporter allowing the detection of a specific target after applying the substrate.

- ▶ Large-Scale Homologous Recombination Approaches in Mice
- ▶ Immunochemical Methods, Localization

Alkanethiol

Definition

Alkanethiol is a hydrophobic organic compound that consists of an alkyl chain and a sulfhydryl group on one end, general formula: $\text{CH}_3-(\text{CH}_2)_n-\text{SH}$.

► [Surface Plasmon Resonance](#)

Alkyltransferases

BERND KAINA, MARKUS CHRISTMANN
Department of Toxicology, University of Mainz,
Mainz, Germany
kaina@uni-mainz.de

Synonyms

O⁶-Alkylguanine-DNA Alkyltransferases; AGT;
O⁶-Methylguanine-DNA Methyltransferase; MGMT

Definition

In prokaryotic and eukaryotic cells a group of DNA repair proteins, termed DNA alkyltransferases, have evolved that are capable of repairing DNA by damage reversal. The repair proteins remove methyl, ethyl and even larger alkyl groups from the O6 position of guanine and O4 position of thymine. *E. coli* alkyltransferases can also remove methyl groups from the phosphate in the DNA backbone. Repair occurs in a single-step reaction, by direct alkyl group transfer from DNA to a cysteine residue in the active centre of the repair molecule. Thereby the alkyltransferase becomes inactivated and guanine (and thymine) in the DNA is restored. As the repair enzyme is used up during the repair reaction, alkyltransferase repair proteins are considered as suicide enzymes. Since O⁶-alkylguanine is a highly pro-mutagenic and pro-carcinogenic DNA lesion, alkyltransferases play an important role in cellular defence against alkylating mutagens and carcinogens.

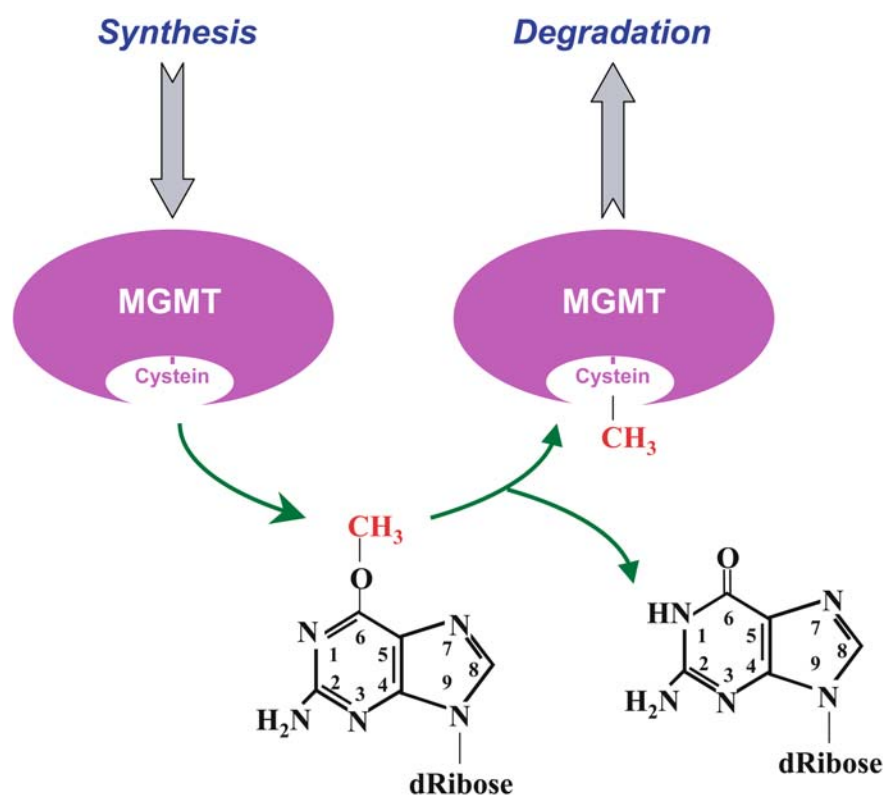
Characteristics

In all organisms, the DNA is subject to harmful damage by alkylation, notably methylation. This occurs by endogenous methylating species such as methyl

adenosine, by metabolic processes and the exposure to environmental carcinogens (e.g. nitrosamines in food and tobacco smoke). Alkylation can occur on 15 sites in DNA. One of the minor alkylation products, O⁶-alkylguanine, amounting to less than 8% of total alkylations, is repaired by O⁶-methylguanine-DNA methyltransferase (MGMT). Since MGMT also removes larger alkyl groups from O6 of guanine, MGMT is also referred to as O⁶-alkylguanine-DNA alkyltransferase (AT or AGT). With the increasing chain length of the alkyl group, the repair reaction becomes less efficient. Alkyltransferases also remove methyl groups from the O⁴-position of thymine. Both O⁶-methylguanine and O⁴-methylthymine are mispairing lesions that give rise to mutations. As already mentioned, methylation of DNA is performed by endogenous cellular reactive species. Therefore, the repair of DNA methylation lesions is presumably the principal role of alkyltransferases. Since O⁴-methylthymine is much less formed than O⁶-methylguanine, its repair by alkyltransferases might be less important than the repair of O⁶-methylguanine.

DNA alkyltransferases belong to the group of DNA repair enzymes that repair DNA by damage reversal. This occurs in a one step reaction without excising the alkylated base from DNA. During the repair reaction, the methyl (alkyl) group from O⁶ of guanine and O⁴ of thymine is transferred to an internal cysteine residue within the active center of the alkyltransferase (Fig. 1). This results in the restoration of guanine and thymine within genomic DNA. Besides methyl groups, larger alkyl groups, which include ethyl, propyl and chloroethyl groups, can also be removed from the O6 position of guanine by alkyltransferases (1). Benzyl groups are also subject to transfer to the enzyme. The efficiency of the transfer reaction decreases with the size of the alkyl adduct. The alkyl group transfer leads to irreversible inactivation of the alkyltransferase and targets it for ►[ubiquitination](#) and proteasome-mediated degradation. Therefore, alkyltransferases are characterized as suicide enzymes. Due to the stoichiometry of the repair reaction, the repair capacity of a cell is determined by the amount of preexisting alkyltransferase molecules and their re-synthesis.

DNA alkyltransferases are ubiquitously distributed in prokaryotic and eukaryotic organisms. Many of them have been cloned; the genes of some of them are summarized in Table 1. Interestingly enough, alkyltransferases have not yet been discovered in plants, and it remains enigmatic whether they exist there or not. In the bacteria *E. coli*, two alkyltransferases were found, Ada and Ogt. Ogt is constitutively expressed at low levels (less than 50 molecules per cell), whereas Ada is normally not expressed but can be induced to high



Alkyltransferases. Figure 1 Methyl group transfer from the O6 position of guanine to MGMT.

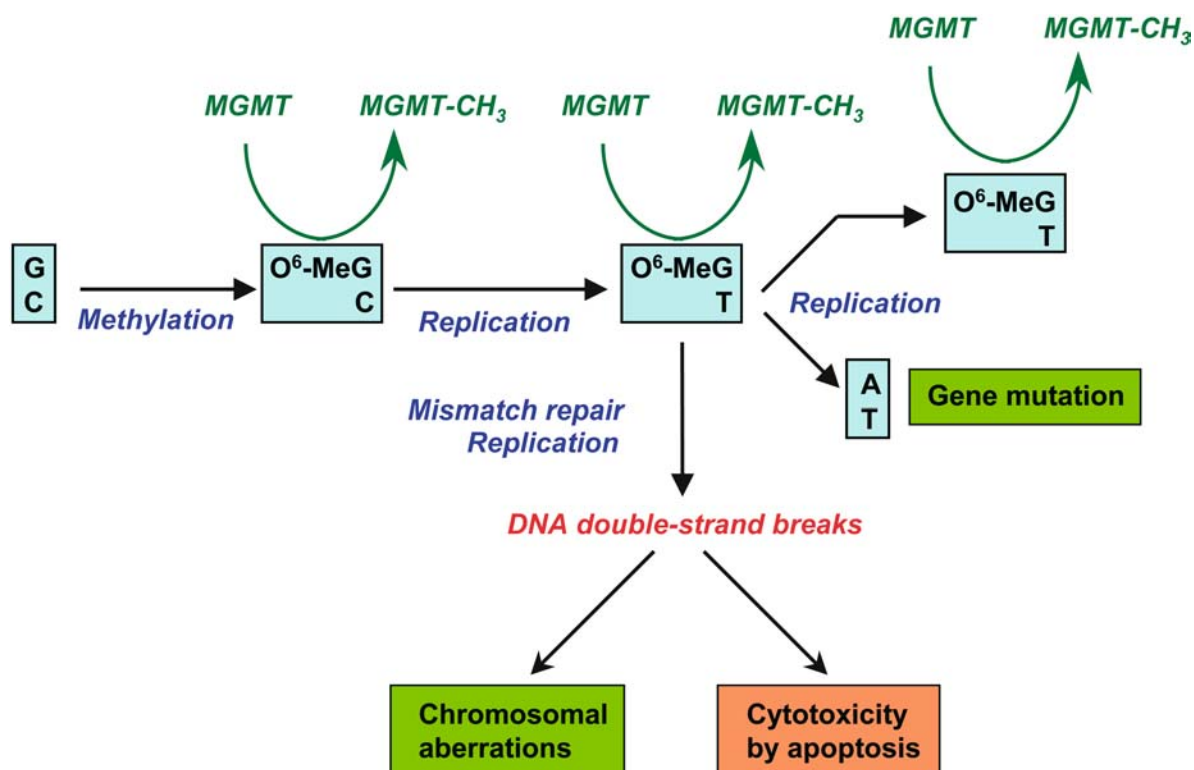
Alkyltransferases. Table 1 DNA-Alkyltransferases

Gene	Organism	Inducibility
<i>Ada</i>	<i>Escherichia coli</i>	yes
<i>Ogt</i>	<i>Escherichia coli</i>	no
<i>Ada</i>	<i>Salmonella typhimurium</i>	yes
<i>Ogt</i>	<i>Salmonella typhimurium</i>	no
<i>adaA</i>	<i>Bacillus subtilis</i>	yes
<i>adaB</i>	<i>Bacillus subtilis</i>	yes
<i>mgt1</i>	<i>Sacharomyces cerevisiae</i>	no
<i>Mgmt</i>	<i>Rattus norvegicus</i>	yes
<i>Mgmt</i>	<i>Homo sapiens</i>	?

expression levels (>1000 molecules per cell) following exposure of the bacteria to methylating agents. This provokes adaptation of cells to simple alkylating agents, and Ada plays a key role in this phenomenon

called adaptive response. During adaptation of *E. coli*, four genes *ada*, *alkA*, *alkB* and *aidB* are upregulated (2). Unlike other DNA alkyltransferases, the Ada protein contains two active sites with reactive cysteine residues at position 69 and 321, respectively. Cysteine 321 binds methyl groups from O⁶-methylguanine and O⁴-methylthymine, whereas cysteine 69 binds methyl groups from the phosphate, thus repairing the methyl phosphotriesters that are formed in higher amounts than the O-methyl purines upon methylation of DNA. Binding of alkyl groups to cysteine 69 leads to a conformational change of the Ada protein. Thereby, Ada is able to bind to the promoter of the *ada* operon and to activate the corresponding genes that are involved in the adaptive response (i.e. *ada* itself, the glycosylase *alkA*, the iron dependent dioxygenase *alkB* and a gene with yet unknown function, *aidB*). The Ogt protein of *E. coli* contains only one active center, which corresponds to Cys 321 of Ada. It does not repair methyl phosphotriester and thus resembles the human MGMT protein.

Human cells contain only one alkyltransferase species. The human MGMT gene is located at the chromosomal position 10q26. It consists of one non-coding and four



Alkyltransferases. Figure 2 Cytotoxicity and mutagenicity induced by O⁶MeG and protection mediated by MGMT.

coding exons with a size of about 145 kb. The gene encodes an mRNA with 866 nucleotides, which codes for a protein containing 207 amino acids and a molecular weight of 24 kDa.

Protection Mediated by MGMT

MGMT protects cells from the cytotoxic, genotoxic (i.e. clastogenic and SCE-inducing), mutagenic and carcinogenic effects of alkylating agents. This is because O⁶MeG lesions in DNA mispair during replication with thymine, which results in GC to AT transition mutations (Fig. 2). These transition mutations are the main cause for the carcinogenic potential of simple alkylating agents. Cytotoxicity and genotoxicity of O⁶MeG requires an additional DNA repair mechanism, termed **mismatch repair**, to operate on O⁶MeG-thymine mispairs (3). Mismatch repair removes thymine from O⁶MeG-T mispairs. Due to the mispairing properties of O⁶MeG, thymine is again inserted, which results in a futile erroneous repair cycle. This may lead to single-strand DNA repair patches that block DNA replication, and finally result in cytotoxic and genotoxic DNA double-strand breaks (4). O⁶MeG is also a powerful apoptotic lesion. Due to the pleiotropic consequences of O⁶MeG, it is evident that

DNA repair by MGMT is most important in the cellular defence against the various effects provoked by simple alkylating agents.

Cells that lack MGMT are highly sensitive to O⁶-alkylating agents. MGMT transfection provokes expression-dependent protection against alkylation-induced cell kill, gene mutations, SCEs and chromosomal aberrations (5). MGMT knockout mice are viable, show a high frequency of spontaneous tumors, and are sensitive to treatment with O⁶-alkylating agents. Conversely, MGMT over-expressing mice show a reduced frequency of tumors in liver, thymus and skin. MGMT over-expression in the skin of mice specifically protects against tumor initiation and tumor progression, without affecting tumor promotion in two-stage skin carcinogenesis experiments.

Regulation of MGMT

MGMT was the first mammalian DNA repair gene shown to be inducible by genotoxic stress. The MGMT promoter represents a classical promoter without TATA and CAAT boxes, but contains a 59 bp enhancer element located at the first exon/intron boundary. Induction of MGMT was shown *in vivo* in the rat liver upon ionising radiation exposure, and *in vitro* upon

exposure of rat hepatocytes to different genotoxins. Besides genotoxic stress, MGMT can also be induced by glucocorticoids that activate the MGMT promoter via a glucocorticoid responsive element. Negative regulation of the MGMT promoter was shown upon transfection with wild-type p53. On the other hand, p53 is required for the induction of the MGMT gene in response to genotoxic stress, at least in the rat and mouse model systems. Whether MGMT gets upregulated in human tissues in response to genotoxic stress is still an open question. With respect to transcription factors, MGMT was also shown to be regulated by AP-1. Posttranslational regulation of MGMT has been demonstrated, which includes phosphorylation and ubiquitination. The latter provokes degradation of the alkylated protein.

An important pathway for the regulation of MGMT gene activity is the promoter methylation (related to the formation of 5-methylcytosine in ►CpG islands). Hypermethylation of the *mgmt* promoter provokes transcriptional silencing. The *mgmt* promoter contains several guanine- and cytosine-rich regions, consisting of ten hexanucleotide motives (CCGCCC). These CpG-islands are positioned between -249 and -103 and between position +107 and +196 of the gene. They represent the CpG methylation hot spots that are responsible for silencing of the promoter. Under non-methylation conditions of the *mgmt* promoter, the transcription start site is flanked by 4 precisely positioned nucleosome-like structures. These structures enable the transcription of the gene. Methylation of the CpG-islands results in heterochromatization, which is accomplished by a re-arrangement and random positioning of the nucleosome. This process of heterochromatization leads to shielding of the transcription start site. MGMT promoter methylation correlates with lack of mRNA expression, loss of MGMT protein and enzymatic repair activity. In contrast to the methylation of the MGMT promoter, methylation of MspI/HpaII sites in the exon-containing regions of the MGMT gene can result in up-regulation of MGMT activity.

Variable Expression of MGMT in Tumors and Normal Tissue

The activity of MGMT has been determined in various human tumor and normal tissue types, including brain, colon, ovary, testis and breast. The studies also showed a high variability in expression notably in tumor tissue (6). High levels of MGMT expression were found in colon cancer, melanoma, pancreatic carcinoma and lung cancer. Brain tumors display, on average, low MGMT expression. Comparison of MGMT activity of tumor versus surrounding normal tissue often shows higher activity in the tumor. For ovarian cancer, MGMT expression correlates with grading and staging.

In gliomas, MGMT activity appears to increase in higher-grade tumors. It also increases during the formation of metastases in skin and lymph nodes.

In normal (non-tumor) tissue, the expression of MGMT is tissue- and cell-type regulated. Thus, in rat and human, liver expresses the highest MGMT level, followed by lung and kidney. The amounts of MGMT protein can also differ between individuals, and in the same individual as a function of time. Thus, a long-term study with peripheral blood mononuclear cells (PBMC) from healthy individuals revealed high inter-individual (7.6-fold) and intra-individual (1.4-fold to 3.5-fold within 42 days) variability of MGMT expression. A comparison on the inter-individual MGMT activity in lung and colon samples also revealed a 4 to 5-fold variation.

MGMT and Resistance to Chemotherapy

Anticancer drugs that belong to the group of methylating and chloroethylating agents target DNA, and act by the formation of O⁶-alkylguanine. Consequently, when MGMT repairs these lesions the killing effect will be counteracted. MGMT must therefore be considered as a very important marker of alkylating drug resistance. In brain tumors, MGMT promoter methylation and activity was shown to correlate with tumor resistance towards chemotherapy with the alkylating drugs carmustine and temozolomide. Since low MGMT activity in tumors appears to correlate with promoter methylation, the methylation status of the MGMT promoter is considered predictive of the success of chemotherapy. Silencing of MGMT by promoter hypermethylation can occur early in human tumorigenesis. It was observed in several tumor types such as lung cancer, bladder cancer, cervix cancer, retinoblastoma and non-small cell lung cancer.

Clinical Applications: MGMT Inhibitors and MGMT Gene Transfer

The finding that MGMT activity has an impact on the sensitivity of tumors led to therapeutic approaches, aimed at inhibiting MGMT activity in the tumor tissue in order to increase the tumor response to chemotherapy. At the same time, strategies were being developed to increase the MGMT level in the normal tissue, in order to protect it against the deleterious side effects of chemotherapy.

A highly specific inhibitor of MGMT, which is in clinical trial, is O⁶-benzylguanine (O⁶-BG). O⁶-BG is a pseudo-substrate for MGMT, which inactivates the protein by direct transfer of the benzyl group to the cysteine of the active site of MGMT. O⁶-BG enhances the cytotoxicity of chloroethylating and methylating agents *in vitro* and *in vivo*. Cancer therapy with the administration of O⁶-BG in combination with alkylating

drugs is currently under investigation. To date, no toxicity of O⁶-BG has been observed in clinical studies, and a 30-fold therapeutic window exists between doses that induced toxicity in animals, and those that depleted MGMT in normal tissue. In pre-clinical studies, it was shown that O⁶-BG, at levels of 1 to 5 μ M, was sufficient to deplete MGMT in tumor cells, and potentiate the cytotoxicity of carmustine and temozolomide. O⁶-BG inhibits human tumor xenograft growth in the nude mouse system. Here, MGMT is depleted within 30 minutes after administration, and depletion is maintained for 6 – 8 hours, after which re-synthesis does occur. Beside O⁶-BG, several other highly efficient MGMT inhibitors have been developed. One of them, O⁶-(4-bromothienyl)guanine, has also currently been proved in a clinical trial.

The protection of stem cells from the mutagenic and cytotoxic potential of chemotherapeutic drugs is highly desired in tumor therapy (7). Myelosuppression is known to be dose limiting during tumor therapy with alkylating agents, which is due to the low amount of MGMT expressed by hematopoietic stem cells. Stem cell toxicity to O⁶-alkylating drugs could be attenuated by *mgmt* gene transfer into blood stem cells. For transduction, mutant variants of *mgmt* are available that harbour a change of amino acids in the active site of the MGMT molecule. This mutant form of MGMT cannot be inhibited by MGMT inhibitors. The two MGMT mutants most commonly used are P140K and G156A. The application of MGMT that is insensitive to O⁶-BG allows the combination of both strategies during cancer therapy: stem cell protection by MGMT gene transfer and inhibition of MGMT in the tumor. In preclinical studies, transduction of hematopoietic progenitor cells with MGMT has been shown to confer protection against O⁶-BG administered together with carmustine or temozolomide. A further promising tool is MGMT inhibitor targeting in order to specifically sensitise the tumor tissue. MGMT gene transfer as a supporting tool in the therapy with O⁶-alkylating drugs is just beginning to enter the clinic field.

References

1. Pegg AE, Dolan ME, Moschel RC (1995) Structure, function, and inhibition of O6-alkylguanine-DNA alkyltransferase. *Prog Nucleic Acid Res Mol Biol* 51:167–223
2. Sedgwick B, Lindahl T (2002) Recent progress on the Ada response for inducible repair of DNA alkylation damage. *Oncogene* 21:8886–8894
3. Christmann M, Tomicic MT, Roos WP, Kaina B (2003) Mechanisms of human DNA repair: an update. *Toxicology* 193:3–34
4. Kaina B, Christmann M (2002) DNA repair in resistance to alkylating anticancer drugs. *Int J Clin Pharmacol Ther* 40:354–367
5. Kaina B, Fritz G, Mitra S, Coquerelle T (1991) Transfection and expression of human O6-methylguanine-DNA methyltransferase (MGMT) cDNA in Chinese hamster cells: the role of MGMT in protection against the genotoxic effects of alkylating agents. *Carcinogenesis* 12:1857–1867
6. Margison GP, Povey AC, Kaina B et al (2003) Variability and regulation of O(6)-alkylguanine-DNA alkyltransferase. *Carcinogenesis* 24:625–635
7. Roth RB, Samson LD (2000) Gene transfer to suppress bone marrow alkylation sensitivity. *Mutat Res* 462: 107–120

ALL

► Acute Lymphoblastic Leukaemia

Allele

Definition

Allele refers to a DNA sequence variation of a gene at a specific locus (location) on a chromosome. In a diploid cell each gene has two alleles, each occupying the same position (locus) on homologous chromosomes. A sequence with two to several alleles in the population is called a polymorphic sequence.

- Cre/Lox P Strategies
- Diabetes Mellitus, Genetics
- Familial Dilated Cardiomyopathy
- Large-Scale ENU Mutagenesis in Mice
- Morpholino Oligonucleotides, Functional Genomics by Gene ‘Knock-down’
- Neurotrophic Factors
- SNP Detection and Mass Spectrometry

Allele-Specific

Definition

Allele-specific designates a suppressor that restores wild-type function only with a specific mutation in the target protein, but not others.

- Protein-Protein Analysis, Suppressor Hunting

Allelic Association

Definition

Allelic association is the situation in which a specific allele of a gene is associated with a disorder at the population level, i.e. allele frequency differences exist between cases and controls (either population or family-based).

► Cleft Lip Palate

Allelic Series

Definition

Allelic series describe different mutant alleles of a gene that cause a range of phenotypes, whereby each one carries a single point mutation within different regions of the same gene.

► Morpholino Oligonucleotides, Functional Genomics by Gene 'Knock-down'

► Mouse Genomics

Allergic Disorders, Genetics

► Atopy, Genetics

Allergic Drug Reactions

► Idiosyncratic Drug Reactions

Allergy

Definition

Allergy is a medical condition in which a normally innocuous environmental antigen provokes an immunological Type I reaction. This is characterised by the production of specific Immunoglobulin E (IgE), raised against the allergen inducing activation of mast cells and release of histamines. In sensitised atopic

individuals, allergen exposure gives rise to the manifestation of hypersensitivity; an exaggerated immune response with various degrees of severity ranging from a mild local inflammation to anaphylactic shock and even death. Most allergies are caused by pollen, dust particles, mould spores, food, latex rubber, insect venom, or drugs.

► Protein Interaction-Phage Display

► Idiosyncratic Drug Reaction

Allostery

Definition

Allostery refers to the ability of a protein to change its conformation (and therefore its activity) at one site, as a result of the binding of a small molecule to a second site, which is not the substrate binding site located elsewhere on the protein.

► Nuclear Hormone Receptors: Regulators of Gene Transcription and Cellular Signaling

Allotetraploid

Definition

Allotetraploid is generally described as an organism having two sets of chromosomes, each from a different ancestral species. As an example, *Xenopus laevis* is considered an allotetraploid organism, although the genomic duplication is not complete, and some of its gene copies are non-functional, presumably due to selection events during the ~ 50 million years since the genome duplication event occurred. As the tetraploid nature of *Xenopus laevis* is not complete, it is also occasionally referred to as "pseudotetraploid."

► Xenopus as a Model Organism for Functional Genomics

Allozygous

Definition

Allozygous refers to a homozygous mutation at a given gene locus, and indicates that the mutations in both alleles are different and of independent origin (in

contrast to autozygous when they are identical by descent).

► [Bloom Syndrome](#)

Alpha-Antitrypsin

Definition

Alpha-antitrypsin is a glycoprotein produced in the liver, which is the major antiprotease in the blood serving mainly to inhibit neutrophil elastase.

► [COPD and Asthma Genetics](#)

Alpha-Glucosidase/ Alpha-Mannosidase

Definition

Alpha-glucosidase/Alpha-mannosidase refers to types of glycosidases that catalyze the hydrolysis of terminal, non-reducing α -linked glucose/mannose residues in a glycan.

► [Glycosylation of Proteins](#)

Alpha-Helix (α -Helix)

Definition

The alpha-helix (α -helix) is one of the different types of protein secondary structure, the other ones being β -pleated sheets and turns. An alpha-helix is a tight helix that results from intrastrand hydrogen bonding of the carboxyl (CO) group of one amino acid, to the amino (NH) group of another amino acid. Seven amino acid residues, (a-g) in the polypeptide chain, form just under two turns of the helix, with the eighth residue (the 'a' residue of the next heptad) falling in line with residue one. The propensity for formation of coiled-coil dimers is associated with the location of hydrophobic residues at the 'a' and 'd' positions, giving a run of hydrophobic residues twisting around the surface of the cylindrical helix and making the long rod structure unstable in the aqueous cytoplasm, unless it twists together with another similar helix (► [Leucine Zipper](#)).

► [Intermediate Filaments](#)

► [Protein Databases](#)

► [Beta-sheet Structure](#)

► [Amino Acids: Physicochemical Properties](#)

Alpha-Mannosidase

► [Alpha-Glucosidase/Alpha-Mannosidase](#)

Alpha-Oxidation

Definition

Alpha-oxidation describes the mechanism by which fatty acids can be degraded via removal of the terminal carboxyl-group as CO₂.

► [Peroxisomal Disorders](#)

Alpharetroviruses

Definition

Alpharetroviruses are simple retroviruses that cause cancers in birds. The avian sarcoma and leukosis virus (ASLV) belong to that group of viruses.

► [Retroviruses](#)

Alpharetroviruses CA (Capsid)

Definition

Alpharetroviruses CA is the capsid protein which is encoded by the gag gene; it is the most abundant protein in the virus particle, forms the viral core, and is visible by electron microscopy.

► [Retroviruses](#)

Alpha-Tubulin / Beta-Tubulin

Definition

α - and β -tubulins associate to heterodimers to form microtubules. Both tubulins have molecular weights of

about 50 kDa and bind GTP. In higher eukaryotes, there are up to 7 isoforms of α - and β -tubulin, respectively, which are encoded by different genes.

► **Cytoskeleton**

Alphoid DNA

Definition

Alphoid DNA defines a family of DNA, unique to the centromere of human chromosomes, characterized by repeating units of 171 base pairs, arranged in tandem arrays of 1,500 to 30,000 copies.

► **Centromeres**

Alternative Splicing

STEFAN STAMM

Institute of Biochemistry, Erlangen, Germany

stefan@stamms-lab.net

Definition

Alternative splicing is the alternate inclusion of ► **exons** or ► **introns** from the primary transcript (► **pre-mRNA**) into mature mRNA. pre-mRNA splicing is a maturation event in which parts of the pre-mRNA sequences are joined and exported into the cytosol as exons and the intervening sequences are removed as introns. The majority of human genes are alternatively spliced, i.e. contain sequences that can be used either as an exon or as an intron. ► **Alternative splicing** creates different mRNA isoforms from a single gene. Mostly, but not always, these mRNA isoforms encode distinct protein products.

Characteristics

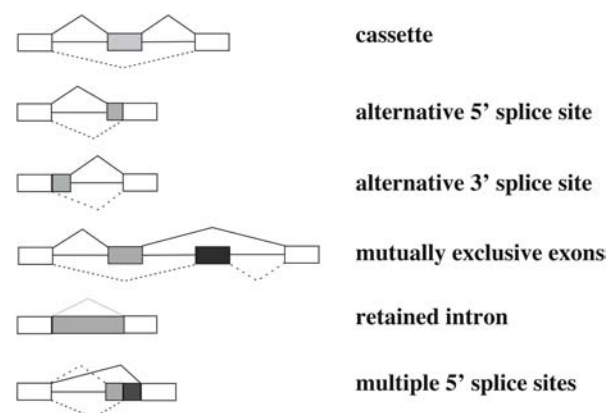
Nomenclature

The removal of an exon from pre-mRNA is termed exon skipping. When an exon is included in mRNA, it is referred to as exon inclusion or exon usage. Several basic types of alternative splicing can be distinguished: alternative cassette exons, retained introns, alternative 5' and ► **3' splice sites** and mutually exclusive exons (Fig. 1). Internal alternative cassette exons are the largest group. However, more complicated patterns such as multiple 5' or 3' splice sites, coordinated usage of internal exons and combinations of the basic types are frequently observed. mRNAs that arise from the

same gene but differ in parts of their sequence are called isoforms. Exons that are always included in mRNA and introns that are always excluded from it are termed constitutive exons and introns. If their usage is different between cells or between mRNA isoforms in a single cell, they are termed alternative. Different mRNA isoforms can also be generated by mechanisms other than alternative splicing, such as alternative promoter usage, RNA editing or alternative polyadenylation. Usually, the most abundant or first discovered isoform is termed the "normal" form and less abundant forms are described as "alternative forms". However, there are no clear rules defining which isoform is normal or alternative.

Abundance of Human Alternative Splicing

An average human gene spans 27 kb and contains a mean of 8.8 exons (average size 145 nt) interrupted by introns (average size 3365 nt) and flanked by 5' and 3' untranslated regions of 770 and 300 nt, respectively. To form mRNA, exons are joined by splicing. On average, they comprise only less than 5% of an average pre-mRNA. The completion of the human genome and comparative analysis with ESTs (expressed sequence tags, corresponding to sequences from actual transcripts) has demonstrated that the majority of genes undergo alternative splicing. Most genes located on chromosomes 21 and 22 undergo alternative splicing, and array analysis indicates that 74% of multi-exon genes are alternatively spliced. The number of mRNAs generated by a gene varies. On average, a human gene generates 2–3 transcripts. However, extreme cases exist; the human *neurexin3* gene can potentially form 1728 transcripts due to alternative splicing at four different sites. In *Drosophila*, the Down syndrome cell



Alternative Splicing. Figure 1 Types of alternative exons. Boxes indicate exons, horizontal lines indicate introns. The alternative exons are shaded in gray. Splicing patterns are indicated by solid lines; alternative splicing patterns as dashed lines.

adhesion molecule DSCAM can potentially generate 38016 isoforms due to alternative splicing. This number is larger than the number of genes present in *Drosophila*. Alternative splicing is observed in all tissues, but it is most prevalent in cells of the immune system and in the brain. It is frequently observed that alternative exon usage changes during development or cell differentiation.

Function

The exact function of the majority of alternative splicing events is not clear. Judging by data published in the literature or obtained by comparison of ►expressed sequence tags (EST) with genomic sequences, about 25–35% of alternative exons introduce frameshifts or stop codons into the pre-mRNA. Since approximately 75% of these exons are predicted to be subject to nonsense mediated decay, an estimated 18–25% of transcripts will be switched off by stop codons caused by alternative splicing and nonsense mediated decay. This process has been termed regulated unproductive splicing and translation. It is currently the function of alternative splicing with the most obvious biological consequences. The biological consequences of protein isoform generation caused by alternative splicing are more diverse and mostly not understood. They include: the generation of soluble rather than membrane bound receptors, the inclusion of novel binding sites leading to alternative exon dependent interaction, the formation of enzymes with altered ligand affinities or catalytic properties, the expression of ion channels with altered electrophysiological properties, the formation of short isoforms that often act as dominant negative forms, the formation of isoforms with different intracellular localization and the formation of proteins with altered pharmacological properties.

The best understood example for the physiological role of alternative splicing is the sex-determination of *Drosophila*. A cascade of splicing regulatory proteins – sex-lethal, transformer and transformer-2 – determines the sex specific alternative splicing of the transcription factor doublesex, which results in male or female specific doublesex isoforms. This cascade of splicing decisions ultimately results in sex specific development. Another well-characterized system is programmed cell death (apoptosis). A number of programmed cell death regulatory genes undergo alternative splicing. The resulting protein isoforms are functionally distinct and sometimes even act antagonistically.

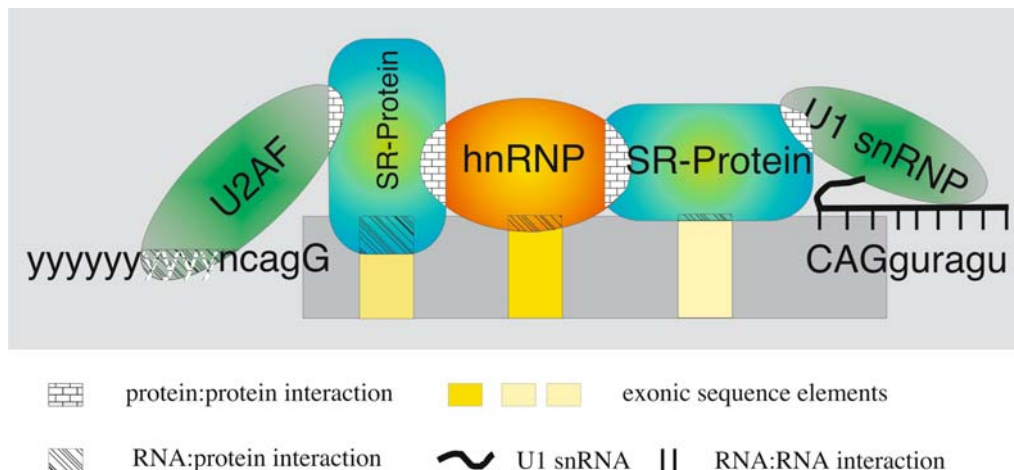
Splice Site Selection

It is still not fully understood how splice sites, especially alternative splice sites, are selected. It is not currently possible to predict accurately which parts of genomic DNA will form exons or alternative exons.

The major problem is that the regulatory sequences that determine an exon, the 5', 3' splice site and the branch point, can deviate significantly from a consensus sequence (3' splice site: $y_{10}ncagG$, ►5' splice site: $CAGgtragt$, branch point: $ynyryay$; intronic sequences: small letters, exonic sequences large letters, invariant nucleotides: bold). 98% of all introns are flanked by invariant GT-AG dinucleotides. The other positions of the splice sites can differ significantly from consensus sequences. 0.76% of introns are flanked by GC-AG dinucleotides and are predominantly alternatively spliced. In a smaller group of introns, these sequences are AT-AC. Alternative exons are often characterized by splice sites that deviate more strongly from the consensus. Since their mutation to the consensus leads to constitutive exons they are referred to as weak splice sites.

Enhancer/Silencer

Since splice sites are so poorly defined, additional sequence elements in the exons (ESE: ►exonic sequence element) or in introns in close proximity (ISE: ►intronic sequence element) are needed for proper recognition. Depending on whether they promote or inhibit the inclusion of an exon, they are named enhancers or silencers (ESE: exonic sequence enhancer, ISE: intronic sequence enhancer). The elements located on the RNA are called cis-elements. They are short (5–8 nt) and can be only described by a loose consensus sequence. Since most exonic elements are in coding regions of genes, flexibility in sequence composition minimizes interference with coding requirements. This ensures that the usage of amino acids is not determined by exonic splicing regulatory sequence elements. Exonic and intronic sequence elements bind to protein factors that are named ►trans-acting factors. The majority of these trans-acting factors belongs to two groups of proteins, ►SR-proteins and hnRNPs. These protein groups contain RNA binding domains and various protein:protein interaction domains, which allow the interaction between individual members of these protein families. As a result, protein-networks on the pre-mRNA can form around exons or introns and aid in their recognition by binding to components of the spliceosome (Fig. 2). The individual interactions between cis and trans-elements involved in splice site selection are weak. Only through several such interactions forming either across an intron or an exon, recognition is achieved. The relative concentration of trans-acting factors varies between cell types and tissues as well as during development. This leads to cell type-specific or developmental stage-specific regulation of alternative exons depending on which proteins assemble on the exon. Due to this combinatorial control, a large number of alternative exons can be regulated by a limited number of regulatory proteins.



Alternative Splicing. Figure 2 Elements responsible for recognizing an exon. The exon is indicated as a gray box. The consensus sequences for the 3' and 5' splice sites are indicated. Capital letters are exonic, small letters intronic. Exonic elements are shown as yellow boxes. Two major groups of proteins bind to exonic elements: SR-proteins and hnRNPs (binding regions hatched). These proteins can interact with each other via several protein:protein interaction domains (region marked with bricks). The 3' splice site is recognized by U2AF (U2 auxiliary factor) that can bind to SR-proteins of exonic enhancer complexes. Components of the constitutive splicing machinery, e.g. U1 snRNP can bind to these proteins as well and define the 5' splice site by base pairing (short vertical lines) between U1 snRNA (thick line) and the 5' splice site. The individual interactions are weak, however in combination they are strong enough to define an exon with high fidelity.

Regulation by External Signals

Alternative splice site selection is often controlled by extracellular factors, e.g. cellular stress, neuronal activity, hormonal stimulation of a receptor and developmental state. Studies performed in animals demonstrated that alternative exon usage changes after physiological stimulation, such as stress evoked by forced swimming or pavlovian fear memory. This indicates that the regulation of alternative exon usage represents a normal adaptation of the cell to environmental changes.

Connection to Other Cellular Processes

Several proteins implicated in alternative splicing also fulfill a role in other aspects of RNA metabolism, such as nuclear export, RNA ►polyadenylation, RNA editing or cytoplasmic mRNA stability. Several proteins regulating alternative splice sites bind to RNA polymerase II or transcription factors. As a result, alternative splice site selection is coupled to transcription and other steps of pre-mRNA processing.

Clinical Relevance

Alternative Splicing in Disease

The splicing process has to be very precise, as the addition or removal of a single nucleotide will alter the reading frame of the mRNA. Normally this process

occurs with high fidelity, meaning that almost no errors are made. Errors in splice site selection can be detrimental for cells and alternative missplicing has been shown to be the cause or consequence of an increasing number of human diseases. Diseases can be caused by mutation in regulatory ►cis elements, and by altering the concentration or structure of trans-acting factors. Mutations in splice sites are the most frequent cause of disease. These mutations can cause exon skipping, intron retention or activation of cryptic 3' or 5' splice sites near an exon. Each of these events changes the protein structure, often causing frameshifts and non-functional mRNAs or proteins. The next class of mutations affects enhancer or silencer regions located in an exon or in the nearby intronic regions. Mutations in exons are often silent and do not change the coding properties of an mRNA. However, since they affect a regulatory pre-mRNA processing element, they induce abnormal splicing. As a result, the mRNA and most frequently the protein structures are changed. Since most exons contain regulatory elements, a significant fraction of disease-causing exonic mutations are probably splicing mutations (Table 1). Mutations in trans-acting factors are far less frequent and complete loss of essential splicing factor function is generally lethal. One exception is retinitis pigmentosa, which is caused by alterations in the splicing factor PRPF31.

Alternative Splicing. Table 1 Some alternative missplicing events associated with human disease caused by mutations in exonic enhancers

Disease	Gene
Acute intermittent porphyria	Porphobilinogen deaminase
Beta-Thalassemia	Beta-globin
Breast and ovarian cancer	BRCA1
Cerebrotenidinous xanthomatosis	CYP 27
Cystic fibrosis	CFTR
FTDP-17	Tau
Immunodeficiency	Adenosine deaminase
Leigh's encephalomyelopathy	Pyruvate dehydrogenase E1
Marfan Syndrome	Fibrillin-1
Menkes disease	MNK
Metachromic leukodystrophy	Arylsulfatase A
Myotonic dystrophy	DMPK
Neurofibromatosis type I	NF-1
Occipital horn syndrome	ATP7A
Sandhoff disease	Beta-hexaminidase
Spinal muscular atrophy	SMN2
Tyrosinemia, Type I	Fumarylacetoacetate hydrolase

Frequently, alternative splicing patterns are changed in diseases without obvious mutations, indicating alternations in trans-acting factors (Table 2). The best known examples are changes in splicing patterns of many genes associated with neoplasia and metastasis. In models of mouse mammary gland tumorigenesis the ratio of various SR-proteins changes and correlates with altered splicing patterns, e.g. the use of variable exons of the hyaluronic acid receptor CD44. Changes in trans-acting factors can be caused by indirect effects. Myotonic dystrophy type I is caused by a CTG expansion in the 3' untranslated region of the myotonic dystrophy kinase. These additional repeats sequester CUG binding proteins that normally regulate alternative splicing of other pre-mRNAs (cardiac troponin T, insulin receptor, muscle-specific chloride channel and myotubularin-related), which causes aberrant splicing of these pre-mRNAs resulting in the disease. Finally, alleles carrying mutations in regulatory elements of the pre-mRNA can have a different phenotype depending on the genetic

Alternative Splicing. Table 2 Some human diseases associated with changes in the ratios of alternative spliced products

Disease	Gene
Breast cancer	MDM2
Melanoma	Bin1
Prostate cancer	FGFR-2
Schizophrenia	N-CAM
Schizophrenia	NMDA R1
Schizophrenia	GABA-A
Sporadic amyotrophic lateral sclerosis	EAAT2
Sporadic amyotrophic lateral sclerosis	NOS
Wilms tumor	WT1

background, indicating that alternative splicing can be a modifier of disease severity.

Therapeutic Approaches

Different therapeutic approaches have been tested in models of diseases caused by missplicing. These include chemically modified oligonucleotides, ribozymes, expression of trans-acting regulatory factors and application of low molecular weight drugs. Oligonucleotides have been used to correct missplicing seen in beta-thalassemia, Duchenne muscular dystrophy and acetylcholine esterase missplicing. Several of these agents are now undergoing clinical trials.

- ▶ COPD and Asthma Genetics
- ▶ DNA Chips
- ▶ Full Length cDNA Sequencing
- ▶ Transposons

Further Reading

- Maniatis T, Reed R (2002) An extensive network of coupling among gene expression machines. *Nature* 416:499–506
- Stamm S (2002) Signals and their transduction pathways regulating alternative splicing: a new dimension of the human genome. *Hum Mol Genet* 11:2409–2416
- Black DL (2003) Mechanisms of Alternative Pre-Messenger RNA Splicing. *Annu Rev Biochem* 72:291–336
- Faustino NA, Cooper TA (2003) Pre-mRNA splicing and human disease. *Genes Dev* 17:419–37
- Jeanteur P (ed) Regulation of alternative splicing. *Prog Mol Subcell Biol*, vol 31. Springer, Berlin Heidelberg New York
- Stamm S, Ben-Ari S, Rafalska I et al (2005) Function of alternative splicing. *Gene*. 344:1–20

Collection of alternative exons and their properties, references to Tables 1 and 2 can be found under: <http://www.ebi.ac.uk/asd/index.html>

Alu Repetitive Elements

Definition

Alu repetitive elements are short (about 300 basepair) stretches of DNA with a repetitive sequence, interspersed throughout the genome, which contain restriction endonuclease sites for the AluI restriction enzyme.

►CpG Islands

Aly/ REF

Definition

Aly/REF designates an mRNA export adapter protein that interacts with TAP-NXF1, and which associates with ►pre-mRNA undergoing splicing as part of a complex of proteins that is deposited near exon-exon junctions.

►RNA Export

Alzheimer Dementia

►Alzheimer's Disease

Alzheimer's Disease

ULRICH FINCKH

Laboratory Medicine Dortmund, Dr. Eberhard & Partner, Dortmund, Germany
finckh@labmed.de

Synonyms

AD; Alzheimer dementia; Senile dementia/Alzheimer's type (SDAT)

Definition

Alzheimer's disease (AD) is defined by a clinical history of dementia and a characteristic pattern of

histopathological changes in the brain. These changes include ►senile plaques (SP), ►neurofibrillary tangles (NFT), neuropil threads (NT), and neurodegeneration. Current criteria for the clinical diagnosis of AD were defined by the National Institute of Neurologic and Communicative disorders and Stroke – Alzheimer's Disease and Related Disorders Associations (NINCDS-ADRDA) (1). Criteria for a neuropathology diagnosis of AD were defined by the Consortium to Establish a Registry for Alzheimer's Disease (CERAD) (2). Neuropathological stages (I–VI) are based on the frequency and distribution pattern of NFT and NT (3). A definite diagnosis of AD requires fulfillment of the clinical criteria of probable AD and histopathological evidence (1, 2). Dementia is defined by the loss of cognitive abilities. There exist two types of AD, common sporadic and mostly late-onset AD (►LOAD) and autosomal dominant familial AD (►FAD) with mostly early onset (Table 1).

Characteristics

AD is characterized by progressive dementia with early (<65 years, EOAD) or late onset (>64 years). The duration of clinical disease may exceed 10 years. Clinical manifestation of dementia is preceded by an accumulation of SP, NFTs and NT over many years. The major component of SP consists of extracellularly aggregated β -amyloid ($A\beta$) (►Beta-Amyloid (β -Amyloid)), i.e. a fragment of mostly 40 or 42 ($A\beta_{40}$; $A\beta_{42}$) residues of the amyloid precursor protein (►APP). NFTs and NTs are formed by helical filaments of abnormally phosphorylated microtubule-associated tau proteins. Neuron loss in AD is typically observed first in the entorhinal cortex affecting its afferent and efferent connections with the hippocampus. It is most likely this "isolation" of the hippocampus that causes the characteristic loss of memory functions, an early clinical sign of AD. To date, there exists no specific biomarker that allows a definite diagnosis of common AD before death. Clinical diagnosis of AD based on neuropsychological testing, brain imaging and the exclusion of other known dementing conditions reaches an accuracy of 85–95%.

Epidemiology and Risk Factors

AD is the most frequent cause of dementia (54–74% of all dementias) of the elderly with an ageing-dependent exponentially increasing frequency (4, 5). Less than 5% of sporadic AD arises before 60 years of age. In the groups between 65 and 85 years of age, prevalence and incidence of AD double every 4.5–5 years (Table 2). The higher prevalence of AD in women compared to men may be due to the higher life expectancy of women. Furthermore, there may be an increased risk of AD in association with overweight in women. Proven

Alzheimer's Disease. Table 1 Two types of AD, same histopathology

	LOAD	FAD
Etiology	multifactorial, sporadic	monogenic, autosomal dominant
Dementia onset	>64 ys	<61 ys
Family history	10–20 % positive	positive
% of all AD	>95	<1
Population frequency	$\sim 1 \times 10^{-2}$	$0.5\text{--}1 \times 10^{-4}$

Alzheimer's Disease. Table 2 Mean age-specific prevalence and annual incidence of AD in Europe

Age group	Prevalence (%)		Incidence (%)	
	♂	♀	♂	♀
65–69	0.6	0.7	0.09	0.22
70–74	1.5	2.3	0.3	0.38
75–79	1.8	4.3	0.69	1.03
80–84	6.3	8.4	1.48	2.73
85–89	8.8	14.2	2.42	4.15
90+	17.6	23.6	2.00	6.97
65+	4,4		~ 1	

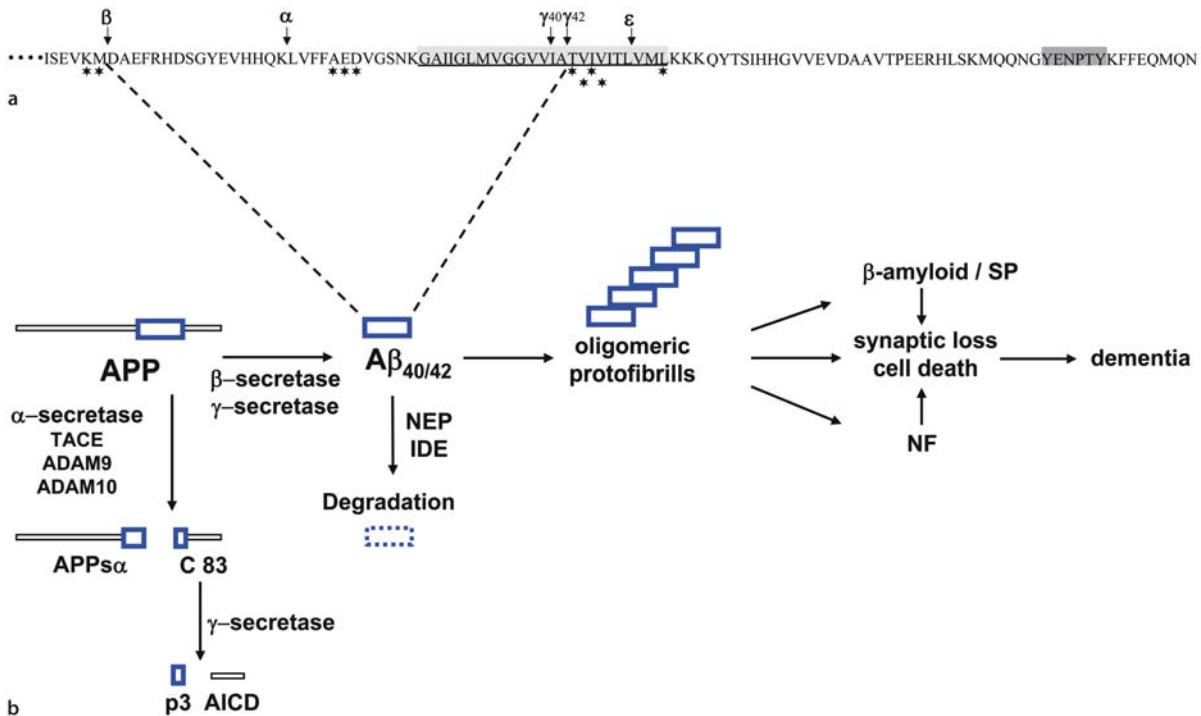
risk factors for sporadic AD are age, education and the $\epsilon 4$ allele of the **▶gene** for apolipoprotein E (Apo E; **▶APOE**) (6, 7, 8). The etiology of sporadic AD is therefore considered to be multifactorial. Mutations causative for FAD, rare monogenic forms of AD, have been described in the genes for APP (**▶APP**), presenilin 1 (**▶PS1**; **▶PSEN1**), and presenilin 2 (**▶PS2**; **▶PSEN2**). FAD contributes to less than 1% of all AD. Typical histopathological features of AD are found in a high proportion of 30–40 year old subjects with trisomy 21 (Down syndrome) or partial trisomy 21 including the chromosomal region harboring *APP* (21q21).

Genetics

Twin and family studies suggest a high proportion of AD is attributable to genetic factors. The relative risk (RR) for AD in first-degree relatives of AD is elevated (RR = 2–6). 50–60% of Caucasian AD patients and 20–30% of non-demented age-matched controls carry one or two *APOE* $\epsilon 4$ alleles. In the population aged 55+, the average life time risk for AD of 8% in males and 20% in females increases to 19% and 33%, respectively in subjects carrying one or two *APOE* $\epsilon 4$ alleles.

10–20% of all AD is attributable to *APOE* $\epsilon 4$. The frequency of the three common alleles of *APOE* ($\epsilon 2$, $\epsilon 3$, $\epsilon 4$) varies between populations. $\epsilon 4$ is considered to represent the archetypical form. $\epsilon 3$ being the most prevalent allele in all populations studied is considered the normal allele, whereas $\epsilon 2$ confers protection and $\epsilon 4$ risk with respect to AD. Compared to $\epsilon 3$ homozygotes, the odds ratio (OR) for AD in Caucasians reaches values of 1.8–3 in $\epsilon 4$ heterozygotes, 6–15 in $\epsilon 4$ homozygotes and ~ 0.5 in $\epsilon 4$ negative carriers of the $\epsilon 2$ allele. The odds ratios may differ with ethnic or geographic background, with $\epsilon 4$ being the risk allele in all populations studied. Since a substantial proportion of elderly non-demented subjects carries *APOE* $\epsilon 4$ and a significant proportion of sporadic AD patients does not carry $\epsilon 4$ alleles, risk alleles of other genes are assumed to exist. To date, no other gene has been shown to be associated unambiguously with sporadic AD.

FAD due to mutations in *APP*, *PSEN1* or *PSEN2* is inherited as an autosomal dominant disorder. To date, databases list 133 different pathogenic mutations in *PSEN1*, 16 in *APP*, and seven in *PSEN2*. No nonsense or loss of function mutation has been found in these three genes. All mutations found to date – mostly missense mutations – are associated with the expression of altered proteins. In patients with suspected FAD based on a positive family history of EOAD and screened for mutations in the three genes, the mutation detection rate is 46–71%. This suggests that there exist yet unknown FAD genes. The functions of APP, PS1 and PS2 are largely unknown. Structural and experimental data suggest multifunctional roles for these proteins in signal transduction, protein processing, apoptosis and calcium homeostasis and as cell surface molecules. APP is a single-pass transmembrane (TM) protein. Structural modeling of the presenilins, highly homologous members of a protein family functionally related to signal peptide peptidases (SPP), predicts eight TM domains. *app* knockout mice (*app*^{-/-}) are fertile, display reduced synaptic plasticity and develop hippocampal gliosis accompanied by



Alzheimer's Disease. Figure 1 (a) Amino acid sequence of the C-terminal part of APP flanking the transmembrane (TM) domain (gray, underlined) and the major secretase cleavage sites β , α , γ_{40} , γ_{42} , and ϵ . The γ -secretase complex cleaves at variable sites between residues 39 and 50 downstream from the β -site. The major cleavage products $A\beta_{40}$ and $A\beta_{42}$ are usually secreted and may aggregate (see part b). The intracellular C-terminal γ -secretase cleavage product of APP, AICD (APP intracellular domain) may act as a transcription factor by interacting via a YENPTY motif (dark gray) with Fe65 and Tip60. Positions with known pathogenic mutations are labeled with an asterisk. (b) Alternative processing of the C-terminal part of APP via non-amyloidogenic α -secretase or amyloidogenic β -secretase pathways. α -Secretase cleaves APP between the β - and γ -sites thus liberating soluble neurotrophic APPs α . After cleavage of the remaining TM-spanning C-terminal fragment C 83 by γ -secretase, non-amyloidogenic peptide p3 and AICD are released. Candidates for α -secretase are TACE (tumor necrosis factor- α converting enzyme; synonym: ADAM17), ADAM9, and ADAM10. Members of the ADAM family share common features like a secretory signal peptide sequence, a disintegrin and a metalloproteinase domain. $A\beta_{42}$, the major amyloidogenic product of consecutive cleavages of APP by β - and γ -secretases, may be degraded by neprilysin (NEP) or insulin degrading enzyme (IDE), a neutral metalloproteinase. Non-degraded $A\beta$ may form submicroscopic oligomeric protofibrils with potentially neurotoxic features. Activation of microglia by these protofibrils may also mediate neurotoxic inflammatory processes. Early in $A\beta$ -associated neuronal degeneration there is an accumulation of neurofibrils (NF) consisting of abnormally phosphorylated tau protein. NFs may accelerate cell death. $A\beta$ protofibrils further aggregate and form the major component of senile plaques (SP). Dementia is associated with synaptic loss, axonal degeneration and cell death.

learning deficiencies. *psen1*^{-/-} embryos display lethal malformations of the skeletal system and CNS. *psen2*^{-/-} mice develop a minor pulmonary disorder and are fertile. *psen1*^{-/-} plus *psen2*^{-/-} embryos resemble the phenotype known from notch 1 deficiency. Presenilins are essential components of the γ -secretase complex. There is complete loss of γ -secretase activity in *psen1*^{-/-} *psen2*^{-/-} cell lines. The γ -secretase complex cleaves APP within its TM domain thus liberating $A\beta$. Mutations in APP cluster in the TM domain downstream of the major γ -secretase cleavage sites and close to the other secretase cleavage sites (Fig. 1a). Mutations

in *PSEN1* are distributed over a broad range of codons with some preference for residues located in TM domains and the first hydrophilic loop (HL1). Mutations in *PSEN2* are very rare and are associated with a high phenotype variability.

Trisomy 21, APOE $\epsilon 4$ and mutations in any of the three known FAD genes are all associated with an abnormally elevated ratio of $A\beta_{42}/A\beta_{40}$ in tissue or supernatant of primary or transgenic cell lines (for a more detailed review on the molecules involved in AD see ref. 8) (Table 3). Compared to $A\beta_{40}$, $A\beta_{42}$ is highly amyloidogenic. Thus, genetic data point to a central

Alzheimer's Disease. Table 3 Genes involved in AD-related pathology

Gene	Protein	Locus	disorder	risk factor	Presumed pathomechanism
<i>APOE</i>	Apo E	19q13.2	LOAD, sporadic AD	$\epsilon 4$ allele	unknown; $A\beta_{42}/A\beta_{40}$ \uparrow
<i>APP</i>	APP	21q21	Down Syndrome	surplus copy of <i>APP</i>	<i>APP</i> expression \uparrow ; $A\beta_{42}/A\beta_{40}$ \uparrow
			FAD	missense mutations	altered APP metabolism or processing; $A\beta_{42}/A\beta_{40}$ \uparrow
<i>PSEN1</i>	PS1	14q24.3	FAD		

role of APP processing in the pathogenesis of AD. Experimental data suggest presenilin is the catalytic subunit of the γ -secretase complex. To date it is not understood, how the high diversity of mutations in the presenilins and in *APP* is associated with a similarly altered or elevated γ -secretase activity causing the abnormally elevated ratio of $A\beta_{42}/A\beta_{40}$.

Cellular and Molecular Regulation

APP is processed alternatively by non-amyloidogenic α -secretase or amyloidogenic β -/ γ -secretase pathways. Candidate molecules for physiological α -secretase are TACE, ADAM9, and ADAM10 (Fig. 1). α -Secretase cleaves APP between the sites for β - and γ -secretase thus generating neurotrophic APPs α and preventing the generation of $A\beta$. The initial and rate limiting step of amyloidogenic APP processing is performed by β -secretase, the β -site APP cleaving enzyme (BACE), which cleaves APP at the amino terminus of the $A\beta$ domain. The cleaved ectodomain is secreted (APPs β). The remaining TM-spanning part of APP is further processed by the γ -secretase complex. Known components of this protease complex are presenilin, nicastrin, \blacktriangleright PEN-2, and \blacktriangleright APH-1 (Fig. 2). Experimental missense mutations in nicastrin are associated with an elevated production of $A\beta$, whereas the deletion of one of its domains abolished $A\beta$ generation. The membrane-associated γ -secretase cleavage of APP is called regulated intramembranous proteolysis (RIP). This cleavage is not site specific, i.e. it produces $A\beta$ peptides ranging in size from 39–49 residues with the major forms being $A\beta_{40}$ and $A\beta_{42}$. γ -Secretase activity is modulated either directly or indirectly by glycogen synthase kinase 3 (GSK3). $A\beta$ is secreted; whereas the C-terminal part of APP called APP intracellular domain (AICD) may have intracellular signaling functions and act as a transcription factor. AICD interacts with Fe65 and Tip60 *via* the YENPTY motif, which is located close to its C-terminus. Additional interaction partners of AICD with a yet unknown biological function are known.

$A\beta$, particularly $A\beta_{42}$ may form soluble protofibrils that aggregate to insoluble oligomers. Most probably these

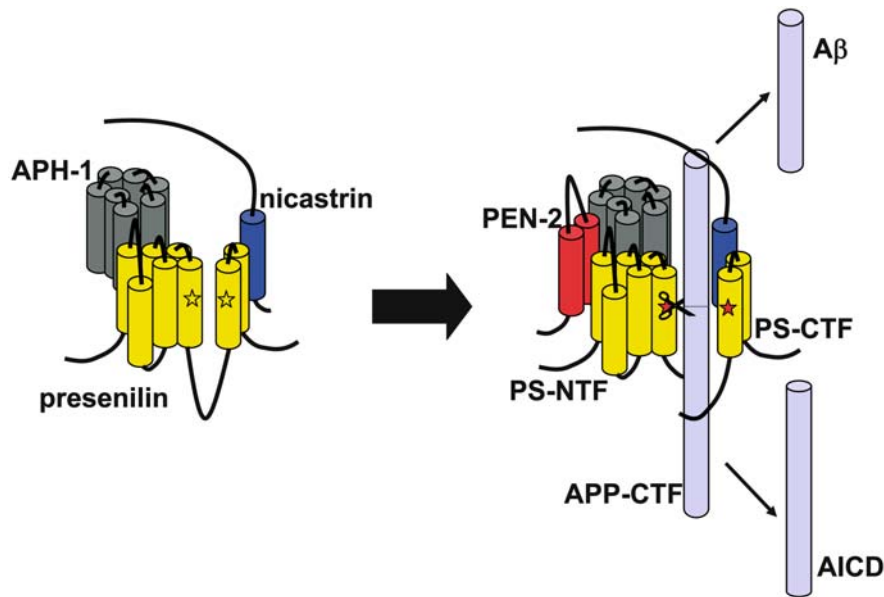
are submicroscopic molecular forms that mediate the neurotoxic effects of $A\beta$ associated with neurite dystrophy, synaptic loss and activation of astrocytes and microglia. SPs, one of the histopathological hallmarks of AD, have a diameter of 5–50 μ m and contain a dense core of fibrillary $A\beta$. Experimental data from transgenic mice suggest that $A\beta$ is involved in the induction of NFT formation in the neurons affected by loss of synaptic connections. This suggests that submicroscopic forms of $A\beta$ may induce neurodegeneration recognizable by NFTs before the formation of microscopically visible $A\beta$ plaques. NFTs are formed by abnormally phosphorylated paired helical filaments of tau, a microtubule-associated protein. Mutations in the gene for tau (*MAPT*) are known in autosomal dominant frontotemporal dementia linked to chromosome 17 (FTDP-17) but not in FAD.

Clinical Relevance

AD is a severe, progressing dementing disease leading in its final stages to complete helplessness and immobilization. The consequences of AD are devastating for the patients affected as well as for their families. Due to its very high and continuously increasing frequency, AD will be one of the major causes of death in the near future and one of the major economic and psychosocial burdens of general health care – if no significant progress is achieved in treatment or prevention. Within the next 50 years the incidence of AD will reach the level of that of cancer, the second most prevalent cause of death. Compared to non-demented age-matched subjects mortality in AD is more than doubled. The median survival time after first contact of a patient with a physician due to signs of dementia ranges from 3.1–3.5 years. Major causes of death in AD are secondary to immobilization or malnutrition and may depend on nursing quality. To date, there exists no curative treatment for AD.

Therapy

In AD, neuron loss leads to a cholinergic deficit, which is associated with dementia. Pharmacological inhibitors of acetylcholine esterase (\blacktriangleright ACH E) may be used to



Alzheimer's Disease. Figure 2 Molecular components and mode of activation of the γ -secretase complex. In the inactive complex, presenilin holoprotein is stabilized by APh-1 and nicastrin. Interaction with PEN-2 leads to cleavage of presenilin into its C- and N-terminal fragments (PS-CTF, PS-NTF) and activation of the γ -secretase complex. Transmembrane (TM) domains of APh-1, nicastrin, presenilin and PEN-2 are represented by cylindrical stretches. The locations of the catalytically essential aspartate residues of PS-NTF and PS-CTF are indicated by asterisks. The scheme suggests that the γ -secretase complex may cleave APP (APP-CTF) within its TM domain only after a preceding cleavage of APP by BACE (β -secretase) or an α -secretase (Fig. 1). If cleavage by BACE was the preceding step, γ -secretase may release amyloidogenic A β . The remaining fragment AICD (APP intracellular domain) may act as a transcription factor. Experimental data show a higher molecular weight of the γ -secretase complex than suggested by the scheme and indicate the presence of dimeric presenilin molecules in the complex, which has an overall unknown topology. Interestingly, none of the known pathogenic presenilin mutations is associated with a loss of γ -secretase activity. Functional or structural disturbance of presenilin dimers by presenilin missense mutations may be a mechanism underlying the dominant mode of action of these mutations.

reduce the physiological degradation of acetylcholine. In mild to moderate AD, progression of dementia may be delayed up to 1–2 years by second-generation inhibitors of **AChE** (donepezil, rivastigmine, galantamine) in a dose-dependent manner. 10–15% of the patients however, are non-responders. Neuron loss or malfunction may be related to tonic or excitotoxic NMDA receptor activation by its ligand, the neurotransmitter glutamate. Memantine, a moderate affinity NMDA receptor antagonist has been shown in clinical studies to slow down disease progression in moderate to severe AD. Possibly due to their different modes of action, i.e. cholinomimetic activity in AChE inhibitors and neuroprotection by memantine, the combined use of both drugs has been successful in a clinical trial. AChE inhibitors and memantine are both used for symptomatic treatment of AD and are of transient benefit. Other drugs of possible benefit in symptomatic treatment may be cholesterol lowering and nonsteroidal anti-inflammatory drugs. Intensive research activities are directed towards disease-modifying, curative or

preventive drugs. Promising strategies include active or passive immunization against A β or pharmacological inhibition of β - or γ -secretase. Preclinical and clinical studies are under way and there is hope that treatment of AD may be improved significantly in the near future.

Genetic Counseling

Molecular genetic diagnostics in suspected FAD requires careful genetic counseling. The detection of a fully penetrant FAD-causing mutation in a patient predicts a 50% recurrence risk in each of the patient's children and a 25% risk in grandchildren. This means that young adults (the children of FAD patients) may ask for presymptomatic testing, thereby possibly getting knowledge of a 50% risk in their children. Attitudes towards presymptomatic testing may change if there is prophylactic or preventive treatment available. Genotyping of *APOE* is not recommended in routine diagnostics, neither for patients nor presymptomatically since it does not allow an unambiguous

prognosis. Sometimes, *APOE* genotyping is considered for purposes of differential diagnostics in patients with dementia. The genotype of the patient may have an impact on the patient's relatives since e.g. all children of an *APOE* $\epsilon 4/\epsilon 4$ homozygous patient will be *APOE* $\epsilon 4$ positive. Therefore, prior to molecular genetic testing, genetic counseling including possible concerns of the patient's relatives is required.

► *In Vivo* Imaging of Transgenic Mice with Fluorescent Protein Expression

► Neurotrophic Factors

► Wnt/Beta-Catenin Signaling Pathway

References

1. McKhann G, Drachman D, Folstein M et al (1984) Clinical diagnosis of Alzheimer's disease: report of the NINCDS-ADRDA Work Group under the auspices of Department of Health and Human Services Task Force on Alzheimer's Disease. *Neurology* 34:939–944
2. Mirra SS, Heyman A, McKeel D et al (1991) The Consortium to Establish a Registry for Alzheimer's Disease (CERAD). Part II. Standardization of the neuropathologic assessment of Alzheimer's disease. *Neurology* 41: 479–486
3. Braak H, Braak E (1991) Neuropathological staging of Alzheimer-related changes. *Acta Neuropathol (Berl)* 82:239–259
4. Lobo A, Launer LJ, Fratiglioni L et al (2000) Prevalence of dementia and major subtypes in Europe: A collaborative study of population-based cohorts. *Neurologic Diseases in the Elderly Research Group. Neurology* 54:4–9
5. Fratiglioni L, Launer LJ, Andersen K et al (2000) Incidence of dementia and major subtypes in Europe: A collaborative study of population-based cohorts. *Neurologic Diseases in the Elderly Research Group. Neurology* 54:10–15
6. Farrer LA, Cupples LA, Haines JL et al (1997) Effects of age, sex, and ethnicity on the association between apolipoprotein E genotype and Alzheimer disease. A meta-analysis. *APOE and Alzheimer Disease Meta Analysis Consortium. Jama* 278:1349–1356
7. Lindsay J, Laurin D, Verreault R et al (2002) Risk factors for Alzheimer's disease: a prospective analysis from the Canadian Study of Health and Aging. *Am J Epidemiol* 156:445–453
8. Selkoe DJ (2001) Alzheimer's disease: genes, proteins, and therapy. *Physiol Rev* 81:741–766

Amadori Rearrangement

Definition

Amadori rearrangement is the isomerization of a double bond into a ketose derivative, by which a more stable structure is generated. Oxidative deamination of the ϵ -group of lysyl residues leads to the generation of an aldehyde group, which with amino groups of hydroxy-lysyl of an adjacent chain, can react to a Schiff-base.

AME

► Apparent Mineralocorticoid Excess

AMH / MIS

Definition

AMH/MIS stands for anti-müllerian hormone/müllerian inhibiting substance. It is a member of the TGF- β superfamily that regulates regression of the Müllerian duct in males.

► Receptor Serine/Threonine Kinase

AMHR-II

Definition

Anti-müllerian hormone Type II receptor.

► Receptor Serine/Threonine Kinase

Amide Bonds

Definition

An amide bond is a chemical bond formed between two molecules when the carboxyl group of one molecule

reacts with the amino group of the other molecule, releasing a water molecule (H_2O).

►PNA Chips

Amide-Linkage

Definition

Amide-linkage (N-acylation) characterises a stable amide-bond between the COOH -group of the fatty acid and $\alpha\text{-NH}_2$ -group, usually a glycine residue at the N-terminus of proteins. An amide-type linkage can be experimentally distinguished from a thioester-bond, by treatment of the acylated protein with hydroxylamine. This compound cleaves ester-linked fatty acids from the protein, whereas amide-linked fatty acids are not affected.

►Fatty Acid Acylation of Proteins

Amino Acids: Physicochemical Properties

EPHRAIM KATCHALSKI-KATZIR, RONI KASHER,
MATI FRIDKIN

Departments of Biological Chemistry and Organic
Chemistry, Weizmann Institute of Science, Rehovot,
Israel

ephraim.katzir@weizmann.ac.il

roni.kasher@weizmann.ac.il

mati.fridkin@weizmann.ac.il

Definition

Amino acids are the building blocks of all peptides and proteins and determine many of their important properties. This paper describes the physicochemical properties of amino acids and tries to shed light on their role as the constituents of living cells, the proteins.

α -Amino acids are molecules that possess an amino group ($-\text{NH}_2$) and a carboxylic group ($-\text{COOH}$) attached to the same tetrahedral carbon atom, called an α -carbon (Fig. 1a). The α -amino acids differ from one another by distinct R-groups (referred to as side chains) that are attached to the same α -carbon. Table 1 gives the structures and names of the 20 most common amino acids found in proteins. Those marked 'e' in Table 1 are the essential amino acids, which are not

available to animals by biosynthesis and must therefore be provided in the diet.

Characteristics

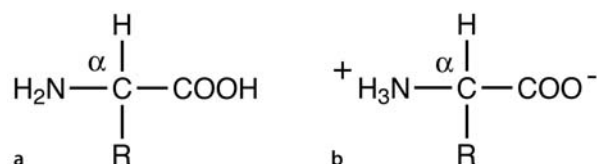
In aqueous solution at neutral pH the carboxylic group of amino acids is dissociated into a carboxylate and a proton ($-\text{COOH} \rightarrow -\text{COO}^- + \text{H}^+$) and the amino group, which functions as a base, is protonated ($-\text{NH}_2 + \text{H}^+ \rightarrow -\text{NH}_3^+$). Therefore, the predominant form of amino acids consists of a carboxylate anion ($-\text{COO}^-$) and an ammonium cation ($-\text{NH}_3^+$) and is termed the dipolar ion form or the zwitterionic form, as shown in Fig. 1b.

The physical properties of amino acids are consistent with their dipolar ionic structure: They are non-volatile crystalline solids, which melt or decompose at fairly high temperatures, in contrast to amines and carboxylic acids. Amino acids possess high dielectric constants and high dipole moment, reflecting the discrete separation of positive and negative charges in their dipolar ionic forms. The high dipole moment accounts for the fact that amino acids are insoluble in non-polar solvents such as benzene, petroleum-ether and ether, but are appreciably soluble in water.

Classification of Amino Acids

The side chain (R group) of amino acids determines their chemical and physical characteristics. There are two broad classes of amino acids, categorized according to the hydrophilic or hydrophobic nature of the side-chain. Hydrophobic chemical groups tend to be dissolved in non-polar organic solvents, whereas hydrophilic groups tend to be dissolved more readily in water.

Knowledge of the hydrophobic or hydrophilic character of the amino acid side chains may be used to identify the chemical type of a given protein or a specific region of a protein. The hydrophobic or hydrophilic character of a solute is referred to as ►hydrophathy. The hydrophobic character of a protein segment is defined by a sum of the hydrophobic contributions of all of its constituent amino acids. The relative polarity of each of the 20 common amino acids has been estimated experimentally by several methods, usually by measuring the free energy change induced



Amino Acids: Physicochemical Properties.

Figure 1 Chemical structure (a) and dipolar-ion form (b) of an amino acid. R represents the side chain group.

Amino Acids: Physicochemical Properties. Table 1 Natural amino acids

Name ^a	Three-letter abbreviation	One-letter symbol	Chemical structure (side chain group is colored red)
Alanine	Ala	A	$\begin{array}{c} \text{H}_2\text{N}-\text{CH}-\text{COOH} \\ \\ \text{CH}_3 \end{array}$
Arginine ^b	Arg	R	$\begin{array}{c} \text{H}_2\text{N}-\text{CH}-\text{COOH} \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{NH}_2 \\ \\ \text{C}=\text{NH}_2 \\ \\ \text{NH}_2 \end{array}$
Asparagine	Asn	N	$\begin{array}{c} \text{H}_2\text{N}-\text{CH}-\text{COOH} \\ \\ \text{CH}_2 \\ \\ \text{C}=\text{O} \\ \\ \text{NH}_2 \end{array}$
Aspartic acid	Asp	D	$\begin{array}{c} \text{H}_2\text{N}-\text{CH}-\text{COOH} \\ \\ \text{CH}_2 \\ \\ \text{C}=\text{O} \\ \\ \text{OH} \end{array}$
Cysteine	Cys	C	$\begin{array}{c} \text{H}_2\text{N}-\text{CH}-\text{COOH} \\ \\ \text{CH}_2 \\ \\ \text{SH} \end{array}$
Glutamic acid	Glu	E	$\begin{array}{c} \text{H}_2\text{N}-\text{CH}-\text{COOH} \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{C}=\text{O} \\ \\ \text{OH} \end{array}$
Glutamine	Gln	Q	$\begin{array}{c} \text{H}_2\text{N}-\text{CH}-\text{COOH} \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{C}=\text{O} \\ \\ \text{NH}_2 \end{array}$

Amino Acids: Physicochemical Properties. Table 1 Natural amino acids (Continued)

Name ^a	Three-letter abbreviation	One-letter symbol	Chemical structure (side chain group is colored red)
Glycine	Gly	G	$\begin{array}{c} \text{H}_2\text{N}-\text{CH}-\text{COOH} \\ \\ \text{H} \end{array}$
Histidine ^e	His	H	$\begin{array}{c} \text{H}_2\text{N}-\text{CH}-\text{COOH} \\ \\ \text{CH}_2 \\ \\ \text{Imidazole ring} \end{array}$
Isoleucine ^e	Ile	I	$\begin{array}{c} \text{H}_2\text{N}-\text{CH}-\text{COOH} \\ \\ \text{CH}-\text{CH}_3 \\ \\ \text{CH}_2 \\ \\ \text{CH}_3 \end{array}$
Leucine ^e	Leu	L	$\begin{array}{c} \text{H}_2\text{N}-\text{CH}-\text{COOH} \\ \\ \text{CH}_2 \\ \\ \text{CH}-\text{CH}_3 \\ \\ \text{CH}_3 \end{array}$
Lysine ^e	Lys	K	$\begin{array}{c} \text{H}_2\text{N}-\text{CH}-\text{COOH} \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{NH}_2 \end{array}$
Methionine ^e	Met	M	$\begin{array}{c} \text{H}_2\text{N}-\text{CH}-\text{COOH} \\ \\ \text{CH}_2 \\ \\ \text{CH}_2 \\ \\ \text{S} \\ \\ \text{CH}_3 \end{array}$
Phenylalanine ^e	Phe	F	$\begin{array}{c} \text{H}_2\text{N}-\text{CH}-\text{COOH} \\ \\ \text{CH}_2 \\ \\ \text{Benzene ring} \end{array}$

Amino Acids: Physicochemical Properties. Table 1 Natural amino acids (Continued)

Name ^a	Three-letter abbreviation	One-letter symbol	Chemical structure (side chain group is colored red)
Proline	Pro	P	
Serine	Ser	S	
Threonine ^e	Thr	T	
Tryptophan ^e	Trp	W	
Tyrosine	Tyr	Y	
Valine ^e	Val	V	

^a Essential amino acids are marked e^b Essential only in young growing animals

by moving a given amino acid residue from water into a hydrophobic solvent or into gas. Table 2 presents the hydropathy scale of the common amino acids. It was obtained by measuring the free energy change of model compounds transferred from water into gas. In these compounds the α -carbon of the corresponding amino acid has been replaced by hydrogen (1).

A more specific classification of amino acids takes into consideration the chemical nature of their side chains, as outlined in Fig. 2. Group 1 includes amino acids with aliphatic side-chains (Ile, Val, Leu, Ala, Gly), group 2, aromatic side-chains (Phe, Trp, Tyr), group 3, basic (positively-charged) side-chains (Lys, Arg, His), group 4, acidic (negatively-charged) side-chains or their

Amino Acids: Physicochemical Properties.**Table 2** Hydropathy scale of amino acids

Amino acid	Model compound	$\delta G^0(\text{H}_2\text{O} \rightarrow \text{gas})^a$ (kJ/mol)
Leu	Isobutane	-14
Ile	Butane	-13
Val	Propane	-12
Ala	Methane	-10
Phe	Toluene	-1
Met	Ethyl-methyl-sulfide	+2
Cys	Methanethiol	+3
Thr	Ethanol	+18
Ser	Methanol	+19
Trp	3-Methylindole	+20
Tyr	4-Cresol	+21
Lys	Butylamine	+34
Gln	Propionamide	+36
Asn	Acetamide	+38
Glu	Propionic acid	+38
His	4-Methylimidazole	+40
Asp	Acetic acid	+42
Arg	Methyl guanidine	+79

^a The standard free energy for transfer of model compounds for amino acids from water at 25°C and pH 7 to the gas phase (1)

corresponding amides (Asp, Glu, Asn, Gln) and group 5, aliphatic hydroxyl side chains (Ser, Thr). The sixth group consists of proline alone, since it has a secondary amine group, which affects the protein backbone conformation in a unique way and the seventh group consists of Cys and Met, which have sulfur-containing side chains.

A disulfide bond that cross-links between two Cys residues is found in many proteins, particularly extracellular proteins such as insulin, immunoglobulins and antibodies. A disulfide bond is created by oxidation of the thiol (-SH) groups of two Cys residues to form the oxidized -S-S- bond. Disulfide bonds have an important role in protein-structure stabilization.

Rare Amino acids and Non-protein Amino Acids

In addition to the 20 common amino acids, proteins may contain non-common residues formed as a result of post-translational modifications of the common

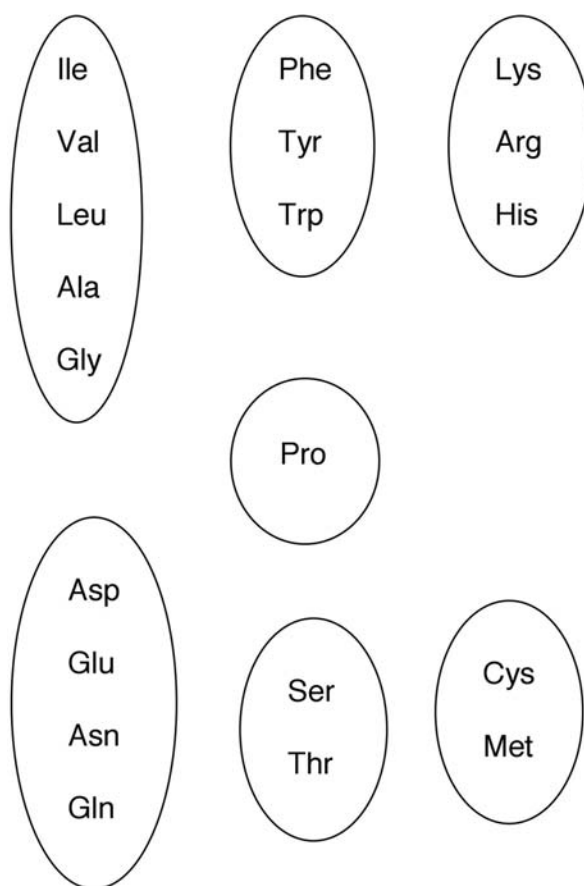
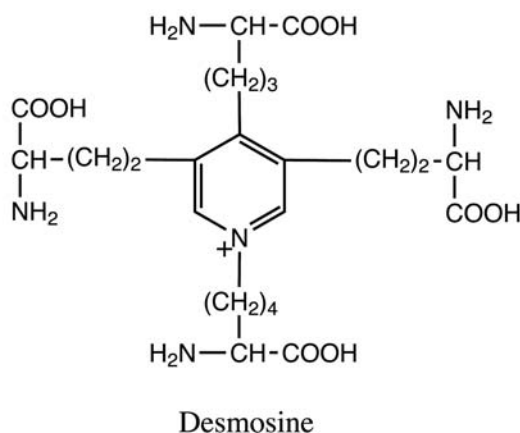
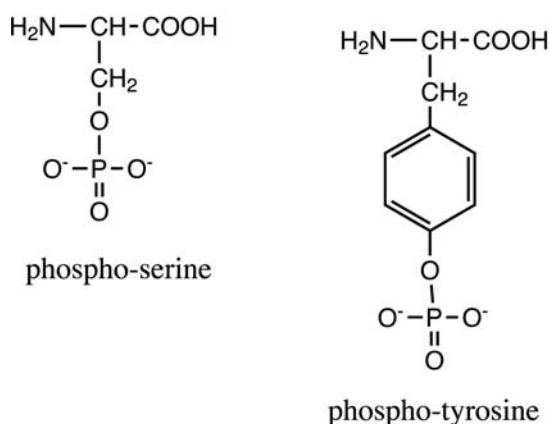
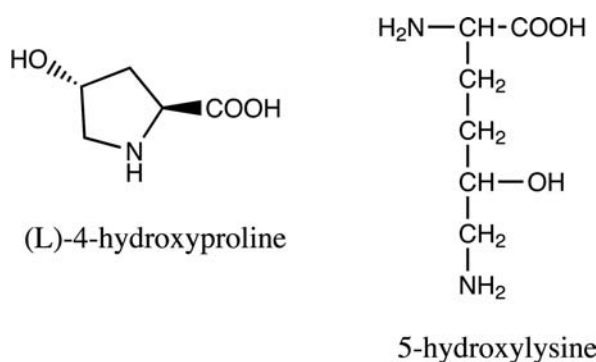
**Amino Acids: Physicochemical Properties.**

Figure 2 Classification of the 20 common amino acids found in proteins, according to the chemical nature of their side-chains.

residues in the protein. Examples are 4-hydroxyproline and 5-hydroxylysine (Fig. 3), which are hydroxy-derivatives of proline and lysine found in the fibrous protein collagen and in some plant proteins. The hydroxyl groups of Tyr and Ser residues of proteins can undergo phosphorylation (Fig. 3), particularly in signaling pathways and controlling metabolic cycles in the cell. Methylation, acetylation and glycosylation can also occur. Desmosine (Fig. 3), found in the fibrous protein elastin, is a derivative of four lysine residues that form a substituted pyridine ring.

In addition to the amino acids that make up proteins, some 300 special amino acids have been found, particularly in plant cells (2). Many of these are β -, γ - or δ -amino acids (Fig. 4), in contradistinction to the α -amino acids, which are the building blocks of proteins. Some non-protein amino acids occur in the (D)-configuration, such as (D)-glutamic acid, found in the cell walls of bacteria.



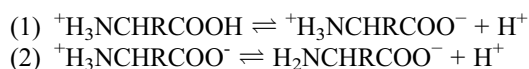
Amino Acids: Physicochemical Properties.

Figure 3 Examples of modified amino acids found in proteins.

Electrostatic Characteristics of Amino Acids

The electrostatic characteristics of the various amino acids are determined by the charge present on the α -amino, the α -carboxy or the ionizable side-chain of the amino acid. The overall charge is determined by the ionization constant of these ionizable groups, as well as by the pH of the aqueous medium in which the amino acids function.

Dissociation of α -ammonium and α -carboxylic groups of an amino acid is shown in equations (1, 2):



Changes in the pH of a solution of an amino acid are accompanied by changes in the electric charge pattern of the amino acid, because of different ionization states. When the pH of a solution is equivalent to the isoelectric point (pI) of an amino acid, the net charge of the amino acid is zero.

The acidity constant of an acid (K_a) is equal to the apparent equilibrium constant of the dissociation reaction of the acid (K'_{eq}). Hence, in the dissociation reaction described in equation (1), K_a is given by

$$K'_{\text{eq}} = K_a({}^\alpha\text{COOH}) = \frac{[\text{H}^+][{}^+\text{H}_3\text{NCHR}\text{COO}^-]}{[{}^+\text{H}_3\text{NCHR}\text{COOH}]}$$

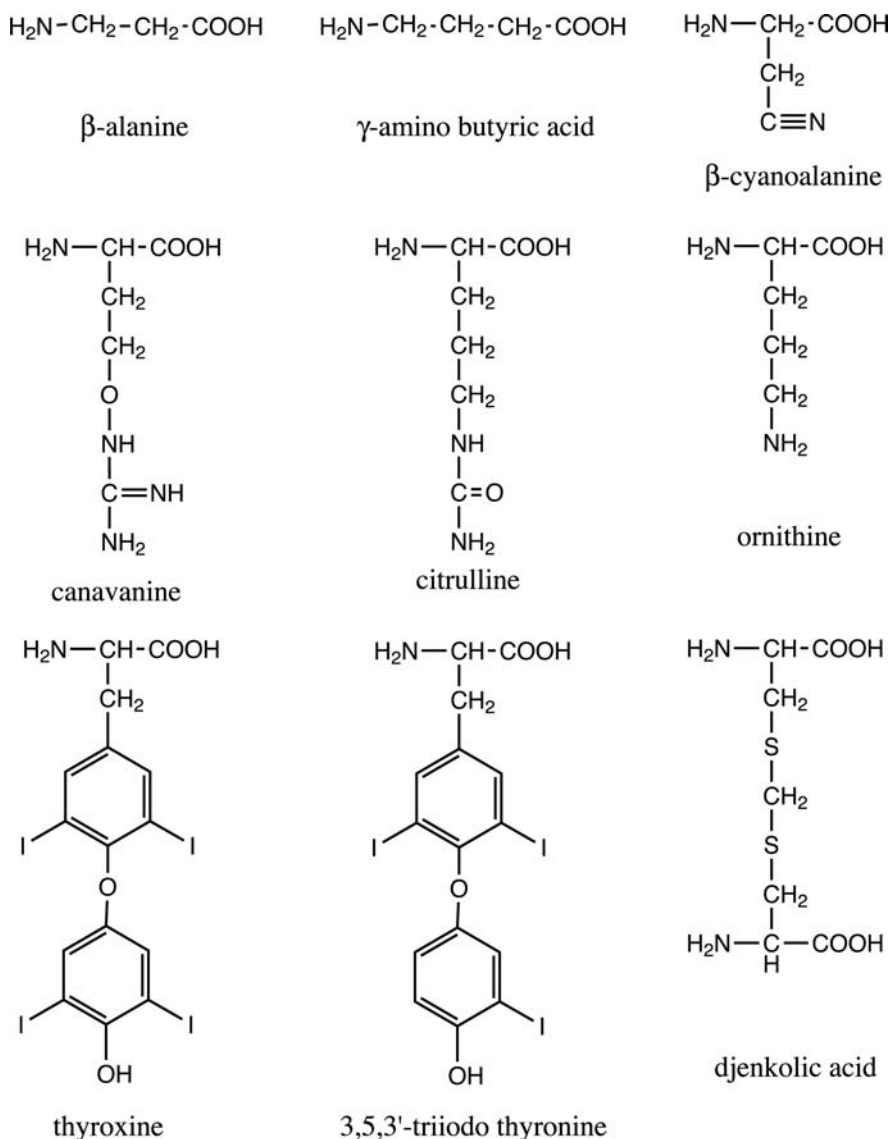
The larger the K_a the stronger the acid. $\text{p}K_a$ equals $-\log(K_a)$, thus the smaller the $\text{p}K_a$ the stronger the acid. Experimentally, it is rather easy to determine the $\text{p}K_a$ values of the various ionizable groups of amino acids by potentiometric titration of these groups with strong base or acid. The $\text{p}K_a$ value of the α -carboxyl group of amino acids is in the range of 1.8–2.4, and the $\text{p}K_a$ of the α -ammonium group is 9.0–11.0. Side chains of amino acids with ionizable groups have different acid strengths, as listed in Table 3.

Optical Characteristics and Stereochemistry of Amino Acids

With the exception of glycine, all amino acids that are obtained by acid or enzymatic hydrolysis of proteins show optical activity, i.e., they can rotate the plane of polarized light. Stereochemical studies of these naturally-occurring amino acids have shown that all have the same configuration about the α -carbon, which is the same as that of L-(–)-glyceraldehyde (Fig. 5). By convention, the two stereoisomers of glyceraldehyde are designated L and D. All stereoisomers of amino acids that are related to L-glyceraldehyde are designated L, whereas those related to D-glyceraldehyde are designated D, regardless of the direction of rotation of plane-polarized light produced by the isomers.

Amino acids with two **asymmetric carbon atoms** (for example isoleucine and threonine) have four stereoisomers. When more than one asymmetric carbon atom exists, it is the configuration about the α -carbon atom that is the basis for configuration assignment.

The **specific rotation**, $[\alpha]^{25}_D$ of some (L)- and (D)-amino acids in water (3) is given in Table 4. It is clear that at pH 7.0 some (L)-amino acids are dextrorotatory (i.e., their rotation of polarized light is positive) and some are levorotatory (negative rotation). Dextrorotatory compounds are designated (+), and levorotatory



Amino Acids: Physicochemical Properties. Figure 4 Examples of non-protein amino acids.

compounds by (-). According to the *RS* system of nomenclature, (rules for classifying the configuration as *R* or *S* (4)), most of the (L)-amino acids have an *S* configuration since the side chain nearly always has lower priority than COOH.

Although only (L)-amino acids are present in proteins, some (D)-amino acids are found in living cells in the cell walls of certain microorganisms and as part of peptide antibiotics such as gramicidin and actinomycin D.

Absorption Spectra of Amino Acids

None of the 20 amino acids found in proteins absorbs light in the visible range; however, amino acids with aromatic side chains absorb light in the ultraviolet

range, tryptophan ($\lambda_{\text{max}} = 279 \text{ nm}$ for the zwitterionic form in water; ►extinction coefficient (ϵ) of $5,580 \text{ M}^{-1} \text{ cm}^{-1}$), tyrosine ($\lambda_{\text{max}} = 275 \text{ nm}$, $\epsilon = 1,405$) and phenylalanine ($\lambda_{\text{max}} = 258 \text{ nm}$, $\epsilon = 195$) (5). Since most proteins contain tyrosine residues, measurement of absorption at 275 nm is a rapid means of estimating the protein content in a solution when the Tyr content is known. Phenylalanine generally contributes little to the absorbance properties of proteins. All amino acids absorb in the far UV range ($<220 \text{ nm}$) owing to their amide bond, peptide bond, and carboxylic groups. Tryptophan fluorescence is very easy to detect and is sensitive to solvent polarity. These properties make tryptophan a useful built-in probe for studying

polypeptide structure and interactions. In proteins, the emission maxima of fluorescence of Trp residues (λ_{em}) vary widely in the range of 308–350 nm, with an excitation wavelength (λ_{ex}) of 279 nm (5).

Chemical Reactions for Amino Acid Identification

The ninhydrin reaction with the α -amino group of amino acids (Fig. 6) is widely used to estimate amino acids quantitatively in very small amounts. Heating an amino acid with two equivalents of ninhydrin yields a blue-colored compound, which is very easy to detect by light absorption.

Dansyl chloride, dabsyl chloride and 1-fluoro-2,4-dinitrobenzene (Fig. 6) react with the amino group of amino acids, to yield derivatives that are stable to hydrolysis. The non-fluorescent compound fluorescamine (Fig. 6) readily reacts with primary amine groups of amino acids, yielding a strongly fluorescent derivative that permits detection of nanogram quantities of amino acids.

Amino Acids: Physicochemical Properties.

Table 3 pK_a values^a of side chains of some amino acids

Amino acid	pK_a value of the side chain (25°C)	Chemical group undergoing ionization
Aspartic acid	3.9	$^{\gamma}\text{COOH}$
Glutamic acid	4.3	$^{\delta}\text{COOH}$
Histidine	6.0	imidazole-NH
Cysteine	8.3	$^{\beta}\text{SH}$
Tyrosine	10.9	aromatic-OH
Lysine	10.8	$^{\epsilon}\text{NH}_3^+$
Arginine	12.5	guanidine-NH ₂

^a pK_a values depend on temperature and ionic strength of the solution (9)

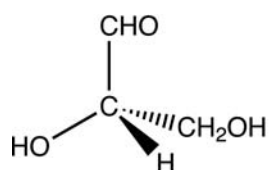
Analysis of Amino Acid Content and Determination of Amino Acid Sequence of Polypeptides and Proteins

The amino acid content of a given protein is determined after acid hydrolysis and chromatographic separation by **ion-exchange chromatography** of the resulting amino acid mixture. The amino-acid sequence of a protein is usually determined by the Sanger or the Edman degradation techniques. In Sanger's procedure, 1-fluoro-2,4-dinitrobenzene is used to label the N-terminal amino acid residue by coupling it to the α -amino group and consequently hydrolyzing the whole protein. The N-terminal modified amino acid is stable to acid hydrolysis and can be identified (Fig. 7). In the Edman degradation procedure the amino terminal residue is labeled and removed from a

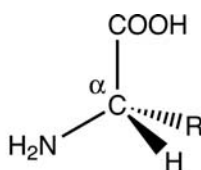
Amino Acids: Physicochemical Properties.

Table 4 Specific rotation, $[\alpha]^{25^\circ}_D$ of some (L)- and (D)-amino acids in water at pH 7 (2)

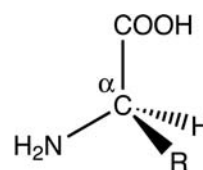
Amino acid	$[\alpha]^{25^\circ}_D$
(L)-Ala	+1.8
(D)-Ala	-1.8
(L)-Asp	+5.0
(L)-Glu	+12.0
(D)-Glu	-12.0
(L)-Ile	+12.4
(L)-Leu	-11.0
(L)-Lys	+13.5
(D)-Lys	-13.5
(L)-Phe	-34.5
(L)-Pro	-86.2
(L)-Ser	-7.5
(L)-Trp	-33.7



L-Glyceraldehyde

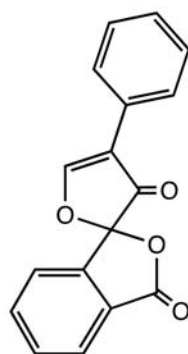
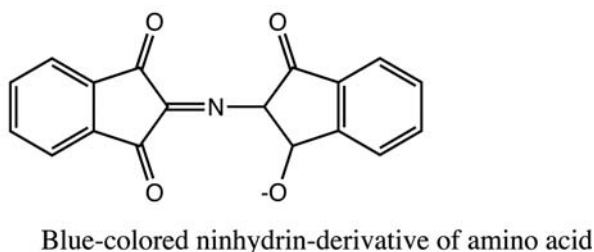
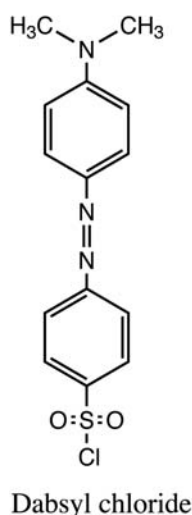
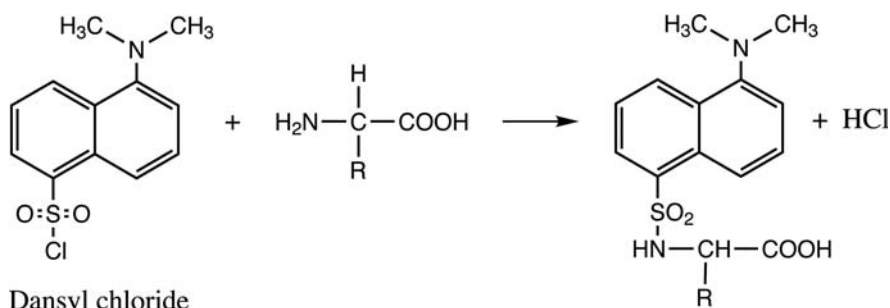


L-Amino acid



D-Amino acid

Amino Acids: Physicochemical Properties. Figure 5 Absolute configuration of (L)-amino acid, D-amino acid and (L)-glyceraldehyde. The extended heavy line represents a bond oriented towards the viewer outside the plane of the page. The elongated dashed triangle represents a covalent bond directed beyond the page.



Amino Acids: Physicochemical Properties. Figure 6 Reagents widely used to assist in the quantitative detection of amino acids.

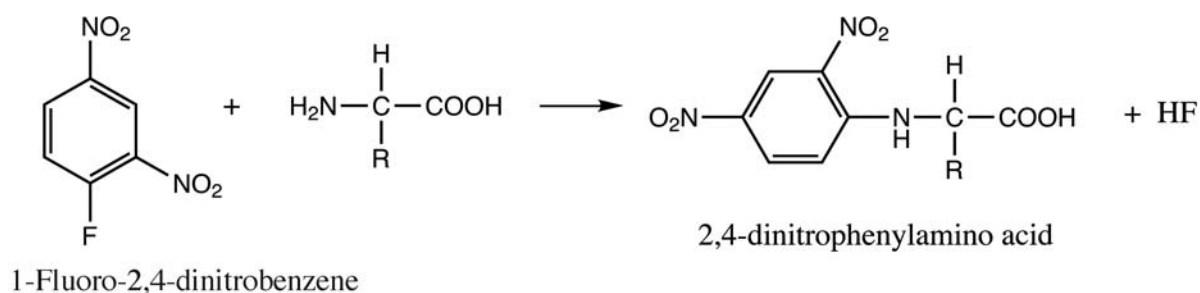
polypeptide. The protein or polypeptide containing a free α -amino group is reacted with phenyl isothiocyanate, which removes the amino terminal residue as a phenyl thiocarbonyl derivative, which is identified, while leaving all other peptide bonds intact (Fig. 7). The remaining shortened polypeptide or protein is then reacted again with phenyl isothiocyanate and the next amino acid is identified. This procedure is repeated until the entire sequence is determined. Using present day automatic sequence machines, the amino acid sequence can be determined with less than 1 μ g protein.

Peptides, Structure and Conformation

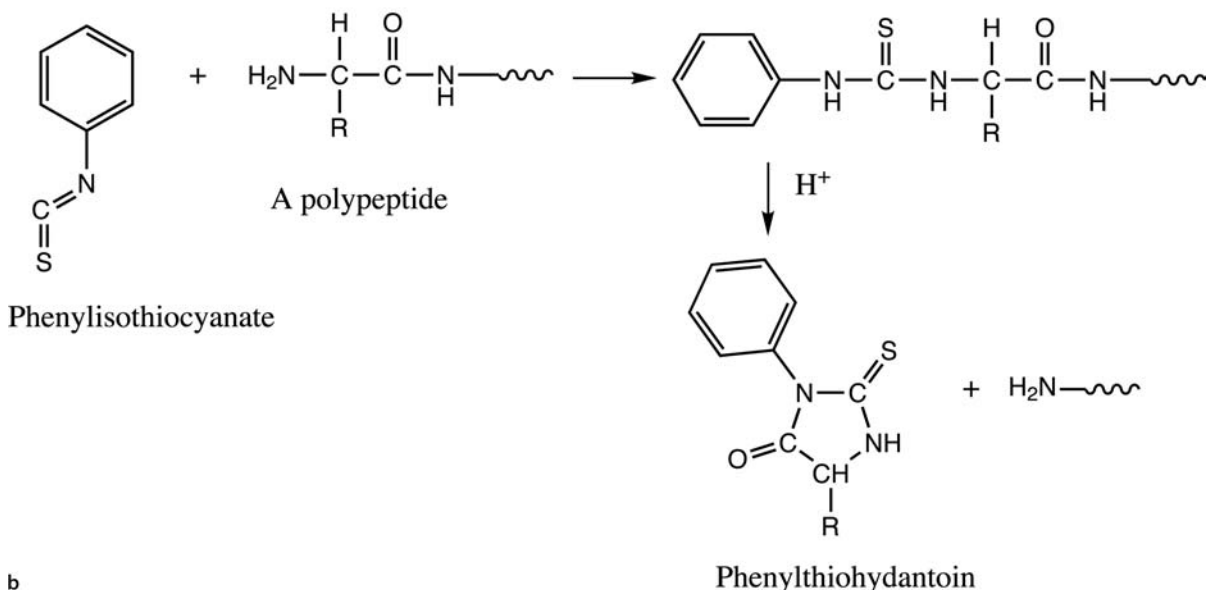
Amino acids in proteins are linked covalently by peptide bonds. The peptide bond is the amide bond

formed by linkage of the carboxyl group of the first amino acid with the amino group of the second amino acid, while releasing water (Fig. 8). When many amino acids are joined by peptide bonds, they form a linear peptide chain.

By convention, the amino end is taken to be the beginning of the polypeptide chain, so that when the sequence of amino acids in a polypeptide is presented, i.e. its primary structure, it starts with the amino terminal residue. Thus in the tripeptide Ala-Gly-Trp, Ala is the amino terminal residue and Trp is the carboxyl terminal amino acid. The number of amino acids in the peptide chain determines its name. Thus the term “di-peptide” indicates that the peptide consists of two amino acid residues.

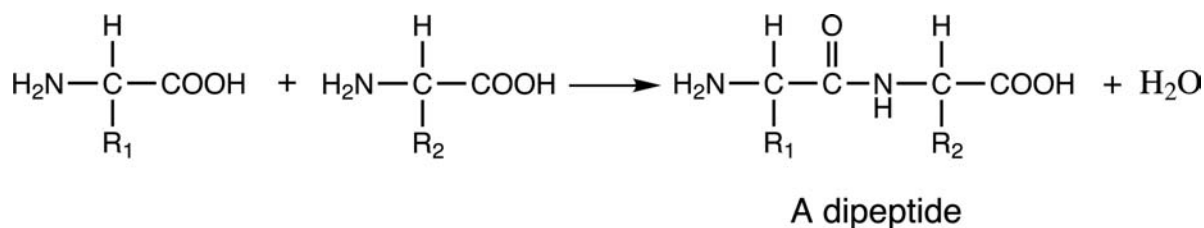


a



b

Amino Acids: Physicochemical Properties. Figure 7 Reactions used in the detection of the amino acid sequence in peptides and proteins: (a) Sanger procedure and (b) Edman degradation.

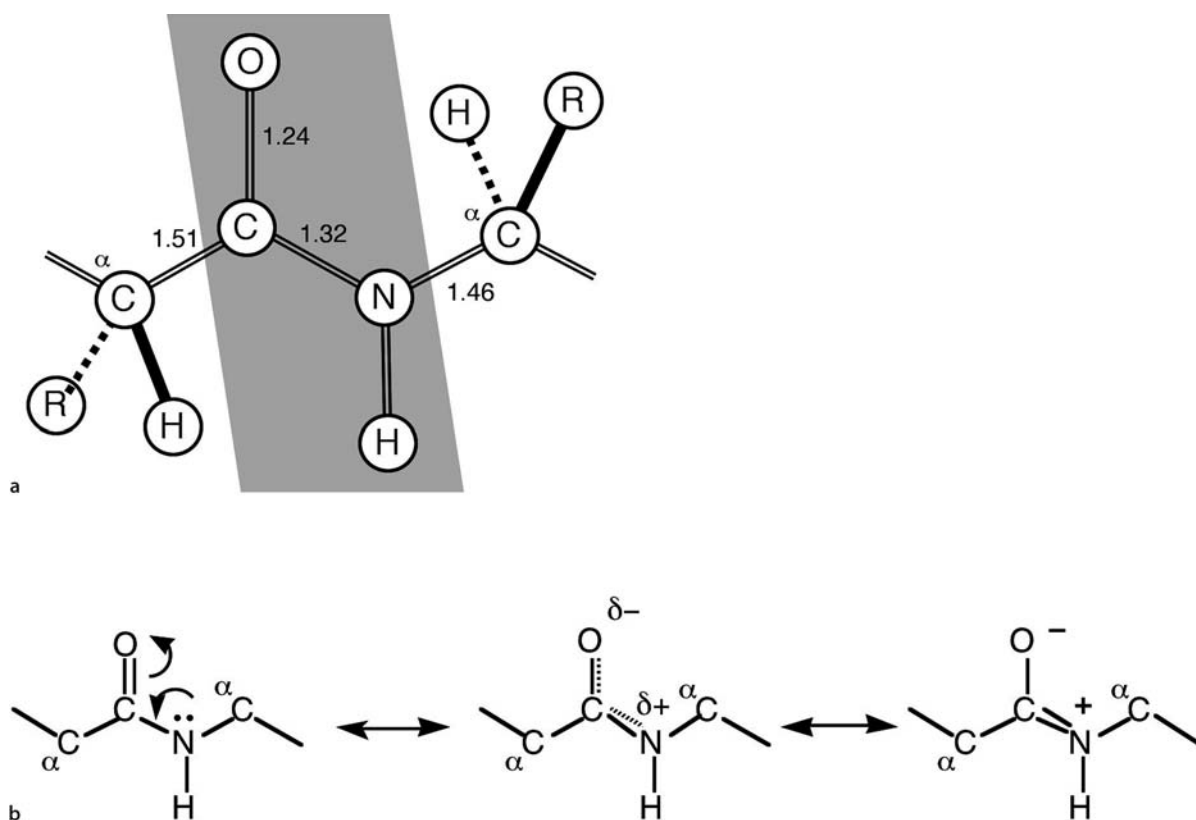


Amino Acids: Physicochemical Properties. Figure 8 Formation of the peptide bond between two amino acids, to form a dipeptide.

The Peptide Bond is Rigid and Planar

In the late 1930s, Linus Pauling and Robert Corey began x-ray crystallographic studies of the precise structure of amino acids and peptides. One of their important findings was that the peptide bond is rigid and planar (6). The hydrogen of the α -amino group is nearly always trans (opposite) to the oxygen of the

carbonyl group (Fig. 9). There is no freedom of rotation about the bond between the carbonyl carbon and the nitrogen atom of the peptide unit, because this link represents a characteristic partial double bond. In contrast, $^{\circ}\text{C}-\text{NH}$ and $^{\circ}\text{C}-\text{CO}$ are pure single bonds and possess a large degree of rotational freedom on either side of the peptide unit.



Amino Acids: Physicochemical Properties. Figure 9 Planar and rigid characters of the peptide group CO-NH. (a) A gray parallelogram marks the plane of the peptide group in a polypeptide. Standard bond distances are marked in Å. Note that the C-N bond in the peptide group is shorter than the α C-N bond, indicating a partial double-bond character of the peptide bond. (b) The electronic resonance of the peptide bond is presented.

The α -helix and β -pleated Sheet – Secondary Structures

Pauling and Corey proposed the periodic polypeptide structures, known as the α -helix and the β -pleated sheet. The α -helix is a rod-like structure. The tightly coiled polypeptide main chain forms the inner part of the rod and the side chains extend outward in a helical array (Fig. 10). The structure of the α -helix is stabilized mainly by intra-molecular **hydrogen bonds** (Fig. 10). The α -helices found in proteins are right handed. The number of amino acid residues per turn is 3.6 and the pitch of the helix is 5.4 Å.

Ten or more α -helices can entwine to form a cable. Such helical coiled-coils are found in **keratin** in hair, in **myosin** and tropomyosin in muscle, in epidermin in skin and in fibrin in blood clots. The helical cables in these proteins form stiff bundles of fibers serving a mechanical role in structures that need support.

Another structural element in proteins is the β -pleated sheet. A polypeptide chain in a β -pleated sheet is almost fully extended (Fig. 11).

Adjacent chains in a β -pleated sheet can run in the same direction (parallel β -sheet) or in opposite directions

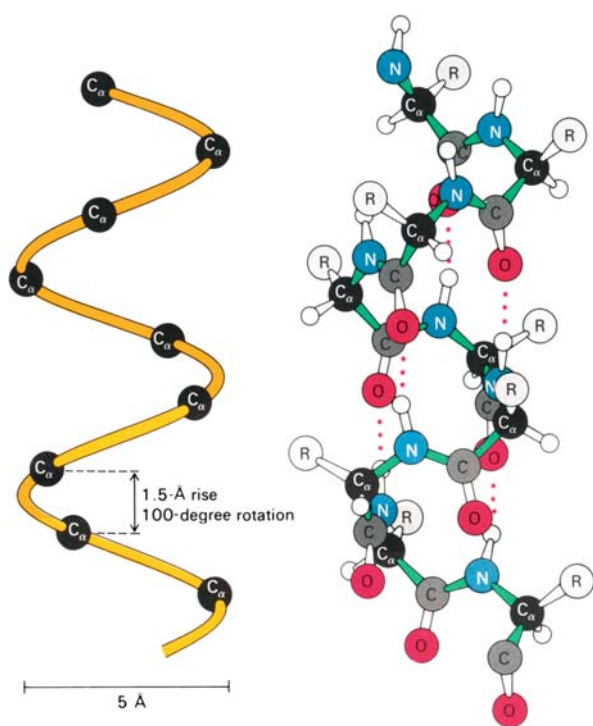
(anti-parallel β -sheet). Silk fibroin, for example, consists almost entirely of stacks of the anti-parallel β -sheets.

β -turns

Most proteins have compact, globular shapes due to numerous reversals of the direction of their polypeptide chains. Many of these chain reversals are accomplished by a common structural element called the β -turn.

Methods Used to Determine the Tertiary Structure of Proteins

The geometry of secondary structures of polypeptides (α -helix, β -sheet) was first determined experimentally by X-ray diffraction analysis. Complete three-dimensional structures of proteins are determined by X-ray diffraction analysis of a single crystal and by **nuclear magnetic resonance** (NMR) analysis. To date, the three-dimensional structure of approximately 25,000 proteins had been elucidated by these techniques.



Amino Acids: Physicochemical Properties.

Figure 10 Models of a right-handed α -helix showing (left) only the α -carbon atoms and (right) the entire helix. Oxygen atoms of the carbonyl group are presented in red, nitrogen atoms are in blue and $^{\circ}\text{C}$ atoms are in black. The side-chain R groups are presented in colorless circles. The hydrogen bonds are shown as red dotted lines, combining the carbonyl oxygen with the hydrogen of the NH group. Both atoms are constituents of the peptide bond.

Poly- α -amino Acids – the Simplest of Protein Models

During the 1950s, linear polymers of amino acids, termed poly- α -amino acids, were synthesized by polymerization of the corresponding N-carboxy-amino acid anhydrides in solution in the presence of suitable catalysts. Linear poly- α -amino acids such as poly-Gly, poly-(L)-Ala, poly-(L)-Phe and poly-(L)-Glu, represent a rather sharp distribution of analogous homologous peptides of different chain lengths (7).

Under appropriate conditions, linear poly- α -amino acids, in the solid state and in solution, were found to acquire conformations of α -helix and β -parallel and anti-parallel pleated sheets or to exist as random coils (see therapeutic uses of poly-amino acids in the section on Clinical Relevance).

Knowledge of the physical and chemical properties of poly- α -amino acids played a major role in the work that led in 1961 to the cracking of the **genetic code** (see next section).

Protein Denaturation and Folding

Moderate changes in a protein's environment can bring about structural changes that affect its function. A change in its three-dimensional structure sufficient to cause loss of function is called denaturation.

Most proteins can be denatured by heat. Heat-stable proteins of thermophilic bacteria have evolved to function at the temperature of hot springs ($\sim 100^{\circ}\text{C}$). Proteins can also be denatured by extremes of pH, by certain miscible organic solvents such as alcohol or acetone, by certain solutes such as urea and guanidine hydrochloride or by detergents.

The Genetic Code

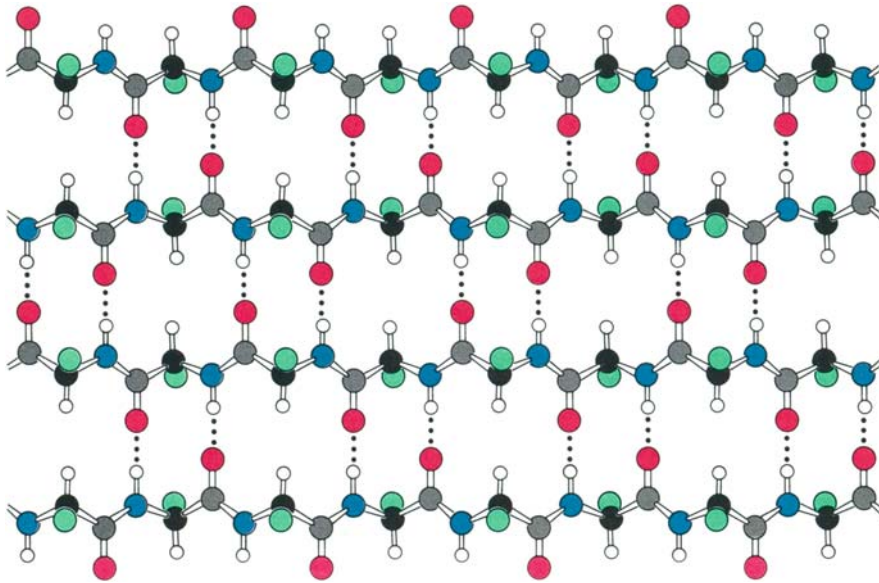
Polypeptide chains in living cells are synthesized by ribosomes, which handle and control peptide bond formation and the amino acid sequence in the peptide by a sophisticated enzymatic translation procedure. The amino acids appear in the form of their corresponding amino acyl-tRNA esters, whereas the amino acid sequence within the peptide is determined by messenger RNA. Messenger RNA represents a complementary nucleotide sequence to that of a DNA fragment bearing the gene to be expressed. Based on theoretical investigations as well as on a great variety of experimental results, the genetic code could be deduced. Each of the 20 protein amino acids is coded by a corresponding set of three nucleotides arranged in a specific order. The resulting genetic three-letter code dictionary is represented in Fig. 12.

The genetic code is a degenerative one, since all of the amino acids except tryptophan and methionine have more than one set of three nucleotides (codon). Moreover, it seems that the third nucleotide of each codon is less specific than the first two.

Finally, one should mention a conspicuous feature of the code, i.e. three of the 64 triplets (UAG, UAA, UGA) do not code for any known amino acid. These triplets are signals of the termination codon of the polypeptide chains formed.

Clinical Relevance

As was described above, some amino acids (marked 'e' in Table 1) are the essential amino acids, which are not available to animals by biosynthesis and must therefore be provided in the diet. Most amino acids in cells are in proteins, however, amino acids have additional functions. Ornithine and citrulline (Fig. 4) are key intermediates in the biosynthesis of arginine and in the urea cycle. β -Alanine (Fig. 4) is the precursor of the vitamin pantothenic acid and γ -amino-butyric acid is a transmitter of nerve pulses. The well-known thyroid hormones, thyroxine and 3,5,3'-triiodothyronine (Fig. 4) are derivatives of diiodo-tyrosine. Fungi and higher plants contain an extraordinary variety of amino acids. Some of them, such as canavanine, djenkolic



Amino Acids: Physicochemical Properties. Figure 11 Model of a polypeptide arranged in a β -pleated sheet structure. The side chains (shown in green) are above and below the plane of the sheet. All other atoms are represented in the same color as in Fig. 10.

		Second position									
		U		C		A		G			
First position of the codon (5' end)	U	UU U	Phe	UC U	Ser	UA U	Tyr	UG U	Cys	Third position (3' end)	
		UU C	Phe	UC C	Ser	UA C	Tyr	UG C	Cys		
		UU A	Leu	UC A	Ser	UA A	End	UG A	End		
		UU G	Leu	UC G	Ser	UA G	End	UG G	Trp		
	C	CU U	Leu	CC U	Pro	CA U	His	CG U	Arg		
		CU C	Leu	CC C	Pro	CA C	His	CG C	Arg		
		CU A	Leu	CC A	Pro	CA A	Gln	CG A	Arg		
		CU G	Leu	CC G	Pro	CA G	Gln	CG G	Arg		
	A	AU U	Ile	AC U	Thr	AA U	Asn	AG U	Ser		
		AU C	Ile	AC C	Thr	AA C	Asn	AG C	Ser		
		AU A	Ile	AC A	Thr	AA A	Lys	AG A	Arg		
		AU G	Met	AC G	Thr	AA G	Lys	AG G	Arg		
	G	GU U	Val	GC U	Ala	GA U	Asp	GG U	Gly		
		GU C	Val	GC C	Ala	GA C	Asp	GG C	Gly		
		GU A	Val	GC A	Ala	GA A	Glu	GG A	Gly		
		GU G	Val	GC G	Ala	GA G	Glu	GG G	Gly		

Amino Acids: Physicochemical Properties. Figure 12 The genetic code. The third nucleotide of each codon is less specific than the first two. The codons are read in the 5' \rightarrow 3' direction.

acid and β -cyanoalanine, are toxic to other forms of life.

The nature of amino acids, as well as their content and sequence in proteins, is of paramount importance in determining protein characteristics. (a) Knowledge of the amino acid sequence makes it possible to identify the protein; (b) it also enables the elucidation of the mechanisms involved in various specific activities of proteins such as enzyme catalysis, antigen binding by antibodies and ligand binding by receptors and (c) it enables detection of physiological disorders. Amino acid sequence determination of proteins represents a branch of molecular pathology, an emerging area of medicine. Alterations in amino acid sequence can produce abnormal function and disease. Fatal diseases such as sickle-cell anemia can result from a change in a single amino acid in a protein.

Four inherited neurodegenerative diseases are linked to the abnormally extended repeats of glutamine residues near the N-termini of affected proteins (see the essay '►Repeat Expansion Diseases', in particular the section dealing with polyglutamine disorders). These diseases include ►Huntington's disease (HD), spinal and bulbar muscular atrophy (SBMA), also known as ►Kennedy's disease, spinocerebral ataxia type I (SCAI), and dentatorubral-pallidoluysian atrophy (DRPLA) (8). All four diseases become more severe the longer are the glutamine repeats.

►Multiple sclerosis is a chronic inflammatory disease of the central nervous system in which infiltrating lymphocytes, predominantly T-cells and macrophages, cause damage to the myelin sheath. Sela and Amon discovered that Cop-I, a basic copolymer of (L)-alanine, (L)-glutamic acid, (L)-lysine and (L)-tyrosine at residue molar ratios of 6.0, 1.9, 4.7 and 1.0 respectively, prevents myelin destruction in humans. As a matter of fact, Cop-I is widely used to reduce the number of crises that usually affect multiple sclerosis patients.

There is evidence that a denatured conformational structure of the prion protein (Prion^{sc}) causes Alzheimer's disease and related diseases in humans, as well as in cattle and other animals (8).

►Proteases and Inhibitors

►3D Structure Determination by X-Ray

References

- Kyte J (1995) Structure in protein chemistry. Garland, New York, pp 189–193
- Nelson DL, Cox MM, Lehninger AL (2000) Lehninger principles of biochemistry, 3rd edn. Worth, New York
- Stryer L, Tasker R, Rhodes C (1995) Biochemistry, 4th edn. Freeman, New York
- Solomons G, Fryhle C (2000) Organic chemistry, 7th edn. John Wiley & sons, New York, p 195
- Guilbault GG (1990) Practical fluorescence, Revised and expanded, 2nd edn. Marcell Dekker, New York, pp 618–628
- Pauling L, Corey RB, Branson HR (1951) The structure of proteins: Two hydrogen-bonded helical configurations of the polypeptide chain. *Proc Nat Acad Sci USA* 37:205–211
- Katchalski-Katzir E (1996) Synthesis, physicochemical and biological properties of poly- α -amino acids – the simplest protein models. *Acta Biochimica Polonica* 43:217–226
- Harris D (2004) Mad cow disease and related spongiform encephalopathies. Springer, Berlin Heidelberg New York
- Edsall JT, Wyman J (1958) Biophysical Chemistry. Academic Press, New York

AML

►Acute Myeloid Leukaemia

AMN

►Adrenomyeloneuropathy

AMP

Definition

Adenosine 5'-monophosphate.

►DNA Ligases

Amphetamine

Definition

Amphetamines are a class of drugs that cause increased activity, decreased appetite, and a general sense of well-being. After an initial "rush," the state of high agitation can lead to violent behavior. From its parent drug, amphetamine, methamphetamine ("speed," "meth," and "chalk"; "ice," "crystal," "crank," and "glass") was developed, which has a similar chemical structure but has more pronounced effects on the central nervous system. Methamphetamine is a powerful addictive stimulant. The effects can last 6 to 8 hours.

►Addiction, Molecular Biology

Amphipathic

Definition

Amphipathic designates a molecule having both hydrophilic and hydrophobic regions.

► [Leucine Zipper Transcription Factors: bZIP Proteins](#)

mediating specific signs of fear and anxiety. It is a part of the limbic nervous system at a strategic location between inputs from the senses and memory. The amygdala represents a target for treating anxiety disorders, post-traumatic stress disorder (PTSD), panic disorder, obsessive/compulsive disorder and phobias.

► [Addiction, Molecular Biology](#)

► [Brain](#)

Amplification

Definition

The term amplification is used in quite different senses, including even intramolecular magnification of a small movement in a protein to a larger movement at the surface, e.g. near the active center through solid body movement of a helix. It is usually reserved for when the single activated protein (e.g. the receptor) induces the generation of many copies of an interacting downstream component. Examples of such “particle” amplification include the generation of many copies of active G-protein by GDP/GTP exchange catalysed by a single receptor, or the enzymatic hydrolysis of cyclic GMP generating many molecules of 5'GMP by an effector. Note that the driving force in both cases is the free energy difference between GTP and GDP.

► [Photoreceptors](#)

Amyloid

Definition

Amyloid is a generic term that describes fibrillar aggregates that have a common structural motif: the β -pleated sheet conformation. These aggregates exhibit specific tinctorial properties, including the ability to emit a green birefringent glow after staining with Congo red, and the capacity to bind the fluorochrome thioflavin S. There are more than a dozen human diseases of different etiology characterized by the extracellular deposition of amyloid. The amyloid fibrils are usually composed of proteolytic fragments of normal or mutant gene products. So far there are over 16 different proteins involved in amyloid deposition in distinct tissues.

► [Alzheimer's Disease](#)

► [Beta-Amyloid](#)

► [Defective Protein Folding Disorders](#)

Amplified Fragment Length Polymorphic Markers

Definition

Amplified fragment length polymorphic markers (ALFPs) that can be amplified in a PCR, which show length polymorphism in gels.

► [Mutagenesis Approaches in Medeka](#)

Amyloidosis

Definition

Amyloidosis is a metabolic disease associated with amyloid fibril deposition in organs such as heart, kidney and peripheral nerves. There are known to be 17 amyloidogenic proteins. Amyloid can be stained with Congo red. Familial amyloidosis is caused by accumulation of mutant proteins such as transthyretin, apoA-I gelsolin, lysozyme or fibrinogen.

► [High-HDL Syndrome](#)

► [Defective Protein Folding Disorders](#)

Amygdala

Definition

The amygdala is a brain region that contains brain circuits related to emotions. It is intimately connected to the hypothalamus, brainstem, and other brain areas

Amyotrophy

► [Muscle Atrophy](#)

Analytical Ultracentrifugation

PETER SCHUCK

Division of Bioengineering and Physical Science, OD,
National Institutes of Health, Bethesda, USA
pschuck@helix.nih.gov

Definition

Analytical ultracentrifugation is a technique based on the application of a centrifugal force to a solution and the real-time observation of the resulting spatial redistribution of the dissolved molecules or particles. This permits the study of the thermodynamic and hydrodynamic properties of purified macromolecules, such as nucleic acids and proteins, in dilute solution and the characterization of protein sizes and shapes, as well as ►protein interactions.

Description

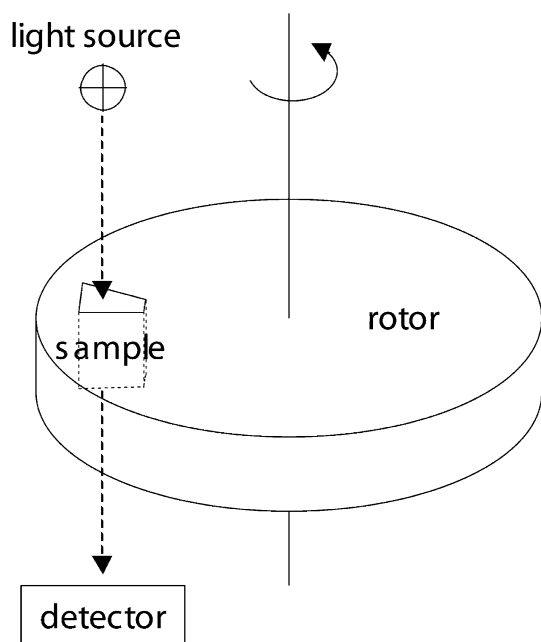
The most frequent application of modern analytical ultracentrifugation is the analysis of protein interactions, such as the characterization of protein oligomeric states and the study of reversible protein complexes, protein-ligand interactions and protein-nucleic acid interactions. This includes ligand-induced conformational changes in proteins, the size-distribution of protein complexes, protein filamentation and the study of detergent-solubilized membrane proteins. It is also applied in a wide variety of other areas, ranging from the study of the solution conformation of macromolecules at low resolution to the analysis of the nucleotide composition of genomes in evolutionary studies and the analysis of protein interactions in crowded media. Examples of current biotechnological applications are the size-distribution and aggregation behavior of protein pharmaceuticals and the study of small molecule binding to protein targets. It is also an important tool in several areas of chemistry.

Analytical ultracentrifugation is a classical method of physical biochemistry, developed in the 1920s by Svedberg (1, 2). Historically, for several decades it has been used as a standard tool to measure the molar mass of purified proteins, protein complexes and nucleic acids in dilute solution. For example, it provided the first demonstration that proteins are macromolecules by measuring the molar mass of hemoglobin and it was a central tool in the discovery of semiconservative replication by Meselson and Stahl. New instrumentation in the 1990s combined with computational advances greatly facilitated and extended the application to protein interactions.

While the basic idea of analytical ultracentrifugation is simple and it provides macromolecular information on biochemical systems rooted in first principles, a wealth of important knowledge has been accumulated regarding the technical implementation of ultracentrifuges, the theoretical foundation of ultracentrifugation in thermodynamics and physical chemistry of macromolecules and the mathematical analysis of ultracentrifugation experiments (3). Many variations of ultracentrifugal approaches have been described in the literature and the following is a short presentation of only the main concepts of the most commonly used configurations.

The principal difference between an analytical ultracentrifuge and a conventional preparative ultracentrifuge is an optical system that permits the observation of the dissolved macromolecules under study during the centrifugation and the real-time measurement of the concentration gradients which evolve in a radial direction as a response to the applied gravitational force. To allow for the optical detection, the sample is contained in an assembly that is optically transparent perpendicular to the plane of rotation. The optical systems are triggered by the revolution of the rotor, so that data are acquired only during the short intervals when the sample assembly is aligned with the optical light path (Fig. 1). The two main modern detection systems are a dual-beam UV/VIS spectrophotometer, which records the absorbance profiles at selected wavelengths and scans the sample in a radial direction and a ►laser interferometer that images the solution column and records the refractive index gradients caused by macromolecular redistribution. Typical samples are 100–400 µl of purified macromolecules or mixtures at concentrations of usually at least 0.05 mg/ml and usually up to 7 samples located at different angular positions in the rotor are studied simultaneously in one centrifugation run.

The purpose of the analytical ultracentrifuge is to observe the sedimentation process, i.e. the free fall of the dissolved macromolecules or particles in the absence of any matrix or supporting solvent density gradient. The gravitational force acting on the particles is proportional to their mass, reduced by the mass of the displaced solvent (following ►Archimedes principle). This requires knowledge of the ►partial specific volume of the macromolecules and in some cases involves the consideration of hydration and preferential solvent interactions. For particle densities smaller than the solvent, buoyancy will cause flotation instead of sedimentation. Opposing the gravitational force is the hydrodynamic friction, which is related to the shape of the macromolecule (e.g. its ►Stokes radius). The ratio of the resulting velocity of the sedimenting particles to the applied centrifugal field is a characteristic constant,

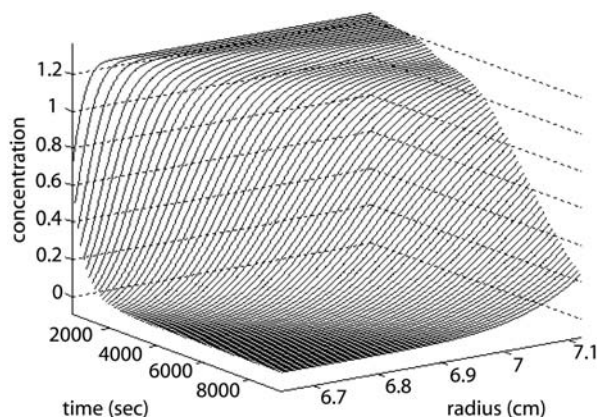


Analytical Ultracentrifugation. Figure 1 Schematic view of the rotor, the sample, and the detection system for measuring the macromolecular concentration distribution.

termed the ►sedimentation coefficient s . It is commonly measured in units of Svedberg, abbreviated S, which is related to the SI units by $1 \text{ S} = 10^{-13} \text{ s}$. Simultaneously with the sedimentation process, diffusion and possible chemical reactions take place. Since the extent of diffusion is also dependent on macromolecular shape, the diffusion constant D is related to the sedimentation coefficient s via the molar mass M , as expressed by the Svedberg equation

$$\frac{s}{D} = \frac{M(1 - \bar{v}\rho)}{RT}$$

(with \bar{v} denoting the partial-specific volume of the macromolecule, ρ the solvent density, R the gas constant, and T the absolute temperature). Typically, at high centrifugal fields of 200,000 g for the study of medium sized proteins, the macromolecular migration caused by sedimentation is of similar magnitude to the diffusional transport, so that the concentration profiles exhibit a moving sedimentation boundary that is broadened by diffusion (Fig. 2). Experiments that are aimed at observing the evolution of the concentration profiles with time under such conditions are termed ►sedimentation velocity experiments. Their duration is usually several hours.



Analytical Ultracentrifugation. Figure 2 Time-course of the macromolecular concentration distributions in a typical sedimentation velocity configuration. The sedimentation boundary migrates away from the center of rotation and is broadened by diffusion. In the plateau region, a second smaller sedimentation boundary is visible that migrates with a higher sedimentation coefficient.

At lower centrifugal fields, diffusion is predominant and the concentration distributions evolve more slowly and assume shapes with broader features, approaching an equilibrium state in which the sedimentation is effectively balanced by diffusion throughout the entire solution column (Fig. 3). This condition is termed ►sedimentation equilibrium. In ordinary configurations with 3–5 mm solution columns it is attained after one to two days.

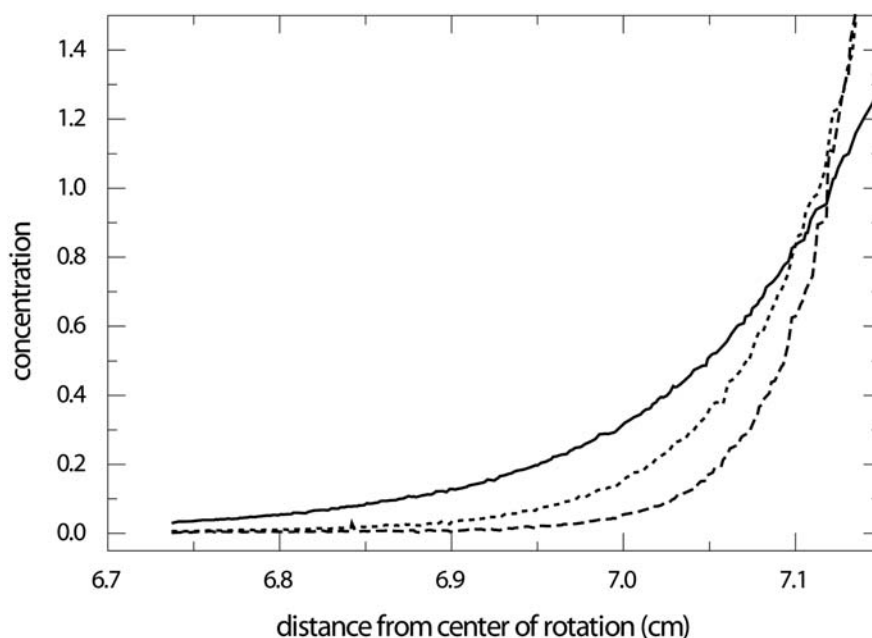
Sedimentation equilibrium is characterized by exponential concentration distributions with a steepness determined by the molar mass:

$$c(r) = c(b) \exp \left\{ M \frac{(1 - \bar{v}\rho)\omega^2}{2RT} (r^2 - b^2) \right\}$$

(where the radius r denotes the distance from the center of rotation, b the radial position of the bottom of the sample column, and ω the rotor angular velocity).

By adjusting the rotor speed, sedimentation velocity and sedimentation equilibrium conditions can be achieved for a very large range of sizes, permitting the study of particles such as small molecules with molar masses of a few hundred Dalton, up to macromolecules, macromolecular complexes and small μm sized organisms.

Conceptually, the key to the exceptional utility of sedimentation experiments in the study of reversible protein interactions is that despite the gravitational separation, larger species always remain surrounded by the smaller species during the entire experiment. This is in contrast, for example, to the typical configuration in



Analytical Ultracentrifugation. Figure 3 Concentration distributions as observed in a sedimentation equilibrium experiment. Shown are experimental distributions obtained at three different rotor speeds.

chromatography, where different species in an initial mixture are separated and isolated during their migration. In sedimentation velocity the faster sedimenting protein complexes continuously migrate through the slower sedimenting free protein species and can therefore remain at all times in an [▶association-dissociation equilibrium](#). Similarly, in sedimentation equilibrium the complexes are localized (due to a steeper exponential) on average closer to the bottom of the solution column, but also the free species increase exponentially in concentration towards the bottom of the solution, such that chemical equilibrium is maintained simultaneously with the mechanical centrifugal equilibrium (and [▶mass action laws](#) are fulfilled at each position throughout the solution column).

As a consequence, reversible protein interactions can be studied. The oligomeric states of [▶protein self-association](#) can be elucidated and the respective equilibrium constants be determined. Likewise, the stoichiometry and affinity of complexes formed between two or more different protein components interacting in reversible hetero-associations can be studied. Affinities can typically range from $K_D \sim 10$ nM to mM (dependent on the molar mass and extinction properties of the reacting macromolecules). Although interactions mediated by specific molecular recognition sites are the focus of most studies, other interactions, such as those derived from protein-solvent interactions, steric repulsion of molecules at high

concentration, charge interactions or physical entanglement of fibers may be observed and characterized as well. Examples of such applications include studies on protein crystallization conditions, interactions in crowded media as a model for the intracellular environment and interactions of amyloid fibers.

Sedimentation Velocity

Sedimentation velocity experiments frequently utilize the strongly size-dependent hydrodynamic separation of macromolecules to study their size and shape, to characterize size distributions and to detect macromolecular complexes through their higher sedimentation velocity.

In principle, the sedimentation coefficient can be determined simply by measuring the displacement of the sedimentation boundary with time, for example by observing the position of the boundary midpoint or by integrating the concentration profiles and assessing the macromolecular depletion with time. Data transformation methods have been developed that are better adapted to the problem of size-distributions. High-resolution distributions can be obtained by boundary modeling with [▶Lamm equation](#) solutions (see below). For a single sedimenting species, the sedimentation coefficient reveals information on both mass and shape:

$$s = \left(\frac{M}{N_A} \right)^{\frac{2}{3}} \frac{1}{(f/f_0)} \frac{(1 - \bar{\nu} \rho)}{6\pi\eta} \left(\frac{4\pi}{3\bar{\nu}} \right)^{\frac{1}{3}}$$

where N_A denotes the Avogadro constant, η denotes the solvent viscosity and where the ►frictional ratio f/f_0 is a molecular constant that measures the shape asymmetry relative to a compact sphere of the macromolecular mass and density. If the molar mass is known, which is frequently the case from the known macromolecular composition or from mass spectrometry, f/f_0 gives information about the solution conformation. This can be compared with theoretical models of the solution structure, derived from low-resolution ►bead models, molecular shapes observed in electron microscopy or from known crystal structures. Information on the dynamics of the solution conformation may be obtained by detecting changes in f/f_0 dependent on solution condition and ligand binding. Conversely, a shape asymmetry f/f_0 may be assumed (possible values fall in a close range of commonly 1.2–1.5 for folded globular proteins, values <1 can be excluded as they are theoretically impossible and values of 2 or larger usually are associated with very extended structures or unfolded chains), which can be useful to obtain an estimate the molar mass of the macromolecular species. For a sedimenting mixture, the sedimentation coefficient derived from the time-course of macromolecular depletion represents a weight-average sedimentation coefficient s_w of the mixture. For reversibly interacting macromolecules, s_w will change with the concentration and composition of the loading mixture. The analysis of this dependence can reveal the binding energy (►affinity constants) between the macromolecular components (through the analysis of the ►binding isotherm with mass action law models). The weight-average sedimentation coefficient can also be obtained by integration of the differential sedimentation coefficient distribution $c(s)$ (see below).

Modern analysis techniques aim to utilize the evolution of macromolecular concentration distribution over the entire solution column, which represents a very rich source of additional information on macromolecular diffusion, molar mass, size-distribution of mixtures and the thermodynamics and kinetics of interactions. This is based on the equation for sedimentation and diffusion in the sector-shaped solution column

$$\frac{\partial c}{\partial t} = \frac{1}{r} \frac{\partial}{\partial r} \left[rD \frac{\partial c}{\partial r} - s\omega^2 r^2 c \right]$$

which was derived by Lamm in 1929 and was the subject of significant theoretical interest for several decades, but became widely applicable to routine analysis in the 1990s.

In the analysis of single sedimenting species, the information on diffusion from the boundary spread allows the direct determination of the molar mass. It also allows extending the sedimentation velocity analysis to configurations that do not produce clear

sedimentation boundaries, such as for small molecules. Extensions to the macromolecular sedimentation in self-forming density gradients from sedimenting cosolutes have been described, as well as corrections for the solvent compressibility.

For sedimenting mixtures, the mathematical modeling of the experimentally observed sedimentation profiles as a superposition of solutions of the Lamm equation permits the deconvolution of the diffusion from the sedimentation process and the determination of high-resolution differential sedimentation coefficient distributions $c(s)$ (4). It also reveals a weight-average frictional ratio f/f_0 that can be exploited to estimate the molar mass distribution of the sedimenting mixture. In practice, this typically leads to baseline-resolved peaks for stable oligomeric species (for $M_w > 10$ kD).

The Lamm equation analysis can be modified to account for chemical reactions between the sedimenting species and this can be utilized to study protein self-association and heterogeneous protein interactions. For such systems, the sedimentation velocity patterns at different loading concentrations exhibit shapes that are characteristic for the association scheme and the type of complexes formed. In addition, the global analysis of the sedimentation data from different loading concentrations allows determining the binding energy, as well as the low-resolution hydrodynamic shape f/f_0 of the reversibly formed complexes. Information on the binding kinetics and the stability of the complexes can also be extracted.

Sedimentation Equilibrium

Sedimentation equilibrium experiments are frequently applied where the determination of the molar mass of protein complexes and/or the measurement of affinity constants is of primary interest. The equilibrium state is only dependent on thermodynamic properties and is not affected by the dynamic processes from reaction kinetics or hydrodynamics. The steepness of the concentration profiles is directly related to the molar mass of the sedimenting species. For interacting systems the local concentration gradient is related to the local weight-average molar mass. Because a gradient spanning a large range of concentrations is established as a result of the centrifugation, ranging from very dilute conditions closer to the center of rotation to very high concentrations at the bottom of the solution column, information equivalent to an entire binding isotherm is gained from a single experiment. This permits the determination of the affinity constants for interacting systems.

The separation of different species in the exponential equilibrium distributions is not as strong as the hydrodynamic separation in sedimentation velocity, which generally imposes higher requirements on

sample purity for the sedimentation equilibrium experiments and can make the identification of the stoichiometry of the complexes more difficult. When studying interactions between components of dissimilar absorption spectrum, multi-wavelength detection can be applied to unravel the individual concentration profiles of each component. Frequently, the redistribution of a macromolecular mixture is observed for a series of loading concentrations and different centrifugal fields. This permits more stringent characterization of the molar mass of the populated species and their association behavior, through global mathematical analysis and mass conservation techniques.

More information on both modern sedimentation velocity and sedimentation equilibrium analytical ultracentrifugation can be found in (5) and the references cited therein

References

1. Svedberg T, Pedersen KO (1940) The ultracentrifuge. Oxford University Press, London
2. Schachman HK (1959) Ultracentrifugation in Biochemistry. Academic Press, New York
3. Fujita H (1975) Foundations of ultracentrifugal analysis. John Wiley & Sons, New York
4. Schuck P (2000) Size distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and Lamm equation modeling. *Biophysical Journal* 78:1606–1619
5. Lebowitz J, Lewis MS, Schuck P (2002) Modern analytical ultracentrifugation in protein science: a tutorial review. *Protein Science* 11:2067–2079

Anaphase

Definition

Anaphase describes a stage of mitosis, when the chromatids are split and become daughter chromosomes that move to opposite spindle.

- ▶ Centromeres
- ▶ Mitosis

Anaphase Movement

Definition

Anaphase movement refers to the initial movement of chromosomes to opposite poles of the spindle during anaphase.

- ▶ Centromeres

Anaphase Promoting Complex (APC)/Cyclosome

Definition

Anaphase Promoting Complex (APC/C) is a large multi-subunit ▶E3–ligase, which plays an essential role in initiating anaphase because it triggers sister chromatid separation. This is achieved by targeting proteins responsible for the progression through the cell cycle (e.g. securins and cyclins) for degradation by the 26S proteasome. Several APC activators (e.g. cdc20 or cdh1) can modulate the activity and specificity of the APC. Among those proteins targeted by APC/C are mitotic cyclins and the so called “cohesions” (which attach sister chromatids). Signalling of kinetochores to APC/C is required for destruction of the cohesions. If these are degraded, sister chromatid separation proceeds.

- ▶ Cell Division
- ▶ Ubiquitination

ANAs

Definition

ANAs or anti-nuclear autoantibodies are usually found in SLE (Systemic Lupus erythematoses) and potentially also in rheumatoid arthritis and other rheumatic diseases, and represent antibodies that target nuclear antigenic components. These antibodies are believed to play an important role in affecting the pathology seen in these diseases.

- ▶ SLE Pathogenesis Genetic Dissection

Anchorage Dependence/Independence

Definition

Anchorage dependence of survival, growth etc. describes the need for cells to attach to a solid substrate in order to exert the activities indicated. Anchorage independence describes the property of transformed cells to form aggregates/colonies in semi-solid agar medium without adherence to the substrate. Normal adherent cells undergo a special form of cell death (called anoikis) under these conditions.

- ▶ Cancer Invasion and Metastasis: Cellular, Molecular and Clinical Aspects
- ▶ Ras Signalling

Androgen Insensitivity

Definition

Androgen insensitivity refers to reduced effects of substances with masculinizing activities such as testicular hormones.

► Repeat Expansion Diseases

Androgen Receptor (Gene)

Definition

The androgen receptor gene encodes a protein with the capability to combine specifically with androgen(s), transforming the hormonal signal into adequate stimuli for the target cells.

► Chromosome 21 Disorders

► Repeat Expansion Disease

Aneuploidy

Definition

Aneuploidy designates a chromosomal abnormality, arising when a diploid genome has either gained or lost some or all chromosomes from a diploid set.

► Centromeres

► Chromosome 21 Disorders

► Telomerase

► X-Chromosome Inactivation

Angelman Syndrome

Definition

Angelman syndrome is a relatively frequent disorder (1:15,000 newborns) caused by functional loss of ► **imprinted** genes on the paternal chromosome 15q11–q13 (in 70% by maternal 15q11–q13 deletion, 4% paternal uniparental disomy, 8% imprinting center mutation on the maternal allele, and 8% mutation of the UBE3A on the maternal allele). Patients show severe mental retardation, seizures, ataxia, and paroxysms of inappropriate laughter (“happy puppet syndrome”).

► Microdeletion Syndromes

Angiofibroma

Definition

Angiofibroma is a hamartoma (benign tumor) consisting of red to pink papules or nodules typically in a malar distribution, which first appear between the ages of 2 and 6 years, and progress to a variable extent during puberty. Histologic findings are dermal fibrosis and angiogenesis with occasional large glial appearing cells. Commonly seen in tuberous sclerosis.

► Tuberous Sclerosis

Angiogenesis

KARL H. PLATE

Neurological Institute, Clinical Center of the Johann Wolfgang Goethe University, Frankfurt, Germany
karl-heinz.plate@kgu.de

Definitions

► **Angiogenesis** is defined as sprouting of capillaries from pre-existing vessels. ► **Vasculogenesis** describes the *in situ* differentiation of endothelial precursor cells during development. ► **Bone marrow** derived vasculogenesis is defined as mobilization of endothelial precursor cells from the bone marrow and their integration and differentiation into sites of adult blood vessel formation. Somatic stem cell derived vasculogenesis is defined as mobilization of precursor cells from differentiated organs and their integration and transdifferentiation into growing blood vessels. ► **Lymphangiogenesis** describes the formation of the lymphatic system during development and in pathological conditions. ► **Vascular morphogenesis** describes the formation and maturation of functional blood vessels, e.g. the interaction of vascular endothelial cells with pericytes and smooth muscle cells.

Characteristics

All tissues in developing and adult organisms need a functional blood supply in order to receive sufficient amounts of oxygen and nutrients, which are necessary for their proper function. Blood vessels start to form early in embryonic development (e.g. E 8.5 in the mouse) before the heart starts to beat. All blood vessels are lined by a monolayer of endothelial cells, which provide a non-coagulant surface for the blood. In the brain, endothelial cells regulate the influx of extracellular

molecules to the brain parenchyma and thus constitute the morphological basis of the blood brain barrier. During development, endothelial cells form tube-like structures that are covered by pericytes and a basal lamina. Larger vessels form by apposition of contractile smooth muscle cells.

In the embryo, tissues are vascularized by vasculogenesis and angiogenesis. Both processes are almost absent in the adult organism. Transient and tightly controlled (physiological) angiogenesis in adult tissues occurs during the female reproductive cycle and during wound healing. In contrast, pathological angiogenesis is characterized by the persistent proliferation of endothelial cells and is a prominent feature of diseases such as proliferative retinopathy, rheumatoid arthritis and psoriasis. In addition, many tumors are able to attract blood vessels from neighboring tissues. Tumor-induced angiogenesis requires a constitutive activation of endothelial cells. These endothelial cells dissolve their surrounding extracellular matrix, migrate toward the tumor, proliferate and form a new vascular network, thus supplying the tumor with nutrients and oxygen and removing waste products.

Recent reports suggest that in adult tissues bone marrow derived ►[vasculogenesis](#) and somatic stem cell derived vasculogenesis may occur, but the extent of these processes compared to angiogenesis appears limited.

Molecular Interactions

Several molecules have been described to be essential for angiogenesis, vascular development and lymphangiogenesis.

VEGF's

The vascular endothelial growth factor family consists of ►[VEGF](#) (also called VEGF-A), VEGF-B, VEGF-C,

VEGF-D, VEGF-E and the related placenta growth factors PlGF-1 and -2. VEGF's bind differentially to three high affinity tyrosine kinase receptors VEGFR-1, -2 and -3 (Table 1). In addition, neuropilin-1 and -2 receptors play a role in modifying VEGF receptor function.

The vascular endothelial growth factor (VEGF)-VEGF receptor system is necessary for early vascular development. Although VEGF, a hypoxia-regulated secreted protein, is expressed by many cell types in the developing mouse, its correct spatial, temporal and quantitative expression appears to be a prerequisite for normal vascular development. Heterozygous inactivation of the *VEGF* gene in mice leads to early embryonic lethality, suggesting that early vascular development is critically dependent on VEGF availability. VEGF binds to two tyrosine kinase receptors, VEGFR-1 (flt-1) and -2 (KDR/flk-1). VEGFR-2 is essential for the generation of endothelial cells and blood cells and for endothelial cell proliferation, migration and survival. In mice, knockout of VEGFR-2 leads to embryonic lethality around embryonic day 8 with no endothelial cells or blood cells being formed. VEGFR-1 may act as a decoy receptor for VEGF and negatively regulate embryonic angiogenesis. General knockout of VEGFR-1 in mice resulted in embryonic lethality due to a disorganized vascular system and an overgrowth of endothelial cells. In contrast, a tyrosine kinase specific VEGFR-1 knockout showed no significant vascular phenotype suggesting that during development, the major VEGFR-1 function is the binding of VEGF and not signal transduction. VEGFR-1 however has been shown to regulate pathological angiogenesis by binding the VEGF homologue placenta growth factor (PlGF). VEGF also induces vascular permeability, but it is unclear which receptor transmits this signal.

Angiogenesis. Table 1 Angiogenic growth factors and receptors

Receptor	VEGFR-1	VEGFR-2	VEGFR-3	TIE1	TIE2
Putative functions	monocytes/macrophages: migration endothelial cells: negative regulation of VEGF? PlGF: stimulation of adult angiogenesis	vasculogenesis hematopoiesis angiogenesis permeability	assembly of large vessels permeability lymphangiogenesis	Vascular remodeling	sprouting angiogenesis vascular remodeling vascular permeability
Ligand(s)	VEGF	VEGF	VEGF-C	?	Angiopoietin-1
	VEGF-B	VEGF-C	VEGF-D		Angiopoietin-2
	PlGF	VEGF-D			Angiopoietin-3
		VEGF-E			Angiopoietin-4

VEGFR-3 (flt-4) serves mainly as a receptor for VEGF-C and -D and appears to have a major function in lymphangiogenesis, both during development and in pathological conditions. Both VEGF-C and -D have been shown to promote ►[lymphangiogenesis](#) and lymphatic metastasis in tumors. Possibly angiopoietin-2 (see below) cooperates in this process by acting as an additional stimulator for lymphangiogenesis.

TIE Receptors and Angiopoietins

Angiopoietins (Ang) 1-4 bind to Tie-2 receptors primarily expressed on endothelial cells. Angiopoietins 1 and 3 and 2 and 4 respectively, have opposing effects on the Tie-2 receptor but this may also be cell type dependent. Ang-1 and -3 stimulate Tie-2 phosphorylation in endothelial cells, whereas Ang-2 and -4 inhibit receptor phosphorylation. Unlike mice deficient in VEGF or VEGF receptors, mouse embryos lacking Ang-1 or Tie-2 develop a normal primary vascular structure but display defects in vascular maturation and vascular sprouting. The major function of Tie-2 appears to be vessel maturation/stabilization by regulating pericyte/endothelial cell interaction. Ang-2 by blocking Tie-2 signaling may lead to vascular destabilization by pericyte drop-off. Ang-2 action sensitizes the endothelial cell for angiogenic stimuli such as VEGF. If VEGF is not available in sufficient amounts, the destabilized vessels regress.

Ang-1 appears to be a major regulator of vascular permeability by stabilizing the vessel wall. The orphan receptor Tie-1 appears to have little function in early vascular development, but plays an important role in vascular morphogenesis as suggested by knockout studies. Whether independent ligands for Tie-1 exist or whether angiopoietins may, under certain circumstances, also bind to Tie-1 is unclear. In addition to angiopoietins and Tie receptors, platelet-derived growth factor receptor β has been shown to be involved in vascular maturation by recruitment of pericytes.

Ephrins

Eph tyrosine kinase receptors and eph ligands are cell bound molecules with a well-known function in neural development. Ephrin-B2 and Eph B4 receptors however have also been shown to play essential roles in vascular development. Studies in knockout mice suggest that ephrin-B2 and Eph B4 play an important role in establishing venous *versus* arterial identity and possibly later on in the development of arteries.

Regulatory Mechanisms

The growth of the vascular system correlates well with body mass and with the metabolic demands of tissues and organs. Hypoxia appears to be a major regulator of angiogenesis and possibly also of vasculogenesis in development. Hypoxia leads to accumulation and

nuclear translocation of ►[Hypoxia Inducible Factors](#) (HIF-1 and HIF-2) that transactivate a set of genes, including stimulators of angiogenesis such as VEGF. VEGF is formed by alternative splicing in five different isoforms (VEGF 121, 145, 165, 189, 204) of which only the three smaller isoforms are efficiently secreted. VEGF binds to VEGFR-1 with a ten-fold higher affinity than to VEGFR-2, but only VEGFR-2 shows significant tyrosine phosphorylation. After ligand binding, VEGFR-2 activates various intracellular signaling pathways in endothelial cells, including the PI3-kinase/Akt, PLC/PKC and ras/MEK/erk pathways.

Evidence suggests that VEGFR-2 can be induced by VEGF in a homologous fashion. The ets-1 and HIF-2 transcription factors appear to cooperate in the hypoxic regulation of VEGFR-2. VEGF is also regulated by MAP-kinase dependent mechanisms, independently of hypoxia. Low pH and low glucose as well as genetic alterations (loss of function of the von Hippel Lindau tumor suppressor gene, VHL) may also activate the HIF pathway and stimulate angiogenesis.

In tumors, hypoxia and hypoxia-independent mechanisms have been shown to trigger angiogenesis. In many cancer types, VEGF is specifically up-regulated in perinecrotic tumor cells, suggesting that hypoxia is a major driving force for ►[tumor angiogenesis in vivo](#). Other factors that may up-regulate VEGF in cancer cells include p53 mutations, epidermal growth factor receptor over-expression, PTEN mutations and ras and src mutations.

In patients with ►[VHL-\(von Hippel-Lindau\) Disease](#), a hereditary cancer syndrome characterized by highly vascularized tumors in the brain, retina and kidney and cysts in the pancreas, kidney and epididymis, VEGF is constitutively expressed in cells deficient in a functional VHL protein. VHL forms a multimeric complex with HIF-1, necessary for proteosomal degradation of the complex. In cells with VHL loss of function, HIF-1 accumulates, leading to VEGF expression and induction of angiogenesis.

The importance of VEGF for tumor angiogenesis and tumor growth has been shown by numerous studies inhibiting the VEGF function in animal models.

Autocrine inhibition of Tie-2 by endothelial Ang-2 expression appears to be a prerequisite for angiogenesis. Possibly, inhibition of Tie-2 by Ang-2 leads to loosening of endothelial cell / pericyte contacts and renders endothelial cells accessible to inducers such as VEGF. Neither the endothelial cell specific expression of Ang-2 nor its regulation is completely understood. Possibly VEGF and hypoxia up-regulate Ang-2, Ang-1 and Tie-2, but this regulation may be dependent on the cell type. In tumors, in contrast to Ang-2, Ang-1 and also Tie-2 appear to be constitutively expressed, suggesting that Ang-2 is the major regulator of the Tie-2 pathway.

In lymphangiogenesis, both Ang-1 and Ang-2 act as stimulators, suggesting that in certain circumstances both ligands can activate the Tie-2 receptor. The VEGF homologues VEGF-C and -D have been shown to bind to VEGFR-3 and to VEGFR-2 and to be major drivers of lymphangiogenesis. The regulation of VEGF-C, -D and VEGFR-3 is poorly understood but hypoxia does not appear to be involved in their regulation.

- [Angiogenesis](#)
- [Breast Cancer](#)
- [Hypoxia Inducible Factors](#)
- [Mesenchymal Cells](#)
- [Rheumatoid Arthritis](#)
- [Vasculogenesis](#)
- [Wnt/Beta-Catening Signaling Pathway](#)

References

1. Risau W (1997) Mechanisms of angiogenesis. *Nature* 386:671–674
2. Yancopoulos GD, Davis S, Gale NW et al (2000) Vascular-specific growth factors and blood vessel formation. *Nature* 407:242–248
3. Hanahan D (1997) Signaling vascular morphogenesis and maintenance. *Science* 277:48–50
4. Plate KH (2001) From angiogenesis to lymphangiogenesis. *Nat Med* 7:151–152
5. Ferrara N (2002) VEGF and the quest for tumour angiogenesis factors. *Nat Rev Cancer* 2:795–803

Angiogenesis Factors

Definition

Angiogenesis factors are a heterogenous group of molecules, able to stimulate endothelial cell proliferation *in vitro* and/or angiogenesis in *In Vivo* assays.

- [Angiogenesis](#)

Angiogenesis-Dependent Diseases

Definition

Many disorders in humans (inflammation, ischemia, hypoxia neoplasia) are characterized by the onset of angiogenesis, e.g. the formation of a new vascular system. Inhibition of angiogenesis is thought to be a therapeutic option for treatment of these diseases.

- [Angiogenesis](#)

Angiomyolipoma

Definition

Angiomyolipoma is a hamartoma (benign tumor) consisting of a variable mixture of aberrant vessels, smooth muscle, and fat, that is typically small but can grow to > 5 cm diameter, occurring most commonly in the kidney but also in other abdominal sites. Commonly seen in tuberous sclerosis.

- [Tuberous Sclerosis](#)

Angiotensin Converting Enzyme

Definition

Angiotensin converting enzyme is an enzyme that converts the molecular substrate angiotensin I into angiotensin II, which is able to constrict blood vessels and increase blood pressure.

- [ACE-Inhibitor](#)
- [Diabetes Mellitus, Genetics](#)
- [Hypertension](#)

Ångström (Å)

Definition

Ångström defines a unit of length defined as 10^{-10} meters, or 0.1 nanometers. This unit is commonly used in crystallography because individual chemical bonds have lengths between 1 and 2 Å.

- [3D Structure Determination by X-Ray](#)

Animal Cap

Definition

Animal Cap designates the multi-potent cells that constitute the roof of the blastocoel (the animal hemisphere) in amphibian blastula-staged embryos. These cells may be dissected from the embryo and cultured.

- [Xenopus as a Model Organism for Functional Genomics](#)

Animal Model

Definition

Animal model describes an animal with a disease resembling a human disease, which is then used to study the development and progression of the disease, and possible treatments.

- ▶ [Cell Polarity](#)
- ▶ [Transgenic and Knock-out Animals](#)

Anionic C-terminal Carboxylate

Definition

The carboxylic acid on the C-terminus of a protein exists in its ionized state and is hence anionic.

- ▶ [Protein Prenylation](#)

Ankyrin Repeat

Definition

Ankyrin repeat is one of the most common, modular, protein-protein interaction motifs, mainly from eukaryotes. Ankyrin repeats are tandemly repeated modules of about 33 amino acids and occur in a large number of functionally diverse proteins.

- ▶ [NFkB Pathway](#)

Anlage

Definition

Anlage describes the condensation of chondrocytes that represent the future bone in shape and relative size.

- ▶ [Bone Disease and Skeletal Disorders, Genetics](#)

Annotation

Definition

Annotation refers to the description of local features in genomic sequences. In sequence databases,

annotations are text fields containing information about a biosequence, which are added to sequence databases.

- ▶ [Chromosome 21 Disorders](#)
- ▶ [Protein Databases](#)

Anomalous Diffraction, Anomalous Scattering

Definition

Electron motions induced by incoming radiation re-emit radiation, thereby scattering it out of phase with the incoming radiation. However, if the incoming energy is near an electron energy transition, the induced electron motion may be retarded by (resonant) excitation of an electron into the higher energy state, thereby shifting the phase of the scattered radiation. Without anomalous scattering, the diffraction intensities from a crystal before and after a 180° rotation are identical; with anomalous scattering, differences arise that can be exploited in structure solution.

- ▶ [3D Structure Determination by X-Ray](#)

Anomalous Scatterer

Definition

Any atom becomes an anomalous scatterer if the wavelength of the exciting radiation (e.g. X-rays) is near its absorption edge. By resonance effects the wavelength remains unchanged (elastic scattering), but the amplitude is changed and the phase is retarded. The scattering effect is nearly independent from the scattering angle. Metal ions, lanthanides, and actinides are sources of high anomalous signals at their corresponding absorption edges.

- ▶ [MAD Phasing](#)

Anomalous Scattering

- ▶ [Anomalous Diffraction, Anomalous Scattering](#)

Anomer

Definition

Anomer denotes stereoisomers of a monosaccharide, which differ only in configuration at the anomeric carbon of the ring structure (i.e. the C-1 position in a cyclic hemiacetal).

► [Glycosylation of Proteins](#)

ANT1 (Adenosine Nucleotide Transporter 1)

Definition

ANT1 (adenosine nucleotide transporter 1) is a transporter protein located in the mitochondrial membrane. It is important for transporting nucleotides.

► [Mitochondrial Myopathies](#)

Antagonist

Definition

The term antagonist refers to a drug or ligand that binds to a target receptor and accordingly blocks the binding of an agonist, usually in a competitive way. In consequence, the normal response of the receptor to ligand binding is obstructed.

► [Cytokine Receptors](#)

► [Structure-Based Drug Design](#)

ANT-C

► [Antennapedia Complex](#)

Antennapedia Complex

Definition

Antennapedia complex (ANT-C) refers to the cluster of homeotic genes that specify the identities of anterior segments in *Drosophila Melanogaster*.

► [Homeobox Transcription Factor](#)

Anterior Horn Cells

Definition

Anterior horn cells are lower motor neurons present in the anterior horn of the spinal cord. Upper motor neurons reside in the cerebral cortex and send signals to the lower motor neurons.

► [Spinal Muscular Atrophy](#)

Anthracyclines

Definition

Anthracyclines are natural anticancer products acting as chemotherapeutics by intercalation into the DNA double helix and inhibiting the enzyme topoisomerase. Thus, DNA repair, replication and transcription are prevented. Commonly used anthracyclines include doxorubicin, daunorubicin, epirubicin, and mitoxantrone.

► [Multi-Drug Resistance](#)

Antibody

Definition

Antibodies are secreted forms of immunoglobulins. They are involved in the humoral immune response. Antibodies are the major product of terminally differentiated B-lymphocytes (plasma cells) and recognize antigens through their antigen binding sites, and trigger the host's immune response by the Fc portion of the molecule. Engineered antibodies are increasingly used for the treatment of human disease.

► [Autoimmune Diseases](#)

► [Autoimmunity](#)

► [Immunoglobulins](#)

► [Immunochemical Methods, Localization](#)

► [Proteomics in Microfluidic Systems](#)

Antibody Array

Definition

An antibody array (antibody microarray) is an analysis system for profiling protein expression in biological

samples, and allows the comparison of the relative abundance of hundreds of proteins. For detection of the specific antigens in complex samples, the specific antibodies are immobilized on a surface in a spatially addressable manner.

- Immunological Methods, the Use of Antibodies to Discover the Function of Novel Gene Products
- Protein Microarrays as a Tool for Protein Profiling and Functional Proteomics

Antibody Chip

- Protein Microarrays as a Tool for Protein Profiling and Functional Proteomics

Antibody Cross Reactivity

Definition

Antibody cross reactivity describes the ability of an antibody to react with or bind an antigen that did not stimulate its production.

- Protein Microarrays as Tool for Protein Profiling and Identification of Novel Protein-Protein Interactions

Anticipation

Definition

Anticipation designates to the apparent occurrence of a hereditary disease at a progressively earlier age in successive generations and is often accompanied by severely pronounced disease course (cave: artefacts arising from the ease of identification of succeeding cases).

- Huntington's Disease
- Repeat Expansion Diseases

Anticodon

Definition

Anticodon designates the three nucleotide sequence in a tRNA that is complementary to a triplet of mRNA (codon), which specifies a certain amino acid.

- tRNA

Anti-Craving Substances

Definition

Anti-craving substances are drugs that have been registered for relapse prophylaxis in weaned alcoholics in various European countries (acamprosate) and the United States (naltrexone). Acamprosate and naltrexone most likely reduce ethanol abuse through different neuronal mechanisms. Acamprosate acts as a weak glutamate (NMDA)-receptor antagonist and modulates the number of specific glutamate (NMDA)-receptor subunits. Naltrexone acts as a morphine antagonist and displaces morphine from its receptor.

- Addiction, Molecular Biology

Antigen

Definition

The term antigen denotes any foreign molecules, e.g. proteins, polysaccharides, liposaccharides or other small compounds, which trigger the generation of antibodies when introduced into the host organism. It can be recognized by antibodies produced by B-lymphocytes or by the T cell antigen receptor.

- Autoimmune Diseases
- Autoimmunity
- Camel as a Model for Functional Genomics
- Immunological Methods, the Use of Antibodies to Discover the Function of Novel Gene Products
- Monoclonal Antibodies

Antigen Expression

Definition

Antigen expression defines the production of proteins that can be used as markers to identify the cell type.

- Microarrays in Pancreatic Cancer

Antigen Receptors

Definition

Antigen receptors are proteins that are expressed on lymphoid cells that specifically recognize antigens.

- DNA Recombination

Antimüllerian Hormone

Definition

The Antimüllerian hormone is secreted by the testicles and prevents development of the uterus, fallopian tubes and cephalad vagina in males. It is also called müllerian inhibiting substance and müllerian inhibition factor (►AMH/MIS).

►SRY – Sex Reversal

one directing protein synthesis. Antisense technology is used to selectively turn off production of certain proteins. Biologically, short synthetic ssDNA (conventional antisense) or dsRNA (siRNA) are most active, as they shut off gene expression by activating endonucleases (Rnase H or dsRNase, respectively) that hydrolyze the target RNA.

►Apoptosis

►Apoptosis, Regulation and Clinical Implications

►Lung

►Morpholino Oligonucleotides, Functional Genomics by Gene ‘Knock-down’

►Spinal Muscular Atrophy

Anti-Nuclear Autoantibodies

►ANAs

Antiparallel

Definition

Antiparallel designates that the two strands of DNA are chemically arranged in opposite directions, with the structure P–5'-sugar–3'-P in one strand opposing the structure P–3'- sugar–5'-P in the other strand. Antiparallelism and base-pairing rules determine the sequence and orientation of the complementary strand.

►Replication Fork

Antisense RNA

Definition

An antisense transcript is produced by the transcription of the DNA strand complementary to the coding strand. Antisense transcripts overlap, at least partially, with the transcript from the coding DNA strand. These RNAs are usually non-coding. They could play a role in the regulation of monoallelic gene expression. The best known examples include, the random monoallelically expressed Xist/Tsix transcript pair on the chromosome X, or the imprinted Igf2r/Air transcripts on the human chromosome 6.

►Catalytic RNA

►Genomic Imprinting

Antiretroviral Agents

Definition

Antiretroviral Agents (ARVs) are inhibitors of retroviral enzymes, receptors, or other important targets in the HIV–1 life-cycle, which are used therapeutically to suppress viral replication and delay the emergence of drug resistance.

►Reverse Transcriptase

Antisera

Definition

Antisera are animal or patient sera that contain specific antibodies, usually against an antigen that had previously been introduced by a natural infection or an intentional inoculation (immunization).

►Immunological Methods, the Use of Antibodies to Discover the Function of Novel Gene Products

►Proteomics in Human-Pathogen Interactions

Antisense (Oligonucleotides)

Definition

A piece of DNA producing a mirror image (“anti-sense”) messenger RNA that is opposite in sequence to

AOM

►Acousto-Optical Modulator

Aorta

Definition

The aorta is the largest blood vessel in the human body, receiving blood that is usually pumped from the left ventricle of the heart into circulation.

- ▶ Heart
- ▶ Marfan Syndrome

AOTF

- ▶ Acousto-Optical Tunable Filter

AP Endonuclease

Definition

AP endonuclease is an enzyme that cleaves 5' to an

- ▶ AP site to generate a single strand break in DNA.
- ▶ Base Excision Repair

AP-1

Definition

AP-1 (activator protein-1) comprises of a collection of dimeric complexes composed of members of the Jun, Fos and ATF protein families of transcription factors.

- ▶ Jun/Fos
- ▶ Leucine Zipper Transcription Factors: bZIP Proteins

APC

- ▶ Adenomatous Polyposis Coli

APC

- ▶ Anaphase Promoting Complex (APC)/Cyclosome

APC/C

Definition

APC/C stands for anaphase promoting complex/cyclosome. It is a large ubiquitin ligase complex that is activated by accessory subunits (Cdc20 and Cdh1) and controls the degradation of proteins, including mitotic cyclins and securin.

- ▶ Cell Cycle – Overview

APCI

Atmospheric pressure chemical ionisation.

- ▶ SNP Detection and Mass Spectrometry

APH-1

Definition

APH-1 is the human homologue of *C. elegans* anterior pharynx-defective protein. APH-1 stabilizes the γ -secretase complex, and has γ -secretase inhibiting characteristics that may be overcome by presenilin enhancer-2 (PEN-2).

- ▶ Alzheimer's Disease

Apheresis

Definition

Apheresis describes the process of removing a specific component from blood and returning the remaining components to the donor.

- ▶ Familial Hypercholesterolemia

Apical Ectodermal Ridge

Definition

A specialised epithelial structure positioned at the border between dorsal and ventral limb ectoderm. It is

responsible for the continuous proliferation of the underlying mesenchymal cells that leads to the outgrowth of the limbs along the proximal-distal axis (shoulder to finger tip).

- ▶ [Limb Development](#)
- ▶ [Notch Pathway](#)

Apical Junctional Complex

Definition

The apical junctional complex is a specialized junctional structure that is found in simple epithelia comprising the tight junction and the zonula adherens at the apical region of the lateral cell membrane.

- ▶ [Cell Polarity](#)
- ▶ [Tight Junctions](#)

Apical Membrane

Definition

Apical membrane is the plasma membrane domain of an epithelial cell that faces the lumen of a cavity or tube, or the outside of the organism.

- ▶ [Microvilli](#)
- ▶ [Tight Junctions](#)

APL

Definition

- ▶ [Acute Promyelotic Leukaemia](#)

APOE

Definition

APOE stands for the human gene for Apo E (apolipoprotein E).

- ▶ [Alzheimer's Disease](#)

APO-2 ligand

- ▶ [TNF-Related Apoptosis Inducing Ligand \(TRAIL, APO-2 ligand\)](#)

apoE Null (Mutant, Knockout) Mice

Definition

ApoE null (mutant, knockout) mice are genetically modified mice that are deficient in the ApoE gene.

- ▶ [Tangier Disease](#)

Apolar

Definition

This term is used in different contexts. It is used – as acronym to "polar" – to describe amino acids or stretches of amino acids with uncharged side chains. The term is also used to refer to the net absence of directionality in the assembled intermediate filaments. Both ends of the filament are biochemically the same as subunits assembled as antiparallel strands during polymerisation. This is in contrast to both actin and tubulin polymers, whose polarised addition of subunits during assembly allows regulation of assembly/disassembly at specific ends of the filaments – this cannot happen in intermediate filaments.

- ▶ [Intermediate Filaments](#)
- ▶ [Amino Acids: Physicochemical Properties](#)

Apolipoprotein

Definition

Apolipoprotein is the protein component of a lipoprotein particle.

- ▶ [Familial Hypercholesterolemia](#)

Apoptosis

Definition

Apoptosis is a form of highly regulated cellular self-destruction, whereby a cascade of signals eventually leads to the breakdown of the cellular DNA and to subsequent cell death (also known as programmed cell death). Apoptosis is characterized by distinct morphologic changes in the nucleus and cytoplasm (nuclear condensation and packaging of cell contents). When cells enter into apoptosis, they lose contact with their neighbours, shrink, and their nuclear DNA is

fragmented. At the end, cells break into vesicles that are engulfed by other cells. Cells with massively damaged DNA are prevented from continuing through the cell cycle by the induction of apoptosis, thus maintaining tissue homeostasis. Apoptosis occurs without causing inflammation and damage to neighbouring cells. In contrast to uncontrolled cell death after tissue injury (►necrosis), apoptosis is an evolutionarily conserved pathway for the elimination of cells during normal physiological processes.

- Apoptosis, Regulation and Clinical Implications
- Cancer Invasion and Metastasis: Cellular, Molecular and Clinical Aspects
- Cardiac Signaling: Cellular, Molecular and Clinical Aspects
- Cap-Independent Translational Control
- Chromosomal Instability Syndromes
- Defective Protein Folding Disorders
- Desmosomes
- High-Throughput Approaches to the Analysis of Gene Function in Mammalian Cells
- Inflammatory Response
- Limb Girdle Muscular Dystrophies
- Morbus Wegener
- Neurotrophic Factors
- Peutz-Jeghers Syndrome
- Senescence
- Telomerase
- TNF Receptor/Fas Signaling Pathways
- Tumor Suppressor Genes

Apoptosis, Regulation and Clinical Implications

HANS-UWE SIMON¹,
UWE ZANGEMEISTER-WITKE²

¹Department of Pharmacology, University of Bern,
Bern, Switzerland

²Department of Oncology, University of Zurich,
Zurich, Switzerland
uwe.zangemeister@USZ.ch

Definition

►Apoptosis is the most common form of physiological cell death and serves to eliminate cells that are (a) genomically altered and therefore potentially harmful to the body, (b) senescent or (c) become replaced by functionally different cell types. It is defined by morphological criteria, including chromatin condensation, nuclear fragmentation, plasma membrane blebbing, organelle disruption, loss of adhesion and

rounding (in adherent cells) and cell shrinkage. Apoptosis is required for normal development and maintenance of tissue homeostasis and its dysregulation is associated with various pathological conditions. For instance, there is often too little apoptosis in cancer and chronic inflammation and there is usually too much apoptosis in neurodegeneration and acquired immune deficiency syndrome (►AIDS).

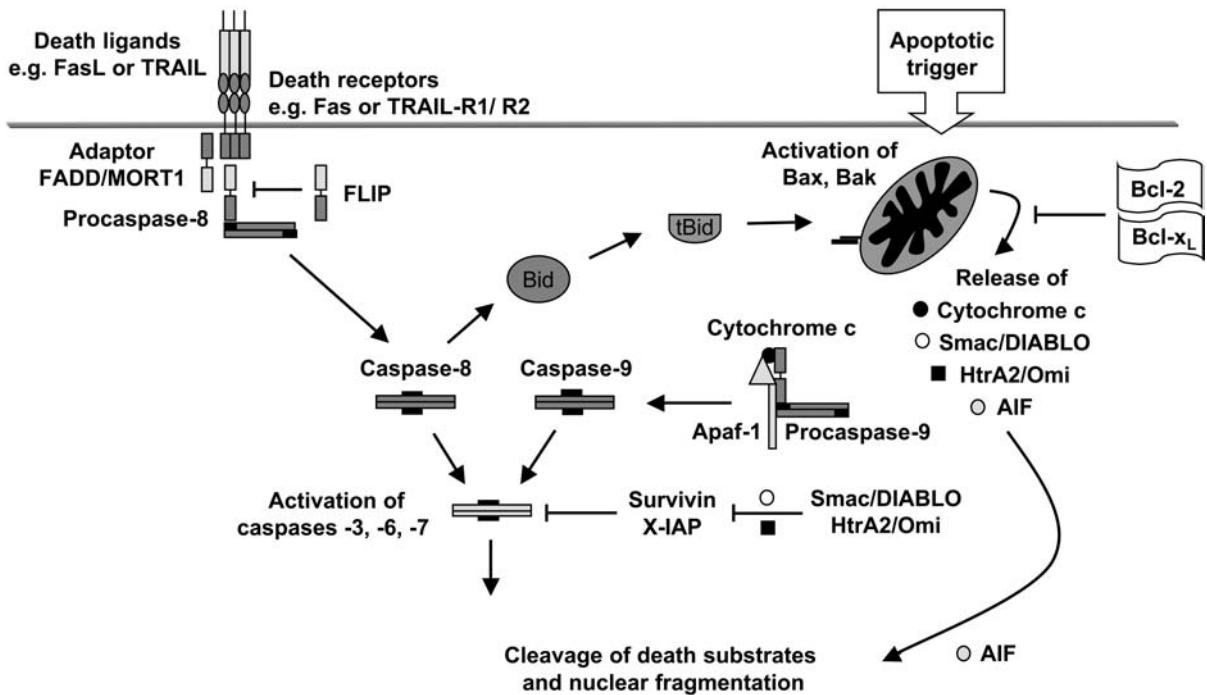
Characteristics

Apoptosis is characterized by the following biochemical features: high molecular weight and/or internucleosomal DNA fragmentation (DNA laddering), ►phosphatidyl serine externalization and intracellular proteolysis. Some of the plasma membrane changes are responsible for the rapid recognition of apoptotic cells by phagocytes (neighboring cells and resident tissue macrophages) before the release of intracellular pro-inflammatory substances into the surrounding tissue occurs. Therefore, in contrast to ►necrosis, which represents a common pathological form of cell death, apoptosis does not induce an inflammatory response. There are various apoptotic stimuli that are partially specific for distinct cell types. The susceptibility to apoptosis induction is modulated by several factors, such as the cellular environment, cell type-specific characteristics in gene expression, the differentiation state or the cell-cycle phase. However, once initiated, the cascade of morphological and biochemical events by which apoptosis proceeds is largely stereotypic and entails common mechanisms which are described below. Despite its emerging complexity in mammalian cells, apoptosis has also been described in lower organisms such as the nematode *C. elegans* and identified as an evolutionarily highly conserved mechanism, suggesting that it is crucial for species survival.

Molecular Interactions

The core cell death machinery is a complex structure of biochemical pathways, which includes death receptors, aspartate-specific cysteine proteases (►caspases), ►mitochondria, the ►Bcl-2 family and the family of ►inhibitors of apoptosis protein (IAP) as the most important regulators (Fig. 1). In principle, there are two alternative pathways that initiate apoptosis. One is mediated by death receptors on the cell surface (extrinsic pathway) (1); the other is regulated by mitochondria (intrinsic pathway) and triggered by stress responses such as genotoxic damage, hypoxia, free radicals or growth factor withdrawal (2). In both pathways, caspases are activated that execute cell death by cleaving cellular substrates to produce the structural and morphological changes described above.

To activate the extrinsic pathway, death receptors of the ►tumor necrosis factor (TNF)/nerve growth factor



Apoptosis, Regulation and Clinical Implications. Figure 1 Apoptotic signaling pathways can be divided into two main routes, the death receptor pathway (extrinsic) and the mitochondrial pathway (intrinsic). Both pathways involve the activation of initiator caspases such as caspase-8 and caspase-9 and converge at the level of effector caspase (caspases-3, -6 and -7) activation. The pathways can be blocked at several levels. In many cells, a cross-talk between the two pathways is provided by Bid, a BH3 only pro-apoptotic member of the Bcl-2 family.

(NGF) receptor family such as ▶Fas, TNF and ▶TRAIL receptors recruit, upon binding of their cognate ligands, molecules of the ▶death inducing signaling complex (DISC). Recruitment of pro-caspase-8 to the DISC results in its autoactivation and active caspase-8 then cleaves the BH3-only family member of the Bcl-2 family Bid to produce its truncated form tBid. Although the direct activation of pro-caspase-3 by caspase-8 may occur in cells expressing large amounts of caspase-8 (type I cells), in the vast majority of cells mitochondria are required for the efficient execution of death (type II cells). tBid either alone or in combination with other molecules induces oligomerization of the pro-apoptotic Bcl-2 family members Bax and/or Bak, thereby stimulating mitochondria to release a plethora of pro-apoptotic factors including ▶cytochrome c, ▶Smac/DIABLO, HtrA2/Omi and apoptosis-inducing factor (AIF), as a consequence of increasing outer membrane permeability. This critical step is blocked by anti-apoptotic Bcl-2 family members, including Bcl-2 and its homologues ▶Bcl-x_L, Bcl-w, Mcl-1 and A1. To this machinery further complexity is added by the presence of the BH3-only proteins Bim, Bmf or the p53 regulated Noxa and PUMA, which serve as sentinels that assist

the pro-apoptotic functions of Bax and Bak. Thus, at the mitochondria the extrinsic and the stress-induced intrinsic pathways converge into a complex switch-board where the life-death decision is made.

Upon release from mitochondria into the cytosol, cytochrome c initiates the formation of a complex known as the “▶apoptosome” by binding to an adapter molecule called Apaf-1, which further recruits ATP to activate pro-caspase-9. Activated caspase-9 cleaves the effector pro-caspases-3, -6 or -7, which then induce proteolysis of the final death substrates. Activation of effector caspases can be blocked by the IAP family members c-IAP1, c-IAP2, X-linked IAP and survivin, which bind to and inhibit the processing of death proteases. Although the presence of IAPs serves as a safeguard mechanism to avoid unwanted or premature induction of apoptosis, it is clear that their activities in turn must also be negatively regulated under conditions where progression of apoptosis is compelling. For this purpose, the “inhibitors of inhibitors” Smac/DIABLO and HtrA2/Omi, which bind to and inactivate IAPs, are released from activated mitochondria in conjunction with cytochrome c.

Normal cell turnover through properly regulated apoptosis is particularly important to maintain the

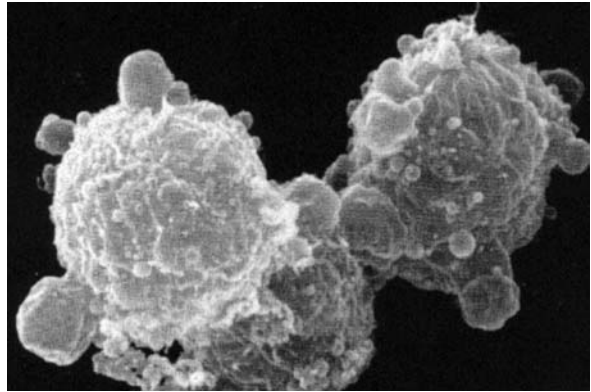
function of the immune system and other complex physiological processes in higher organisms. A look at the molecular basis of many diseases reveals an apoptotic component that either accounts for the disease or contributes to disease progression. The growing knowledge of the apoptotic machinery together with technological developments and advances in combinatorial chemistry and gene therapy have paved the way for the more effective treatment of diseases associated with deregulated apoptosis. The recent progress in apoptosis research and its implication for disease management and treatment are discussed below.

Clinical Implications

Cancer

Failure of proliferating tissues to maintain normal cell turnover may result in cell accumulation, neoplastic cell expansion and tumor formation. Moreover, defects in apoptosis signaling create a permissive environment for mutations and genetic instability. These alterations allow cells to disobey ►[cell-cycle checkpoints](#), facilitate growth factor and oxygen-independent survival and support anchorage-independent growth, e.g. during metastasis. Anticancer agents and ionizing radiation activate the apoptosis machinery in cells, a process that constitutes the main part of their cytotoxic effects (Fig. 2). Since apoptosis induced by anticancer agents converges into a final common death pathway, disruption of this mechanism may confer resistance to a broad spectrum of treatments. In addition, immune cells induce tumor cell apoptosis by binding with their death ligands to cognate death receptors on the tumor cell surface or by triggering the granzyme B/perforin pathway. Hence, apoptosis resistance also enables tumor cells to escape destruction by the immune system and recent evidence suggests that tumor cells can counterattack infiltrating immune cells (2).

Apoptosis resistance in cancer is mediated by diverse mechanisms that take effect at distinct points in the death pathway, resulting in its disruption. Many of the underlying molecules as well as their cancer-specific alterations have been identified and provide targets for rationally designed intervention approaches (3). In principle, they separate into two major classes: (a) (candidate) oncogenes with anti-apoptotic activity which are frequently over-expressed or amplified due to transactivation or posttranscriptional modulation, and (b) (candidate) tumor suppressor genes with pro-apoptotic potential which are underrepresented due to chromosomal deletions, negative transcriptional regulation, mutations or epigenetic events. In addition, as shown for TRAIL, death receptor signaling can be attenuated by the expression of inactive decoy receptors on the cell surface (1). Important molecular players and their alterations in cancer are shown in Table 1.



Apoptosis, Regulation and Clinical Implications.

Figure 2 Raster scanning electron micrograph of lung cancer cells undergoing drug-induced cell death. The extensive blebbing of the plasma membrane occurs as a consequence of cytoskeletal changes and represents a hallmark of apoptosis.

A number of apoptosis modulating therapeutics have shown efficacy in preclinical models, and, although only a few have finally reached the stage of clinical testing, this is significant progress, given that the field as a whole is only about a decade old. A strategy, which has already met with clinical success, is the use of ►[antisense](#) technology to down-modulate the expression of anti-apoptotic proteins or of protein kinases located in survival pathways or that posttranslationally activate anti-apoptotic proteins (reviewed in 4). Alternatively, strategies can be envisaged that enhance pro-apoptotic signaling in cells by transferring BH3 peptides or small molecule BH3 mimetics to inhibit Bcl-2/Bak interaction, by use of synthetic peptides derived from Smac/DIABLO to antagonize the effect of X-linked IAP or by transfer of inducible caspase genes. The death ligand TRAIL has attracted attention as a tumor-selective agent that spares normal tissues from cytotoxicity, albeit its clinical use is still disputed, due to the issue of potential hepatotoxicity. Which of these new approaches will eventually prove to be safe and effective in the clinic has to be awaited.

Bacterial and Fungal Infections

Neutrophils represent the most common granulocyte subtype in blood, and play an important role in anti-bacterial and anti-fungal defense. Apoptosis of neutrophils is an important mechanism that maintains appropriate neutrophil numbers under physiological conditions, since approximately 10^{11} mature neutrophils are generated every day in normal adults. Delayed neutrophil apoptosis has been associated with several acute and chronic inflammatory diseases (e.g. pneumonia, sepsis, cystic fibrosis) and appears to be largely mediated by excessive production of granulocyte

Apoptosis, Regulation and Clinical Implications. Table 1 Examples of major (candidate) oncogenes and tumor suppressor genes in cancer that negatively impact the apoptosis machinery

Oncogenes (gain of function)		Tumor suppressor genes (loss of function)	
Name	Type of alteration	Name	Type of alteration
Bcl-2	Over-expression (t14;18 translocation in follicular lymphoma)	Bax	Inactivation mutation; lack of transcriptional activation in p53 deficient cells
Bcl-x _L	Over-expression	p53	Inactivation/deletion mutations at various hot spots
Mcl-1	Over-expression	p14 ^{ARF}	Promoter hypermethylation
A1	Over-expression	PUMA, NOXA	Lack of transcriptional activation in p53 deficient cells
c-FLIP	Over-expression	Apaf-1	Promoter hypermethylation
Survivin	Expressed in tumors but rarely in normal adult tissues	Caspase-2	Over-expression of antagonistic short isoform
Mdm2	Prolonged half-life due to p14 ^{ARF} silencing	Caspase-3	Deletion mutation
PKCα	Over-expression	Caspase-8	Promoter hypermethylation
PKA-R1α	Over-expression	TRAIL-R1/R2	Inactivation mutation; decoy receptors
Clusterin	Over-expression	Fas	Inactivation mutation; promoter hypermethylation
HSP70	Over-expression	PTEN	Inactivation mutation
DNA methyl-transferase	Over-expression	DAPK	Promoter hypermethylation

colony-stimulating factor (G-CSF) and granulocyte-macrophage colony-stimulating factor (GM-CSF). The induction of neutrophil apoptosis during the resolution of a neutrophilic inflammatory response can be mimicked *in vitro* by culturing the cells in the absence of sufficient concentrations of survival factors, a process called spontaneous apoptosis. Most studies have been performed on purified blood neutrophils, aiming to understand the molecular events that control apoptosis in these cells (5). ▶Tyrosine kinases, phosphoinositide 3-kinase (▶PI3K), components of the mitogen-activated protein kinase (▶MAPK) pathways, the cyclic adenosine monophosphate (cAMP)/protein kinase A (PKA) pathway and protein kinase C (PKC) pathways all play a role. An early anti-apoptotic event is the phosphorylation of Bad, which leads to its translocation into the cytosol and its inability to interact with anti-apoptotic Bcl-2 family members. Later on further anti-apoptotic effects occur by modulation of the expression of apoptosis-regulating molecules of the Bcl-2 and IAP families.

Mechanisms that control the accumulation of neutrophils at sites of inflammation probably involve all events that limit the synthesis of neutrophil survival factors in inflammatory and structural cells. Intracellular inhibitors of cytokine signaling expressed in neutrophils are the Src homology domain 2 (SH2)-containing tyrosine phosphatase-1 (SHP-1) and cytokine-inducible SH2 (CIS) containing protein 1. Evidence for a functional role of SHP-1 as an inhibitory phosphatase activated by death receptors has been obtained by comparing neutrophils from normal and SHP-1-deficient mice. Inhibition of GM-CSF-mediated survival by simultaneous activation of Fas receptors was significantly reduced in SHP-1-deficient neutrophils, indicating that SHP-1 is important in limiting anti-apoptotic signals. Such a mechanism could largely explain why SHP-1-deficient mice demonstrate persistent neutrophilic inflammation. Further support for a role of SHP-1 in regulating neutrophil numbers was obtained by the observation that this phosphatase is over-expressed in patients with severe neutropenia.

Increased enzymatic SHP-1 may block cytokine effects important for both anti-apoptosis and differentiation, resulting in low neutrophil numbers. Moreover, decreased enzymatic SHP-1 correlates with inhibition of neutrophil apoptosis, since FMLP mediates PKC-mediated SHP-1 inactivation as well as delayed apoptosis in these cells.

In sepsis, accelerated T cell apoptosis can cause a condition called immune paralysis followed by death of the patient. Caspase inhibitors have been successfully applied as therapeutic agents to prevent immune paralysis and death in experimental models of sepsis.

Viral Infections

Viral infections can cause apoptosis in the infected cell, a scenario considered to be a cellular defense mechanism to prevent viral propagation. Apoptosis is often mediated by cytotoxic T cells, which activate caspases *via* Fas receptor ligation or delivery of granzyme B into the infected target cell. Too much apoptosis can result in organ dysfunction (e.g. hepatitis).

Some viruses are able to prevent apoptosis of the host cell. They either stimulate the host cell to synthesize anti-apoptotic proteins (e.g. Epstein-Barr virus, Sindbis virus) or they express their own anti-apoptotic genes which have great structural and functional similarities to host genes (e.g., baculoviruses, poxviruses). The anti-apoptotic genes involve FLIP-, Bcl-2- or IAP-like genes.

Allergic Inflammation

Eosinophils are prominent effector cells in many chronic allergic and parasitic inflammatory responses. Like other leukocytes, they are constantly generated in the bone marrow. To maintain cellular homeostasis, aged eosinophils must die by apoptosis and become removed in the absence of inflammation. Interleukin-5 (IL-5) appears to be a specific survival factor for eosinophils, at least within the human system. Therefore, and not surprisingly, eosinophilia and high IL-5 expression levels have often been associated, especially in chronic allergic disorders such as bronchial asthma and atopic dermatitis. The phenomenon of delayed eosinophil apoptosis has been demonstrated in nasal polyp explants as well as in purified blood eosinophils from patients with atopic dermatitis and some patients with the hypereosinophilic syndrome. Moreover, it appears that the severity of asthma negatively correlates with the amount of eosinophil apoptosis in the airways. These findings suggest delayed eosinophil apoptosis as a hallmark of tissue eosinophilia (6).

Besides eosinophils, T cells also have an increased life span at the site of allergic inflammation. IL-2, IL-4 and IL-15 are important survival factors that prevent apoptosis even in activated and Fas-sensitive T cells. Both T cells and eosinophils are believed to induce

apoptosis in epithelial cells by different mechanisms and the resulting epithelial damage may largely contribute to the symptoms seen in asthma and atopic dermatitis.

Autoimmune Inflammation

Rheumatoid arthritis (RA) is a frequent autoimmune disease in which patients suffer from chronic inflammatory synovitis that is dominated by the presence of macrophages, neutrophils, lymphocytes and synovial fibroblasts. Synovial macrophages are highly resistant towards apoptotic stimuli (7). TNF- α stimulation appears to participate in the anti-apoptotic effect, since it results in increased expression of (a) FLIP, (b) the Bcl-2 family member A1 and (c) the Bcl-2 family member Mcl-1 in these cells. Therefore, anti-TNF- α therapies might partially act *via* the promotion of macrophage apoptosis. Synovial fibroblasts have also been identified as important contributors to the pathogenesis of RA and apoptosis of these cells is rare. The relative insensitivity of synovial fibroblasts to undergoing apoptosis might be the consequence of high expression of FLIP and X-linked IAP, as well as inactivation of p53. Moreover, it has been demonstrated that the synovial fluid of RA patients delays neutrophil apoptosis and interferon- β and adenosine have been identified as the key mediators responsible for this effect. Although defective Fas ligand expression on activated synovial T cells has been reported, this finding may not be of pathogenic relevance, since therapies with depleting monoclonal antibodies directed against T cells have not been effective.

Patients with **systemic lupus erythematosus (SLE)** have increased levels of soluble Fas, which competitively inhibit the Fas ligand-Fas interactions. The resulting decrease in Fas-mediated apoptosis may contribute to the accumulation of autoimmune cells in this disorder. In patients with leukopenia or thrombopenia, autoantibodies that induce massive apoptosis (often in conjunction with the complement system) may cause the disease. In other circumstances, the expression of growth and differentiation factors might not be sufficient to prevent apoptosis of hematopoietic precursor cells. In autoimmune endocrine diseases, such as type 1 diabetes, β cells of islets of Langerhans in the pancreas undergo increased apoptosis due to attacks of autoreactive Fas ligand-expressing CD4⁺ and CD8⁺ T cells. There are many more autoimmune diseases for which published reports on deregulated apoptosis exist.

One example is the autoimmune lymphoproliferative syndrome (ALPS) in which dominant mutations in the Fas receptor, Fas ligand or other genes in the Fas pathway are found (8). ALPS is clinically characterized by massive lymphadenopathy and splenomegaly, which typically presents in childhood. These symptoms

are often accompanied by autoimmune hemolytic anemia, thrombocytopenia and other autoimmune manifestations. In addition, patients have markedly increased numbers of circulating double-negative T cells (CD4⁺CD8⁻). ALPS patients also have increased numbers of circulating B cells and polyclonal hypergammaglobulinemia.

Acquired Immunodeficiency Syndrome (AIDS)

Aberrant apoptosis has been identified as the major mechanism underlying the excessive immune cell depletion associated with the pathogenesis of AIDS. Infection with the human immunodeficiency virus (HIV) is associated with increased apoptosis of CD4⁺ T cells and the development of AIDS has been directly correlated with this process. CD4 acts as a receptor for viral attachment, thus facilitating HIV infection of CD4 expressing cells. How the apoptosis machinery responds to intracellular HIV and whether this is part of a cellular defense mechanism to prevent viral propagation is unclear. Strikingly, most T cells undergoing apoptosis are not HIV-infected. Recent studies suggest that depletion of CD4⁺ T cells is driven by a general immune activation process, leading to activation-induced death in this cell population (9).

Neurodegeneration

Developmental neuronal death occurs at least partially through the process of apoptosis, which modulates neuronal number, connectivity and neuronal volume. The role of apoptosis in neurodegenerative disorders, however, is less clear. Enhanced apoptosis signaling has been discussed in ►Alzheimer's disease and ►Parkinson's diseases, spinal muscular atrophy, amyotrophic lateral sclerosis, ►Huntington's disease and ischemic stroke, but recent evidence suggests that other forms of programmed cell death are also involved. In addition, it is unclear why aberrant cell death leading to disease is more frequently associated with neurodegeneration than with other degenerative processes. It is possible that neurons contain a death program that is more readily triggered than in other cells, e.g. neurons in the central nervous system are particularly susceptible to pro-apoptotic stimuli such as intracellular calcium or free radicals. Alternatively, excessive cell death may occur in various tissues of the organism but degeneration of the neuronal system is more serious, due to the low capacity for regeneration (10).

Neuronal death is accompanied by the activation of caspases and has been detected in Alzheimer's disease (caspases-3, -6 and -9), amyotrophic lateral sclerosis and Parkinson's disease (caspases-3, -8 and -9). In models of Alzheimer's disease it was shown that neuronal cell cultures lacking either caspase-2 or caspase-12 are resistant to β -amyloid-induced death, whereas, on the other hand, fragments of β -amyloid

down-regulate Bcl-2 and up-regulate Bax expression. In addition, using substrate inhibitors, dominant-negative constructs of caspases, death receptor adaptor molecules or over-expression of X-linked IAP, it was demonstrated that both the death receptor and the mitochondrial death pathways are distinctively implicated in the various neurodegenerative diseases.

The complexity and the myriad intertwining pathways involved in neuronal demise represent a fundamental challenge for the design of effective therapies. In principle, the blockade of pro-apoptotic pathways represents a valid strategy to impede neurodegeneration. For example, the drug selegiline, which has historically been used for the treatment of Parkinson's disease because it indirectly enhances dopamine signaling, was found to also up-regulate the expression of the survival factors Bcl-2 and Bcl-x_L. Another approach is to block the death pathway further downstream at the level of effector caspases, using competing peptides, small molecule inhibitors, dominant-negative constructs of caspases or inhibition of gene expression by antisense technologies. Whether similar approaches are useful for the management of slowly progressing neurodegenerative conditions, where neuronal function fails before cell death occurs, remains to be investigated.

Osteodegeneration

Osteoporosis and osteoarthritis are two degenerative diseases of the musculoskeletal system that frequently occur in elderly individuals. Increased apoptosis is seen ►in osteocytes in osteoporosis and in ►chondrocytes in osteoarthritis patients. Among other factors, the lack of hormones (e.g. estrogen, androgens), which are physiological inhibitors of apoptosis in these cells, might play a role.

Ischemic Injury

Myocardial infarction and stroke are common disorders associated with necrosis and increased apoptosis. These diseases result from an acute loss of blood flow. The central area of ischemia is characterized by necrotic cell death, whereas further distal from this area cell death is delayed and a consequence of apoptosis. Additional tissue injury frequently occurs during reperfusion. This is due to increases in oxygen radical production and mobilization of calcium that both trigger apoptosis. Clearly, the use of agents that can block apoptosis during reperfusion would be beneficial for these patients.

References

1. Ashkenazi A (2002) Targeting death and decoy receptors of the tumor necrosis factor superfamily. *Nat Rev Cancer* 2:420–430
2. Igney FH, Krammer P (2002) Death and anti-death: tumor resistance to apoptosis. *Nat Rev Cancer* 2:277–288

3. Nicholson DW (2000) From bench to clinic with apoptosis-based therapeutic agents. *Nature* 407:810–816
4. Jansen B, Zangemeister-Wittke U (2002) Antisense therapy for cancer—the time of truth. *Lancet Oncol* 3:672–683
5. Simon HU (2003) Neutrophil apoptosis pathways and their modifications in inflammation. *Immunol Rev* 193:101–110
6. Simon HU (2000) Eosinophil apoptosis—pathophysiology and therapeutic implications. *Allergy* 55:910–915
7. Pope RM (2002) Apoptosis as a therapeutic tool in rheumatoid arthritis. *Nat Rev* 2:527–535
8. Siegel RM, Chan FKM, Chun HJ, Lenardo MJ (2000) The multifaceted role of Fas signaling in immune cell homeostasis and autoimmunity. *Nat Immunol* 1:469–474
9. Grossman Z, Meier-Schellersheim M, Sousa AE et al (2002) CD4+ T-cell depletion in HIV infection: Are we closer to understanding the cause? *Nat Med* 8:319–323
10. Stefanis L, Burke RE, Greene AL (1997) Apoptosis in neurodegenerative disorders. *Curr Opin Neurol* 110:299–305

Apoptosome

Definition

The apoptosome is a cytoplasmic complex of molecules consisting of pro-caspase-9, Apaf-1 (apoptotic protease activating factor), and cytochrome c, leading to the activation of caspase-3 in the presence of either ATP or dATP.

Activation of the effector caspase-3 is a pivotal event in apoptosis induction. A variety of cellular stresses induce caspase-3 activation by triggering the release of pro-apoptotic mitochondrial proteins, among others cytochrome c. Cytosolic cytochrome c in turn initiates the formation of a caspase-3 activating multi protein complex named the apoptosome. The apoptosome consists of six or seven molecules of each of the scaffold protein apoptosis promoting factor 1 (Apaf-1), cytochrome c, dATP and caspase-9. So, formation of the apoptosome comes along with oligomerization and activation of caspase-9, which in turn is able to stimulate caspase-3 processing.

- [Apoptosis, Regulation and Clinical Implications](#)
- [TNF Receptor/Fas Signaling Pathways](#)

APP/App

Definition

APP stands for amyloid precursor protein and its gene (*App*).

- [Alzheimer's Disease](#)

Apparent Mineralocorticoid Excess

Definition

AME is a mendelian hypertensive syndrome. The clinical signs are volume dependent salt sensitive hypertension, tendency to hypokalemia and metabolic alkalosis, low renin and low aldosterone values, responsiveness to both thiazides and spironolactone despite absence of aldosterone or any abnormal mineralocorticoid products and resemblance to licorice gluttony. The disease is caused by mutations in the renal-specific isoform gene for 11 β -hydroxysteroid dehydrogenase coding for an enzyme which is responsible for converting cortisol to cortisone. In the distal renal tubule this step is crucial for protecting the mineralocorticoid receptor which has the same affinity for cortisol as it does for aldosterone.

- [Mendelian Forms of Human Hypertension and Mechanisms of Disease](#)

Appositional Growth

Definition

Appositional growth is defined as growth of bone or a cartilage element due to the addition of new matrix by ► [osteoblasts](#) in bone, or ► [chondroblasts](#) in cartilage from the outside.

- [Bone and Cartilage](#)

APR

- [Acute Phase Reaction](#)

Aptamer

Definition

Aptamers are specific RNA or DNA oligonucleotides that are isolated from combinatorial synthetic libraries that consist of 15–60 nucleotides. Aptamers can adopt a vast number of three dimensional shapes. Due to this

property, aptamers can be produced to bind tightly to a specific molecular target. As an extraordinary diversity of molecular shapes exist within the universe of all possible nucleotide sequences, aptamers may be obtained for a wide array of molecular targets, including most proteins, carbohydrates, lipids, nucleotides, other small molecules or complex structures such as viruses. Aptamers are powerful tools for analytic, diagnostic and therapeutic applications. Aptamers are easy and cheap to synthesize, although their polyanionic character limits the number of useful applications.

- ▶ Catalytic RNA
- ▶ Monoclonal Antibodies
- ▶ Protein Microarrays as Tool for Protein Profiling and Identification of Novel Protein-Protein Interactions

Apurinic or Apyrimidinic (AP) Sites

Definition

AP site (apurinic or apyrimidinic site; commonly referred to as abasic site) describes sites in the DNA with an intact sugar phosphate backbone, but a lacking purine or pyrimidine residue.

- ▶ Base Excision Repair

AQP2

- ▶ Aquaporin-2

Aquaporin

Definition

Aquaporins are membrane water channels that play critical roles in controlling the water content of cells. These channels are widely distributed in all kingdoms of life, including bacteria, plants, and mammals. More than ten different aquaporin proteins have been found in the human body, and several diseases, such as congenital cataracts and nephrogenic diabetes insipidus, are connected to the impaired function of these channels.

Aquaporin-2

Definition

Aquaporin-2 (AQP2) is the vasopressin sensitive water channel that regulates water reabsorption in the distal nephron collecting duct. Inherited AQP2 mutations that disrupt folding, lead to nephrogenic diabetes insipidus by targeting newly synthesized protein for degradation in the endoplasmic reticulum (ER).

Arabinogalactan

Definition

Arabinogalactan is a plant-derived polysaccharide abundant in arabinose and galactose.

- ▶ Glycosylation of Proteins

Arachnodactyly

Definition

Arachnodactyly refers to long, thin fingers.

- ▶ Marfan Syndrome

Archimedes Principle

Definition

The net gravitational force on a particle suspended in solution is determined by the mass of the particle minus the mass of the displaced solvent.

- ▶ Analytical Ultracentrifugation

ARE

- ▶ AU-Rich Elements

ARF

Definition

ARF is a small, conserved GTPase of the ras superfamily involved in regulating the recruitment of several sets of coat proteins to membranes, namely COP I, the adaptor complexes AP-1, AP-3 and AP-4, and ►GGAs. The cycle of GTP-binding and GTP hydrolysis on ARF is controlled by guanine nucleotide exchange factors (ARF GEFs) and GTPase activating proteins (ARF GAPs), which have a more restricted localization than ARF itself.

- INK4a / ARF
- Vesicular Traffic

Armadillo

Definition

Armadillo is a *Drosophila melanogaster* protein, and the homologue of β -catenin, which is a component of the Wnt/Wg signaling pathway, and whose family members are present in cell-cell junctions. When allowed to accumulate and translocate to the nucleus, armadillo binds DNA binding proteins TCF/Lef and activates gene transcription. It is important in axis patterning during development.

- Desmosomes
- Wnt/Beta-Catenin Signaling Pathway

Arp2/3 Complex

Definition

Arp2/3 complex is a complex of seven proteins that nucleates the formation of new actin filaments branching off existing filaments.

- Rho, Rac, Cdc42

Array

Definition

Array defines a systematic arrangement of objects. In the case of microarrays for genomic analyses, the

arrangement is either cDNA clones or oligonucleotides (based on accurate sequence information) spotted at high density on a membrane, suitably coated glass slide or other appropriate material employed in hybridization based techniques to assay RNA or DNA abundance.

- ES Cell Differentiation as a Model System for Functional Genomics
- Protein Microarrays as a Tool for Protein Profiling and Functional Proteomics

ARS

- Autonomously Replicating Sequence

ARVs

- Antiretroviral Agents

Arylsulfatase A

Definition

Arylsulfatase A is a lysosomal enzyme that catalyses the degradative conversion of sulfatide to galactocerebroside. Inborn deficiency of arylsulfatase A causes metachromatic leukodystrophy.

- Polyadenylation

Ashkenazi

Definition

Ashkenazi is a term derived from the Hebrew *Askenaz* ("German"). It refers to Jews who lived in the Rhine valley and neighboring France before they migrated to Slavic countries in Eastern Europe, mainly Poland, Lithuania, Ukraine, and Russia, after they were expelled by crusades in the 11th – 13th century.

- Bloom Syndrome

Asn-Linked Glycosylation

Definition

Asn-linked glycosylation is a co-translational protein modification, catalyzed by the dolichyl-diphospho-oligo-saccharide:protein-L-asparagine oligopolysaccharidotransferase complex (EC2.4.1.119), which mediates covalent attachment of the oligosaccharide Glc2Man9GlcNAc2 to acceptor asparagine residues (Asn-X-Ser/Thr) in polypeptides.

► [Methylation of Proteins](#)

Assembly

Definition

Assembly describes the joining of individual nucleotide sequences via the alignment of identical nucleotide bases. Overlapping nucleotide sequences can be used to generate a contiguous consensus sequence of a DNA molecule.

► [Shotgun Libraries](#)

Assisted Reproduction

Definition

Assisted reproduction refers to the use of assistive technologies such as artificial insemination, microsort, *in vitro* fertilization and preimplantation genetic diagnosis to achieve a pregnancy. Often used in cases of infertility or to avoid a known genetic disorder in a family.

► [Duchenne Muscular Dystrophy](#)

Association

Definition

Association describes the statistical fact that people with a disorder have a higher frequency of certain alleles and nearby markers of factors, with increasing susceptibility

than unaffected individuals. Phenotype and genotype appear together more than expected by chance.

► [Atopy Genetics](#)

► [Cleft Lip Palate](#)

► [Genetic Predisposition to Multiple Sclerosis](#)

Association Studies

Definition

Association study refers to a genetic analysis method, which in its simplest form, tests whether a specific allele is more or less common in a test population compared to a control group. Association studies have been proposed to identify the genetic determinants of complex traits such as obesity or hypertension. DNA markers (usually a large number of ► [SNPs](#)) from selected populations of affected and non-affected (control) individuals are genotyped. A significant frequency change of typed markers indicates an association between these markers and the complex trait under investigation. Unfortunately, the reproducibility of most association studies to our understanding of a disease and its genetic determinants has been modest. Criteria for “high quality” association studies have been published (see for instance: Freely associating (1999) *Nat Genet* 22:1–2). These criteria are: large sample size, small P values, biological plausibility, functional significance, independent replication in several populations, confirmation in family-based studies and high odds ratios and/or high attributable risk. One difficulty also lies in the recruitment of ideal controls.

► [Diabetes Mellitus, Genetics](#)

► [COPD and Asthma Genetics](#)

► [SNP Detection and Mass Spectrometry](#)

► [Mendelian Forms of Human Hypertension and Mechanisms of Disease](#)

Association-Dissociation Equilibrium

Definition

Association-dissociation equilibrium describes the equilibrium between the formation and disruption of transient molecular complexes of finite lifetime, which governs the equilibrium populations of unbound molecular species and molecular complex.

► [Analytical Ultracentrifugation](#)

Aster

Definition

Aster refers to a structure that is made of microtubules. It forms during the prophase of the mitosis. Due to their typical “sunburst” appearance they are called an aster. They form near the centrosomes.

► [Cell Division](#)

Asthma

Definition

Asthma is a chronic inflammatory disorder of the respiratory tract associated with (o)edema (fluid in the intracellular space) and the influx of inflammatory cells into the walls of the alveolar cells.

► [Common \(Multifactorial\) Diseases](#)
 ► [COPD and Asthma, Genetics](#)

Astral Microtubule

Definition

Astral microtubules are microtubules of the mitotic spindle, emanating from each spindle pole to form asters.

► [Mitotic Spindle](#)

Astrocyte

Definition

Astrocytes represent the major subtype of neuroglia in the central nervous system. Their name was originally coined because of their star-shaped morphology. Astrocytes provide neurons with nutrient and growth factors, contribute to the blood-brain barrier, regulate ion homeostasis of the extracellular space, and control synaptic transmitter levels. In recent years it has been

established that astrocytes sense neuronal activity and are involved in signal transmission.

► [Glial Cells](#)
 ► [In Vivo Imaging of Transgenic Mice with Fluorescent Protein Expression](#)

Asymmetric Carbon Atom

Definition

A tetrahedral carbon atom that is connected to four different ligands. An asymmetric carbon atom is a chiral center, and hence is optically active.

► [Amino Acids: Physicochemical Properties](#)

Asymmetric Cell Division

Definition

Asymmetrical cell division is a cell division that produces unequal descendants. The asymmetry may be achieved by polarity within the dividing cell, the orientation of the mitotic spindle and in special cases by the segregation of a “differentiation or stem cell determinant” into the two daughter cells.

► [Cell Division](#)

Asymmetric Unit

Definition

The minimal subunit of a crystal from which the unit cell can be generated *via* the crystal symmetry operations. The asymmetric unit can be made up of one or more copies of one or more molecules, but all are distinguishable at least in terms of relative crystal packing interactions.

► [X-Ray Crystallography—Basic Principles](#)

Definition

The mathematical relationship ($n\lambda = 2d \sin \theta$) that relates the spacing d of a diffraction grating, the wavelength λ of diffracted light, the order n of the reflected wave and the diffraction angle θ . In a three dimensional crystal, the grating spacing for first order

($n = 1$) reflections corresponds to the spacing between Miller planes.

► [X-Ray Crystallography—Basic Principles](#)

Asynchronous Replication

Definition

The two alleles of imprinted or other monoallelically expressed genes do not replicate at the same time during the mitotic cycle. This phenomenon is called asynchronous replication. In contrast, duplication of the two alleles of biallelically expressed genes usually occurs at the same time in the mitotic cycle.

► [Genomic Imprinting](#)

lymphomas and leukaemia, but about half of the patients have problems of the immune system as well. The ATM protein plays a key role in the detection and cellular response to DNA double-strand breaks.

► [Bloom Syndrome](#)

► [DNA-Repair Mechanisms](#)

Ataxia Telangiectasia-Like Disease

Definition

Ataxia telangiectasia like disease (AT-LD) is a disease with similar phenotypes compared to AT, but is caused by truncating mutations in the MRE11 gene.

► [DNA-Repair Mechanisms](#)

Ataxia

Definition

Ataxia describes the uncoordinated or unsteady movement of an affected person due to the brain's failure to regulate the body's posture and regulate the strength and direction of limb movements. Ataxia is usually a consequence of disease in the brain, specifically in the cerebellum. Ataxia may appear as a congenital defect, or following a simple viral infection such as chicken pox. It may also appear following encephalitis, head trauma, and diseases affecting the central nervous system or spinal cord. Appearance as a genetic disorder, or as a toxic reaction to drugs, medications, alcohol or environmental toxins, is also possible.

► [Prader Willi and Angelman Syndromes](#)

Ataxia Teleangiectasia Mutated Gene

Definition

Ataxia Teleangiectasia Mutated gene is a tumour suppressor gene that activates p53 through phosphorylation in response to DNA damage. Mutations or chromosomal deletions of the ATM locus are associated with a poor prognosis in chronic B-cell lymphocytic leukaemia (B-CLL).

► [Leukemia](#)

Atelosteogenesis

Definition

Atelosteogenesis refers to a group of chondrodysplasias, defined by a characteristic phenotype, consisting of tapering/hypoplastic distal long bones.

► [Bone Disease and Skeletal Disorders, Genetics](#)

Ataxia Telangiectasia (AT)

Definition

Ataxia telangiectasia (MIM 208900) (also called Louis-Bar syndrome) is an inherited autosomal recessive disease caused by mutations in the ► [ATM](#) gene (Ataxia telangiectasia mutated gene). Affected persons have a predisposition to develop tumors, particularly

AT-LD

► [Ataxia Telangiectasia-Like Disease](#)

ATM

► Ataxia Teleangiectasia Mutated Gene

Atomic Force Microscopy

FILIPP OESTERHELT

Institute for Physical Chemistry II, AG Nanoanalytics,
Heinrich-Heine University, Düsseldorf, Germany
filipp.oesterhelt@uni-duesseldorf.de

Definition

The invention of the atomic force microscope (AFM) by Binnig, Quate and Gerber in 1986 opened up a new view of the world of small structures. In contrast to the previously invented ► [scanning tunneling microscope](#), the AFM allowed the imaging of non-conducting surfaces with molecular to atomic resolution for the first time. This was made possible by using the force acting on a pointed tip-probe as feedback parameter, while the tip was scanned across the sample.

For AFM imaging almost no sample preparation is needed. The sample surfaces can be imaged in any atmosphere and also in liquids. Thus, the AFM soon developed to a powerful standard technology that is used to investigate all types of biological surfaces. Single molecules such as nucleic acids and proteins and their supramolecular assemblies can be visualized adsorbed on solid surfaces. But whole living cells can also be imaged. Membrane proteins can be analyzed in their native environment. When imaging whole native cell membranes, the interaction and spatial arrangement of membrane proteins can be investigated. The small probe-sample interaction allows repeated scanning of the same area. This makes it possible to investigate dynamic effects from the cellular level down to single molecules. Beside this wide field of biological applications, the ability of the AFM to image any surface with high resolution also leads to broad applications in materials research and quality control.

Description

To measure the force acting on the pointed probe, it is mounted on the end of an elastic spring cantilever. Any forces acting on the probe lead to a deflection of the cantilever spring. To read minute deflections without touching the cantilever, the light pointer principle is used. A laser beam is focused on the end of the cantilever and the position of the reflex is detected with

a position sensitive photodetector. With this method, it became possible to measure deflections in the Å (10^{-10} meter) range and forces down to a few piconewton (10^{-12} newton), respectively.

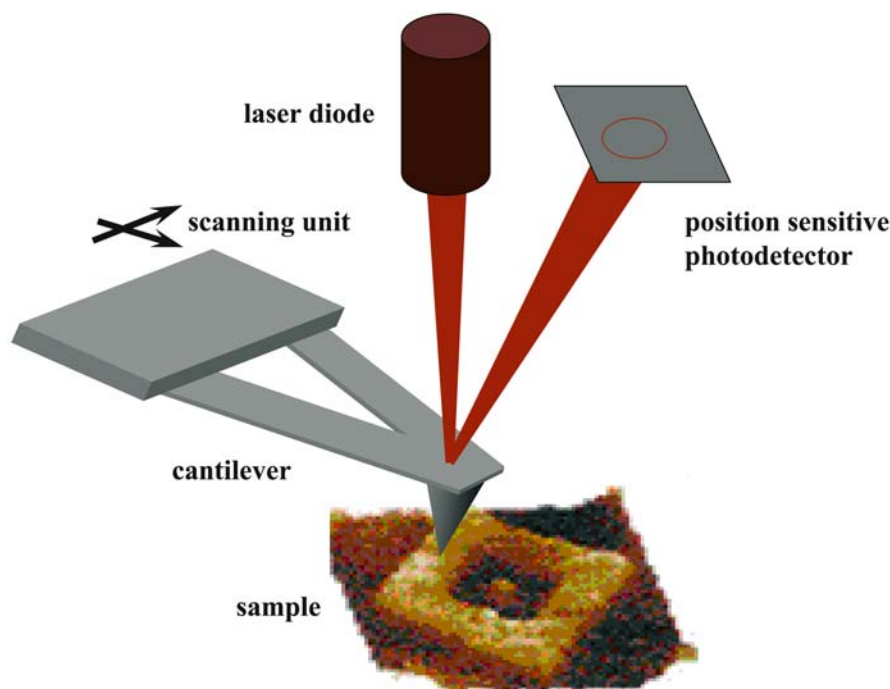
The first AFMs used diamond splinters as probes. Today a variety of probes are commercially available. Produced by semiconductor etching technologies, they are made of crystalline silicon or silicon nitride having tip radii of a few nanometers. Tips with high aspect ratios, needed for imaging narrow and deep structures, are produced by growing secondary tips on top of the probe by ► [electron beam deposition](#) (EBD). Another alternative is carbon nanotubes that are attached to or catalytically grown on the probe. These tips provided for the first time an atomically well-defined probe.

The forces acting on the tip can be of any nature. This allows imaging of many different surface properties with nanometer resolution. When the probe is scanned across the surface, the repulsive force upon contact is used to measure the height profile of the surface. To minimize the probe induced sample damage that can occur when the tip hits higher structures from the side, different modes were introduced. Called tapping mode, intermittent contact or non-contact mode, they all oscillate the tip and measure the change in the oscillation amplitude or frequency, which gives a measure for the surface probe distance without touching the surface.

But advantage can also be taken of high surface-tip interactions. Scanning the tip across the surface with intensive contact, a lateral deflection of the cantilever can be measured, indicating the friction force between tip and surface.

Upon touching the surface with the tip, the cantilever is deflected away from the surface. The characteristics of this indentation dependent bending contain information about the softness of the surface. Performing a so-called force curve (approaching the tip to the surface, indenting and retracting again) at each position on a surface, an image of the surface softness can be obtained (called force volume imaging). Upon retraction of the tip, adhesive forces between tip and surface can also be measured. Depending on the chemical composition of the tip, hydrophobic, hydrophilic and other chemical interactions can be measured with high spatial resolution. Biomolecular interactions between a functionalized tip (e.g. with antibodies) and a protein bearing surface can be used in the same way to localize proteins on the surface due to their specific interaction. These procedures are usually called adhesion imaging. Magnetic forces can also be used to produce an image contrast. In the semiconductor industry, magnetic forces are used to characterize magnetic bits of data storage devices.

Beside its ability to image surfaces with nanometer resolution, the AFM is a spring scale that allows the



Atomic Force Microscopy. Figure 1 The atomic force microscope consists of three main parts. A scanning unit that scans either the probe or the sample with sub nanometer accuracy, the probe itself, mounted on a spring cantilever which serves as a sensitive force sensor and a deflection detection system that uses the light pointer principle to achieve sub nanometer sensitivity. This simple setup makes it possible to use the instrument in any atmosphere including liquids.

measurement of forces down to the piconewton range. This made it possible for the first time to measure the interaction force of single receptor-ligand pairs like biotin-avidin or any antibody-antigen interactions. A molecule clamped between the tip of the AFM and the surface can be loaded with different forces while measuring its change in length. These experiments, called force spectroscopy, revealed for the first time the characteristic mechanical properties of single polymer molecules, which could be correlated with their chemical structure. Proteins clamped between tip and surface, unfold at typical forces and show a characteristic unfolding pattern, which reveals the presence of different stable sub domains.

Clinical Applications

Due to its ability to work under biologically relevant conditions, the AFM is applied in all the areas of medical relevance such as pharmacology, biotechnology, microbiology, molecular biology and genetics. In pharmaceutical research, the application of force microscopy allows the characterization of small particles that play an important role in pharmaceutical forms of drugs. Force measurements between individual particles make a quantitative analysis of the interaction forces and energies possible (1).

Force microscopy can be applied to image living cells. This can reveal their three-dimensional structure as well as the consistency of the cell. Force volume imaging allows the visualization of the cytoskeleton and its dynamic growth and degradation. Long-term dynamic changes such as biomineralization or cell movement can be studied. Non-contact imaging and force pulling experiments on diatoms (2) revealed the characteristics of natural adhesives that are present on their surfaces.

To grow on surfaces, cells have special requirements with respect to surface roughness and chemical nature. This is an important issue especially for implants, where cells have to settle on the surface to guarantee its biocompatibility. The AFM is well suited and is often used to characterize the surface roughness and its change in the aging process of artificial bone materials to optimize them for cell adhesion. Specific cell-surface adhesion forces can be measured directly. Cells, attached to the end of a tipless cantilever can be brought in contact with the surface and kept there for a defined time. Upon retraction, the holding force and dynamics of the rupture process can be measured (3). Analysis of the cell morphology is an important tool in drug discovery. Besides measuring cell growth, atomic force microscopy can also be applied to measure time



Atomic Force Microscopy. Figure 2 An AFM contact-mode image of conjoined NRK daughter cells. The image was recorded about 30 min after cytokinesis. Dynamically and spatially-resolved changes in cortical stiffness that play a role in cell division can be observed with the force microscope. Scale bar 20 μm . (From: Matzke R, Jacobson K and Radmacher M (2001) Direct, high-resolution measurement of furrow stiffening during division of adherent cells. *Nat Cell Biol* 3:607–610).

resolved cell volume shifts. Oberleithner *et al.* used this technique to investigate the influence of hormone exposure on cells. They analyzed the dynamic changes by repeated imaging and calculating the volume shifts between cytosol and cell nucleus (4). In the same way, dynamic processes like degradation of lipid membranes

and outer cell walls by enzyme digestion can be monitored in real time.

Another important field where the AFM became an important tool is virology. The molecular structure of small viral particles and their distribution on living cell surfaces can be characterized on living cells at different stages of the cell-cycle. The structure of the viral particles themselves can also be investigated. Images of the human immunodeficiency virus (►HIV) and HIV-infected lymphocytes showed the protein surface structure and the thickness of its outer protein shell (5).

Therapeutic Consequences

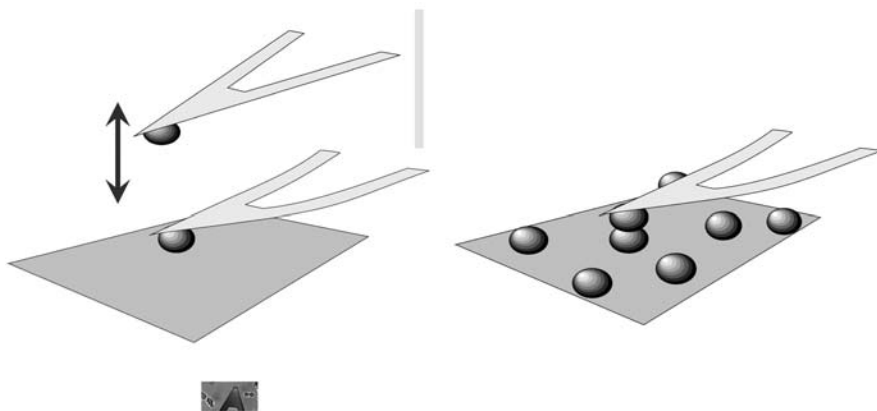
Force microscopy will have an effect on therapeutics *via* its influence on medical research. The AFM is a tool that opens up new possibilities for investigation of biological samples. Research on living cells and the possibility of observing dynamic effects in real time as well as the possibility of analyzing the smallest surface structures under native conditions will greatly aid the search for new therapies.

►Two-dimensional Crystallization of Membrane Proteins

►Nuclear Pore Complex

References

1. Hooton JC, German CS, Allen S *et al* (2003) Characterization of particle–interactions by atomic force microscopy: effect of contact area. *Pharm Res* 20:508–514
2. Gebeshuber IC, Kindt JH, Thompson JB *et al* (2003) Atomic force microscopy study of living diatoms in ambient conditions. *J Microsc* 212:292–299
3. Benoit M, Gaub HE (2002) Measuring cell adhesion forces with the atomic force microscope at the molecular level. *Cells Tissues Organs* 172:174–189



Atomic Force Microscopy. Figure 3 Single cells can be picked up by the force microscope using a tipless cantilever previously coated to enhance cell attachment. Subsequently cell-surface or cell-cell interaction forces can be measured by allowing the cell to make contact with the surface of interest or by positioning the cell on top of another cell. From (4).

4. Oberleithner H, Schneider SW, Albermann L et al (2003) Endothelial cell swelling by aldosterone. *J Membr Biol* 196:163–172
5. Kuznetsov YG, Victoria JG, Robinson Jr et al (2003) Atomic force microscopy investigation of human immunodeficiency virus (HIV) and HIV-infected lymphocytes. *J Virol* 77:11896–11909

Atopy, Genetics

SABINE HOFFJAN

Department of Human Genetics, Ruhr-University of Bochum, Bochum, Germany
sabine.hoffjan@ruhr-uni-bochum.de

Synonyms

Allergic Disorders, Genetics

Definition

Atopic disorders, including atopic dermatitis (AD), allergic rhinitis and asthma, are characterized by the development of immunoglobulin E (IgE) antibodies to ubiquitous, nonpathogenic allergens. Almost one-third of the population worldwide is affected by allergic symptoms today and the incidence of all three disorders has been on the rise over the past few decades. A genetic component for allergic diseases has long been suggested since they tend to cluster in families and twin studies have shown higher **▶concordance rates** for monozygotic compared with dizygotic twins. However, unraveling the genetic background of atopy has proven to be much more difficult than identifying genes for **▶Mendelian disorders**.

Characteristics

Atopic disorders belong to the group of **▶complex genetic disorders**, which means that no simple Mendelian pattern of inheritance can be identified as for single-gene diseases (e.g. **▶cystic fibrosis**). Complex disorders are multifactorial, with genetic as well as environmental factors contributing to disease development. Environmental factors influencing asthma/atopy development include, for example, smoke exposure, respiratory tract infections, and exposure to allergens and endotoxin in early childhood. Further, complex disorders are multigenic, which means that more than one gene influences susceptibility. Results from genome-wide screens (see below) suggest that probably many different genes are involved in risk for atopic disease, each gene contributing little to the overall atopy risk. Thus, identifying an “atopy gene”

with satisfying statistical significance can be a difficult task. Genetic studies on atopy are further influenced by the fact that expression of atopic phenotypes is variable and even members of the same family might show very different phenotypes (e.g. measurable IgE antibodies without clinical symptoms in some individuals; rhinitis, AD, asthma or often combinations of all three in other individuals). Even completely healthy individuals may carry the susceptibility gene, because in complex diseases the **▶penetrance** is usually reduced. An additional problem is the existence of so-called **▶phenocopies**, i.e. phenotypically identical diseases that are due to different underlying causes (genetic/environmental). Lastly, there is growing evidence that **▶gene-gene** as well as **▶gene-environment interactions** play an important role in asthma susceptibility. However, despite these problems, characteristic for the genetics of complex disorders, considerable progress in atopy research has been made over the last several years.

Cellular and Molecular Regulation

Positional Cloning

One approach to identifying disease-causing genes is to perform a **▶genome-wide screen**, in which several hundred markers spaced evenly throughout the genome are typed in families with the disease phenotype. Markers that are located close to the disease gene will be co-inherited (linked) with the disease. After a linked region has been found, **▶positional cloning** is performed to identify the susceptibility gene in the linked region. Alternatively, known **▶candidate genes** in the region of interest can be screened for variation directly, called **▶positional candidate cloning**. This approach is rather cost- and time-consuming, but can potentially identify all genes with detectable effects on asthma susceptibility, even if the gene function is unknown before.

Sixteen genome-wide screens for atopic disorders have been completed to date (Table 1). Because of the variability of atopic phenotypes, some studies use intermediate phenotypes [e.g. bronchial hyperresponsiveness (BHR), positive skin prick test (SPT) or total or specific IgE levels] instead of a dichotomous disease phenotype. Overall, the results differ quite substantially between study populations, but there are some chromosomal regions that are rather consistently linked to atopic phenotypes in many studies (Table 2). Among these are the **▶cytokine cluster region** on chromosome 5q, the human leukocyte antigen (HLA) region on chromosome 6p and regions on chromosomes 12q, 13q and 17q.

Recent reports of the first asthma genes identified by positional candidate cloning provide evidence that this approach can be successful. The **▶ADAM33** gene was identified as the asthma susceptibility gene in a linked

Atopy, Genetics. Table 1 Genome-wide screens for atopy phenotypes

Reference	Study population	Phenotypes
Daniels et al., 1996	Australian	<i>BHR, atopy score, skin test index, total serum IgE, eosinophil count, asthma</i>
Wjst et al., 1999	German	<i>Asthma, BHR, peak flow, total serum IgE, RAST, eosinophil count</i>
Ober et al., 2000	Hutterites	<i>Asthma, BHR, SPT to 14 allergens</i>
Dizier et al., 2000	French	<i>Asthma, SPT, BHR, total serum IgE, eosinophil count</i>
Yokouchi et al., 2000	Japanese 1	<i>Mite-sensitive atopic asthma</i>
Lee et al., 2000	European	<i>AD</i>
CSGA*, 2001	American (Caucasian, African-American, Hispanic)	<i>Asthma, total serum IgE</i>
Laitinen et al., 2001	Finnish	<i>Asthma</i>
Xu et al., 2001	Chinese	<i>FEV₁, FVC, BHR, total serum IgE, eosinophil count, SPT to cockroach and house dust mite</i>
Haagerup et al., 2001	Danish 1	<i>Allergic rhinitis</i>
Cookson et al., 2001	British	<i>AD, asthma, total serum IgE</i>
Bradley et al., 2002	Swedish	<i>AD, specific IgE, extreme AD, severity score of AD</i>
Koppelman et al., 2002	Dutch	<i>Total serum IgE, specific IgE, SPT to 16 allergens, eosinophil count</i>
Haagerup et al., 2002	Danish 2	<i>Asthma, total serum IgE, RAST</i>
Yokouchi et al., 2002	Japanese 2	<i>Seasonal allergic rhinitis, orchard grass-specific RAST, total serum IgE</i>
Hakonarson et al., 2003	Icelandic	<i>Asthma</i>

*Collaborative Study on the Genetics of Asthma

AD, atopic dermatitis; BHR, bronchial hyperresponsiveness; FEV₁, forced expiratory volume; FVC, forced volume capacity; IgE, immunoglobulin E; RAST, radio-allergo sorbent test; SPT, skin prick test

region on chromosome 20p13 (1). ►**ADAM** (a disintegrin and metalloproteinase domain) proteins are surface proteins involved in cell-cell adhesion as well as protein cleavage. *ADAM33*, which is expressed in bronchial smooth muscle cells, has been suggested to play a role in airway remodeling characteristic of chronic asthma, but functional studies have not been done yet. However, the associated variants in this gene were intronic rather than coding, and ►**haplotypes** rather than single nucleotide polymorphisms (►**SNPs**) conferred risk for asthma, with different haplotypes associated in different populations. Thus, the overall importance of this gene in asthma development might be rather modest.

Similarly, positional cloning of a quantitative trait locus on chromosome 13q14 that influences IgE levels and asthma recently identified the *PHF11* gene as another asthma/atopy gene (2). The gene product contains two

PHD (plant homeodomain) zinc fingers, which suggests that it might be involved in regulation of transcription. Three ►**SNPs** in *PHF11* were found to have independent effects on IgE levels, and these SNPs are located in introns and the 3' untranslated region of the gene, similar to the findings for *ADAM33*. Ongoing positional cloning efforts worldwide are likely to discover many more susceptibility loci, each of which has a small effect on the overall asthma or atopy risk.

Candidate Gene Studies

Candidate gene studies focus on genes with known functions that suggest they might be implicated in disease susceptibility. For atopic disorders, these are particularly genes encoding cytokines or chemokines and their receptors, IgE receptor, or transcription factors (Table 3). Genetic variation in these candidate

Atopy, Genetics. Table 2 Most consistently identified chromosomal regions in atopy genome-screens: regions showing genome-wide significance for linkage (p -value ≤ 0.000049 or a LOD ≥ 3.3) in at least one study or suggestive evidence for linkage ($p \leq 0.002$ / LOD ≥ 1.9) in at least three studies. Regions were identified by markers showing evidence for linkage that were ≤ 20 cM apart. Criteria for genome-wide significance and suggestive evidence for linkage are used according to Lander and Kruglyak (1995)

Chromosome	cM from pter [^]	Study population ⁺	Phenotypes
1p	4–20	Hutterites	asthma ($p = 0.0002$)
		Japanese 2	seasonal allergic rhinitis, total IgE ($p > 0.002$)
		Danish 2	asthma (LOD = 2.02)
2p*	10	Chinese	slope BHR ($p = 0.00002$)
2q*	173–210	Hutterites	SPT cockroach ($p = 0.00004$)
		German	total IgE ($p = 0.0016$)
		Dutch	total IgE (LOD = 1.96)
3p	52–68	Hutterites	asthma ($p = 0.0004$)
		Japanese 2	total IgE ($p < 0.001$)
		Swedish	AD ($p < 0.001$)
3q*	143	European	AD ($p = 0.0000084$)
5q*	105–172	Japanese 1	mite-sensitive asthma ($p = 0.0000013$)
		Hutterites	asthma ($p < 0.0009$)
		Dutch	total IgE (LOD = 2.73)
		Japanese 2	total IgE ($p < 0.001$)
		Danish 2	asthma (LOD = 2.20), total IgE (LOD = 2.12)
6p	34–60	Australian	eosinophils ($p < 0.0001$), atopy ($p < 0.005$), total IgE ($p < 0.05$)
		CSGA (Caucasians)	asthma ($p = 0.003$, LOD = 1.91)
		German	total IgE ($p = 0.0012$), RAST ($p = 0.0011$), eosinophils ($p = 0.0005$), asthma ($p = 0.0081$)
		Japanese 1	mite-sensitive asthma ($p = 0.0009$)
		Danish 2	asthma (LOD = 2.41)
7p	50–67	Australian	BHR ($p < 0.0005$), total IgE ($p < 0.005$), eosinophils ($p < 0.05$)
		Finnish	IgE, asthma ($p = 0.02^a$)
		French	eosinophils ($p = 0.002$)
7q*	98–109	Dutch	total IgE (LOD = 3.36), SPT aeroallergens (LOD = 1.04)
11q*	58–60	Australian	skin test index ($p < 0.00005$), total IgE ($p < 0.005$)
		CSGA (African American)	asthma (LOD = 2.00)
12q	111–134	French	eosinophils ($p = 0.0003$)
		Japanese 1	mite-sensitive asthma ($p = 0.001$)
		Dutch	total IgE (LOD = 2.46, $p = 0.0004$)
		Japanese 2	total IgE ($p < 0.001$)

Atopy, Genetics. Table 2 Most consistently identified chromosomal regions in atopy genome-screens: regions showing genome-wide significance for linkage (p -value ≤ 0.000049 or a LOD ≥ 3.3) in at least one study or suggestive evidence for linkage ($p \leq 0.002$ / LOD ≥ 1.9) in at least three studies. Regions were identified by markers showing evidence for linkage that were ≤ 20 cM apart. Criteria for genome-wide significance and suggestive evidence for linkage are used according to Lander and Kruglyak (1995) (Continued)

Chromosome	cM from pter [^]	Study population ⁺	Phenotypes
13q	6–45	Hutterites	asthma (p = 0.0006)
		Japanese 1	mite-sensitive asthma (p = 0.0004 , p = 0.001)
		Dutch	total IgE (LOD = 2.28), SPT (LOD = 1.27)
		Australian	atopy (p < 0.001)
		French	eosinophils (p = 0.002)
		Swedish	severity score of AD (p < 0.00074)
14q(*)	75–96	Hutterites	asthma (p = 0.0001 TDT)
		Icelandic	asthma (LOD = 2.66^b)
17q	62–117	French	SPT (p = 0.001), asthma ($p = 0.003$)
		Dutch	eosinophils (LOD = 1.97), SPT mite (LOD = 1.21)
		British	AD (p = 0.0004)
		Swedish	severity score of AD (p < 0.00074)

[^]Genetic (cM) distance based on the Marshfield map (<http://research.marshfieldclinic.org/genetics/>)

⁺See table 1 for references

*regions showing genome-wide significant evidence for linkage

^a after adding additional marker $p = 0.0001$

^b after adding additional markers LOD = 4.00

AD, atopic dermatitis; BHR, bronchial hyperresponsiveness; IgE, immunoglobulin E; SPT, skin prick test

genes, mostly coding SNPs leading to changes in the protein structure, is tested for ►association with atopic phenotypes. Most ►association studies are done in a case-control study design, by comparing allele frequencies in unrelated patients with unrelated controls. A significantly higher frequency of a certain allele in cases compared with controls suggests that either there is a causal relationship between variant and disease or that the variant is in ►linkage disequilibrium (LD) with the true disease-causing variant. LD is the nonrandom association of alleles at linked loci in populations and will usually only be detected over relatively small distances, indicating that the disease-causing variant must be located close to the marker. The advantages of candidate gene studies are that they do not require families and thus are easier to conduct and more cost-effective. Further, they could potentially identify variants with rather modest effects on susceptibility. However, these studies are limited to genes of known function and the present knowledge about physiological pathways important for atopy. Further, the results of association studies can be hampered by small sample sizes and population substructure.

Out of the many candidate genes that have been investigated so far, some have been rather consistently associated with atopic phenotypes across studies. These results will be reviewed in more detail.

HLA-DRB1

The human leukocyte antigen (HLA) class II genes, HLA-DR and DQ on chromosome 6p21, are intriguing atopy candidate genes because the main function of class II molecules is to present antigen to the T lymphocyte receptor. Thus, the very first genetic association studies with atopy focused on the role of these genes in specific IgE responses, and since the 1980s, many groups have reported an association of the *HLA-DRB1* locus with specific IgE responses to a wide variety of antigens, including ragweed pollen, birch pollen, rye grass pollen, cat dander, mold, olive pollen, *parietaria*, and house dust mites. More recent studies found associations of this locus with total serum IgE, atopy and asthma. Although the associations and phenotypes vary quite substantially between studies, overall most studies consistently imply that *HLA-DRB1* is an atopy susceptibility locus. However, there

Atopy, Genetics. Table 3 Candidate genes that have been associated with atopy and asthma-related phenotypes in at least two independent studies

Gene	Gene name	Chromosomal location
<i>IL10</i>	Interleukin 10	1q31
<i>CTLA4</i>	Cytotoxic T lymphocyte-associated 4	2q33
<i>CCR5</i>	Chemokine, CC motif, receptor 5	3p21
<i>IL4</i>	Interleukin 4	5q31
<i>IL13</i>	Interleukin 13	5q31
<i>CD14</i>	Monocyte differentiation antigen CD14	5q31
<i>IL12B</i>	Interleukin 12B	5q31–33
<i>ADRB2</i>	Beta-2 adrenergic receptor	5q32–34
<i>LTC4S</i>	Leukotriene C4 synthase	5q35
<i>HLA-DRB1</i>	Major histocompatibility complex, class II, DR beta 1	6p21
<i>TNF</i>	Tumor necrosis factor	6p21
<i>LTA</i>	Lymphotoxin alpha	6p21
<i>TAP1</i>	Transporter, ATP-binding cassette 1	6p21
<i>PAFAH</i>	Platelet-activating factor acetylhydrolase	6p21
<i>CFTR</i>	Cystic Fibrosis transmembrane conductance regulator	7q31.2
<i>NOS3</i>	Nitric oxide synthase 3	7q36
<i>NAT2</i>	N-acetyltransferase 2	8p22
<i>CC16</i>	Clara cell protein 16 kD, uteroglobin	11q12–13
<i>FCER1B</i>	High affinity IgE receptor, beta chain	11q13
<i>GSTP1</i>	Glutathione S-transferase P1	11q13
<i>STAT6</i>	Signal transducer and activator of transcription 6	12q13
<i>IFNG</i>	Interferon gamma	12q21
<i>NOS1</i>	Nitric oxide synthase 1	12q24
<i>IL4RA</i>	Interleukin 4 receptor alpha chain	16p12
<i>RANTES</i>	Regulated upon activation, normally T-expressed and secreted	17q11–q12
<i>EOTAXIN</i>	Eotaxin	17q21
<i>ACE</i>	Angiotensin-1 converting enzyme	17q23
<i>TGFB1</i>	Transforming growth factor beta 1	19q13.1

are several other interesting candidate genes in the linked region on chromosome 6p21, including the tumor necrosis factor alpha (*TNF*) and lymphotoxin-alpha (*LTA*) genes both of which have been associated with atopy-phenotypes in several studies, although negative results have been reported as well. It is also possible, if not likely, that variation in another gene in

the HLA region confers atopy risk and is in LD with the reported variants.

Genes in the Interleukin (IL)-4/IL-13 Pathway

IL-4 and IL-13 are important T helper type 2 (Th2) cytokines that mediate inflammation in asthmatic airways. The genes encoding both cytokines are

adjacent and located in the cytokine cluster region on chromosome 5q31, which has been linked to atopy or asthma in five genome screens (see above). Their receptors share a common α -chain, encoded by the *IL4RA* gene on chromosome 16p12. All three genes have rather consistently been associated with atopy-related phenotypes.

IL13

Two functional SNPs have been described in the *IL13* gene: a promoter polymorphism at position –1112 that alters regulation of IL-13 production and increases binding of nuclear proteins to this region and an amino acid exchange in exon 4 (Arg110Gln). The –1112 SNP was associated with asthma, BHR and SPT in two different Dutch populations, and Arg110Gln was associated with total IgE levels, atopic dermatitis, asthma, and atopy in seven different studies. There are only a few studies that failed to replicate an association of this SNP with atopic phenotypes. Thus, *IL13* is likely to be an atopy susceptibility locus in the 5q31 region, even though no study to date has shown that variation in *IL13* accounts for the evidence of linkage to this region (3).

IL4

The *IL4* gene is another interesting candidate gene in the 5q31 region, but association results are not as consistent as for *IL13*. A functional SNP at position –509 in the promoter region has been associated with total and specific IgE, childhood asthma, asthma severity, AD, and lung function. However, negative results were reported for childhood asthma, atopic asthma, and AD and the promoter SNP did not account for the evidence of linkage to the 5q cytokine cluster. A second promoter region SNP at position +33 was associated with total IgE levels in one study but not with AD in another.

IL4RA

Ten amino acid substitutions have been described in coding regions of the *IL4RA* gene on chromosome 16p12, three of which were shown to be functional (Ile50Val, Ser478Pro and Gln551Arg). Association studies on single SNPs have been conducted in many populations with conflicting results; while overall most studies reported an association of *IL4RA* variation with asthma-related phenotypes, the associated alleles differed quite substantially between populations. Therefore subsequent studies evaluated multiple SNPs simultaneously in the form of haplotypes and found haplotypes to be more informative than single SNPs and significantly associated with atopy- or asthma-related phenotypes (4). Overall, associations between coding region variation and asthma or atopic

phenotypes in at least nine independent studies indicate that the *IL4RA* gene is an atopy susceptibility locus.

ADRB2

β_2 -adrenoceptor agonists that activate β_2 -adrenoceptors on the bronchial smooth muscle cells are the most widely used bronchodilator drugs in asthma. Since long-term treatment results in down-regulation and desensitization of receptors in some, but not all asthma patients, it was proposed that genetic variation in the gene encoding the β_2 -adrenoceptor (*ADRB2*) gene might influence drug response. Two common coding SNPs, Arg16Gly and Gln27Glu, have been shown to be functional, with the 16Gly allele leading to increased, and the Glu27 allele to decreased receptor down-regulation. Most case-control studies to date have not found an association between either of these variants and asthma *per se*, but rather implicate *ADRB2* as a disease-modifying gene (e.g. associations with steroid-dependent asthma, nocturnal asthma, asthma severity, and lung function). Clinical trials investigating the role of these polymorphisms in long-term drug response gave conflicting results. However, Drysdale et al. recently examined haplotypes comprised of 13 SNPs in promoter and coding regions of *ADRB2* and reported associations between certain haplotypes and *in vivo* bronchodilator response as well as *in vitro* expression levels, indicating that haplotypes might have a greater predictive power than individual SNPs as pharmacogenetic loci (5). Overall, there is evidence that variation at the *ADRB2* locus might indeed influence drug response, but the relative importance of this effect on treatment outcomes and the specific SNPs or haplotypes that influence response remain to be determined.

CD14

CD14 is the primary receptor for lipopolysaccharide (LPS, or endotoxin). LPS has attracted much attention recently because there is growing evidence that exposure to LPS in the first years of life might prevent allergic sensitization by stimulating the early development of a Th1-like immune response. Association studies with the *CD14* gene on chromosome 5q31 suggest that variation in this gene might interact with the environment (i.e. endotoxin levels) in conferring risk for atopic diseases. A functional SNP in the promoter region of this gene (–159C/T) was identified, and the –159T allele was associated with reduced levels of total serum IgE in the Tucson Respiratory Study (Arizona, USA) as well as in the Dutch population. However, the same allele was associated with increased risk for atopy in the Hutterites, a farming population, and neither allele showed an association with asthma phenotypes in a German allergy study or in the Icelandic population. Vercelli recently suggested an

intriguing explanation for these discrepant findings; it is possible that the amount of endotoxin load to which a population is exposed could modulate the effect of the genotype at this locus (6). While in the Tucson children with an intermediate exposure to LPS, the T allele seems to be protective by increasing soluble CD14 levels and lowering IgE levels, this protective effect might not be seen in a population that has a very low endotoxin load (German), and the opposite effect might be seen for a population that has constant contact with livestock and LPS like the Hutterites. This is an interesting example of a gene-environment interaction that needs to be confirmed in additional populations.

Clinical Relevance

Identifying genetic variation that influences the development of atopy could have several important clinical implications. First, knowledge about genetic susceptibility could help to identify individuals at risk for atopic disease at a very early age and develop specific prevention strategies. Second, since treatment response to certain pharmaceuticals might be influenced by genetic variation (as in the example of *ADRB2*, see above), specific treatments for genetically different individuals could be developed. Lastly, a better understanding about the complex genetic and molecular background that underlies atopic diseases could lead to the invention of new diagnostic and therapeutic strategies in the future (7).

References

1. Van Eerdewegh P, Little RD, Dupuis J et al (2002) Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness. *Nature* 418:426–430
2. Zhang Y, Leaves NI, Anderson GG et al (2003) Positional cloning of a quantitative trait locus on chromosome 13q14 that influences immunoglobulin E levels and asthma. *Nat Genet* 34:181–186
3. Vercelli D (2002) Genetics of IL-13 and functional relevance of IL-13 variants. *Curr Opin Allergy Clin Immunol* 2:389–393
4. Ober C, Leavitt SA, Tsalenko A et al (2000) Variation in the interleukin 4-receptor alpha gene confers susceptibility to asthma and atopy in ethnically diverse populations. *Am J Hum Genet* 66:517–526
5. Drysdale CM, McGraw DW, Stack CB et al (2000) Complex promoter and coding region beta 2-adrenergic receptor haplotypes alter receptor expression and predict in vivo responsiveness. *Proc Natl Acad Sci USA* 97:10483–10488
6. Vercelli D (2003) Learning from discrepancies: CD14 polymorphisms, atopy and the endotoxin switch. *Clin Exp Allergy* 33:153–155
7. Hoffjan S, Epplen JT (2005) The genetics of atopic dermatitis: recent findings and future options. *J Mol Med* 83(9):682–692

ATP

Definition

Adenosine 5'-triphosphate (ATP) is the main source of energy in the cell. It is used to provide most of the cellular processes with energy. ATP is primarily produced within the mitochondria, but the glycolysis that takes place in the cytoplasm is of importance for ATP production.

► DNA Ligases

► Mitochondrial Myopathies

ATPase

Definition

ATPase is an enzyme that hydrolyzes ATP, releasing energy for other cellular activities.

► Splicing

ATP-Binding Cassette (ABC) Transporters

Definition

ABC Transporters comprise of a large protein family of transmembrane proteins that contain an ATP-binding cassette (ABC). ABC proteins are low capacity-but high affinity transporters able to transport substrates against a concentration gradient of up to more than 10,000 fold. Hydrolysis of ATP is required for substrate transport. ABC transporters are mainly either import or export pumps, bi-directional ABC proteins appear to be rare exceptions. Substrates include lipids, sugars, amino acids, peptides, proteins, metals, (in)organic ions, toxins, and antibiotics. Some ABC transporters are able to pump anticancer drugs out of the cells, thus making them drug resistant.

► Biological membranes

► High-HDL Syndrome

► Multi-Drug Resistance

► Tangier Disease

ATP-Dependent Nucleosome Remodelling

► Chromatin Remodelling

Atrophia Gyrate

Definition

Atrophia gyrate is a rare autosomal recessive disorder with clinical signs similar to those seen in Retinitis pigmentosa (RP), however, is characterized by sharply defined Garland shaped areas of chorioretinal atrophy. An elevation of the plasma ornithine level, caused by an ornithine aminotransferase deficiency, is involved in this disease.

► [Retinitis Pigmentosa](#)

Attention Deficit Hyperactivity Disorder

Definition

Attention deficit hyperactivity disorder (ADHD) is a condition affecting children and adults, which is characterised by problems with attention, impulsivity and overactivity. Today, it affects between 4–6 percent of schoolage children and between 2–4 percent of adults.

► [Fragile X Syndrome](#)

Attenuation

Definition

The term attenuation describes a mode of transcription regulation at the transcription elongation level, which is frequently used for feedback control of metabolic pathways in bacteria.

► [Transcription Elongation](#)

AUG Translational Start Site

Definition

AUG translational start site is the first translated codon within the gene/mRNA strand, and the position from which peptide synthesis starts.

► [Morpholino Oligonucleotides, Functional Genomics by Gene ‘Knock-down’](#)

AU-Rich Elements

Definition

AU-rich elements (ARE) are short regions rich in A and U nucleotides, which are located in the 3' untranslated regions of short-lived mRNAs, including many encoding cytokines, proto-oncoproteins, and growth factors.

► [RNA Stability](#)

Aurora Kinases

Definition

Aurora kinases represent a novel family of serine/threonine kinases. They have been identified as key regulators of the mitotic cell division process, and play a role in cell transformation and tumorigenesis. Three members of this kinase family have been identified so far, referred to as Aurora-A, Aurora-B and Aurora-C kinases. They are close homologues of the prototypic yeast Ip11 and Drosophila aurora kinases, which are known to be involved in the regulation of centrosome function, bipolar spindle assembly and chromosome segregation processes. Expression of Aurora kinases is elevated in human cancer cells.

► [Chromosome Condensation](#)

Autoantibody

Definition

Autoantibody refers to an antibody that is directed against a self-molecule.

► [Autoimmune Diseases](#)

Autoantigen

Definition

Autoantigen is a self-molecule (nucleic acid, protein, glycolipid etc.) that is recognized by an autoantibody or an autoreactive T lymphocyte.

► [Autoimmune Diseases](#)

Autocorrelation

Definition

Autocorrelation defines a mathematical measure to compare a single signal with itself at different times with respect to the information it carries. In fluorescence correlation spectroscopy (FCS) analysis, autocorrelation functions have distinct decay times reflecting characteristic temporal processes, e.g. the diffusion of a molecule through the measurement volume.

► FCS

► Molecular Dynamics Simulation in Drug Design

Autocrine

Definition

Autocrine refers to the action of small inducible polypeptides (e.g. cytokines in the immune system) that act on the same cells which produce them.

► Growth Factors

► Rheumatoid Arthritis

Autogenous Translational Control

Definition

Autogenous translational control describes a mechanism by which the protein product encoded by an mRNA, regulates the amount of its own synthesis, by altering the efficiency of its own translation.

► Translational Frameshifting, Non-standard Reading of the Genetic Code

Autoimmune Diseases

WESTLEY H. REEVES

Division of Rheumatology & Clinical Immunology,
Center for Autoimmune Diseases, University of
Florida, Gainesville, Florida, USA
whreeves@ufl.edu

Definition

The immune response to foreign ►antigens is generated in two steps: 1) an early (hours to days) ►innate immune response during which molecules expressed by the foreign invader (e.g. bacteria, virus, fungus, protozoa) but not by the host's cells are recognized as "dangerous" by pattern receptors on antigen presenting cells and 2) an adaptive immune response (days to weeks after antigen exposure) mediated by antigen-specific ►T and ►B lymphocyte receptors. The immune system has evolved to discriminate self from non-self; it must be tolerant to self-antigens while at the same time responding to and eliminating foreign (non-self) antigens. ►Autoimmunity (the abnormal production of self-reactive T cells and/or ►autoantibodies produced by B lymphocytes) may evolve into a state of ►autoimmune disease (tissue injury mediated by these abnormal cells).

Characteristics

There are close to 100 different types of autoimmune disease (Table 1). Autoimmune diseases are a major cause of serious chronic illnesses, affecting up to 3% of the general population (1). Nearly 75% of afflicted patients are women, making autoimmune disorders one of the 10 leading causes of death for women under the age of 65 (2). Autoimmune disease may affect nearly any organ system. Sometimes it affects a single organ, such as the thyroid gland (organ-specific autoimmune disease). In systemic autoimmune diseases such as ►rheumatoid arthritis or ►systemic lupus erythematosus, many organs are involved. The societal costs can be enormous. ►Rheumatoid arthritis afflicts 2.1 million Americans (1.5 million women and 600,000 men) and costs roughly \$6000 per patient annually in direct medical costs and indirect costs such as absence from work. Lupus affects 500,000 Americans and costs an estimated \$13,000 per patient, translating to an annual cost of \$6.5 billion.

Systemic Autoimmune Diseases

The systemic autoimmune diseases are primarily diseases of women (Table 1). The most common is rheumatoid arthritis, which has a female to male ratio of ~3:1. Rheumatoid arthritis is characterized by a chronic, symmetric and erosive synovitis in association with other less frequent manifestations, such as subcutaneous nodules, lung nodules, vasculitis, and hematological abnormalities. Pathologically, the disease is associated with ectopic lymphoid tissue (pannus) consisting of lymphocytes, macrophages, and fibroblasts. Although autoreactive T cells are thought to mediate the arthritis, characteristic autoantibodies are produced as well (Table 2), including rheumatoid factor and autoantibodies to citrulline-modified proteins.

Autoimmune Diseases. Table 1 Some Examples of Autoimmune Diseases

Disease	Organ	Prevalence per 100,000 ¹	Female: male ratio	HLA association (relative risk)
Organ specific autoimmune diseases				
Addison's disease	Adrenal gland	5	9:1	N/A
Autoimmune hepatitis	Liver	0.4	9:1	B1, B8, DR3, DR4
Diabetes (Type I)	Pancreas (islets)	192	1:1	DR4 and DR3 (3)
Goodpasture's disease	Kidney	0.05	1:1	DR2 (16)
Graves disease	Thyroid	1152	9:1	DR3 (4)
Hashimoto's thyroiditis	Thyroid	792	9:1	DR5 (3)
Multiple sclerosis	Nervous system (white matter)	58	2:1	DR2 (10)
Myasthenia gravis	Nervous system (cholinergic nerve endings)	5	2:1	DR3 (3)
Pemphigus	Skin	N/A	N/A	DR4 (14)
Pernicious anemia	Stomach (parietal cells)	151	2:1	DR5 (5)
Rheumatic fever	Heart	1	N/A	N/A
Thrombocytopenic purpura	Platelets	N/A	3:1	N/A
Vitiligo	Skin	400	1:1	N/A
Systemic autoimmune diseases				
Biliary cirrhosis	Bile ducts and other organs	N/A	N/A	N/A
Systemic lupus erythematosus	Multiple	24	9:1	DR3 (6)
Polymyositis	Muscles and other organs	5	2:1	N/A
Rheumatoid arthritis	Joints and other organs	860	3:1	DR4 (4)
Scleroderma	Multiple	4	9:1	
Sjogren's syndrome	Exocrine glands and other organs	14	9:1	DR3 (6)

¹ USA, 1996 estimate (1)

N/A, not available

Systemic lupus erythematosus (SLE) is the prototype of human immune complex disease. It causes inflammation of the kidneys, skin, joints, blood, brain and other organs, occurs mainly in young women (female to male ratio 9:1 in the 15–45 year old age group) and affects African-Americans, Native American, and Hispanics much more frequently than Caucasians. Immunologically, it is characterized by the production of antinuclear antibodies reactive with double stranded DNA, Sm antigen, and other antigens. Scleroderma (systemic sclerosis) is characterized by fibrosis of the skin and internal organs, including the

lungs, intestine, and other visceral organs as well as vascular disease (Raynaud's phenomenon) and kidney failure. It is associated with autoantibodies against DNA topoisomerase I, fibrillarin (U3 ribonucleoprotein), and RNA polymerases I and III. Although polymyositis is primarily an autoimmune inflammatory disease of skeletal muscle, it is frequently complicated by lung disease, vasculitis, and skin involvement (dermatomyositis). It is frequently associated with autoantibodies against the enzymes that charge transfer RNAs with the appropriate amino acid. These autoantibodies predict high a risk factor

for developing lung involvement. Sjogren's syndrome is a systemic autoimmune disease affecting the salivary and lacrimal glands and other organs that is associated with a 44-fold increased risk of lymphoma. It is associated with autoantibodies against the Ro (SS-A) and La (SS-B) antigens.

Organ-specific Autoimmune Diseases

There are many organ-specific autoimmune diseases, some of which are listed in Table 1. The thyroid gland is most commonly affected, and autoimmunity in this organ can lead to either ►**hyperthyroidism** (Graves' disease) or hypothyroidism (e.g. Hashimoto's thyroiditis). Graves' disease is induced by stimulatory (agonistic) autoantibodies that bind to the active site of the thyroid stimulating hormone receptor (TSHR), leading to the increased production of thyroid hormone. The IgG autoantibodies can be transferred transplacentally to the fetus, resulting in transient neonatal hyperthyroidism. Hashimoto's thyroiditis is a lymphocytic infiltration of the thyroid gland that occurs most frequently in middle-aged women. It leads to the formation of ectopic lymphoid tissue resembling rheumatoid pannus, followed by destruction of the gland and ultimately hypothyroidism. Autoantibodies against thyroid peroxidase, thyroglobulin, and other thyroid components are typical, but are not thought to be involved directly in the pathogenesis of the hypothyroidism. Type 1 diabetes is caused by immune-mediated destruction of the insulin producing cells in the islets of Langerhans. The onset of Type 1 diabetes is typically in childhood. As in Hashimoto's thyroiditis, the islet cell destruction is thought to be mostly T cell-mediated, even though the disease is associated with autoantibodies against a variety of islet cell antigens, most notably glutamic acid decarboxylase (GAD), islet cell cytoplasmic antigens, and insulin. Myasthenia gravis is caused by receptor antagonistic autoantibodies, which block the acetylcholine receptor (AChR), causing muscular weakness and fatigue. Immunization of experimental animals with AChR from the electric organs of *Torpedo* species causes a similar syndrome. Like Graves' disease, transplacental passage of IgG autoantibodies results in transient neonatal myasthenia gravis. Pemphigus vulgaris is an autoantibody-mediated autoimmune disease of the skin characterized by intra-epidermal blistering. It is caused by autoantibodies against desmoglein 3, a key desmosomal antigen. The titer of these autoantibodies correlates with the severity of the disease, which can be fatal.

Autoantibodies in the Early Detection of Autoimmune Disease

Autoimmunity may evolve to a state of autoimmune disease over a period of years. There is evidence in

healthy individuals that the production of disease-specific or associated autoantibodies (Table 2) frequently predates the onset of autoimmune disease. In some cases, the appearance of specific autoantibodies precedes the disease onset by years or decades. For example, anti-mitochondrial antibodies can precede the onset of biliary cirrhosis by decades and anti-Sm or anti-DNA antibodies may appear months or years before the onset of clinical manifestations of lupus. In a population at high risk for rheumatoid arthritis (Pima Indians), the rheumatoid factor titer is predictive of disease risk. Thyroid peroxidase antibodies are highly predictive of the development of hypothyroidism, and in family members of probands with type I diabetes, 30–60% of those with one of the diabetes-related antibodies develop the disease within 5–10 years (3). This pre-clinical phase of autoimmune disease argues for the screening of healthy, at-risk individuals. The early detection of autoimmune disease will probably involve both genetic screening and screening for early biological markers, such as autoantibodies.

Genetics

Autoimmune disease is influenced by genetic, environmental, and random (stochastic) factors (4). The risk of autoimmune disease is higher among family members of affected individuals than in the general population, consistent with a genetic component. The relative risk to siblings of affected probands *vs.* that of the general population (λ_s = disease prevalence in siblings of affected individuals ÷ disease prevalence in the general population) gives an indication of the importance of genetic factors. In most autoimmune diseases, the relative risk is between 5 and 50-fold higher in the siblings of affected probands (Table 3). MHC-linked genes account for a portion of this effect (Tables 1, 3). Studies of monozygotic *vs.* dizygotic twins also illustrate the importance of heredity. If the ►**concordance** rate for monozygotic and dizygotic twins is about the same, the genetic effect is small. Concordance rates are typically 15–30% for monozygotic twins *vs.* 2–5% for dizygotic twins (Table 3). It has proved difficult to identify mutations conferring susceptibility to autoimmune diseases because a) autoimmune disorders involve multiple genes, each with only a small effect, and b) autoimmune “diseases” actually are “syndromes” that may be caused by more than one gene. Even in inbred lupus-prone (NZB X NZW) F1 mice, 10 or more susceptibility loci may contribute to disease severity in an additive fashion, consistent with a ►**threshold liability model**. The same may be true in humans. Thus, the genetics of autoimmune disease are complex. Certain mutations, such as deficiency of the early complement components or Fc γ receptor gene polymorphisms in lupus, are

Autoimmune Diseases. Table 2 Some autoantibodies associated with autoimmune diseases

Disease	Antigens recognized by associated autoantibodies
Organ specific autoimmune diseases	
Autoimmune hepatitis	Liver-kidney microsomes (cytochrome P450 2D6), smooth muscle, soluble liver antigen
Diabetes (Type I)	Islet cells, glutamic acid decarboxylase
Goodpasture's disease	Glomerular basement membrane: type IV collagen, $\alpha 3(\text{IV})\text{NC1}$
Graves' disease	Thyroid stimulating hormone receptor
Hashimoto's thyroiditis	Thyroglobulin, thyroid peroxidase
Myasthenia gravis	Acetylcholine receptor
Pemphigus	Desmoglein 3
Thrombocytopenic purpura	Platelet glycoprotein IIIa
Systemic autoimmune diseases	
Biliary cirrhosis	Mitochondrial antigens
Polymyositis	Transfer RNA synthetases, signal recognition particle
Rheumatoid arthritis	Immunoglobulin (rheumatoid factor), citrullinated peptides
Scleroderma	DNA topoisomerase I, fibrillarin, RNA polymerases I/III
Sjogren's syndrome	Ro (SS-A) and La (SS-B) antigens
Systemic lupus erythematosus	Double stranded DNA, Sm antigen

Autoimmune Diseases. Table 3 Role of genetics in selected human autoimmune diseases

Disease	Concordance rates		Sibling risk/population risk (λ_s)	
	Dizygotic	Monozygotic	Overall	Attributable to MHC
Graves' disease	0%	36%	15	N/A
Type 1 diabetes	5%	33%	15	2.4
Multiple sclerosis	3.5%	21–40%	20	2.4
Rheumatoid arthritis	4%	12–15%	8	1.6
SLE	2%	24%	20	N/A

Compiled from the following references: Vyse TJ, Todd JA. Genetic analysis of autoimmune disease. *Cell* 1996; 85:311–318; Jarvinen P, Aho K. Twin studies in rheumatic diseases. *Semin Arthritis Rheum* 1994; 24(1):19–28; Brix TH, Christensen K, Holm NV, Harvald B, Hegedus L. A population-based study of Graves' disease in Danish twins. *Clin Endocrinol (Oxf)* 1998; 48(4):397–400; Redondo MJ, Fain PR, Eisenbarth GS. Genetics of type 1A diabetes. *Recent Prog Horm Res* 2001; 56:69–89; Kahana E. Epidemiologic studies of multiple sclerosis: a review. *Biomed Pharmacother* 2000; 54(2):100–102.

strong candidates for autoimmune susceptibility genes. Other loci have been mapped by genome-wide marker screening. Interestingly, there may be overlap in the susceptibility genes contributing to several clinically distinct autoimmune diseases (5). Linkage data

compiled from genome-wide scanning in a variety of systemic and organ-specific autoimmune diseases suggest that more than half of the linkages may map non-randomly into 18 chromosomal clusters, possibly explaining the occurrence of several autoimmune

disease in a given individual or family. For instance, Hashimoto's thyroiditis is associated with a variety of other autoimmune conditions, both organ-specific (e.g. type I diabetes, pernicious anemia, autoimmune hepatitis, and ►Addison's disease) and systemic (e.g. lupus, rheumatoid arthritis, and Sjogren's syndrome). Similarly, families with more than one systemic autoimmune disorder are not unusual. Of course, familial clustering also can sometimes be explained by shared environmental influences. The environment certainly plays a role in triggering disease in the susceptible host, as illustrated by the initiation of disease by ultraviolet light in a subset of lupus patients. The role of environmental factors is emphasized further by the induction of a murine lupus-like syndrome by the hydrocarbon pristane.

Genetic Screening in the Early Detection of Autoimmune Disease

The strong familial clustering exhibited by many autoimmune diseases suggests that genetic screening may ultimately hold much promise in the early detection of individuals at risk. Unfortunately, due to the genetically complex nature of these disorders, progress in this area has been slow. Genetic testing is an area of active investigation, but generally is not yet available.

Molecular Diagnostics

The molecular cloning of cDNAs encoding many of the relevant ►autoantigens recognized by disease-associated autoantibodies has greatly facilitated the development of molecular diagnostics in autoimmunity. Molecular cloning has frequently been accomplished using human autoimmune sera to screen expression libraries followed by sequence determination. These antigens can be expressed in bacteria or tissue culture cells. By attaching a purification tag, such as a 6-histidine sequence, sufficient quantities of purified recombinant autoantigens can be isolated for a variety of immunodiagnostic tests. Synthetic peptides carrying self-►epitopes can also be used. For detecting autoreactive T-cells, which recognize short peptides in the context of major histocompatibility antigens, peptides are ideal. But this approach is not widely employed due to the technical difficulty of detecting T-cell responses. Autoantibody assays are more widespread. Peptides may not be the optimal antigens for detecting autoantibody responses. Although examples exist of linear B-cell epitopes, typical B cell epitopes are conformational and/or discontinuous sequences. Full-length recombinant autoantigens therefore are often preferable for ►antibody immunoassays, though they are not without problems, such as the absence of antigenically important post-translational modifications (e.g. phosphorylation, methylation, or glycosylation). Recombinant

autoantigens have been used in enzyme linked immunosorbent assays for many years. More recently, multiplexed fluorescent bead assays (e.g. Luminex) have been developed. Although at present used primarily for cytokine assays (6), they should be readily adaptable to autoantibody testing and will permit the rapid screening of sera against a battery of autoantigens. Carrying this principle a step further, proteome arrays made up of hundreds of protein autoantigens and/or peptides are being developed (7). These newer approaches raise hope that screening for multiple autoantibodies will become a cost-effective means of identifying individuals at the pre-clinical stage of autoimmune disease when therapy aimed at preventing evolution into autoimmune disease may be most likely to succeed. Advances in this area may come more quickly than advances in genetic screening for autoimmune disease, at least in the near-term. The question of whether early intervention can prevent autoimmune disease altogether is timely and is an active area of investigation.

►Desmosomes

►Diabetes Mellitus, Genetics

►Genetic Predisposition to Multiple Sclerosis

►Hyper- and Hypoparathyroidism

►Inflammatory Response

►Protein Interaction-Phase Display

►Protein Microarrays as a Tool for Protein Profiling and Functional Proteomics

References

1. Jacobson DL, Gange SJ, Rose NR, and Graham NM (1997) Epidemiology and estimated population burden of selected autoimmune diseases in the United States. *Clin Immunol Immunopathol* 84:223–243
2. Walsh SJ and Rau LM (2000) Autoimmune diseases: a leading cause of death among young and middle-aged women in the United States. *Am J Public Health* 90:1463–1466
3. Atkinson MA and Eisenbarth GS (2001) Type 1 diabetes: new perspectives on disease pathogenesis and treatment. *Lancet* 358:221–229
4. Vyse TJ and Todd JA (1996) Genetic analysis of autoimmune disease. *Cell* 85:311–318
5. Becker KG, Simon RM, Bailey-Wilson JE, Freidlin B, Biddison WE, McFarland HF, and Trent JM (1998). Clustering of non-major histocompatibility complex susceptibility candidate loci in human autoimmune diseases. *Proc Natl Acad Sci USA* 95:9979–9984
6. Fulton RJ, McDade RL, Smith PL, Kienker LJ, and Kettman Jr. (1997) Advanced multiplexed analysis with the FlowMetrix system. *Clin Chem* 43:1749–1756
7. Robinson WH, Di Gennaro C, Hueber W., Haab BB, Kamachi M., Dean EJ, Fournel S, Fong D, Genovese MC, de Vegvar HE et al (2002) Autoantigen microarrays for multiplex characterization of autoantibody responses. *Nat Med*. 8:295–301

Autoimmune Thyroid Disease

►Thyroid Disorders, Genetic Basis

Autoimmunity

Definition

Autoimmunity is characterized by the production of autoantibodies (by ►B lymphocytes) or self-reactive T lymphocytes resulting in an adaptive immune response against a self-molecule (protein nucleic acid, carbohydrate or lipid).

►Autoimmune Diseases

►Genetic Predisposition to Multiple Sclerosis

Autoinflammatory Disease

Definition

Autoinflammatory disease is a chronic inflammatory process, as in rheumatic diseases or atherosclerosis.

►Inflammatory Response

►Autoimmune Diseases

Automated High Throughput Functional Characterization of Human Proteins

DIETER LINK

Department Robotics & Screening, Xantos
Biomedicine AG, Munich, Germany
d.link@xantos.de

Synonyms

Cellular proteomics; high throughput functional cell-based screening; genome functionalization by arrayed cDNA transduction – GFACT; expression screening

Definition

Automated High Throughput Functional Characterization of Human Proteins

High throughput functional characterization of proteins is the analysis of gene function in mammalian cell cultures. Technically it means the automated transfection and over-expression of single ►cDNAs – or alternatively for knock down purposes ►RNAi constructs – in cell cultures grown in microplates. Subsequently the modulated cellular phenotype of interest is monitored, using a high throughput ►cell-based assay. The robotics platform is designed to handle huge amounts of parallelized transfection, expression and read-out processes. This allows analysis of large cDNA libraries comprising whole ►transcriptomes and their related ►proteomes with reasonable time and cost. Cell cultures modeling aspects of human diseases could be used to identify cDNAs likely to be involved in disease processes and thus to define novel drug targets.

Characteristics

Transfecting cDNAs into cell cultures is old hat to cellular biologists. The rather new feature of automated experimentation is that no attention has to be paid to the identity of a cDNA, before its gene product exerts a function in cells. The guiding interest of screening lies in a predefined cellular phenotype and the effect of a modulating protein on it.

The automation allows screening of the phenotypic consequences of protein expression of about 100,000 clones a month with a single robotics infrastructure.

The capability provided to test all cDNAs known and available makes it an exquisitely comprehensive approach. The ►Human Genome Project has led to the identification of an estimated 30,000 different human genes, though this number does not account for numerous ►splice variants. Still growing ►cDNA collections currently comprise up to 35,000 different cDNAs (1). ►cDNA libraries that contain all expressed genes of a certain cell type or tissue, might easily reach a million cDNA clones. Though there is large redundancy in cDNA libraries, at least some hundreds of thousands of clones have to be screened to find rare transcripts in a cell-based experiment.

Despite the growing information on the sequence of the human genome and the possibility predicting new genes and gene products, there is only poor annotation of function(s) to this expanding list of genes. Lack of functional information does not necessarily mean that the specific task of a protein, e.g. exhibiting an enzymatic activity or a being structural protein, is not apparent. Frequently, information as to its involvement in a certain cellular phenotype is not known. Automated cell-based assays reveal such functional roles by looking for modifications of a cellular phenotype.

Cell-based assays are *in vitro* systems that can be set up to reflect aspects of an *in vivo* situation either in healthy or diseased individuals.

Other highly parallelized techniques have been developed recently to describe disease conditions from a genomic perspective. Differential expression analyses between healthy and diseased individuals describe clusters of concertedly regulated genes/proteins, e.g. ►nucleotide-microarrays or protein arrays. Changed expression of a single gene in a disease condition could be interpreted as increased or reduced cellular function of this gene.

Clear and decisive information about function, interdependence and hierarchy of genes and their related transcripts and proteins cannot easily be obtained from these differential expression analyses. They require testing of numerous candidate genes in rather cumbersome follow-up analyses. The difficulty in straightforward interpretation is a serious bottleneck in identifying genes that are causative for a disease and its progression. This bottleneck hampers the definition of new drug targets.

The automated high throughput screening of proteins, as described here, could cope with the comprehensiveness of array-technologies. Furthermore it provides insight into the function of the gene products. The shortening of the list of candidate genes enabled by expression analyses is of paramount importance. Ultimately confirming animal experiments need be carried out with only a very few of the most promising candidates (2).

The modeling of a disease relevant cellular phenotype in an *in vitro* cell culture model is central to a disease-oriented strategy of cellular screening.

The cell-based model is a package of the cDNA source, type of cell culture and phenotype of interest. Transfectability issues and the development of a high throughput amenable assay are the key technical hurdles that have to be overcome.

Technical Considerations for Set-up of a Disease-relevant Screening

Choice of cDNA source

As noted above, two different sources of cDNA are to hand. First, cDNA collections, where each cDNA is represented singly and already known by sequence. They are built up to collect all available cDNAs in expression vectors usually regardless of their tissue origin. Second, cDNA libraries that comprise all transcribed genes of certain cells or tissues – from healthy or diseased individuals. But they exhibit redundancy due to numerous mRNA copies of a gene per cell. Smaller subcollections of cDNAs are conceivable. They might comprise only secreted proteins (3), G-protein coupled receptors, kinases or proteases from the outset. Here, sequence information is a must to

select by sequence-similarities (e.g. protein-family domains) or by consensus-sequences (e.g. leader sequences of secreted proteins). This inevitably bears the risk of missing cDNAs that slip under the threshold of sequence comparison or prediction algorithms. Unknown genes/transcripts are not included at all. High throughput analysis of cDNA libraries supplemented by a cDNA collection can most effectively compensate for avoiding limited cDNA sub-collection approaches that might fail to be comprehensive.

Naturally a protein could be characterized not only by its (over-) expression but also by its functional knock down. The transfection of antisense-RNAs or small-interfering double stranded RNAs (►siRNAs) leads to the elimination of a protein by destroying or blocking the transcript and thereby its translation. Depending on the protein turnover rate, consequences for a monitored cellular phenotype might occur. Approaches to build up siRNA collections – vector based or pure ►dsRNA moieties – to exploit the RNAi-effect, are currently under intense development. But collections will yield sufficient comprehensiveness for true genome-wide screenings only in the near future.

Arrayed cDNAs and Tracked Transfections into Cell Cultures

As depicted in Fig. 1a, cDNA expression vectors are arrayed on microplates (from the 96- to the 1536-well formats). Vector types commonly used are viruses or eukaryotic expression plasmids, the latter propagated and stored in bacteria, from which they are individually prepared before transfection. Transfections are usually carried out as transient transfections, where cells are allowed to express the encoded protein for some time, before the changed cellular phenotype is assessed (Fig. 1a). Incubating monitor cells with supernatants of easily transfected producer cells (see legend to Fig. 1 and an example below) is valuable not only as a resort for poorly transfectable cells, but also permits focusing on the secreted part of the proteome.

Each well gets only one type of cDNA during transfection. Transfection efficacy is crucial as it mainly determines the sensitivity of the assay. Single transfections *per se* improve sensitivity for cDNA related effects compared to pooled transfections. Pooled transfections of several different cDNAs in one cell culture vessel were frequently done in the pre-automated era. There, a single cDNA caused a phenotype in only some of the transfected cells leading to poor sensitivity in the subsequent assay. Furthermore cDNAs had to be obtained rather clumsily from the selected cells to determine their sequence. Non-selectable functions as for instance ►apoptosis, where the transfected vector inevitably gets lost in a dying cell, were impossible to screen at all. In the automated system, the ►laboratory informatics management

system (LIMS) ensures tracking of each vector by machine-readable barcodes through the whole screening cascade. In order to obtain sequence information for a cDNA-library hit, it can be backtracked to the originating DNA well, isolated and sequenced.

Cell-based Assays

The type of transfected cells is preset by the screening objective and only limited by the effort of cell culture and transfection methodology. To model a human condition as closely as possible, cell cultures of human origin are preferred over those from other species. Assays interrogate the cellular phenotype of interest and in the following some examples may illustrate the rationales.

Disease conditions, like tumorigenesis, might be investigated by screening for proteins that enhance *in vitro* tumor cell viability, proliferation or metastasis, hence driving tumor establishment, growth or spreading. Proteins found to exert this function would have to be inhibited to ameliorate the condition. In contrast, if death of tumor cells were the criterion, functional proteins would be active principles for therapeutic intervention. From a pharmaceutical point of view, the inhibition of protein function is still easier to achieve than its activation, which would normally mean the induction of expression. Secreted proteins binding to cellular receptors exerting a desired function are an exception.

More complex cellular interactions could be modeled as well. For instance, all tumors rely on enhanced supplies of oxygen when growing and this is supplied by the vasculature. Tumors tend to get control over the process of vessel formation by directly stimulating vessel growth. A tumor cell derived cDNA-library might be specifically screened for potentially novel, secreted endothelial cell proliferation factors. The cell-based assay would be a supernatant assay (see also legend to Fig. 1a). A producer cell line would be transfected with the cDNAs and the supernatants would be transferred to human endothelial cells. Their rate of proliferation would be monitored after a couple of days. Revealing the proliferative activity of a factor might provide a basis for intervention in tumor growth by developing inhibitory antibodies.

► **Homogeneous assay** formats, i.e. simple mix and read assays, are applied whenever possible to minimize handling steps by robots. Favored read-out principles are therefore colorimetric, fluorescent or luminescent detection methods determining the cell number, enzymatic activity of reporter genes or the staining patterns of specific cellular structures or marker proteins. Usually modern standard microplate reader devices could easily be integrated into a robotics and informatics infrastructure. A recent trend is the

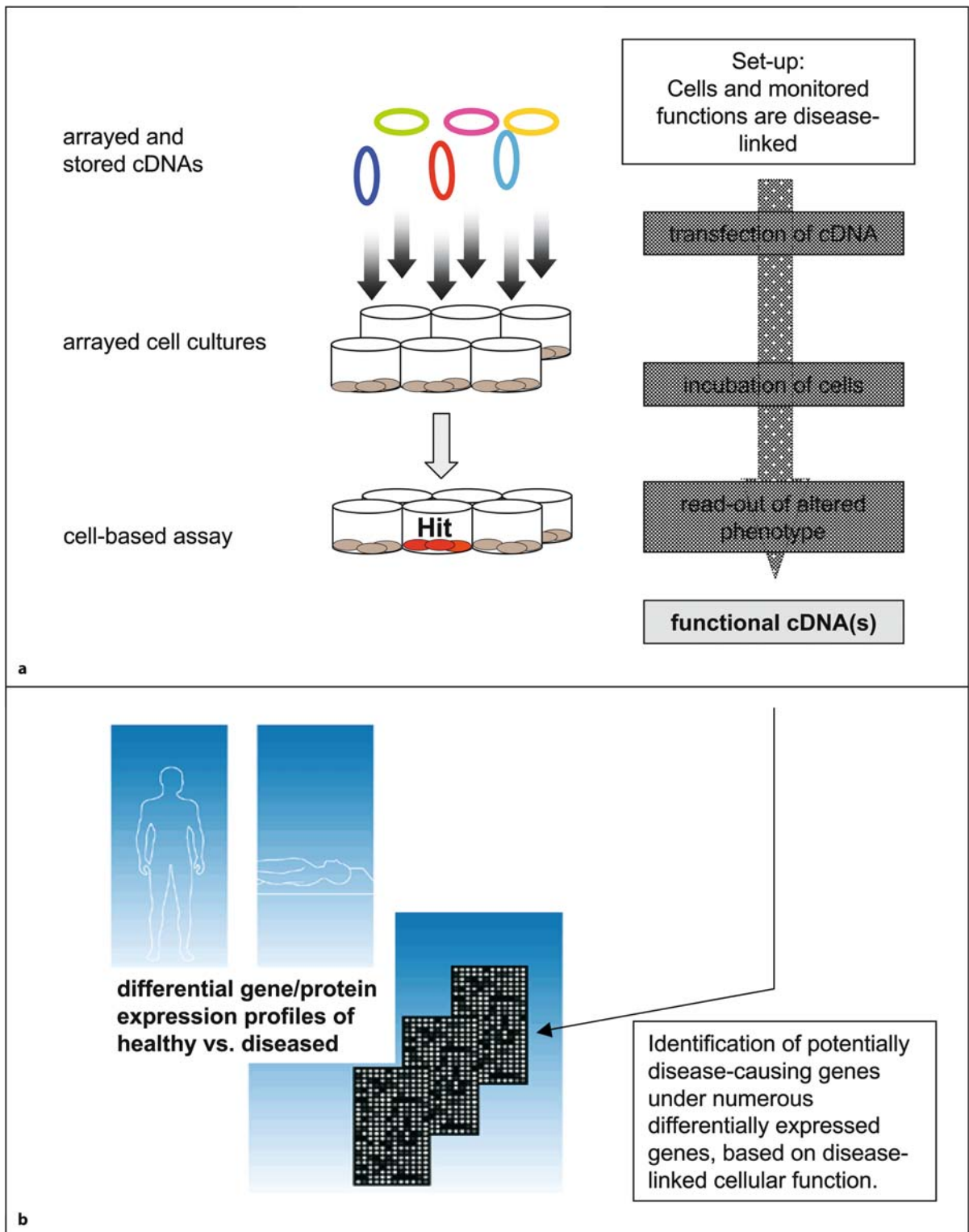
application of ► **high content screening** methodologies mainly by using highly automated imaging technologies. This automated microscopy allows monitoring of more than one parameter or object in one read-out step (4, 5). These options might be used to assess the same cellular phenotype with more accuracy or different phenotypes or objects at the same time. Current improvements in automation technologies reach these goals without considerable loss of speed.

In any case, assays have to be specifically adapted from the bench to automation. Samples (here cDNAs) will not usually be tested in triplicate. Thus, assays should exhibit a sufficient dynamic range of measurement. To discriminate between positive hits and negative ones, a measurement threshold is set. Primary hits exceed this threshold by definition. Subsequent retestings have to confirm this finding. The occurrence of false positive clones is expected since any results will yield a normal distribution in the course of multiple measurements. A small number of measurements therefore will always exceed the threshold without qualifying the hit as truly positive. To minimize such false positive events, the Z' -factor, a quantitative parameter describing the capability of an assay to discriminate between positive and negative hits could be calculated and used for assay comparison (6).

Clinical Relevance

Functional cellular screenings are located rather early in the drug-development chain. But, as shown above, they allow proceeding very fast to decisive animal experiments.

Any results of clinical sample analyses – e.g. protein level determinations in body fluids, protein expression analyses in cells and tissues with array or histological techniques or even pharmacogenomic data connecting inherited mutations with disease conditions could be interpreted in the light of novel augmented functional data (Fig. 1b). Novel hypotheses based upon the functional findings could also be obtained by comparison with expression data from clinical sample material. A close collaboration of screening institutions and clinical research is obviously desirable. Cellular functions with disease relevance should be determined in cooperation for the design of screening experiments. Clinicians could assist in making relevant cell cultures available for screening purposes or they could provide cDNAs derived from normal or diseased tissues. Screening itself will mainly be performed in the biotech or pharmaceutical industry, since it takes a considerable investment in robotics and human resources to set up the necessary screening infrastructure. Again, there is inevitably a need to prove therapeutic concepts, like inhibition or activation of function, in animal experiments prior to the entry into clinical phases.



Automated High Throughput Functional Characterization of Human Proteins. Figure 1 High throughput functional analysis of protein function and how to find and identify genes causing a disease or its progression. (a) Rationale and sequence of handling steps during cell-based screening. Given that the type of cell cultures and the cellular function under investigation are set up to model aspects of a disease, the screen validates hits as genes potentially involved in disease etiology. Single arrayed cDNAs are transfected automatically into cell

Further Applications with Clinical Relevance

The technology described here is not only able to deliver proteins with new cellular functions; de-orphanisation is an interesting mode of application of this technology as well. Finding a chemical drug's target, a receptor's ligand or a ligand's receptor is sometimes challenging (2). Such problems might be solved by analyzing the transcriptome of a cell type, where the drug or protein is known to exert effects. Assays would take advantage of an inhibited or enabled cellular function or more directly of binding studies using labeled drug or protein.

If the protein's function is causative for disease etiology and has to be blocked for therapeutic reasons, the rational would be to look for inhibitors. These might be antibodies in the case of secreted or extracellularly exposed proteins. But, with a high throughput cell-based assay performed successfully, the same assay could be used to screen for low molecular weight (LMW) compounds. These chemicals will show up directly in a cell culture assay as inhibitors of the cDNA exerted function. Finally optimization of a protein with regard to improved or altered activity, stability or specificity could be achieved by screening the respective cDNAs, in which experimentally random mutations had been introduced by mutagenesis. Numerous different mutated clones could be screened in a short time to select for the altered but desired protein properties in an appropriate cell-based assay.

Finally, a list of selected companies performing human cell-based high throughput screenings of cDNAs is: Galapagos Genomics, Mechelen/Netherlands, www.galapagosgenomics.com; Human Genome Science, Rockville, MD/USA, www.hgsi.com; Kalypsys LaJolla, CA/USA, www.kalypsys.com; Xantos Biomedicine AG, Munich/Germany, www.xantos.de.

References

1. Enssle C, Kazinski M, Koenig-Hoffmann K et al (2003) High throughput genome-wide cellular screening PharmaGenomics 1:19–23 (see also www.xantos.de)
2. Chanda SK, Caldwell JS (2003) Fulfilling the promise: drug discovery in the post-genomic era. Drug Discov Today 8:168–74
3. Chen C et al (2003) An integrated functional genomic screening program reveals a role for BMP-9 in glucose homeostasis. Nat Biotechnol 21:294–301
4. Hertzberg RP, Pope AJ (2000) High throughput screening: new technology for the 21st century. Curr Opin Chem Biol 4: 445–51
5. Beske OE, Goldbard S (2002) High-throughput cell analysis using multiplexed array technologies Drug Discov Today 7 (No.18 Suppl):131–135
6. Zhang JH, Chung TD, Oldenburg KR (1999) A simple statistical parameter for use in evaluation and validation of high throughput screening assays. J Biomol Screen 4:67–73

Autonomic Nervous System

Definition

The autonomic nervous system (ANS) is part of the peripheral nervous system, which conveys involuntary motor innervation to viscera and glands. The ANS is divided into three parts, the sympathetic nervous system, the parasympathetic nervous system and the enteric nervous system.

► Neural Crest Cells and their Derivatives

cultures grown on microplates. Transfected cells are allowed to express the cDNA and incubated to let the cellular phenotype change. Finally, depending on the kind of cell-based assay chosen, a phenotype is read out from each well. If the measurement exceeds the predefined threshold, the respective cDNA clone is termed a primary hit. The corresponding cDNA clone can be obtained from the stored cDNA library in order to get sequence information and to repeat the assay several times to dissect truly positive clones from false positives (see text). Variant approaches (not shown): In a supernatant-screen transfected cells and monitored cells are different. The supernatant of transfected cells is transferred to similarly arrayed monitor cell cultures, regardless of their transfectability. Such supernatant-screens allow focusing exclusively on the secreted part of the proteome (for an example see text). Naturally the whole automated process could also be used to define protein functions by experimentally knocking down a gene transcript rather than over-expressing it. This would be done for instance by expressing or adding short double stranded RNAs exerting a functional knock down based on the *RNAi* principle. (b) Synergism with other disease linked data. Differential expression profiles list numerous changes of transcript or protein levels in cells or tissues of affected versus unaffected individuals. The value of such results could be drastically increased by the data from a functional screen. Comparison of data allows identification of those genes that are potentially causative for a disease or its progression. Shortening the list of candidate genes from expression analyses is of paramount importance. Ultimately confirming animal experiments could be carried out with only a very few and the most promising candidates.

Autonomous Transposable Element

Definition

Autonomous transposable element denotes a mobile DNA element encoding enzyme(s) necessary for its proliferation in the host genome. All other transposable elements are called non-autonomous.

- ▶ Repetitive DNA
- ▶ Transposons

Autonomously Replicating Sequence

Definition

Autonomously replicating sequences (ARSs) have been identified in budding yeast as specific DNA sequences acting as replication origins. The ORC can bind ATP-dependently to an ARS through the 11-bp ARS consensus sequence, found within the A element of ARS1. These ARS-regions are not generally transferable to other yeast genera, mostly due to instabilities and possible integration at one or more unspecific chromosomal loci.

- ▶ DNA Replication Initiation
- ▶ Recombinant Protein Expression in Yeast

Autophosphorylation

Definition

Autophosphorylation designates phosphorylation of tyrosine residues in the same protein.

- ▶ Growth Factors

Autosomal

Definition

Autosomal refers to any of the chromosomes other than the sex-determining chromosomes (i.e., the X and Y) or the genes on these chromosomes.

- ▶ Bone and Cartilage

Autosomal Dominant (Inherited Disorder)

Definition

The term autosomal dominant describes a trait or disorder in which the phenotype is expressed in those who have inherited only one copy of a particular gene mutation, except from sex chromosomes (heterozygotes). For example, in MODY, inheritance of one copy of a mutant gene is usually enough to cause this autosomal inherited disorder in a child.

- ▶ Diabetes Mellitus, Genetics
- ▶ Tumor Suppressor Genes

Autosomal Dominant Hypercholesterolemia

- ▶ Familial Hypercholesterolemia

Autosomal Dominant Polycystic Kidney Disease

- ▶ ADPKD

Autosomal Recessive (Disorder)

Definition

The term autosomal recessive describes a trait or disorder that requires the presence of two copies of a mutated gene at a particular locus in order to express observable phenotype; specifically refers to genes on one of the 22 pairs of autosomes (non-sex chromosomes). The mutation may hereby occur in a homozygous or compound heterozygous state, and only in this case cause the autosomal recessive disorder.

- ▶ Familial Hypercholesterolemia
- ▶ Mitochondrial Myopathy
- ▶ Neurotrophic Factors

Autosomal Recessive Spinal Muscular Atrophy

► Spinal Muscular Atrophy

Autosome

Definition

Autosome describes any chromosome that is not a sex chromosome; they exist in somatic cells, normally in pairs.

► Diabetes Insipidus, a Water Homeostasis Disease

Auxotrophic

Definition

If an essential catabolic gene in an organism, e.g. bacteria or yeast, is made inoperable, the organism is not viable without addition of this catabolic product; it is auxotrophic. Upon replacement by an intact gene or a foreign gene coding for the same function, it is possible to complement this deficiency.

► Recombinant Protein Expression in Yeast

Avascular

Definition

Avascular means that no blood vessels are present.

► Bone and Cartilage

Avidin

Definition

Avidin is a 70 kDa protein found in egg white. It has a very high affinity for biotin, and is therefore

often used in coupling procedures for bioanalytical purposes.

► Surface Plasmon Resonance

Avidity/Affinity

Definition

Avidity is defined as the binding strength of multimeric or clustered molecules. The binding strength of a single molecule is called affinity.

► Focal Complexes/Focal Contacts

Axial Element

Definition

Following DNA replication in the premeiotic S-phase, the chromosomes start to condense and a proteinaceous core, called axial element, forms between the sister chromatids of unpaired chromosomes at ► Leptotene/ ► Zygotene.

Axin/Conductin

Definition

Axin and conductin (also termed axin2, axil) are scaffold proteins of the β -catenin destruction complex, and serve as negative regulators of Wnt signaling. Both proteins share 45% identical amino acids and have similar binding domains for other proteins. They bind to APC with an N-terminal RGS (regulator of G-protein signaling) domain, to GSK3 β and β -catenin with central domains, and to the Wnt signaling component dishevelled via a C-terminal DIX domain. Axin/conductin facilitate phosphorylation of β -catenin by GSK3 β and hence promote degradation of β -catenin.

► Colorectal Cancer

► Wnt/Beta-Catenin Signaling Pathway

Axis Formation – Formation and Function of the Dorsal Organizer

MASAHIKO HIBI

Center for Developmental Biology (CDB) RIKEN,
Laboratory for Vertebrate Axis Formation, Hyogo,
Japan
hibi@cdb.riken.go.jp

Definition

During vertebrate development, the embryonic axes, consisting of the dorso-ventral (DV), anterior-posterior (AP), and left-right (LR) axes, are established through a series of inductive signals that are followed by the concerted movement and differentiation of groups of cells. Axis formation is the process that generates these body axes.

Characteristics

Roles of the Dorsal Organizer in the Formation of Dorso-ventral and Anterior-posterior Axes

The dorsal **▶organizer** plays a major role in the process of axis formation (1). In 1924, Spemann and Mangold observed that equatorial cells derived from the dorsal lip of an early newt gastrula grafted onto the ventral side of an isochronic host, elicited the development of a secondary neural tube with a well-formed AP axis, as well as the formation of a secondary notochord, somites, and, in some cases, gut. In these embryos, the graft differentiated into notochord, prechordal plate, floor plate and somites, but most parts of the secondary axis, including the neural plate and somites, were derived from host cells. The dorsal equatorial cells having this activity were named the “Spemann organizer,” the “Spemann-Mangold organizer,” or simply the “dorsal organizer.” The dorsal organizer is a cell population that can generate a complete axis, in part by self-organizing, but primarily by emitting inductive signals that are involved in forming the DV and AP axes. While the original work was carried out in amphibians, an organizer region has been identified in all vertebrate embryos, including **▶zebrafish** (embryonic shield), chick (Hensen’s node), and mouse (node) (1, 2, 3). The organizer and its derivatives are also involved in the establishment of the LR axis.

Left-right Axis Formation

LR asymmetry is observed morphologically as the looping of the heart and gut, and positioning of the associated organs, such as blood vessels, lung, spleen, liver, and pancreas. Studies in model animals have shown that the LR axis is established through several

steps (4). Although there are several reports showing that LR asymmetry is initiated at the beginning of development, it is believed that the breaking asymmetry, which takes place near the organizer derivatives during or after gastrulation, contributes to the LR axis formation. In the mouse, cells on the node (the mouse organizer) have a monocilium that rotates in a counterclockwise direction and generates a leftward flow of the extra-embryonic fluid (nodal flow). The nodal flow has been shown to play a pivotal role in biasing the left-side signal on the left side. The left-side signal is transferred to the left lateral plate mesoderm (LPM), and the signal is expanded within the left LPM, resulting in asymmetric development (cell movements and differentiation) of LPM-derived tissue and associated organs.

Mechanisms that Generate the Dorsal Organizer: Dorsal Determinants

Specification of DV polarity is initiated soon after fertilization in amphibians and teleosts. In *Xenopus*, the sperm enters anywhere in the animal hemisphere of the egg. During the first cell cycle, the egg cortex rotates about 30° relative to the underlying cytoplasm along a meridian determined by the sperm entry point in a process called, “cortical rotation.” During this process, parallel microtubule arrays are formed below the cortex, with their plus ends oriented toward the future dorsal side. Although the rotation is limited to 30°, small particles, such as membrane-bound organelles, are transported about 60° along the microtubule arrays. In zebrafish, although the sperm enters at the animal pole and no cortical rotation ensues, a microtubule array arises in the vegetal region. Disruption of this microtubule array in *Xenopus* or zebrafish by ultraviolet (UV) irradiation, low temperature or the microtubule-depolymerising drug nocodazole leads to a defect in the organizer formation. Removal of the vegetal cortex material in *Xenopus* and zebrafish before the microtubule array formation results in embryos that display dorsal deficiencies. Transplantation of the vegetal cortical materials can generate a secondary axis. These studies indicate that dorsal determinants are located at the vegetal pole of amphibian and teleost eggs, and transferred to the dorsal blastomeres shortly after fertilization, leading to the dorsal organizer formation.

Wnt Signalling

Although the molecular nature of the dorsal determinants remains unknown, **▶Wnt signalling** is believed to function downstream of the dorsal determinants and to be involved in the formation of the dorsal organizer. Wnt signalling was recently categorized into two cascades: the canonical and non-canonical pathways. The canonical Wnt pathway induces the nuclear

accumulation of β -catenin and activates transcription through a complex of β -catenin and the transcription factors Tcf and Lef. The non-canonical pathway leads to the up-regulation of cytoplasmic calcium and protein kinase C, and/or the activation of small G proteins and the JNK protein kinases. Activation of the canonical pathway in *Xenopus* and zebrafish leads to ectopic induction of the dorsal organizer, and inhibition of this pathway leads to defects in the dorsal axis formation. It is still controversial at what points in the canonical pathway it is activated by the dorsal determinants. However, ►**Dishevelled** and GBP (Frat), which participate in the Wnt canonical pathway, are reported to move in a microtubule-dependent manner. A nuclear accumulation of β -catenin on the future dorsal side is observed at the midblastula stage in both *Xenopus* and zebrafish, and marks the prospective organizer region and ►**Nieuwkoop centre**. The Nieuwkoop centre is a conceptual domain that can induce the organizer non-cell-autonomously, and it resides in the dorsal-vegetal blastomere and dorsal yolk syncytial layer in *Xenopus* and zebrafish, respectively. β -catenin-mediated zygotic transcription has been shown to be required for the organizer induction. The Wnt non-canonical pathway functions to inhibit the canonical pathway that leads to dorsal axis formation. Inhibition of the calcium pathway in *Xenopus* or a mutation in the zebrafish *wnt5a* gene leads to the expansion or ectopic formation of dorsal structures, suggesting a role for the antagonistic interaction between the canonical and non-canonical pathways in early dorsal specification.

Targets of the Wnt Canonical Pathway

Several genes are reported to be direct targets of the canonical pathway during the early development of *Xenopus* and zebrafish. These include the homeobox genes *siamois* and *twin*, and the *nodal*-related TGF- β cytokine genes, *Xnr3*, 5, and 6 in *Xenopus*, and the homeobox gene *bozozok/dharma* and the *nodal*-related gene *squint* in zebrafish. These gene products are involved in the non-cell-autonomous induction at the Nieuwkoop centre and cell-autonomous establishment of the dorsal organizer. In *Xenopus*, *Twin* directly binds the promoter region of the organizer-specific gene *goosecoid*, and activates its transcription. *Xnr5* and 6 are expressed in dorso-vegetal blastomeres in a manner dependent on β -catenin as well as the T-box protein *VegT*. *Xnr5* and 6 are thought to bind serine/threonine kinase receptors and induce the expression of organizer genes. *Xnr3* functions as an inhibitor of BMP signaling, although no homologues of *Xnr3* have been reported in other vertebrate species. In zebrafish, *Squint* functions similarly to *Xnr5* and 6, and is involved in forming the dorsal organizer and mesoderm. *Bozozok* functions as a transcriptional repressor that interacts with a co-repressor, *Groucho*, and

represses the expression of the ventrally expressed homeobox genes, *vox*, *vent* and *ved*. *Vox*, *Vent* and *Ved* function as transcriptional repressors and decrease the expression of the organizer genes. De-repression of the organizer genes by *Bozozok* is involved in restricting the organizer domain in the dorsal region.

Early Axis Formation in Mouse

Analyses of mutant mice have revealed that Wnt signalling is involved in the organizer formation and early axis formation. β -catenin-deficient mice display strong abnormalities in axis formation, abnormal specification and movement of anterior visceral endoderm and defects in the formation of mesoderm and head structure. Deficiencies in *Axin1* and *APC*, which regulate the phosphorylation and degradation of β -catenin, lead to axis duplication. These reports indicate a conserved role for Wnt signalling in early axis formation. *Nodal* is required for the organizer formation in mouse. Homologues of *Siamois*, *Twin*, and *Bozozok* have not been reported in mammals. The functions of repressor-type homeobox genes, including *bozozok*, *vox*, *vent*, and *ved*, in the organizer formation and axis formation remain to be elucidated.

Molecular Interactions

Molecular Nature of the Organizer

Many genes have been identified that are expressed and function in the organizer (1, 2).

Transcription Factors (*goosecoid*, *lim1*, *foxa2/hnf3 β* /axial, and *floating head/Xnot*)

goosecoid encodes a repressor-type ►**homeobox** gene. It is expressed first in the organizer region and later in the prechordal plate. *Goosecoid* has been shown to be involved in the migration of the prechordal plate. Due to its sequence similarity to *Bozozok*, *Goosecoid* has also been suggested to function as a repressor for ventral factors, such as *Vox*, *Vent*, and *Ved*. *Lim1* is a transcriptional activator that contains both a LIM domain and a homeodomain. *Lim1* functions with *Ldb1*, which interacts with the LIM domain of *Lim1*, forming a tetramer (two *Lim1* and two *Ldb1* proteins). *lim1* is expressed throughout the mesoderm, and strongly in the dorsal organizer region and the visceral endoderm in the mouse; it is later expressed in the pronephron and a subset of neurons. *Lim1* is involved in the regulation of signalling molecules expressed in the organizer, and a *lim1* deficiency in mouse shows defects in head formation. *Foxa2* is a Fork head-type transcriptional activator that is required for the formation of the organizer and organizer-derived tissues, such as the notochord and floor plate. *foxa2* is also expressed in the endoderm. *floating head/Xnot*

is a homeobox gene that is expressed in the superficial layer of the organizer and later in the notochord. Mutations in the *floating head* gene in zebrafish lead to loss or reduction of the axial structures, including the notochord and floor plate, and instead the midline cells are fated to become paraxial mesoderm. Floating head is thought to function to repress the genetic programming for formation of the paraxial mesoderm, in which the T-box gene *spadetail* (a homologue of ►*Xenopus VegT*) is involved.

BMP Inhibitors (chordin, noggin, follistatin, cerberus, ogon/sizzled, and twisted gastrulation)

Chordin, Noggin, Follistatin, and Cerberus directly interact with ►BMPs and prevent the interaction of BMPs with their receptors. *chordin* and *noggin* are expressed in the organizer in *Xenopus* and zebrafish. *follistatin* is expressed in the *Xenopus* organizer. *cerberus* is localized to the anterior endomesoderm in *Xenopus* and the anterior visceral endoderm in mouse. Cerberus also inhibits Wnt8 and the ►Nodals. The inhibition of BMP signalling by these molecules is involved in the dorsalization of the mesoderm and endoderm, and in the induction of the neuroectoderm. The roles of the antagonistic interactions between BMP and the BMP inhibitors are supported by gain-of-function studies in *Xenopus* and the analysis of zebrafish mutants that display abnormalities in DV patterning. Mutations in the BMPs or BMP signalling components (*bmp2b/swirl*, *bmp7/snail house*, *smad5/somitabun*, or *alk8/lost-a-fin*) lead to the dorsalization and expansion of the ►neuroectoderm (dorsalized mutants). There are only two mutants, *dino* and *ogon*, that have been reported to display ventralized phenotypes. The *dino* locus encodes the *chordin* gene, and the *ogon* locus was recently shown to encode Secreted Frizzled, which is a secreted protein homologous to the ►Wnt receptor, Frizzled. Unlike the other ►BMP inhibitors, *sizzled* is expressed on the ventral side in a manner dependent on ►BMP signalling. Sizzled requires Chordin to inhibit ►BMP signalling and dorsalize embryos. Thus, Sizzled functions as a negative ►feedback regulator of BMP signalling. The precise mechanism by which Sizzled inhibits BMP signalling is not known. Chordin has been shown to be required for the development of dorsal tissues, including the neuroectoderm, in *Xenopus* and zebrafish. The activity of Chordin is modulated by the metalloproteinase Tolloid, which cleaves and inactivates the Chordin protein, and Twisted gastrulation (TSG), which can make a complex with BMP and Chordin.

Wnt Inhibitors (frzb1 and dickkopf1)

The Wnt inhibitors Frzb (sFrp3) and Dickkopf1 (Dkk1) are expressed in the *Xenopus* and zebrafish organizers. Both Frzb1 and Dkk1 inhibit the signalling of Wnt8,

which is expressed in the ventro-lateral marginal blastoderm in zebrafish (the marginal zone in *Xenopus*), and promotes dorsalization (specification of the notochord and prechordal plate) and anteriorization of the neuroectoderm. Frzb inhibits Wnt8 signalling by interacting either with Wnt8 or the Wnt receptor Frizzled, whereas Dkk1 inhibits it by interacting with the Wnt co-receptor LRP (lipoprotein receptor-related protein) 5 and 6. It is reported that Dkk1 interacts with the transmembrane proteins Kremen1 and 2 and thereby promotes the endocytosis of LRP6. *dkk1*-deficient mice show a headless phenotype, implying that Dkk1's role in head formation is conserved. In addition to these secreted molecules, Tcf3 functions as a transcriptional repressor in the absence of Wnt stimulation and inhibits the expression of targets of the Wnt canonical pathway, and Tcf3 is required for head formation in zebrafish. Wnt8 is involved in the posteriorization of neuroectoderm, and the inhibition of Wnt signalling plays a role in the formation of anterior neuroectoderm (5).

Fibroblast Growth Factors (FGFs)

►FGFs are involved in many aspects of development and have roles in mesoderm induction and patterning of the nervous system. *fgf3* and *fgf8* are expressed in a characteristic dorso-ventral gradient within the margin. The misexpression of FGF8 suppresses *bmp* expression (also BMP signalling), leads to dorsalization, and can induce a partial secondary axis. FGFs induce the expression of Sprouty4 and Sef, which function as feedback regulators of FGF signalling. The misexpression of Sprouty4 or Sef ventralizes the embryos and the inhibition of Sef in zebrafish leads to dorsalization, indicating that FGF signalling is involved in the dorsal specification.

Nodal and Nodal Inhibitors

There are three reported *nodal*-related genes in zebrafish (*cyclops*, *squint*, *southpaw*), six in *Xenopus* (*Xnr1-6*), and one in mouse (*nodal*). The *nodals* are expressed in the organizer and mesendoderm (and in the visceral endoderm in mouse). In addition to the organizer-inducing ability, the Nodals have been shown to be involved in the induction of mesoderm and endoderm, the posteriorization of neuroectoderm, and the LR patterning. Nodals bind type I and type II activin-like receptors and thereby activate Smad2 and/or Smad3. Phosphorylated Smad2 and Smad3 make a complex with Smad4 and enter the nucleus, where they bind specific DNA elements either directly or through FoxH1 and regulate gene expression. Nodal regulates the expression of *nodals* by themselves (autoregulation) as well as the feedback inhibitors Lefty1 and Lefty2. Antagonistic interactions between Nodals and Lefty1/2 are proposed to participate in the reaction-diffusion

mechanism that controls cell fates at the proper position. *lefty1* and *cerberus*-like, a mouse homologue of *cerberus*, are expressed in the anterior visceral endoderm. A deficiency in both genes in the mouse leads to ectopic formation of the primitive streak, indicating that Nodal inhibitors function to restrict the domain that generates the mesoderm and organizer. Nodal signalling is also involved in the left-right axis formation (4). *nodal* (*southpaw* in zebrafish) is expressed in the region adjacent to the definitive node. Nodal signalling is considered to be the major left-side signal, which functions first in the node region and then in the left LPM. Within the left LPM, the left-side signal is expanded through the Nodal signalling, in which Nodal activates expression of the homeobox gene *pitx2*, *left2*, and *nodal* itself. In zebrafish, the Nodal signal is transmitted to the left diencephalon and is involved in the LR-asymmetric development of the parapineal body and habenula nuclei. The LR asymmetry is maintained by a midline barrier composed of the notochord and floor plate, which express the Nodal inhibitor Lefty. A downstream target of Nodal signalling, such as *pitx2*, is involved in the induction of the LR-biased morphology.

Prospects

In the past decade, many transcription factors and signalling cascades that are involved in axis formation have been elucidated. However, several important questions remain to be answered. The nature of the dorsal determinant is still uncertain. The interactions between the gene and signalling cascades are not well understood. Cell movements, such as the convergent-extension that occurs during gastrulation and the looping of the internal organs that takes place during LR specification, have important roles in axis formation; however, the molecular mechanism that controls such cell movements is poorly understood. The involvement of additional components in early axis formation also remains an open issue. Future studies will seek to answer these questions.

References

1. Harland R, Gerhart J (1997) Formation and function of Spemann's organizer. *Annu Rev Cell Dev Biol* 13:611–667
2. Hibi M et al (2002) Organizer formation and function. *Results Probl Cell Differ* 40:48–71
3. De Robertis EM, Larrain J, Oelgeschlager M et al (2000) The establishment of Spemann's organizer and patterning of the vertebrate embryo. *Nat Rev Genet* 1:171–181
4. Hamada H, Meno C, Watanabe D et al (2002) Establishment of vertebrate left-right asymmetry. *Nat Rev Genet* 3:130–113
5. Niehrs C (2001) The Spemann organizer and embryonic head induction. *EMBO J* 20:631–637

Axon

Definition

The axon is a long cellular process of the neuron that carries the outgoing action potential from the soma (cell body) to the target neuron, and is thus the primary output station of a neurone. Axons display a high degree of structural diversity from local circuit to projection axons. Axons are usually surrounded by a myelin sheath.

► Glial Cells and Myelination

► Neurons

► Synapses

Axon Guidance

Definition

Axon guidance describes the proper navigation of neuronal extensions to their synaptic targets, which establishes the neuronal wiring of the organism.

► *Caenorhabditis Elegans* as a Model Organism for Functional Genomics

► Neurons

Axon Initial Segment

Definition

The axon initial segment is the thick and unmyelinated part of the axon close to the soma containing a high density of Na⁺ and K⁺ channels. It is thought to be the initiation zone for the Na⁺-dependent action potentials of a neurone.

► Neurons

Axonal Outgrowth

Definition

Axonal outgrowth describes a large presynaptic projection from the nerve cell body.

► Cell Polarity

B Lymphocyte

Definition

B lymphocyte (B cell) refers to a type of immune cell specialized for the production of antibodies.

► [Autoimmune Diseases](#)

β -TrCP

Definition

β -TrCP refers to an ubiquitin ligase that is responsible for adding ubiquitin to β -catenin and many other proteins, tagging them for recognition and destruction by the proteasome

► [Wnt/Beta-Catenin Signaling Pathway](#)

BAC

Definition

BAC (bacterial artificial chromosome) is a cloning vector that allows the propagation of very large exogenous DNA fragments (100 to 300 kb fragments) as large plasmids in suitable host strains. BACs are propagated into recombination deficient strains of E. coli. BACs have been used in many genome-sequencing projects as the principal vector system for the primary cloning of genomic DNA. They are more stable and easier to handle than yeast artificial chromosomes (► [YAC](#)).

► [Hereditary Neuropathies, Motor and/or Sensor](#)

► [Large-Scale Homologous Recombination Approaches in Mice](#)

► [Shotgun Libraries](#)

BAC Transgene

► [YAC/PAC/BAC Transgene](#)

Back-Projection

Definition

This terminus describes the process of using two-dimensional (2D) image data (► [projections](#)) in order to reconstruct a three-dimensional (3D) structure.

► [Cryo-Electron Microscopy: Single-Particle Reconstruction](#)

Bacterial Artificial Chromosome

► [BAC](#)

Bacteriophage

Definition

Bacteriophages include any virus whose host is a bacterium. The group of bacteriophages (or phage) is highly diverse and many hundred different are known. Most bacteriophage are specialised to infect only a small range of bacteria, and the consequence of infection depends on the phage and the host. Generally, phage can be classified into three groups: virulent, temperate and non-lytic phage. Virulent phages induce the lytic cycle and eventually kill their host. Temperate phages can establish a stable, non-lytic and generally

latent relationship (lysogeny) until exogenous triggers activate the lytic cycle. Non-lytic phages can replicate and produce progeny without killing the host but hampering its growth.

- ▶ [Cre/Lox P Strategies](#)
- ▶ [DNA Helicases](#)
- ▶ [Protein Interaction Analysis: Phage Display](#)

Bacteriophage Display

- ▶ [Protein Interaction Analysis: Phage Display](#)

Bacteriophage Lambda

Definition

Bacteriophage lambda is a very well characterised virus that has been a workhorse of molecular biology, whose host is that other great workhorse, *Escherichia coli*. Parts of the bacteriophage lambda sequence form the basis of many cloning vectors.

- ▶ [YAC and PAC Maps](#)

Bacteriophage Surface Display

- ▶ [Protein Interaction Analysis: Phage Display](#)

Bait

Definition

Bait is a fusion protein that is composed of a DNA binding domain and a protein of interest.

- ▶ [Two-Hybrid System](#)

Balancer Chromosome

Definition

A balancer chromosome is a chromosome often having one or more inverted segments. In organisms that have

two sets of chromosomes (diploids), a balancer chromosome acts to suppress recombination with its chromosome homologue. In addition to one or more dominant markers, they usually carry 2–4 recessive markers, and induce lethality as homozygotes. A variety of balancers exist for each chromosome in *Drosophila*.

- ▶ [Cre/Lox P Strategies](#)
- ▶ [Drosophila as a Model Organism for Functional Genomics](#)
- ▶ [Mouse Genomics](#)

Balbiani Ring mRNP

Definition

Balbiani ring mRNP refers to giant puffs of polytene chromosomes in dipteran insects, for example, in larval-salivary-gland cells of the midge *Chironomus tentans*.

- ▶ [Nuclear Pore Complex](#)

Band Keratopathy

Definition

Band Keratopathy is a horizontal opacity of the cornea that begins at the periphery and progresses centrally.

- ▶ [Hyper-and Hypoparathyroidism](#)

Baroreceptor Reflex/Baroreflex

Definition

The baroreceptor reflex is a central reflex mechanism that reduces heart rate following an increase in blood pressure. Each change in blood pressure is changed by baroreceptors in the carotid arteries, which activate the autonomic nervous system to alter the heart rate and thereby readjust blood pressure.

- ▶ [Mendelian Forms of Human Hypertension and Mechanisms of Disease](#)

Barr Body

Definition

Barr body (termed after its discoverer, Barr) is a mass of condensed chromatin in the nuclear periphery of normal female somatic cells, which correspond to inactive X chromosome.

► [X-Chromosome Inactivation](#)

Basal Cell Carcinoma

Definition

The most common malignant neoplasm of Caucasians, usually occurring on sun exposed skin in areas that contain hair follicles. These are slow growing tumors which are locally invasive but seldom metastasize. The tumor is composed of basaloid cells, which resemble normal basal keratinocytes and grow in clusters with peripheral palisading. There are three main types of BCCs, superficial, nodular and sclerosing. Most BCCs arise sporadically, but they also occur in familial form in Gorlin's syndrome, Bazex syndrome and Xeroderma pigmentosum.

► [Skin and Hair](#)

► [Hedgehog Signaling](#)

Basal Ganglia

Definition

Basal ganglia roughly correspond to those areas of the brain that if damaged would disrupt motor abilities and includes the following areas: caudate nucleus, globus pallidus, putamen and the substantia nigra. The exact definition of basal ganglia has changed with increased understanding of the functioning of the brain.

► [Hyper- and Hypoparathyroidism](#)

Basal Promoter

► [Core Promoters](#)

Base Excision Repair

MARIA KARTALOU, LEONA D. SAMSON

Division of Biological Engineering and Center for Environmental Health Sciences, Massachusetts Institute of Technology, Massachusetts, USA
lsamson@mit.edu

Definition

Base excision repair (BER) is one of the mechanisms that cells use to defend themselves when challenged by exogenous and endogenous physical and chemical DNA ► [damaging agents](#) (1). It is involved in the repair of a wide variety of damaged or inappropriate (mismatched) bases. BER substrates include damaged bases that arise spontaneously through hydrolytic deamination, oxidation and alkylation reactions and mismatched bases that arise during ► [DNA replication](#). BER also deals with ► [apurinic/apyrimidinic \(AP\) sites](#) that are formed during spontaneous depurination reactions and after enzymatic and chemical reactions. The persistence of any of these lesions in DNA would increase the mutagenic burden of a cell as these lesions frequently mispair during DNA replication. Finally BER deals with single strand breaks that can be formed after free radical attack; such strand breaks may lead to cell death if left unrepaired.

Characteristics

Mechanism

The first step in BER is catalyzed by a class of enzymes called DNA ► [glycosylases](#) (2). These are typically small monomeric proteins that do not require cofactors for substrate recognition or activity. DNA glycosylases catalyze the cleavage of the N-glycosidic bond between the damaged base and the deoxyribose phosphate backbone, leaving behind an apurinic/apyrimidinic (AP) site. The AP site is processed by an ► [AP endonuclease](#), which hydrolyzes the phosphodiester bond 5' to the AP site, thereby generating a 3'OH end that can serve as a primer for a ► [DNA polymerase](#) and a 5'-deoxyribose phosphate group. The deoxyribose phosphate residue can be excised by a 5'-deoxyribose phosphodiesterase activity. Alternatively, extension from the 3'OH can displace a single stranded DNA flap several nucleotides long and this flap containing the 5'-deoxyribose phosphate can be removed by a flap endonuclease. Sealing of the nick by a ► [DNA ligase](#) completes the repair reaction.

Bifunctional glycosylases can not only catalyze base release, they also possess an additional AP lyase activity that cleaves the phosphodiesterase backbone 3' to the AP site leaving a modified 3' end at the nick that

can be trimmed by an AP endonuclease to leave a 3' OH. The combined activities of a bifunctional DNA glycosylase and an AP endonuclease result in a single nucleotide gap flanked by 3' OH and 5' phosphate ends. The resulting single-nucleotide gap is filled by a DNA polymerase, and a DNA ligase completes the repair reaction.

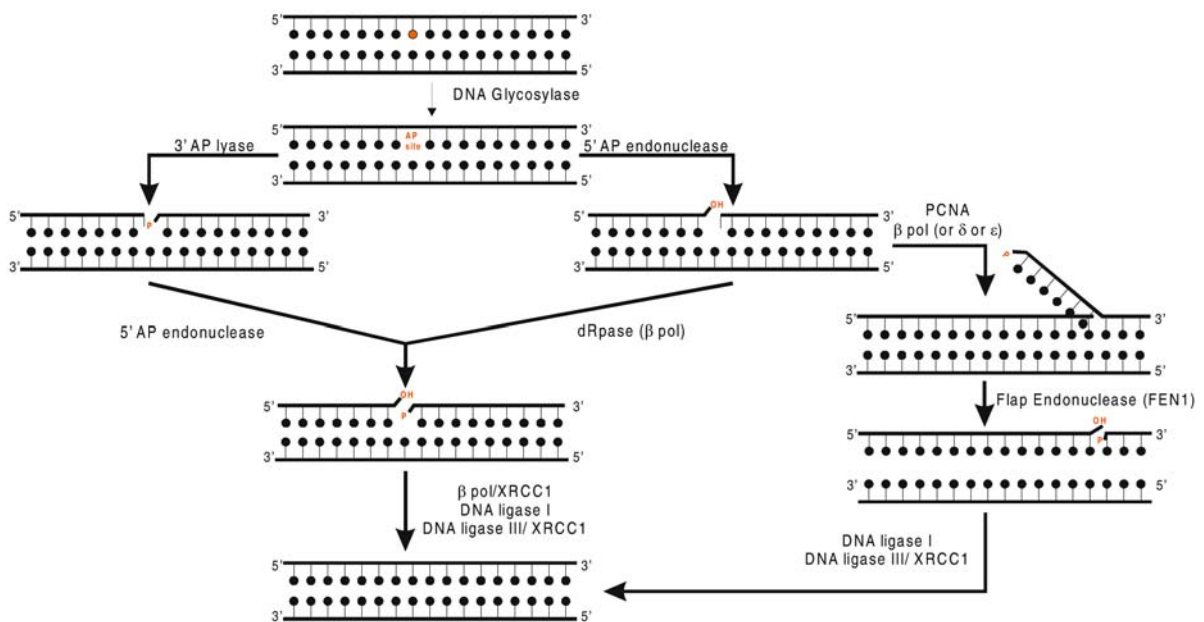
In mammalian cells, AP sites are processed *via* two alternative pathways, the short patch base excision repair and the long patch base excision repair (3). In short patch base excision repair, DNA polymerase β removes the 5'-deoxyribose phosphate and adds a nucleotide followed by DNA ligase I or DNA ligase III/XRCC1 sealing the nick. The short patch base excision repair pathway requires the action of DNA polymerase β and can be reconstituted *in vitro* using human UDG, APE1, DNA polymerase β , XRCC1 and DNA ligase I or DNA ligase III. The long patch base excision repair results in the displacement of 2–13 bases surrounding the AP site and involves the activity of DNA polymerase β or δ/ϵ (pol β or δ/ϵ), flap endonuclease 1 (FEN1) and PCNA. Strand displacement synthesis by pol β or δ/ϵ , generates a flap like structure that can be cleaved by FEN1. PCNA interacts with FEN1 and stimulates its nuclease activity. PCNA can also function as a clamp for pol δ/ϵ . Complete repair occurs when DNA ligase I or DNA ligase III/XRCC1 seals the nick. Although the relative contribution of these two pathways has yet to be elucidated, long patch BER occurs when the deoxyribose phosphate moiety is refractory to excision by pol β . Long patch base excision repair also occurs at oxidized or reduced AP

sites, while normal AP sites are processed by the short patch pathway. Additionally, the type of DNA glycosylase that excises the lesion can also determine which pathway will be followed as single nucleotide gaps generated by the activity of a bifunctional glycosylase are refractory to displacement synthesis. A schematic representation of base excision repair is shown in Fig. 1.

Substrates and Structure of DNA Glycosylases

In humans ten DNA glycosylases have been identified (Table 1). Some glycosylases like the human 3-methyladenine DNA glycosylase (AAG) can remove a variety of chemically and structurally diverse bases from DNA (including normal bases) whereas others like uracil DNA glycosylase (UDG) are only involved in the repair of a limited number of base lesions.

The structure of UDG, T4 endonuclease V, MutY, OGG1, AAG and AlkA DNA glycosylases reveal a novel mechanism of substrate recognition. In the enzyme-substrate complex, the DNA is distorted so that an unpaired nucleotide containing the substrate base is rotated out of its stacked position in the DNA helix and into the active site of the enzyme, thus enabling chemical reactions to occur on atomic centers that would have otherwise been inaccessible. The following model has been proposed for damage recognition. Glycosylases, which have basic DNA recognition surfaces, track along the surface of DNA under the attractive influence of electrostatic forces. While scanning the surface of DNA, the glycosylase encounters a locally destabilized site containing a base



Base Excision Repair. Figure 1 Schematic representation of base excision repair.

Base Excision Repair. Table 1 Base Excision Repair

Name	Acronym	Size (amino acid residues)	Location	Major substrates	<i>E.coli</i> homologue	<i>S. cerevisiae</i> homologue	AP lyase activity
Alkyladenine DNA Glycosylase	AAG (MPG, ANPG)	293	16p	3MeA, 7MeG, 3MeG, εA, Hx,	<i>alkA, tag</i>	MAG1	No
Endonuclease III homologue 1	NTH1	312	16p13.2–p13.3	Tg, 5-OHU, 5-OHC, DHU, Fapy G	<i>nth</i>	NTG1/OGG2 NTG2	Yes
Methyl CpG binding domain 4	MBD4	580	3q21	U or T opposite G at CpG sequences			No
MutY homologue	MYH	535	1p32.1–p34.3	A opposite 8-oxoG	<i>mutY</i>		Maybe
Nei like 1 DNA glycosylase	NEIL1	390	15q25	5-OHU, 5-OHC, DHU, Tg, 8-oxoG, Fapy G, Fapy A	<i>nei, fpg</i>		Yes
Nei like 2 DNA glycosylase	NEIL2	332	4q35	5-OHU, 5-OHC	<i>nei, fpg</i>		Yes
8-Oxoguanine glycosylase 1	OGG1	345	3p25	8-oxoG	<i>fpg (mutM)</i>		Yes
Single stranded selective mono-functional DNA glycosylase 1	SMUG1	270	12q13.1–q14	U			No
Thymine (T:G) mismatch DNA glycosylase	TDG	410	12q24.1	U or T opposite G, ethenocytosine			No
Uracil DNA glycosylase	UDG	313	12q24.1	U, isodialuric acid, 5-OHU, 5-fluorouracil	<i>ung</i>	UNG1	No

Abbreviations: 3MeA, 3-methyladenine; 7MeG, 7-methylguanine; 3MeG, 3-methylguanine; εA, 1,*N*⁶-ethenoadenine; Hx, hypoxanthine; Tg, Thymine glycol; 5-OHU, 5-hydroxyuracil; 5-OHC, 5-hydroxycytosine; DHU, 5,6-dihydrouracil; Fapy, formamidopyrimidine; 8-oxoG, 7,8-dihydro 8-oxoguanine

lesion that is perhaps already in an extrahelical position. The enzyme then causes a kink in the DNA by compressing the lesions surrounding the damaged base, resulting in the flipping out of the nucleotide with the damaged base. Subsequently, the active site is locked on the lesion and catalytic chemistry ensues. Damaged bases might be more susceptible than normal bases to being flipped out of the double helix. Alternatively a DNA glycosylase might flip nucleotides out of the double helix indiscriminately, but only those damaged sites that fit precisely into the substrate-binding site of the enzyme would be subject to hydrolysis.

Mouse Models

The genes for several DNA glycosylases including Aag, Ogg1, Myh, Mbd4, Ung and Nth have been

knocked out in mice (4). To date these mice display normal development and do not have increased cancer incidence. This result is not surprising since there is functional redundancy; DNA glycosylases have overlapping substrate specificity. It seemed likely that combining [▶mutations](#) in glycosylases involved in the repair of the same type of damage, e.g. an *Ogg1*^{-/-} *Myh*^{-/-} mouse, might reveal more dramatic phenotypes. This is indeed the case, mimicking what was found in *E. coli* (7). *mutM* and *mutY* mutant *E. coli* cells each display a moderate [▶mutator phenotype](#) presumably because of their inability to repair 8-oxoG: A mismatches whereas there is a synergistic effect of combining the two mutations; *mutM mutY* double mutant cells have a mutation rate much greater than the sum of the rates for each single mutant.

In contrast to glycosylase deficient mice, mice deficient in subsequent steps of base excision repair (pol β , AP endonuclease, Xrcc1 and ligase I knockout mice) are embryonic lethal. It is unclear whether the embryonic lethality of these mice is due to the accumulation of toxic intermediates, whether BER is essential during development or whether these proteins are also involved in other pathways.

Clinical Relevance

MYH ► Polymorphisms

Polymorphic variants of enzymes involved in base excision repair have been linked with cancer predisposition. Inherited defects of the DNA glycosylase gene *MYH* cause a predisposition to multiple colorectal adenomas and colorectal cancer (5). Individuals with MYH gene mutation are not very efficient at removing A incorporated opposite 8-oxoG and therefore they have an increased frequency of G:C to T:A transversion mutations. Furthermore there has been a polymorphism in an ►intron sequence of the *MYH* gene resulting in alternative splicing of the transcript and reduced synthesis of the MYH protein.

OGG1 Polymorphisms

Polymorphisms of the OGG1 gene have been associated with increased lung and esophageal cancer. The Cys326 allele of the human *OGG1* gene was less capable than the Ser326 allele at suppressing the spontaneous mutation frequency of a *mutM mutY E. coli* strain suggesting that different variants might have different repair activities.

Alzheimer's disease is linked to oxidative stress and reduced OGG1 activity has been observed in cells from Alzheimer's patients. Furthermore, reduced AP endonuclease activity has been observed in brain tissue from amyotrophic lateral sclerosis disease suggesting that polymorphic variants in base excision repair enzymes might play a role in degenerative diseases.

XRCC1 Polymorphisms

The Trp194 variant of the *XRCC1* gene has been associated with reduced risk for lung and breast cancer whereas Gln399 may be associated with increased risk for solid tumors like breast and lung cancer. There are also some studies that correlate the presence of the Gln399 allele with reduced risk for bladder and non-melanoma skin cancer and some studies that do not find any correlation between Gln399 and cancer risk. It remains to be seen what role polymorphic variants at ►codon 399 play in cancer development (6).

►DNA-Repair Mechanisms

References

1. Nilsen H, Krokan HE (2001) Base excision repair in a network of defence and tolerance. *Carcinogenesis* 22(7):987–998
2. Schärer OD, Jiricny J (2001) Recent progress in the biology, chemistry and structural biology of DNA glycosylases. *Bioessays* 23(3):270–281
3. Dogliotti E, Fortini P, Pascucci B et al (2001) The mechanism of switching among multiple BER pathways. *Prog Nucleic Acid Res Mol Biol* 68:3–27
4. Friedberg EC, Meira LB (2003) Database of mouse strains carrying targeted mutations in genes affecting biological responses to DNA damage. Version 5. *DNA Repair (Amst)* 2(5):501–530
5. Al-Tassan N, Chmiel NH, Maynard J et al (2002) Inherited variants of MYH associated with somatic G:C→T:A mutations in colorectal tumors. *Nat Genet* 30(2):227–232
6. Ladiges W, Wiley J, MacAuley A (2003) Polymorphisms in the DNA repair gene XRCC1 and age-related disease. *Mech Ageing Dev* 124(1):27–32
7. Xie Y, Yang H, Cunanan C et al (2004) Deficiencies in mouse Myh and Ogg1 result in tumor predisposition and G to T mutations in codon 12 of the K-ras oncogene in lung tumors. *Cancer Res* 64:3096–3102

Basement Membrane (BM)

Definition

Basement membrane (BM) is a specialised form of extracellular matrix. Ultrastructurally, it can be identified as a dense, about 100–250 nm thick, structure on the basal side of epithelial and endothelial cells, and surrounding Schwann, muscle and fat cells. The subepithelial BM is composed of specialised molecules, Type IV collagen, several laminin isoforms, heparan-sulfate proteoglycan and nidogen, which organise into a complex three dimensional structure. These molecules are tissue and age-specific; they modulate epithelial cell behaviour (proliferation, differentiation, migration) through their binding to cellular integrin receptors. Mutations in the genes encoding laminin isoforms are responsible of pathologies (epidermolysis bullosa, muscular dystrophy). Abnormal synthesis or degradation of BM molecules is involved in tumor progression and invasion.

►Cancer Invasion and Metastasis: Cellular, Molecular and Clinical Aspects

►Cell Polarity

►Extracellular Matrix

►Gut Epithelium

►Hemidesmosomes

►Lung

Base-Pairs (bp)

Definition

Base pairs are pairs of nucleotide bases that constitute a fundamental unit of the DNA double helix. Base-pairs are formed with either adenine (A) and guanine (G), or thymine (T) and cytosine (C) as purine or pyrimidine bases, respectively. The complementary bases A and T, or G and C are connected by hydrogen bonds in the DNA double helix.

- ▶ Centromeres
- ▶ DNA Structure
- ▶ Mouse Genomics

Basic Helix-Loop-Helix

Definition

Basic Helix-Loop-Helix (bHLH) is a structural motif that mediates protein-protein interaction in transcription factor dimers that bind to the DNA with their basic domain. It is characterized by two alpha-helices separated by a loop. The helices mediate dimerization, and the adjacent basic region is required for DNA binding.

- ▶ Basic Helix-Loop-Helix Transcription Factors
- ▶ Muscle Development
- ▶ Transcription Factors and Regulation of Gene Expression

Basic Helix-Loop-Helix Transcription Factors

Definition

Basic Helix-Loop-Helix (bHLH) transcription factors are characterized by the basic helix-loop-helix structural motif, which consists of two alpha-helices separated by a loop. The helices mediate dimerization, and the adjacent basic region is required for DNA binding. These genes are positive or negative regulators of cell fate commitment, by controlling the Notch signalling in adjacent cells.

- ▶ Gut Epithelium
- ▶ Neural Development
- ▶ Muscle Development

Basic-Region-Leucine Zipper

Definition

Basic-region-leucine zipper refers to a class of DNA-binding domains that bind their target site as dimers of uninterrupted helices. The basic-region and leucine zipper region mediate DNA-site specific recognition and dimerization, respectively.

- ▶ Protein/DNA Interaction

Basolateral Domain

Definition

Basolateral domain (or basolateral side) refers to the domain of an epithelial cell that adjoins underlying tissue.

- ▶ Tight Junctions

Batch Entrez

Definition

Batch Entrez is a feature of Entrez (provided by the National Center for Biotechnology Information NCBI) that allows the retrieval of a large number of nucleotide sequences or protein sequences, in a batch mode, by importing a file containing a list of the desired GI or accession numbers.

- ▶ Protein Databases

BC1/BC100

Definition

BC1/BC100 are small dendritic RNAs.

- ▶ Fragile X Syndrome

BCC

- ▶ Basal Cell Carcinoma

B-Cell Epitope

Definition

A small part of an antigen, the antigenic determinants are recognized and bound by the B-cell receptor or antibody. Epitopes recognized by the B-cell receptor/antibody are located on the surface of the antigen

► [Peptide Chips](#)

leukaemia, the occurrence of the BCR-ABL fusion gene is associated with a poor prognosis.

► [Leukemia](#)

Bcl-2/Bcl-xL

Definition

Bcl-2 and its homolog Bcl-xL are members of a large evolutionarily conserved protein family (the Bcl-2 family) of regulators of apoptosis. The Bcl-2 family members contain at least one of the so-called Bcl-2 homology (BH) domains 1 to 4, originally defined in Bcl-2. Dependent on their BH domain architecture, Bcl2 family members exert pro-or anti-apoptotic properties. Bcl-2 and BclxL inhibit apoptosis by preventing the release of proapoptotic factors from mitochondria.

► [Apoptosis](#)

► [Apoptosis, Regulation and Clinical Implications](#)

► [TNF Receptor/Fas Signaling Pathways](#)

BCRP

► [Breast Cancer Resistance Protein](#)

BCYRN1

Definition

BCYRN1 (alternative names BC200 RNA; BC200) is a small cytoplasmic RNA expressed in a highly defined neural and subcellular distribution in primate brain and in cell lines.

► [Fragile X Syndrome](#)

B-CLL

► [Chronic B-Cell Lymphocytic Leukaemia](#)

BDNF

Definition

Brain-derived neurotrophic factor (BDNF) is a cytokine found in the central nervous system, where it is expressed predominantly in hippocampus, cortex, and synapses of the basal forebrain. The synthesis of BDNF is regulated by neuronal activity and specific transmitter systems. BDNF is also expressed by Schwann cells following peripheral nerve lesion, and in muscles. Expression of BDNF is upregulated in denervated muscles.

► [Polyglutamine Disease, the Emerging Role of Transcription Interference](#)

BCR-ABL

Definition

BCR-ABL is a fusion gene resulting from a reciprocal translocation involving chromosomes 9 and 22. The BCR-ABL fusion gene leads to the expression of a constitutively activated non-receptor tyrosine kinase, and is the major pathogenetic event in the process of malignant transformation in chronic myelogenous leukaemia (CML). In patients with acute lymphoblastic

Bead Models

Definition

Bead models are an assembly of spherical units of finite size to approximate the molecular shape; bead models are used in hydrodynamics and scattering techniques to predict frictional and scattering properties from low-resolution molecular contours and mass distributions.

► [Analytical Ultracentrifugation](#)

Beads

Definition

Beads are solid support particles for biomolecule separations of typically 10–1000 nm in size. Beads consist of inert natural or synthetic polymers and are loaded with ligands for binding of biomolecules.

► [Affinity Chromatography and *In Vitro* Binding \(Beads\)](#)

Becker Muscular Dystrophy

Definition

Becker muscular dystrophy refers to a milder form of muscular dystrophy (in comparison to Duchenne muscular dystrophy), with a slower disease progression caused by a genetic defect in the dystrophin gene (gene map locus Xp21.2). It is X-chromosomal-recessively inherited and caused by a deficiency of the protein dystrophin.

► [Duchenne Muscular Dystrophy](#)

Beckwith-Wiedemann Syndrome

Definition

Beckwith-Wiedemann syndrome is characterized by overgrown big tongue, protrusion of the umbilicus and increased risk of kidney tumors.

► [Prader Willi and Angelman Syndromes](#)

BER

► [Base Excision Repair](#)
 ► [DNA Ligases](#)

Bernard-Soulier Syndrome

Definition

Bernard-Soulier Syndrome is an autosomal recessively transmitted disorder characterized by the absence of

glycoprotein Ib/IX (GPIb/IX), which binds von Willebrand Factor (vWF), resulting in defective platelet adherence.

► [Hereditary Hemostatic Defects and Recombinant Proteins for Treatment](#)

Beta Adrenergic Blocker

Definition

Beta adrenergic blockers (β -blockers) are pharmacological agents, which exert their effect by binding to the β -adrenergic receptors of the heart. Medication by beta-blockers reduces the heart's tendency to beat faster by blocking beta adrenergic receptors.

► [Cardiac Signaling: Cellular, Molecular and Clinical Aspects](#)
 ► [Marfan Syndrome](#)

Beta2-Syntrophin/Utrophin Complex

Definition

The beta2 (β 2)-syntrophin/utrophin complex consists of dystrophin-associated and dystrophin-like proteins that interact with ATP binding cassette transporter A-1 ABCA1.

► [Tangier Disease](#)

Beta-Amyloid (β -Amyloid)

Definition

Beta-amyloid is the monomeric or oligomeric soluble or aggregated form (fibrillar or porous) of insoluble A β fragments of the amyloid precursor protein (APP).

► [Alzheimer's Disease](#)
 ► [Amyloid](#)

Beta-Blocker (β -Blocker)

► [Beta Adrenergic Blocker](#)

Beta-Catenin (β -Catenin)

Definition

Beta-catenin (β -catenin) is a multifunctional adapter protein that is involved in adherens junctions at the plasma membrane and Wnt signaling. β -catenin is phosphorylated in the absence of a Wnt signal by the “destruction complex” and degraded by the proteasome. The *Drosophila* homologue is Armadillo.

- ▶ Wnt/Beta-Catenin Signaling Pathway
- ▶ Colorectal Cancer
- ▶ Cell Adhesion

groups in the same strand, in β -sheets the bonds are between adjacent strands. The second β -strand can come from a different region of the same protein or from a different molecule. Therefore, formation of β -sheets can be stabilized by protein oligomerization or aggregation.

- ▶ Amino Acids: Physicochemical Properties
- ▶ Defective Protein Folding Disorders

Beta-Oxidation

Definition

Beta-oxidation is the degradative mechanism in which fatty acids are degraded to acetyl-CoA units.

- ▶ Peroxisomal Disorders

Beta-Sheet Oligomers (β -Sheet Oligomer)

Definition

Beta (β)-sheet oligomers refer to protein structures that contain several monomers of a protein organized in a β -pleated sheet conformation.

- ▶ Defective Protein Folding Disorders

Beta-Sheet Structure (β -Sheet Structure)

Definition

β -sheets are prevalent repetitive secondary structures in folded proteins. β -sheets are formed by alternating peptide pleated strands linked by hydrogen bonding between the NH and CO groups of the peptide bond. While in α -helices the hydrogen bonds are between

Beta-Tubulin

- ▶ Alpha-Tubulin/Beta-Tubulin

bHLH

- ▶ Basic Helix-Loop-Helix

bHLH Transcription Factors

- ▶ Basic Helix-Loop-Helix Transcription Factors

Biallelic

Definition

Biallelic designates the existence of two alleles. A single nucleotide polymorphism (SNP) is usually biallelic or has two alleles.

- ▶ SNP Detection and Mass Spectrometry

Bicistronic

Definition

The term bicistronic refers to an mRNA molecule that encodes two proteins.

- ▶ Transposons

Bijvoet Difference

Definition

Bijvoet pairs are Bragg reflections that are space group-symmetry equivalents to the two members of a Friedel pair, which holds that they have equal amplitude and opposite phase (Friedel pair $|F(hkl)| = |F(-hkl)|$, $\varphi(hkl) = -\varphi(-hkl)$).

They have been measured for the same wavelength. The difference in the measured amplitude of a Bijvoet pair is called a Bijvoet difference. The Bijvoet difference $\Delta F_{\text{Bijvoet}} = |F_{\text{obs}}(\lambda_i, hkl)| - |F_{\text{obs}}(\lambda_i, -hkl)|$ is maximal when measured at λ_{peak} , with maximum fluorescence. A complete dataset of anomalous reflections comprises of all Bijvoet reflection pairs.

► [MAD Phasing](#)

Bile Acids

Definition

Bile acids are oxidation products of cholesterol, with the enzyme cholesterol 7 α -hydrolase as a rate limiting step in their synthesis. In humans, cholic acid (the most abundant) and chenodeoxycholic acid are the most important bile acids. They function in the gut to facilitate solubilization and disposal of cholesterol.

► [Nuclear Hormone Receptors: Regulators of Gene Transcription and Cellular Signaling](#)

Bile Ducts (Choledochus)

Definition

Choledochus stands for the common bile duct (ductus choledochus) after unification of ductus cysticus and ductus hepaticus and flows at proffilla duodeni major into the duodenum.

Biliary Atresia

Definition

Biliary atresia is a hepatic disorder in infants caused by an obstruction of the ► [bile](#) ducts (cholestasis), leading

to the impairment of bile flow, which rapidly progresses to biliary cirrhosis.

► [Microvilli](#)

Bimolecular Process

Definition

Bimolecular process describes a process that involves the association or dissociation of two molecules.

► [Thermodynamic Properties of DNA](#)

Binary GAL4-Expression System

The GAL4 system is a method for directed gene expression that can be used to misexpress genes in specific cell types, or tissues, at different times of development. This system relies on the generation of transgenic lines that carry ‘activator’ or ‘effector’ constructs. Activator lines express the yeast transcription factor, GAL4, under the control of a desired promoter, whereas effector lines contain DNA-binding motifs for GAL4-(UAS) linked to the gene of interest.

Binding Affinity

Definition

Binding affinity comprises of all energetic effects that are responsible for the binding of two molecules, e.g. a ligand and a receptor. In general, the ligand and the receptor are dissolved in water. Thus, the binding process occurs in aqueous solution, and water molecules are highly involved in the binding process.

► [Structure-Based Drug Design](#)

Binding Isotherm

Definition

Binding isotherm refers to the dependence of the population of unbound molecules and complex in

association-dissociation equilibrium on the total concentrations of the interacting molecular components.

► **Analytical Ultracentrifugation**

Bio-Beads

Definition

Bio-Beads are polystyrene beads with small hydrophobic pores that allow detergent molecules to be adsorbed.

► **Two-dimensional Crystallization of Membrane Proteins**

Biochemical Engineering of Glycoproteins

RÜDIGER HORSTKORTE

Institute for Biochemistry and Molecular Biology,
Charité, Berlin, Germany
ruediger.horstkorte@charite.de

Definition

During recent years the human ► **genome** project has produced an enormous quantity of data. This flood of information has been further increased by the development and extended use of ► **DNA chip** technologies. There has been parallel progress in the analysis of the proteome, in particular through improvements in analytical methods.

The main goal of biochemical engineering in the life sciences is to use the enzymatic machinery of cells to achieve the chemical synthesis of novel biomolecules. Such engineered biomolecules may have novel properties and are potentially useful in both science and medicine. The challenge facing biochemical engineering is to obviate the limited potential of the gene by supplying non-DNA-coded products for basic research and industrial production. The present account focuses on the production of novel ► **glycoproteins** by biochemical engineering and outlines some potential applications of this methodology in medicine.

Characteristics

Glycosylation and Sialylation

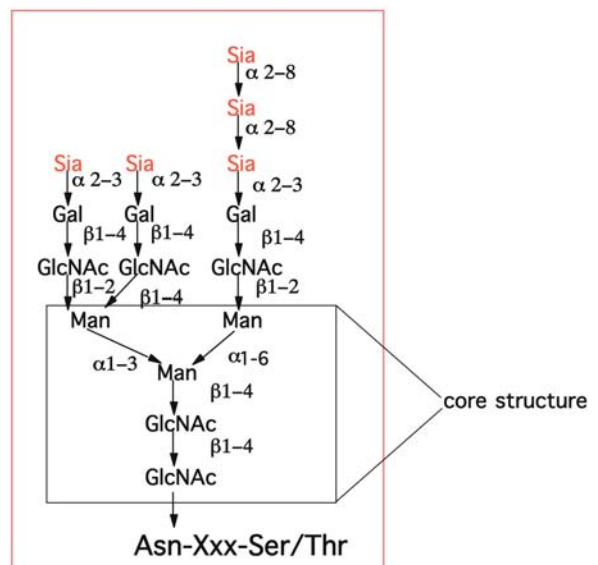
► **Glycosylation of Proteins** represents the most common posttranslational modification of proteins. ► **Glycans** of glycoproteins are synthesized in the

► **Golgi apparatus** by specific glycosyltransferases, which attach nucleotide-activated ► **monosaccharides** to specific sugar residues of glycoproteins. Although these glycans have a common core structure (Fig. 1) the combination of different monosaccharides and different linkage-types allows the cell to create an enormous number of different structures. Note that the terminal monosaccharide of most glycoproteins is a ► **sialic acid**.

Sialic acids represent a family of aminosugars with over 40 members. ► **N-Acetylneuraminic acid** is the most prominent sialic acid in eukaryotes. Different sialic acids possess different highly specific recognition and binding properties for cellular receptors. This structural and functional diversity of sialic acid is exploited by viruses, bacteria and toxins and by the sialoglycoproteins and sialoglycolipids involved in cell-cell recognition.

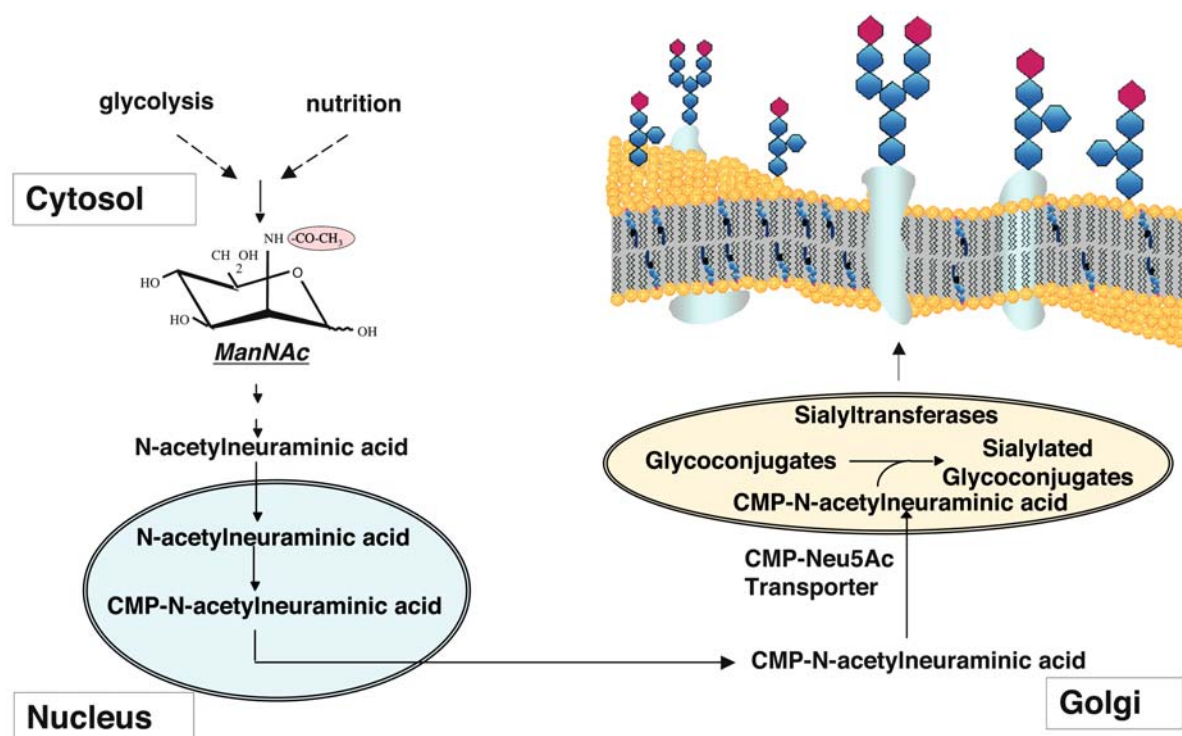
Biosynthesis of Sialic Acid and Biochemical Engineering of Glycoproteins

The physiological precursor of all sialic acids is ► **N-acetyl-D-mannosamine** (ManNAc) (Fig. 2). It has been shown recently that synthetic *N*-acyl-modified D-mannosamines can be taken up by cells and efficiently metabolized to the respective *N*-acyl-modified neuraminic acids *in vitro* and *in vivo* (3). *N*-acyl-modified D-mannosamines successfully employed in this way include *N*-propanoyl- (ManNProp),



Biochemical Engineering of Glycoproteins.

Figure 1 Typical structure of a glycan from a glycoprotein. The common core structure of all glycoproteins is indicated. GlcNAc, N-acetylglucosamine; Man, mannose; Gal, galactose; Sia, sialic acid.



Biochemical Engineering of Glycoproteins. Figure 2 Biosynthesis of sialic acid. The biosynthesis of sialic acid starts in the cytosol. Activation occurs in the nucleus, whereas glycoproteins are built in the Golgi. Finished glycoconjugates are expressed on the cell surface. The structure of the precursor of all sialic acids, N-acetylmannosamine is shown. Its N-acyl side chain, the target of biochemical engineering, is marked in red.

N-butanoyl- (ManNBut)-, *N*-pentanoyl- (ManNPent), *N*-hexanoyl- (ManNHex), *N*-crotonyl- (ManNCrot), *N*-levulinoyl- (ManNLev), *N*-glycolyl- (ManNGc) and *N*-azidoacetyl D-mannosamine (ManNAc-azido) (Fig. 3). All these synthetic, unnatural sialic acid precursors are metabolized by the promiscuous sialic acid biosynthetic pathway in the cytosol and are incorporated into cell surface sialoglycoconjugates where, depending on the cell type, they replace 10–85% of normal, physiological sialic acids (for review, see 4). This method of introducing novel, unphysiological sialic acids into cellular glycoconjugates has been termed biochemical engineering of the side chain of sialic acid.

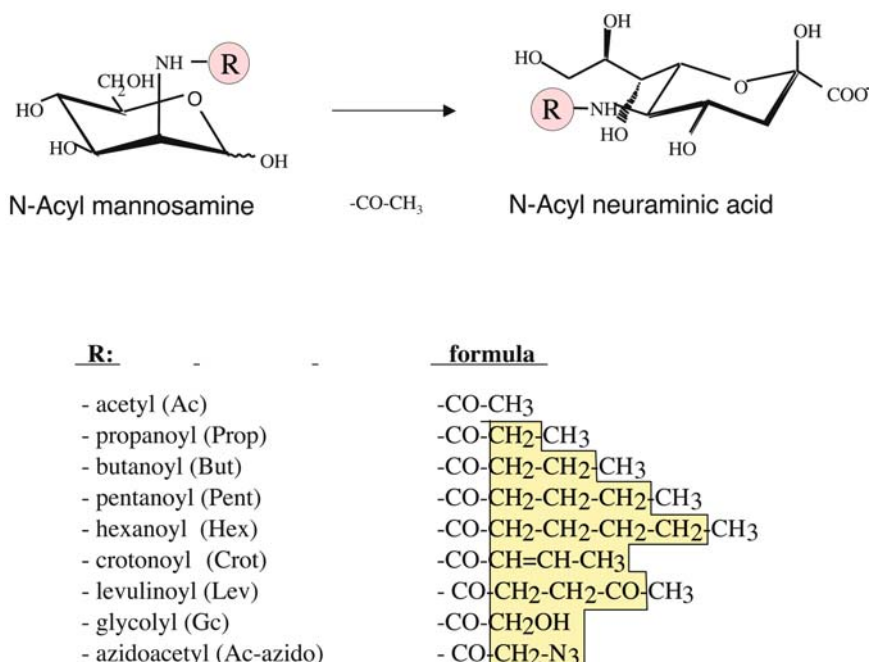
Application of the above mentioned synthetic sialic acid precursors to different biological systems has revealed important and unexpected functions of the *N*-acyl side chain of sialic acids. Furthermore, the introduction of chemically reactive ketone and azido groups into cell surface glycoconjugates using the respective *N*-acyl-modified sialic acid precursors, offers a variety of applications including the generation of artificial cellular receptors for viral gene delivery (6). Novel sialic acid precursors have enabled studies on sialic acid modifications on the surface of living cells

and have improved our understanding of carbohydrate receptors in their native environment. Biochemical engineering therefore provides new tools for studying the biological relevance of sialic acid and for exploiting it as a tag for therapeutic and diagnostic applications.

Clinical Relevance

Engineering of Cell Surface Sialic Acid Interferes with Virus Infections

Recognition of and binding to an appropriate receptor on the surface of the host cell by a virus is the first step of viral infection. Despite the ubiquity of sialic acid on the cell surface, sialylated oligosaccharides are an essential receptor component for many animal viruses from different virus families, such as influenza A and C viruses, Newcastle disease virus and cardioviruses as well as murine and primate polyomaviruses. When 18–64% of the sialic acids on host cells were biochemically engineered by treatment with the respective analogues (usually ManNProp), binding and/or infection by different primate polyomaviruses, which depend on cell surface sialic acids for entry, were markedly altered (5). For human polyomavirus BK the elongation of the *N*-acyl side chain by one methylene group (from *N*-acetyl to *N*-propanoyl) resulted in up to 7-fold



Biochemical Engineering of Glycoproteins. Figure 3 Biochemical engineering. The common precursor and the final sialic acid are shown. The structural differences between the physiological, natural acetyl side chains (Ac) and all engineered analogues are indicated.

enhancement of infection, whereas further elongation to *N*-pentanoyl drastically reduced infection. Binding and infection by the African green monkey B-lymphotropic papovavirus were decreased about 5-fold and more than 10-fold respectively, by incorporation of the *N*-propanoyl side chain. In contrast, the ▶sialidase-independent infection of closely related simian virus 40 was, as expected, unaffected (5).

Immunotargeting of Tumor Cells Expressing Unnatural Polysialic Acids

Immunotargeting of tumor cells by creating vaccines based on cancer-specific cell surface glycoconjugate antigens has long been proposed, but many tumors fail to express unique markers. Polysialylation, e.g. alpha 2–8 linked polysialic acid sequences located on the outer chains of an N-linked oligosaccharide, is a unique and functionally important property of the neural cell adhesion molecule. Its expression is regulated during embryogenesis, with maximal expression in the perinatal phase and restriction in the adult nervous system to plastic regions. However, over-expression of polysialic acid is found on a number of cancers, including small cell lung carcinomas and Wilms tumors, and is thought to promote tumor cell metastasis. Using ManNProp, Jennings and co-workers successfully generated polysialic acid containing *N*-propanoylneuraminic acid. Since a monoclonal antibody to polyNeuProp is available (13D9; 8), this

biochemically engineered polysialic acid expressed on leukemic tumors can be used for an antibody-mediated cell killing. Furthermore, in an *in vivo* study this antibody effectively controlled metastasis of a solid tumor model in mice, which had been given ManNProp. Bertozzi and coworkers have demonstrated that even ManNLev is incorporated into poly sialic acid. Using ManNLev, it is possible to introduce reactive ketone groups onto cell surfaces; this also offers novel possibilities for specifically attacking polysialylated tumor cells. The same group reported recently that ManNBut is a potent inhibitor of polysialylation (7). Application of ManNBut resulted in biosynthesis of the expected *N*-acyl-modified sialic acid (*N*-butanoylneuraminic acid), but this biochemically-engineered sialic acid seems to induce termination of polysialylation. Recent data suggest that biochemically engineered sialic acids specifically inhibit the polysialyltransferase II (2). Future work will have to address the applicability of these strategies to the targeting of polysialic acid-expressing human cancers and assess the potential adverse effects of biochemical engineering.

Activation of Human T-lymphocytes by ManProp

Sialic acids are implicated in the ▶differentiation and maturation of lymphocytes. Preliminary studies revealed that application of ManProp leads to an incorporation of *N*-propanoylneuraminic acid into human T-cells. These biochemically engineered T-cells

show several hallmarks of activation, including proliferation, secretion of interleukin-2 and the expression of the IL-2-receptor α -chain. This stimulation is dose-dependent and in the same range as that observed with the commonly used toxic plant lectins, concanavalin A or wheat germ agglutinin. ManNProp, however, did not induce cytotoxicity, even at high concentrations. The ManNProp-induced stimulation is accompanied by an increase in the peptidase-activity of CD26, a costimulator of lymphocytes. In agreement with this, an incorporation of N-propanoylneuraminic acid into CD26 was observed.

N-Acyl-modified Sialic Acids Stimulate Neuronal Cells Stimulation of Glia Cells

Modified sialic acids have striking biological consequences for glia cells of the mammalian central nervous system. Application of ManNProp to primary glial cells in culture leads to an expression of N-propanoylneuraminic acid in glycoproteins on the cell **▶membrane** (10). This biochemical engineering of the N-acyl side chain of sialic acid stimulated the proliferation of astrocytes and microglia *in vitro*. Furthermore, oligodendrocytes showed increased signs of a non-mature cell stage when ManNProp was applied. Mature oligodendrocytes are the myelin-forming cells in the central nervous system. They develop from oligodendrocyte progenitor cells, which can be immunologically identified by their expression of specific ganglioside-**▶epitopes**, such as the A2B5-epitope. This A2B5 epitope is regarded as a specific marker for a subset of rat oligodendrocyte progenitor cells. Application of ManNProp leads to a dramatic increase in A2B5-expression *in vitro*. Since the A2B5 epitope is considered to be a functional marker of cells of the early oligodendrocyte lineage, ManNProp has to be considered as a potent regulator of the lineage progression of oligodendrocytes at early stages of their development. Only oligodendrocyte progenitor cells and not mature oligodendrocytes are proliferative and migratory. These properties play an important role not only during development but also in the regeneration of the adult nervous system, since they develop constitutively into myelin-forming cells *in vitro* and *in vivo*. Oligodendrocytes are functionally impaired in a number of severe neurological diseases. The most important disease is the loss of oligodendrocytes followed by demyelination in multiple sclerosis. These results underline the important role of biochemically-engineered sialic acid in neural development and regeneration. The molecular mechanism underlying the stimulation of oligodendrocytes might depend on signal transduction events. The incorporation of N-propanoylneuraminic acid, followed by the application of GABA, leads to calcium oscillations in oligodendrocytes. It has been proposed

that biochemical engineering of the N-acyl side chain of sialic acid in conjunction with the activation of GABA-receptors, which are sialylated glycoproteins, modulates, e.g. increases, the intracellular calcium concentration in oligodendrocytes. The prolonged increase in intracellular calcium concentration after calcium oscillations could be responsible for the occurrence of A2B5-positive oligodendrocyte precursor cells.

Further data underlining the potency of biochemically-engineered sialic acid in myelination have been reported by Schnaar and coworkers. They used ManNGcPA as a synthetic precursor and demonstrated the conversion of neuronal sialic acids from N-acetyl- to N-glycolylneuraminic acid. The result was an inhibition of the binding of myelin-associated glycoprotein (MAG) to neuronal cells. MAG is expressed on myelin in the central nervous system and is known to be a potent inhibitor of neurite regrowth after nerve damage. This inhibitory capacity is mediated by the interaction of oligodendrocyte-expressed MAG with gangliosides of **▶neurons**, since MAG is a sialic acid-binding molecule of the siglec family. Thus, a means of interfering with MAG binding to nerve cells might enhance the possibility of post-traumatic nerve regeneration.

In this context it should be noted that in human lung fibroblast cultures also contact inhibition-regulated cell growth is influenced by chemically modified N-acyl mannosamines. By treatment of these cells with ManNProp, ManNBut or ManNPent for 7 days the density-dependent inhibition of growth is abolished.

These examples hopefully demonstrate the potency of synthetic D-mannosamines as therapeutic agents in nerve regeneration.

Stimulation of Neurons

Rat PC12-cells have been widely used as a standard system for the study of neurite outgrowth. These cells respond to nerve growth factor (**▶NGF**) by extending neurites *via* a ras-dependent pathway.

PC12 cells were cultured in the presence of suboptimal concentrations of NGF on poly-D-lysine, collagen I or laminin. The best neurite outgrowth was observed on laminin. In the presence of ManNProp, PC12-cells extended up to over 60% longer neurites on laminin compared to control cultures in the absence of ManNProp. This stimulation of neurite outgrowth was concentration-dependent (1). Furthermore, re-establishment of the perforant pathway was stimulated in brain slices. In addition, several cytosolic proteins with regulatory functions that are differentially expressed after treatment with ManNProp were identified. Since sialic acid is the only monosaccharide activated in the nucleus, we hypothesize that transcription is modulated by the unnatural CMP-N-propanoylneuraminic acid and that sialic acid activation might serve as

a general tool for regulating cellular functions, such as neurite outgrowth (1). This indicates the fundamental role of sialic acids in general and especially the role of the N-acyl side chain of the biochemically engineered sialic acid in neurite outgrowth.

Stability of Glycoproteins Is Increased after Incorporation of *N*-acyl-modified Sialic Acids

The biological ▶half-life time of many glycoproteins is regulated *via* terminal sialic acids. We found that the half-life of the highly sialylated CEACAM1, a member of the ▶immunoglobulin superfamily, is increased by more than 50% by replacement of the N-acetylneuraminic acid by the novel, engineered N-propanoylneuraminic acid. This demonstrates that biochemical engineering not only modulates the structure of cell surface sialic acids, but that biochemical engineering also influences biological stability of defined glycoproteins. Some of the observed effects of ManNProp on cells might be explained by the prolonged expression of specific cell surface receptors.

References

1. Büttner B, Kannicht C, Schmidt C et al (2002) *J Neurosci* 22:8869–8875
2. Horstkorte R, Muhlenhoff M, Reutter W et al (2004) Selective inhibition of polysialyltransferase ST8SiaII by unnatural sialic acids. *Exp Cell Res* 298:268–74
3. Kayser H, Zeitler R, Kannicht C et al (1992) *J Biol Chem* 267:16934–16938
4. Keppler OT, Horstkorte R, Pawlita M et al (2001) *Glycobiology* 11:11R–18R
5. Keppler OT, Stehling P, Herrmann M et al (1995) *J Biol Chem* 270:1308–1314
6. Mahal LK, Yarema KJ, Bertozzi CR (1997) *Science* 276:1125–1128
7. Mahal LK, Charter NW, Angata K et al (2001) *Science* 294:380–381
8. Pon RA, Lussier M, Yang QL et al (1997) *J Exp Med* 185:1929–1938
9. Schauer R, Kelm S, Reuter G et al (1995) In: Rosenberg A (ed) *Biology of the sialic acids*, Plenum Press, New York, pp 7–67
10. Schmidt C, Stehling P, Schnitzer J et al (1998) *J Biol Chem* 273:19146–19152

- ▶ Microarray Data Analysis
- ▶ Microarrays in Colorectal Cancer
- ▶ Microarrays in Pancreatic Cancer
- ▶ Microarrays in Plant Genomics
- ▶ Microarrays in Rheumatoid Disease
- ▶ Chip Technologies: Basic Principles

Biochip Technology

- ▶ Chip Technologies: Basic Principles
- ▶ DNA Chip Technology
- ▶ Gene Chip Technology
- ▶ Microarray Technology

Bioconjugation

Definition

Bioconjugation denotes the covalent coupling of two or more biological molecules.

- ▶ Protein Interaction Analysis: Chemical Cross-Linking

Bioethics

Definition

Bioethics involves the systematic study of human conduct in the area of life sciences, health care, and disease management, examined in the light of moral principles, values, and visions (Encyclopedia of Bioethics).

- ▶ Ethical Issues in Medical Genetics

Biochip

Definition

Biochip designates a small-scale device (microsystem) that is used to analyze organic molecules (DNA, proteins) associated with living organisms.

- ▶ Camel as a Model for Functional Genomics
- ▶ Uncaging and Photoconversion/Activation

Biogerontology

Definition

Biogerontology is a branch of biomedical research aimed at finding the biological determinants of ageing, based on different biological models. These models range from lower organisms like *C. Elegans* to *in vitro* models based on human cells, like the model of

stress-induced premature senescence. It also considers cross-sectional or longitudinal studies of human populations including the very old (centenarians (>100) and supercentenarians (>110)).

► [Proteomics in Ageing](#)

Bioinformatics

Definition

Bioinformatics comprises of a broad area of computational methods that try to handle, store, retrieve, and analyze information about biological systems. Starting with the sequence information on the genetic code and gene products, bioinformatics allows the mining and prediction of properties of biomolecules. Furthermore, bioinformatics provides tools to compare biomolecules and to predict how their function could be modified, e.g. by docking ligands into binding pockets or forming protein/protein complexes.

► [Proteomics in Human-Pathogen Interactions](#)

► [Structure-Based Drug Design](#)

► [Translational Frameshifting, Non-standard Reading of the Genetic Code](#)

Biological Membranes

THOMAS POMORSKI, ANDREAS HERRMANN
Institute of Biology, and Center of Biophysics
and Bioinformatics, Humboldt University Berlin,
Berlin, Germany
thomas.pomorski@rz.hu-berlin.de
andreas.herrmann@rz.hu-berlin.de

Definition

Biological membranes are the basis for highly defined and separated functional units such as a cell and its subcellular compartments. The cell membrane, or plasma membrane, resembles the external boundary of every cell, separating the cytoplasm from the surrounding environment and strictly regulating in- and out-put of material and information. In animal cells, this separation is established by the plasma membrane alone, while in other eukaryotic cells (e.g. plant and yeast cells) and prokaryotic cells (bacteria) an additional cell wall serves as the outermost boundary, providing primarily mechanical support. Eukaryotic

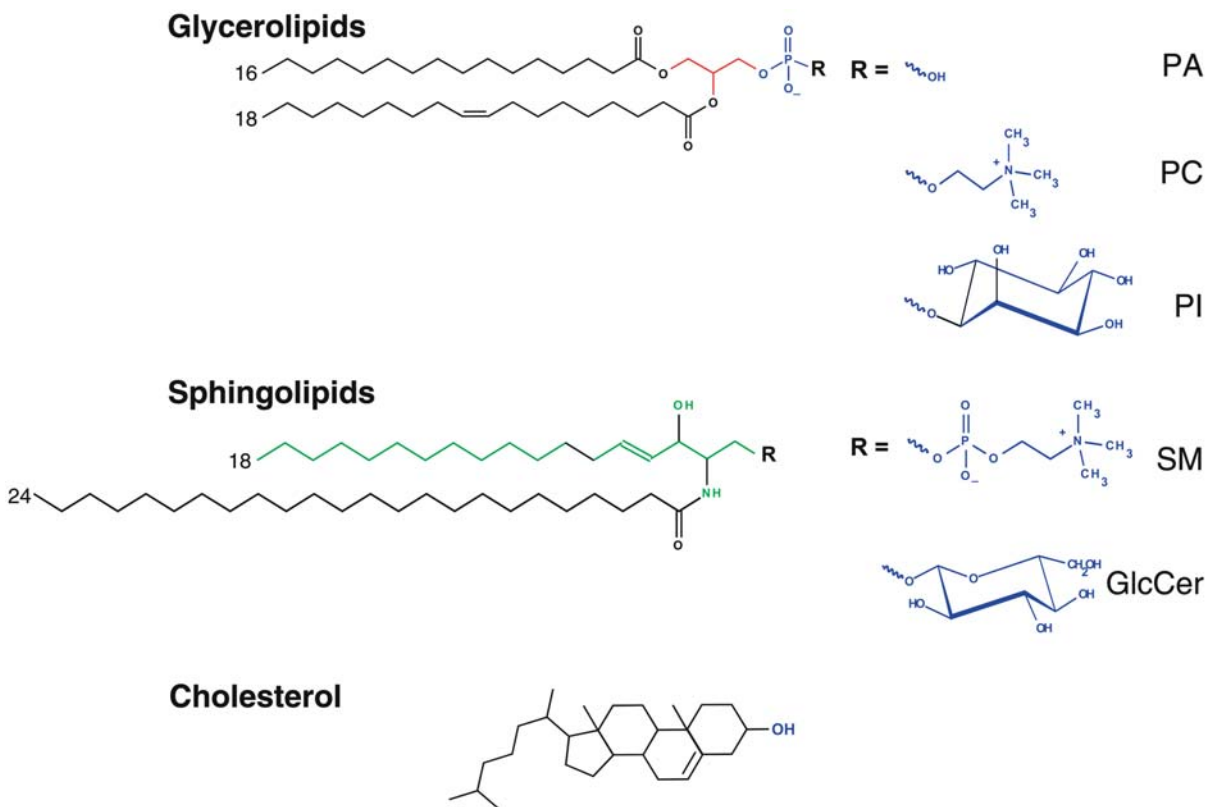
cells contain in addition numerous subcellular membranes that divide the cytoplasm into multiple compartments (organelles), thereby allowing different functions to occur efficiently and simultaneously in different parts of the cell. Among these functions are nutrient and ion transport, oxidative and photosynthetic phosphorylations, signal transduction and electrical excitability.

Almost all biological membranes are organized as bilayers consisting of two leaflets, which are structurally formed by phospholipids. Depending on the (sub) cellular membrane, both leaflets may harbor also other classes of lipids including glycolipids and sterols. Proteins with specific functions important for the structure and stability of membranes as well as for cell homeostasis span (integral membrane proteins), loosely bind to the bilayer (peripheral membrane proteins) or are covalently linked to a fatty acid or a lipid molecule that is embedded in the membrane (lipid-anchored proteins). Biological membranes are highly dynamic structures as regards rotational, lateral and transversal motions of the components within the membrane. In addition, many membrane systems are capable of extruding membrane domains in the form of vesicles and of incorporating vesicles by ► [fusion](#), processes that provide for intracellular transport and secretion and that are consequently central to cell growth and proliferation.

Characteristics

Lipid Composition

The primary building blocks of most membranes are glycerolipids which consist of a glycerol backbone, two fatty acid chains and a polar head group at the *sn*-3 position (Fig. 1) (1, 2). The fatty acid chains are attached to the *sn*-1 (usually saturated, 16 or 18 carbons) and *sn*-2 (usually (poly)unsaturated, 18–20 carbons) positions of glycerol by ester or ether linkages, and they vary widely in length, branching, and degree of unsaturation. The major subclass of glycerolipids and the most important class in mammalian cells are the phospholipids (or phosphoglycerides). Their polar head group consists of a phosphate residue which is (except for phosphatidic acid, PA) esterified by an alcohol such as choline (to form phosphatidylcholine, PC), ethanolamine (phosphatidylethanolamine, PE), serine (phosphatidylserine, PS) or inositol (phosphatidylinositol, PI). The glycosylglycerolipids are another important subclass, in which the *sn*-3 position of glycerol is linked glycosidically to a carbohydrate, e.g. galactose. Sphingolipids constitute a second important lipid class in eukaryotic cells, where they compose 10–20% of the total lipid. The lipid moiety consists of a sphingoid base, which is an amino alcohol (usually of 18 carbons) and a (usually saturated) fatty acid chain

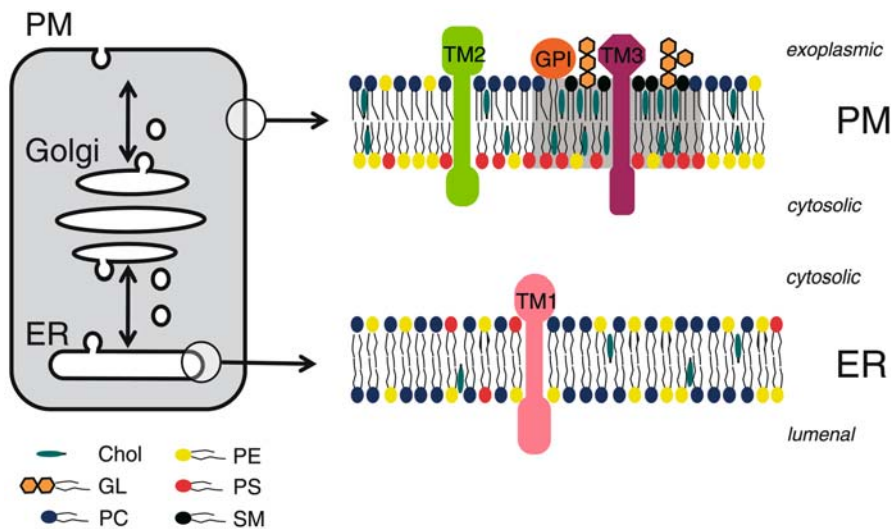


Biological Membranes. Figure 1 Biological membranes are mainly formed from three different classes of lipids. Phosphatidic acids (PA), phosphatidylcholines (PC), and phosphatidylinositols (PI) are examples of glycerolipids derived from glycerol. The fatty acid at the *sn*-1 position has a saturated chain with 16 or 18 carbon atoms. At the *sn*-2 position, the fatty acid is generally longer (at least 18 carbon atoms) and is unsaturated, with one or more *cis* double bonds. Sphingomyelins (SM) and glucosylceramides (GlcCer) are examples of sphingolipids derived from sphingosine. Typically, the amide-linked fatty acyl chains in sphingolipids are long (up to 24 carbons), saturated and often α -hydroxylated. Cholesterol is an abundant sterol in animal membranes. All lipids that are shown in the pictures are amphipathic molecules. The polar head groups are highlighted with blue, glycerol in the glycerol lipids with red and the sphingoid base in the sphingolipids with green.

linked *via* an amide linkage. Sphingolipids are richer in longer fatty acids (22–24 carbons) than glycerolipids and thereby tend to form thicker bilayers. The type of head group attached leads to their classification into sphingomyelins (phosphocholine or phosphoethanolamine head group), cerebrosides (head groups are sugar residues) and gangliosides (oligosaccharides with at least one sialic acid residue). The third class of membrane lipids found in eukaryotes are sterols. The most abundant sterol in animal membranes is cholesterol. It is essential for the properties of biological membranes and is, on the other hand, a precursor for steroid hormones.

In eukaryotes, membrane lipid composition varies dramatically between organelles (1). For example, plasma membranes are typically enriched in sphingolipids, PS and sterols, while the endoplasmic reticulum is depleted in these lipids (Fig. 2). Several lipids exist in

minor amounts. However, they often serve important biological functions and may be localized in specific (sub)cellular membranes. For example, the amount of glycolipids in the plasma membrane is typically about 5% of total lipid. In the apical membrane of polarized cells, however, they are greatly enriched, supporting the stability of the membrane. Likewise, the myelin shells of neuronal axons consist for the most part of sphingomyelin and gangliosides amount to 6% of the brain lipids. Cardiolipin, a diphosphatidylglycerol, is enriched in the inner mitochondrial membrane (~20%), but rare in other membranes. Although the membrane concentration of PIs is comparatively low, e.g. less than 10% in the plasma membrane, these lipids play an essential role in signal transduction. In gram-negative bacteria, the lipid and protein composition of the cell wall membrane (outer membrane) is very different from that of the plasma membrane (inner



Biological Membranes. Figure 2 Cellular membranes contain unique sets of proteins and differ in their lipid compositions and transbilayer lipid distributions in spite of extensive bi-directional membrane trafficking. For example, eukaryotic plasma membranes (PM) are typically enriched in sphingolipids, cholesterol and phosphatidylserine and display transbilayer lipid asymmetry, while the endoplasmic reticulum (ER) is depleted in those lipids and shows a symmetric lipid distribution. In addition, cellular membranes display a lateral heterogeneous organization by the clustering of specific lipids into highly condensed domains (termed lipid rafts). Lipid rafts in the plasma membrane outer leaflet are composed of cholesterol, sphingolipids and phospholipids with saturated fatty acyl chains (indicated as straight lines in lipid tails) while the inner leaflet contains domains of unknown lipid composition. Various domains of different composition may occur within a monolayer and domains in both monolayers may colocalize (gray bilayer). A variety of proteins partition into lipid rafts (gray bilayer), e.g. glycosylphosphatidylinositol-anchored proteins (GPI) and transmembrane proteins (TM3 in contrast to TM1 and TM2). PC, phosphatidylcholine; PE, phosphatidylethanolamine; PS, phosphatidylserine; SM, sphingomyelin; Chol, cholesterol; GL, glycolipid.

membrane). For example, the outer leaflet of the cell wall membrane of *E. coli* harbors solely lipopolysaccharides while the inner leaflet is enriched in PE.

Protein Composition

Biological membranes contain between 20% and 80% (w/w) proteins (1). They are of broad functional variety covering substance exchange and transport (pores, channels, transporters), biochemical reactions including energy perception and conversion (enzymes), signal transduction, immunoreaction, endo- and exocytosis and cell-cell interaction. In addition, membrane proteins may serve to support the stability of membranes (membrane skeleton) and to establish a lateral and transversal heterogeneous organization. Membrane proteins are either anchored to the membrane surface (peripheral membrane proteins) *via* hydrophobic or electrostatic interaction, covalently linked to a fatty acid or to a lipid molecule that is embedded in the membrane (lipid-anchored proteins; e.g. ►GPI proteins) or span the hydrophobic part of the membrane (integral membrane proteins). Integral membrane proteins may cross the membrane several times, for example G-protein linked receptors, ion channels and ATP-dependent transporters. The

transmembrane domains are usually α -helices, but in some cases, such as porins, also β -strands. Membrane protein activity often requires homo- or hetero-oligomerization of proteins. The majority of proteins that are secreted or embedded in the plasma membrane are modified by carbohydrate attachment and called glycoproteins. Glycoproteins on cell surfaces are important for communication between cells, cell adhesion, maintaining cell structure and self-recognition by the immune system.

Molecular Interactions

Biological membranes are typically organized in bilayers (3). The formation of this lamellar structure by lipids does not require energy and is an entropic process driven by so-called hydrophobic force (1). However, maintenance of a specific lipid composition and an asymmetric distribution of lipids between leaflets (see below) requires energy consumption. Non-bilayer organization of lipids as hexagonal structures has been observed in model membranes, depending on the lipid composition and the characteristic shape of lipids as well as other factors, e.g. interaction with cations and temperature. Formation of local non-bilayer structures, which are very likely of transient occurrence, has also

been detected in biological membranes. The significance of these structures is not well understood. They have been suggested to be involved in fusion of biological membranes.

Lateral Organization of Biological Membranes

In model membranes of homogenous lipid composition, lipids can adopt different phases depending on the lipid as well as on various factors like temperature, pressure and ionic strength. For example, the liquid-disordered (or fluid) phase transforms into the ordered gel phase on lowering the temperature, a phase transition. Membranes of lipid mixtures do not usually exhibit phase transitions (1). Typically, biological membranes with several hundred chemically different lipids are in the fluid state. Nevertheless, the degree of interaction and of miscibility of different kinds of lipids leads spontaneously to formation of separate lateral domains of different lipid composition with striking similarities to phases described in model membranes. Most prominent are so-called **rafts** which are found in the plasma membrane and probably also in subcellular membranes of eukaryotic cells (Fig. 2). These domains are essentially enriched in phospholipids carrying long saturated fatty acid residues, sphingolipids and cholesterol and resemble a specific interaction between these lipids (3, 4). Rafts are in a liquid-ordered state in which the lipids are ordered but retain their free rotational and lateral diffusion. The size and stability of these domains are not clear. Dimensions between a few molecules and 300 nm and half times between nsec and sec have been reported. Membrane proteins may selectively redistribute into and enrich in rafts. Apart from redistribution into these domains, non-random lateral distribution of proteins is achieved by other phenomena such as protein oligomerization or interaction with peripheral proteins (membrane skeleton) on the cytoplasmic leaflet. Furthermore, membrane proteins may become depleted in membrane regions of high curvature/bending.

Specific protein-lipid interactions are another factor leading to lateral heterogeneous lipid arrangement. These interactions induce an enrichment of selected lipid species in the immediate vicinity of membrane proteins, forming a lipid shell.

The so-called tight junctions separating the plasma membrane of polarized cells (e.g. epithelial cells) into an apical and a basolateral domain are a characteristic example of protein-mediated membrane domain formation. The two domains differ with respect to protein and lipid composition.

Transversal Distribution of Lipids

Numerous eukaryotic cell types exhibit a non-random, asymmetric distribution of phospholipids across their

plasma membranes and probably also in the late Golgi and endosomal compartments (Fig. 2) (2). For example, cerebrosides, gangliosides and sphingomyelins are found only in the exoplasmic leaflet of the plasma membrane. The enrichment of sphingolipids on the cell surface not only plays an important structural role in protecting cells but is also essential for intercellular recognition and signal transduction. For most mammalian cells, the exoplasmic leaflet of the plasma membrane is especially rich in PC, providing a rather inert surface essential for stability and barrier function, whereas the inner leaflet is rich in the amino phospholipids PS and PE, lipids which promote membrane-membrane interactions and fusion.

Dynamics

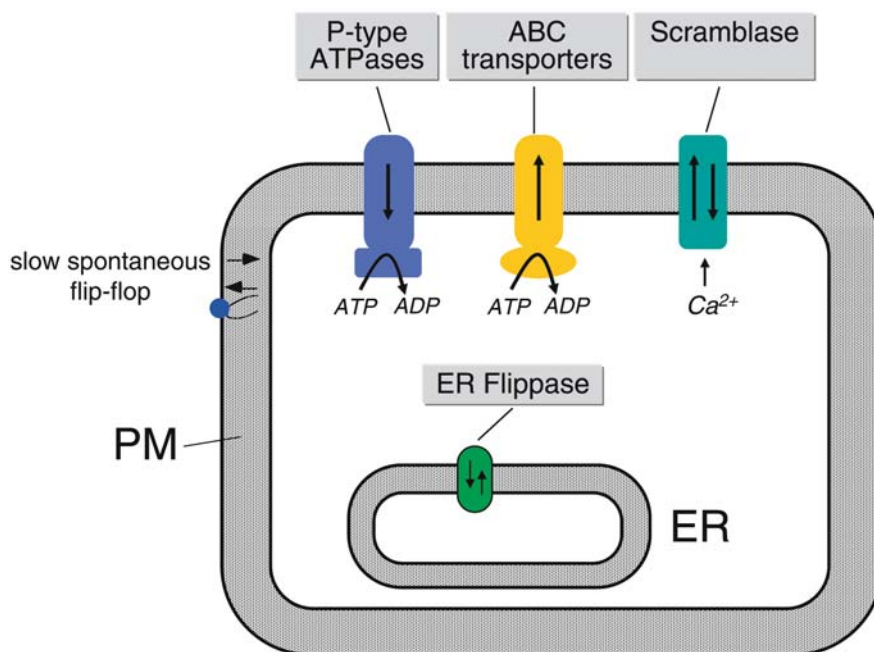
Lipid bilayers are two-dimensional fluids, since membrane lipids and many proteins can easily move within the plane of a membrane leaflet (1, 3). In the case of lipids, this lateral diffusion is very rapid; the lateral diffusion coefficient (the area over which a molecule moves in 1 sec) is about 10^{-7} to 10^{-8} cm²/sec, being slightly lower for liquid-ordered domains. Furthermore, individual lipid molecules rotate very rapidly along their head-to-tail axis with a characteristic time of 10^{-7} sec, and display molecular vibrations with a characteristic time of 10^{-14} sec. Membrane proteins can also move laterally within the membrane. Their lateral diffusion is about two orders of magnitude slower and depends on their size, degree of oligomerization and interaction with peripheral proteins as well as on the physical state of the surrounding lipid phase (for example, entrapment in rafts). Lipids can move between the two leaflets of the membrane bilayer, a process termed transverse diffusion or flip-flop (2). Neutral lipids such as diacylglycerol and charged lipids such as free fatty acids, phosphatidic acid or phosphatidylglycerol in their protonated form can move from one leaflet to the other in seconds or minutes. By contrast, phospholipids with polar head groups and glycolipids with bulky hydrophilic carbohydrate moieties flip spontaneously only slowly across a lipid bilayer, displaying half-times of hours to days depending on the size and charge of the head group. Cholesterol is embedded in the membrane with its polar OH group facing the aqueous phase and the acyl chain pointing towards the bilayer center. Experimental evidence supports rapid flip-flop of cholesterol in PC membranes, red cell membranes and, presumably, in most other cellular membranes.

Regulatory Mechanisms

For the specific function of each biological membrane, distinct sets of lipids and proteins are essential. In many cases, structure, organization and activity of membrane

proteins depend on the presence of specific lipids. In eukaryotes, a selective transport of lipids and proteins between organelles is the central process in the organization of membrane compartments (2). This process is mainly mediated by the budding of transport vesicles from a donor compartment followed by the vectorial trafficking to and fusion with an acceptor compartment. Apart from differences in protein and lipid composition, the various organellar membranes display striking differences in the leaflet distribution and transbilayer movement of their lipids (Figs. 2, 3). For example, the endoplasmic reticulum is the principal site of membrane assembly in eukaryotes and in this aspect similar to the plasma membrane in prokaryotes. Lipid synthesis in these membranes is an asymmetric process that results in the insertion of new lipids into the cytoplasmic leaflet. To ensure balanced growth and thus stability of the membrane, half of the newly synthesized lipids must flip to the opposing leaflet. This process is facilitated by so-called ►flippases, membrane proteins that catalyze a transverse diffusion

of most, if not all, lipid classes in both directions and function independently of metabolic energy (2). The situation is different for the plasma membrane of eukaryotes where flip-flop of phospholipids is constrained and ATP-dependent transporters maintain an asymmetric lipid distribution by moving specific lipids towards (►P-type ATPase family members) (5) or away from (►ABC transporters) the cytosolic leaflet (6). The maintenance of the nonrandom lipid distribution across the plasma membrane is important for cellular functions. Changes in this arrangement generally trigger a physiological event such as the nucleation of coagulation cascades on the surface of activated platelets or the phagocytic clearance of apoptotic cells by macrophages. Local changes in transbilayer ►lipid asymmetry caused by local lipid metabolism, transbilayer movement or domain formation are crucial for processes such as membrane budding and fusion, where membrane curvature plays an important role. Furthermore, the lateral organization of lipids, especially cholesterol and sphingolipids, into



Biological Membranes. Figure 3 Specific proteins (flippases) involved in lipid transbilayer movement are embedded in biological membranes. The activity of these proteins may be unspecific or highly selective (e.g. for the head-group of phospholipids). In the endoplasmic reticulum (ER), bi-directional flippases facilitate rapid flip-flop of lipids and allow them to equilibrate between the two membrane leaflets independently of ATP. This system is unable to accumulate a given lipid in one leaflet, thereby promoting a symmetric lipid distribution across the bilayer. In contrast, flip-flop of phospholipids across the plasma membrane (PM) is constrained owing to high levels of cholesterol and sphingolipids and/or the absence of constitutive bi-directional flippases. Thus, ATP-dependent flippases can maintain an asymmetric lipid distribution by moving specific lipids towards (P-type ATPase family members) or away from the cytosolic leaflet (ABC transporters). Cellular activation triggered by cytosolic calcium can collapse the lipid asymmetry by the transient activity of an ATP-independent scramblase.

rafts has consequences for the recruitment and concentration of proteins, regulation of protein-protein interactions within the membrane and cell signaling functions. Since rafts have also been reported for subcellular membranes, these domains may provide a molecular basis for sorting lipids and proteins into specific vesicular pathways.

References

1. Gennis R (1989) Biomembranes. Molecular Structure and Function. Springer, Berlin Heidelberg New York
2. Sprong H, Sluijs P van der, Meer G van (2001) How proteins move lipids and lipids move proteins. *Nat Rev Mol Cell Biol* 2:504–513
3. Edidin M (2003) Lipids on the frontier: a century of cell-membrane bilayers. *Nat Rev Mol Cell Biol* 4:414–418
4. Munro S (2003) Lipid rafts: elusive or illusive? *Cell* 115:377–388
5. Kühlbrandt W (2004) Biology, structure and mechanism of P-Type ATPases. *Nat Rev Mol Cell Biol* 5:282–294
6. Holland IB, Cole SPC, Kuchler K et al (eds) (2003) ABC proteins: From Bacteria to Man. Academic Press

Biologicals

► Biopharmaceuticals

Biologics

► Biopharmaceuticals

Biomarker

Definition

A biomarker is a biomolecule-based diagnostic marker for a disease, or certain disease stage.

- Mass Spectrometry: SELDI
- Proteomics in Human-Pathogen Interactions

Biomimetic Ligands

Definition

Biomimetic ligands are synthetic protein ligands that mimic natural ligand-protein interactions.

- Affinity Chromatography and *In Vitro* Binding (Beads)

Biopanning

Definition

Biopanning is the procedure of selecting binding partners from phage display libraries. In this affinity-driven selection procedure, specific binders against a target of choice are enriched from a phage display library during consecutive cycles of incubation, washing, amplification and re-selection of bound phage. Multiple rounds of selection are necessary to enrich specifically binding phage particles over a general background of unspecific sticky phage.

- Protein Interaction Analysis: Phage Display

Biopharmaceuticals

Definition

Biopharmaceuticals, biologics and biologicals are synonyms that refer to natural biomolecules that are used in medical therapy and are produced biotechnologically in host organisms or cells. Many biopharmaceuticals are identical to molecules of the human body, e.g. interferon. The largest class of biopharmaceuticals are antibodies.

- Affinity Chromatography and *In Vitro* Binding (Beads)

Biosensors

Definition

Biosensors are detection systems that use reactions that are performed or mediated by biological materials (enzymes, antibodies, cells) to detect different substances. The reaction is generally measured as an electrical, thermal or optical signal.

- Monoclonal Antibodies

Bipartite NLS

Definition

Bipartite NLS (Nuclear Localization Signal) designates a specific sequence within a protein that is responsible for the translocation into the cell nucleus, composed of

two interdependent clusters of basic amino acids separated by a spacer region of 10–12 amino acids.

►NLS (Nuclear Localization Signal)

Bipolar Affective Disorder

MARKUS M. NÖTHEN^{1,2,3}, SVEN CICHON^{1,2,3},
CHRISTINE SCHMÄL⁴, MARCELLA RIETSCHEL⁴

¹Department of Medical Genetics, University of Antwerp, Belgium

²Department of Genomics, Live and Brain Center, University of Bonn, Germany

³Institute of Human Genetics, University of Bonn, Germany

⁴Division of Genetic Epidemiology in Psychiatry, Central Institute of Mental Health, Mannheim, Germany

markus.noethen@uni-bonn.de

sven.cichon@uni-bonn.de

schmael@as200.zi-mannheim.de

rietschel@zi-mannheim.de

Synonyms

Manic depression

Definition

Bipolar affective disorder, also termed “manic depression”, is characterized by an alternation in mood between two poles, mania and depression. This alternation can be rapid, taking place within hours or days, or the episodes of illness may occur at longer intervals, separated by months or years.

The mood in the depressed phase is sad and despondent. Other characteristic features are loss of drive, slowness of thought, feelings of guilt and loss of interest in life. These features are typically accompanied by appetite and sleep disturbance. In contrast, the manic phase is characterized by an expansion of mood, with exaggerated estimation of ability, increased drive, rapid thoughts and speech. Marked irritability is another frequent feature of the manic phase.

The modern classification system DSM IV (1) distinguishes between bipolar I and bipolar II disorder, and the distinction is based on the severity of the manic symptoms. Bipolar I disorder has clear manic symptoms and may include delusions. Bipolar II disorder has less distinct symptomatology, so called hypomanic symptoms. In cases where only depressive symptoms are found, the diagnosis of unipolar depression is applied. The depressive symptoms of unipolar depression and the depressive phases of bipolar disorder

cannot be distinguished clinically. However, course of disease, treatment and formal genetic findings suggest that unipolar depression and bipolar disorder should be considered as separate illnesses.

Characteristics

Bipolar affective disorder is a serious and common illness. The lifetime prevalence of bipolar I disorder is 0.5–1%, and the suicide rate for affected persons is approximately 10–15%. Despite intensive research, the cause of the disorder remains obscure. Thus, although some effective medications are available, treatment remains incomplete at this time.

Formal Genetic Studies

A substantial contribution of genetic factors to the development of bipolar disorder has been suggested through a large number of formal genetic studies (5).

Family Studies

All completed family studies to date, using standard diagnostic methods, have shown that first-degree relatives of patients with bipolar I disorder have an increased risk of developing a bipolar disorder. The relative risk for first-degree relatives is approximately 7. These studies also show that the first-degree relatives have an increased risk of developing unipolar depression. The risk is approximately doubled. It is not possible in individual cases to determine if the prevailing depressive illness is related to the familial genetic loading with bipolar disorder, as unipolar depression is very common in the general population, with a life time prevalence of 5–10%. There are estimations that 70% of unipolar cases among relatives of bipolar patients share a common genetic background. In these families it is noted that unipolar depression occurs twice as commonly in women as in men, while the risk for bipolar disorder is equal for both sexes. The same sex ratio is observed in the general population. Other disorders that are observed in relatives of patients with bipolar I disorder are bipolar II disorder and schizoaffective disorder, so that, as with unipolar disorder, there is a suspicion of overlapping aetiology.

Twin Studies

The ►concordance rate for bipolar disorder for monozygotic twins is on average 50% in comparison with the rate for dizygotic twins, which is approximately 10%. If unipolar disorder is considered, the concordance rate is approximately 80%. The extent of heritability, that is the extent to which genetic individual differences contribute to individual differences in phenotype, is 60–80%.

Adoption Studies

There has only been one adoption study to date from which meaningful conclusions can be drawn concerning

bipolar disorder (6). This identified increased morbidity rates for affective disorders in biological parents of bipolar patients, compared to adoptive parents of affected children and to a control group of adoptive parents of non-affected children.

Specificity of the Genetic Contribution

The fact that schizoaffective disorder is also found in the relatives of bipolar patients may suggest an aetiological overlap with schizophrenia spectrum disorders. Bipolar patients are frequently observed to have co-morbidity with other psychiatric disorders. Early onset disorders have been associated with attention deficit hyperactivity disorder and conduct disorder, while later onset illness has been associated with substance misuse. This observed co-morbidity might be a sign of ►pleiotropic gene effects.

Cellular and Molecular Regulation

The biological mechanisms responsible for the development of bipolar disorder remain largely unknown. Biological psychiatric research has proposed a multitude of possible mechanisms, including a disturbance of neurotransmitters, intracellular transduction or neuroendocrine regulation. Whether these mechanisms are genuinely implicated, and to what extent they are involved, is still unclear. The advantage of the molecular genetic approach is that the identification of vulnerability genes reveals causal factors. This will allow understanding of the function of a given gene product, and gradually reveal the functional context which ultimately produces the clinical phenotypes.

►Candidate Gene Studies

The number of published studies on candidate genes for bipolar disorder is large, and a simple interpretation of the findings is difficult. Many candidate gene studies have the disadvantage of small sample sizes and the lack of a systematic approach for the investigation of specific candidate genes (e.g. low number of investigated polymorphisms and insufficient knowledge of haplotype structure). Furthermore, there are theoretically a multitude of biological mechanisms that could be responsible for the illness, and therefore there are a large number of potential candidate genes to be examined. When a candidate gene is in a chromosomal region, which has been shown to have linkage with the disorder, it naturally has a higher plausibility.

Amongst the most discussed candidate genes are the *serotonin transporter gene (5-HTT)*, the *catechol-O-methyl-transferase gene (COMT)*, and the *brain-derived neurotrophic factor gene (BDNF)* (4). On account of the inconsistency of reported findings, no judgement can be made at this time.

Linkage Studies

The linkage study approach is carried out to localize disposition genes to specific chromosomal regions. Since the mid 1980s, linkage studies in bipolar disorder have been carried out with the help of genetic markers. The value of the first studies lay less in their results than in the insights they gave into the methodological and conceptual weaknesses of transferring the classical approach of linkage studies for monogenic disorders to complex genetic disorders. This understanding gave impetus to the further development of the linkage method, in particular the creation of assumption-free analysis methods using large samples of “nucleus families”, that is families with a minimum of an affected sib-pair and their parents. In recent years these methods have enabled the localisation of chromosomal regions which have a high probability of containing genes which contribute to bipolar disorder: 4p16, 4q35, 8q24, 10q25-q26, 12q23-24, 13q32-q33, 18p11.2-cen, 18q21-q23, 21q22 and 22q12-q13. Although all these chromosomal loci have been found by two or more independent research groups, no locus has been consistently replicated by all of the groups. This reflects on the one hand the high degree of locus heterogeneity and on the other the variability of studies with regard to their phenotype definition, sample sizes and the type and density of genetic markers used.

In recent years two large meta-analyses of all published genome-wide linkage studies have been conducted (2, 7). Since the original studies were not uniform with regard to a number of methodological issues (e.g. diagnostic criteria, genetic markers), by integrating the results the ►meta-analyses had to accept a degree of loss of information. A specific advantage of such meta-analyses, however, lies in their sensitivity for genes with a relatively small but widespread contribution to the development of the illness. These genes may remain undetected in an individual study but may be detectable if results from a large number of studies are combined. It is therefore not surprising that the meta-analyses, in addition to the confirmation of previously implicated loci, also suggested new loci (9p22-p21, 10q11-q22, and 14q24-q32). Interestingly, for some of the chromosomal regions suggested by the meta-analyses, linkage has also been reported with schizophrenia. Since there is a possible aetiological overlap between bipolar disorder and schizophrenia, the genes that have been identified for schizophrenia in these regions are excellent candidate genes for bipolar disorder. Using this strategy, the G72/G30 locus (chr. 13q33) has recently been associated with bipolar disorder. The responsible gene has not yet, however, been unequivocally identified.

Clinical Relevance

Phenotypic Heterogeneity

The results to date of linkage studies suggest that the contribution of an individual gene to the aetiology of bipolar affective disorder is low to moderate. With complex genetic illnesses there has been a series of examples where characteristic illness phenotypes have been helpful in defining aetiologically more homogeneous sub-types. In bipolar affective disorder the following features have been proposed for homogenisation: age at onset, presence of psychotic symptoms, rapid cycling, response to lithium and co-morbidity with other psychiatric disorders (3). The inclusion of these parameters is only in its initial stages within the context of linkage studies. Future studies will consider these parameters in more detail.

The Need for Large Samples

For genes with small to moderate effects, the search for genes is highly reliant on the size of the samples. This relates to both association and linkage studies, particularly in the latter, where small family numbers can dramatically reduce statistical power. Several collaborative efforts to collect large samples are under way. Retrospective meta-analyses, such as the above-mentioned studies, have the disadvantage that the original studies are often not comparable in terms of phenotype definition, selection of genetic markers and so forth, which results in considerable information loss. Prospective collaborative studies, in principle, will be better able to identify genes with small to moderate effects.

The Potential of Isolate Populations

In isolated populations it is hoped that, through founder effects, the contribution of individual genes might be larger and easier to detect. There is, however, no unanimous opinion regarding which population size and history are most appropriate for use in gene identification. In bipolar disorder, genome-wide linkage studies have been conducted in families drawn from isolated populations such as the Old-Order-Amish, the Finns, Central Valley Costa Rican families and families from the Quebec region of Canada. Despite promising linkage findings, this approach has not yet identified a gene. In isolated populations, the ►linkage disequilibrium can extend over a large genomic region, which could be unfavourable for the detection of genes.

Gene-environment Interaction

Better understanding of the interaction between genes and environment may rest with identification of illness genes. The mode of gene-environment interaction is still unknown. It is possible that many gene-environment

interactions are based on additive effects, though more complex interactions are also possible since data from family and twin studies suggest that genetic factors also influence the susceptibility and the tendency for exposure to environmental factors.

References

1. American Psychiatric Association (1994) Diagnostic and Statistical Manual of Mental Disorders, 4th edn. American Psychiatric Association, Washington, DC
2. Badner JA, Gershon ES (2002) Meta-analysis of whole-genome linkage scans of bipolar disorder and schizophrenia. *Mol Psychiatry* 7:405–411
3. Faraone SV, Tsuang MT (2003) Heterogeneity and the genetics of bipolar disorder. *Am J Med Genet (Sem Med Genet)* 123:1–9
4. Jones I, Craddock N (2001) Candidate gene studies of bipolar disorder. *Ann Med* 33:248–256
5. Jones I, Kent L, Craddock N (2002) Genetics of affective disorders. In: McGuffin P, Owen MJ, Gottesman II (eds) *Psychiatric Genetics and Genomics*. Oxford University Press, Oxford, pp 211–245
6. Mendlewicz J, Rainer JD (1977) Adoption study supporting genetic transmission in manic-depressive illness. *Nature* 268:327–329
7. Segurado R et al (2003) Genome scan meta-analysis of schizophrenia and bipolar disorder, part III: Bipolar disorder. *Am J Hum Genet* 73:49–62

Bispecific Antibodies

Definition

Bispecific antibodies react with two different antigens. All natural antibodies only react with one antigen, because both antigen-binding fragments have the same amino acid composition. Bispecific antibodies can be produced by biochemical, cell fusion or recombinant methods.

►Monoclonal Antibodies

Bithorax Complex

Definition

Bithorax complex (BX-C) designates the cluster of homeotic genes that specify the identities of posterior segments in *Drosophila Melanogaster*.

►Homeodomain Transcription Factors

BLAST

Definition

BLAST stands for Basic Local Alignment Search Tool, a program for searching biosequence databases, designed to identify the classification and potential homologs for a given sequence. Different versions allow DNA or protein sequences as input, as well as DNA or protein databases. This tool is maintained by the National Center for Biotechnology Information (NCBI). BLAST locates paths of regional similarity instead of calculating the best overall alignment using gaps. The program then uses a scoring matrix to rank these matches as positive, negative or zero. If the initial match is scored highly, the search is expanded in both directions until the ranking score falls off.

- [Protein Databases](#)
- [Protein Domains](#)
- [RNA Interference in Mammalian Cells](#)

BLAST Link

- [BLink](#)

Blastocyst

Definition

The blastocyst is a multicellular early embryo (approximately 3.5 days post fertilization in mouse) in which the cells destined to form the developing fetus have not yet differentiated into discrete cell types. A blastocyst consists of a hollow ball of trophoblast cells, surrounding a fluid-filled cavity, and an inner cell mass (which is composed of pluripotent stem cells). Gene-modified mice are generated by injection of blastocysts with gene-modified embryonic stem cells.

- [Cre/Lox P Strategies](#)
- [Large-Scale Homologous Recombination Approaches in Mice](#)
- [Hereditary Neuropathies, Motor and/or Sensor](#)
- [Shotgun Libraries](#)

Blastocyst Injection

Definition

Blastocyst injection is a method to make chimeric animals, where embryonic stem cells are injected into the cavity (the blastocoel) of blastocysts and then implanted into pseudopregnant host mice.

- [Large-Scale Homologous Recombination Approaches in Mice](#)
- [Transgenic and Knockout Animals](#)

Blastomere

Definition

Blastomere is a single large (undifferentiated) cell of the early multicellular embryo.

- [Prader Willi and Angelman Syndromes](#)
- [Morpholino Oligonucleotides, Functional Genomics by Gene 'Knock-down'](#)

BLink

Definition

BLink displays the results of BLAST searches that have been done for every protein sequence in the Entrez Proteins data domain. BLink displays the graphical output of pre-computed BLASTP results against the protein non-redundant (nr) database.

- [Protein Databases](#)
- [BLAST](#)

Blocked 5' Structure

Definition

Blocked 5' structure denotes mRNA 5' terminal phosphate groups at the 5' end of mRNA, which are blocked and protected from removal by phosphatases, in the case of cap, by 7-methyl-guanosine.

- [RNA Capping](#)

Blocks

Definition

Blocks are multiply aligned ungapped (►Gap) segments corresponding to the most highly conserved regions of proteins.

►Protein Databases

Bloom Syndrome

EBERHARD PASSARGE

Institute of Human Genetics, University Clinic Essen,
Essen, Germany

eberhard.passarge@uni-essen.de

Definition

Bloom syndrome (►MIM 210900) is a human hereditary disease. It was first described by a dermatologist in New York, David Bloom (1954), under the title “Congenital telangiectatic erythema resembling lupus erythematoses in dwarfs” (1). It is one of a group of disorders characterized by signs of genomic and ►chromosomal instability, an increased rate of ►somatic mutations and an increased risk of developing different types of malignancies. Bloom syndrome can be classified as a ►DNA helicase-deficiency disorder that results from mutations in the *BLM* gene, which encodes a ►DNA helicase of the RecQ type.

Characteristics

Bloom syndrome is unique among human hereditary diseases as it combines clinical features that occur individually in other disorders, but here together result in a rather specific clinical and cellular phenotype. The main clinical features are (1) striking pre- and postnatal growth retardation, (2) a characteristic configuration of the face and head, including facial skin manifestations in most patients, (3) variable, unspecific immune deficiency resulting in increased susceptibility to infections, mainly of the broncho-pulmonary system, (4) areas of hyper- and hypo-pigmentation of the skin, (5) absence of mental retardation, but development of a behavioral pattern of reduced attention span with learning and memory deficits, (6) increased risk of developing various types of malignancies throughout life, often with multiple independent primary tumors at

different ages, (7) hypersensitivity to DNA-damaging chemicals *in vivo* and *in vitro*, (8) male infertility, (9) feeding difficulties in infancy and early childhood, followed by lack of appetite in adults and (10) diabetes mellitus type 1 and type 2 in about 17% of patients over the age of 20. A spontaneous ten-fold increase in the rate of ►sister chromatid exchanges (SCE) in metaphases derived from cultured lymphocytes serves as a specific diagnostic test.

- Data Base:

The Bloom Syndrome Registry has provided a comprehensive database for Bloom syndrome (2, 3, 4, 5).

- Inheritance pattern:

autosomal recessive. No clinical or cellular manifestations in heterozygotes.

- Gene Locus:

on the long arm of chromosome 15 at region 2, band 6.1 (15q26.1).

- Genetic Heterogeneity:

No other gene locus known, but considerable allelic heterogeneity demonstrated by molecular gene mutation analysis.

- Population Genetics:

A characteristic ►founder mutation occurs in populations of ►Ashkenazi Jewish origin with a heterozygote frequency of about 1 in 107 and linkage disequilibrium with linked polymorphic microsatellite loci. The estimated heterozygote frequency in non-Jewish populations is about 1 in 300; parental consanguinity is increased in non-Jewish populations. Bloom syndrome has been observed in different populations of European ancestry, including Latin American countries and in Japan.

- Mouse model:

Homozygous knockout mice (*Blm*^{-/-}) are not viable and die at 13.5 days of embryogenesis (6).

Cellular and Molecular Regulation

Cellular Phenotype

Cultured cells from homozygotes *BLM*^{-/-} grow poorly. Lymphocytes transform with reduced efficiency. Lymphoblastoid cell lines are difficult to establish and require considerably more time than normal cells. Cultured fibroblasts have an extended doubling time and occasionally form anaphase bridges. DNA excision repair is normal.

Metaphase Chromosomes

A high rate of breaks in one or both chromatids is visible in about 30–40% of metaphases. Reunion of homologous chromosomes, evident as characteristic quadriradial figures, occurs in about 1–3% of metaphases. Dicentric chromosomes, acentric fragments

and premature chromosome condensations can be observed in a few percent of cells.

Sister Chromatid Exchanges (SCEs)

An about ten-fold spontaneously increased rate of sister chromatid exchanges is the diagnostic hallmark of Bloom syndrome. It is the only disease with this cellular phenotype, which serves as a specific diagnostic test.

Occasionally in certain patients a few individual metaphases show a normal SCE pattern. This is interpreted as a reversion to the normal phenotype as a result of somatic intragenic recombination when the underlying mutations are ►**allozygous**, i.e., at different sites in the *BLM* gene (compound heterozygous state). This observation was instrumental in identifying the gene (7).

Increased Rate of Somatic Mutations

Bloom syndrome is the prototype of a human mutator phenotype. Somatic mutation occurs at an increased rate in all tissues tested. Direct evidence has been obtained by studying the following loci in lymphocytes from affected persons: HGRT, glycophorin A, MN blood type, HLA (5).

Identification of the Bloom Gene *BLM*

The gene is localized on 15q26.1. It was identified by a novel variant of ►**homozygosity mapping** called somatic crossover point (SCP) mapping (7). This was based on five lymphoblastoid cell lines from *bona fide* patients with a normal rate of SCE. German and

colleagues argued that the reversion to a normal rate of SCEs was the result of intragenic recombination, which rendered the cell involved heterozygous and thereby normalized the rate of SCEs. The objective was to identify the region (the *BLM* region) between the most proximal polymorphous marker locus (in the direction of the centromere) that was constitutionally heterozygous, but had become homozygous in the cell lines with low SCE and the region with loci that always were homozygous and therefore were not subject to Bloom-related intragenic recombination. In this region of 250 kb between marker loci D15S1108 (proximal) and D15S127 (distal), a 4437-bp cDNA sequence was found containing an open reading frame encoding a 1417-amino acid protein. This hybridized to single-copy genomic sequences spanning about 100 kb of genomic DNA. A 4.5-bp transcript was identified from total RNAs prepared from different human cells proliferating *in vitro*.

Structure of the *BLM* Gene

The *BLM* cDNA contains 7 motifs present in most DNA and RNA helicases. From this, and sequence homology observed in cDNA, it can be deduced that the gene encodes a DNA helicase that belongs to the RecQ family of DNA/RNA helicases (see below).

Mutations

The Bloom Syndrome Registry (3, 5) contains 64 different mutations scattered throughout the gene, identified in 124 of 134 affected persons (Table 1). Eighteen mutations were recurrent in unrelated

Bloom Syndrome. Table 1 Summary of Mutations reported in Bloom syndrome, in a sample of 133 patients*

Observation	Number of patients in which mutation found
Mutations present	124
Unique mutations:	64
Amino acid substitutions	10
Small deletions/insertions	21
Nonsense mutations	20
Splice site mutations	9
Deletions of one exon or more	4
Total number of protein truncating mutations	54
Missense mutations	10
Recurrent mutations	18
No mutation	9

* Data reported by German & Ellis, 2002 (4)

persons. Most common are nonsense mutations resulting in truncated protein (21 small insertions/deletions, 20 resulting in a premature stop codon, 10 substitutions in conserved regions, 9 splice site mutations, 4 deletions of a whole exon or more and 10 missense substitutions (of these 5 in a helicase domain). A unique mutation occurs in persons of Ashkenazi Jewish origin, a 6-bp deletion/7-bp insertion in exon 10 at nucleotide 2281 (*BLM*^{Ash}). The next most common mutation allele is C1933T, which results in an exchange of a glycine (Q) for a stop codon (X) at amino acid 645 (Q645X).

Homology of the BLM Protein

Homology comparison of amino acid sequence data bases established that the 1417-amino acid protein encoded by the *BLM* gene belongs to the ▶*RecQ* helicases, a subfamily of DExH box-containing DNA and RNA helicases (5). These occur in such diverse organisms as *E. coli*, yeast (*S. cerevisiae*, *S. pombe*), *Drosophila melanogaster*, *Xenopus laevis* and *Homo sapiens*. In humans, five helicases of this type are known, BLM (1417 amino acids), WRN (1432 amino acids), RECQL4 (1208 amino acids), RECQL (649 amino acids) and RECQL5 (410, 435 and 991 amino acids owing to three RNA splice variants) (6). Of these, three are known to be related to human genetic disorders, BLM in Bloom syndrome, WRN in Werner syndrome (MIM 277700) and RECQL4 in Rothmund-Thomson syndrome (MIM 268400).

Molecular Biology of the BLM Protein

Antibodies to an N-terminal part of BLM raised in rabbits make it appear to be identical to a protein of 180 kD identified by western blot analysis of fibroblast, lymphoblastoid and HeLa cells that is absent in all Bloom syndrome cell lines homozygous for a premature translation-termination mutation (3). Presumably this protein is the *BLM* gene product. This protein is present in the nucleus of all proliferating cells, except those homozygous for the *BLM* mutation. This corresponds to the presence of a nuclear localization signal in the last 100 amino acids of the BLM protein. BLM is present in the nucleus in two types of distribution, as fine granular speckles in varying local concentrations and as bright, punctate foci. The latter have been identified as ▶*PLM* nuclear bodies (reviewed in 5). At metaphase, the BLM protein becomes homogeneously and diffusely distributed throughout the cytoplasm (8). This reorganization correlates with ▶*ATM*-dependent phosphorylation. BLM interacts with RAD51 in the yeast two-hybrid system suggesting that BLM participates in recombination-coupled repair of double strand breaks.

In meiosis BLM is present near paired chromosomes in the zygotene and leptotene stages. This suggests that

lack of functional BLM protein interferes with meiotic prophase and chromosome pairing. This may be the basis of the clinically observed lack of spermatogenesis and male infertility (reviewed in 3).

Clinical Relevance

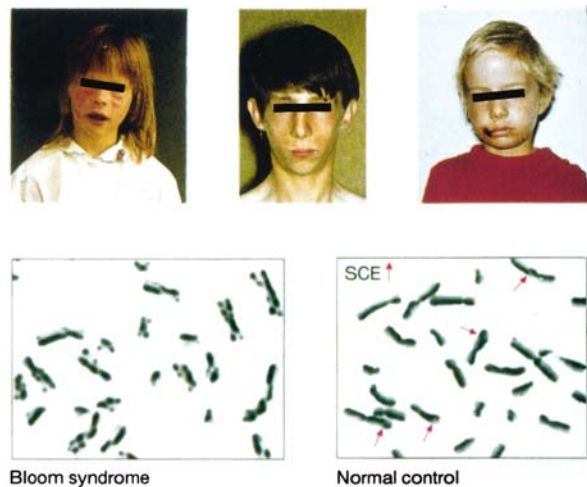
The various clinical features of Bloom syndrome have been reviewed in detail elsewhere (2, 3, 4, 5). Therefore, only the main features will be mentioned here.

Growth Retardation

This is a constant feature that can be considered as obligatory. The mean birth weight is 1800–1900 g at normal gestation. Growth is below the third percentile throughout childhood until an average adult height of about 140–150 cm is reached. No adult patient has been taller than 160 cm. The prepubertal growth spurt does not take place. Growth retardation is the most frequent cause of seeking medical advice.

Craniofacial Phenotype

Until early adulthood patients can be easily recognized by their narrow face and skull (dolichocephaly) (Fig. 1). Later in life this may become less obvious, up to point that the diagnosis cannot be entered based on the facial phenotype (own unpublished observations).



Bloom Syndrome. Figure 1 The clinical and the chromosomal phenotype in Bloom syndrome. The upper row shows three different patients with the characteristic facial phenotype. Below are part of a metaphase with a high rate of sister chromatids in Bloom syndrome (left) and a normal metaphase (right). (Figure from E. Passarge, *Color Atlas of Genetics*, 2nd ed., Thieme Medical Publishers, Stuttgart-New York, 2001).

Skin Manifestations

Most patients develop a sun-sensitive telangiectatic erythema during their second or third year of life. This may be the chief complaint in some patients. Eyelids, ears and the dorsal side of their hands may be involved, but no other parts of the body even when exposed to sun. A few patients do not manifest skin lesions or show just some reddening.

Immune Deficiency

Reduced levels of immunoglobulins are common albeit variable. Common infections, mainly of the bronchopulmonary system, the ears and the gastrointestinal system occur more often and are more prolonged than in unaffected children. In some cases they may be life threatening. Response to vaccinations is usually normal, but immunity to repeat infections is reduced.

Feeding difficulties

This is a common complaint. Vomiting, regurgitation and esophageal reflux are frequent signs. In addition, lack of appetite is a constant feature. A number of patients had to be fed by gastric tube.

Hyper- and Hypo-pigmented Areas of the Skin

This is a medically unimportant, but diagnostically typical sign. The areas of hyper- or hypo-pigmentation range from a few millimeters to decimeters in size are irregularly shaped and range from a few to a dozen or more areas. We interpret them as signs of somatic recombination involving pigmentary skin cells.

Predisposition to Cancer

This is the most impressive feature of Bloom syndrome. In the most recent assessment (5) no fewer than 90 malignant tumors arose in 70 patients. In 13, two or more primary tumors developed. The most common types of cancer are leukemias of different types, lymphomas (Burkitt and others) and carcinomas of the gastrointestinal tract, stomach, naso-oro-pharynx and esophagus, breast and other organs. The mean age at the diagnosis of a malignancy is 24.7 years (range 2–48 years). ► **Cancer** is the documented cause of death in 50 of 60 deceased persons in the Registry. In addition to leukemias, a myelodysplastic syndrome has been diagnosed in 14 registered persons between the ages of 4 and 39 (mean age 22.8). The precise reasons for the remarkably high prevalence of malignancies in Bloom syndrome have not been established. ► **Genomic instability** resulting from a deficient DNA helicase and the increased rate of somatic mutations are assumed to be important factors.

Hypersensitivity to DNA-Damaging Chemicals

Cells homozygous for a mutation in the *BLM* gene are highly sensitive to DNA-damaging agents. This needs

to be considered when initiating chemotherapy in a patient with cancer. The bone marrow is especially vulnerable. Single agent chemotherapy in a dose reduced to about 50% compared to the usual dose should be employed initially.

Male Infertility

Mild to moderate hypogonadism is present at puberty and thereafter. Puberty occurs at about the normal time. Adult males have been observed to be infertile. The testes are small and lack of spermatogenesis has been documented. Females may have reduced fertility. Four women with Bloom syndrome have borne a normal baby.

Diabetes Mellitus

In recent years an increased rate of diabetes mellitus has been observed in Bloom syndrome in about 17% of adult patients. Both insulin-dependent and non-insulin-dependent types have occurred. The basis for this observation is not known.

Diagnosis

Diagnosis is clinically based on the typical history and phenotype. Increased SCE in metaphases of cultured lymphocytes is diagnostic. Diagnostic mutational analysis is usually not necessary.

Management

Management is by diagnostic and genetic counseling. No regular diagnostic hematological evaluation in children, but parental awareness of the increased rate of leukemia and lymphoma is recommended. Adult patients benefit from regular examinations for adenocarcinoma of the gastrointestinal tract, breast and other organs.

► Chromosomal Instability Syndromes

Acknowledgement

I thank Dr. James L. German III. and Dr. Maureen Sanz, New York, for pleasant cooperation.

References

1. Bloom D (1954) Congenital telangiectatic erythema resembling lupus erythematosus in dwarfs. *Am J Dis Child.* 88:754–758
2. German J (1993) Bloom syndrome: A Mendelian prototype of somatic mutational disease. *Medicine* 72:393–406
3. German J, Ellis NA (2002) Bloom syndrome. In: Vogelstein B, Kinzler KW (eds) *The Genetic Basis of Human Cancer*. McGraw-Hill, New York, pp 267–288
4. German J, Passarge E (1989) Bloom's syndrome. XII. Report from the Registry for 1987. *Clin Genet* 35:57–69
5. German J, Passarge E, Sanz MM (2003) Bloom's syndrome. XXI. Report from the Registry for 2002, submitted

6. Levitt NC, Hickson ID (2003) DNA helicase-deficiency disorders. In: Nature Encyclopedia of the Human Genome, Vol 2, pp 96–102
7. Ellis NA, Groden J, Ye T-Z et al (1995) The Bloom's syndrome gene product is homologous to RecQ helicases. Cell 83:655–666
8. Sanz MM, Proytcheva M, Ellis NA et al (2000) BLM, the Bloom's syndrome protein, varies during cell cycle in its amount, distribution, and co-localization with other proteins. Cytogenet Cell Genet 91:217–223

Blotting

Definition

The transfer of proteins from a gel onto a membrane made of nitrocellulose or polyvinylidene difluoride is named blotting.

►Two-Dimensional Gel Electrophoresis

BMP

Definition

Bone morphogenetic proteins (BMPs) are members of the transforming growth factor- β (TGF- β) superfamily of growth factors. BMPs were originally identified as osteoinductive proteins in bone that induce ectopic bone and cartilage formation *In Vivo*. However, they are now known to be multifunctional regulators of cell growth, differentiation, apoptosis, and neurogenesis, and to play important roles during embryonic development

►Receptor Serine/Threonine Kinases

BMPR-IA/-IB

Definition

BMP Type IA receptor, also known as ALK-3, and IB receptor also known as ALK-6 (see ►BMP).

►Receptor Serine/Threonine Kinases

BMPR-II

Definition

BMP Type II receptor (see ►BMP and ►BMPR-IA/-IB).
►Receptor Serine/Threonine Kinases

Boc Chemistry

Definition

Boc chemistry is a peptide synthesis strategy in which the tert-butylmethoxycarbonyl group (t-Boc group) is used as a temporary protecting group for the N-terminus. The t-Boc group is cleaved under acidic conditions. Trifluoroacetic acid is mainly used for this purpose.

Bone and Cartilage

CHRISTINE HARTMANN

Research Institute for Molecular Pathology, Vienna, Austria

hartmann@imp.univie.ac.at

Definition

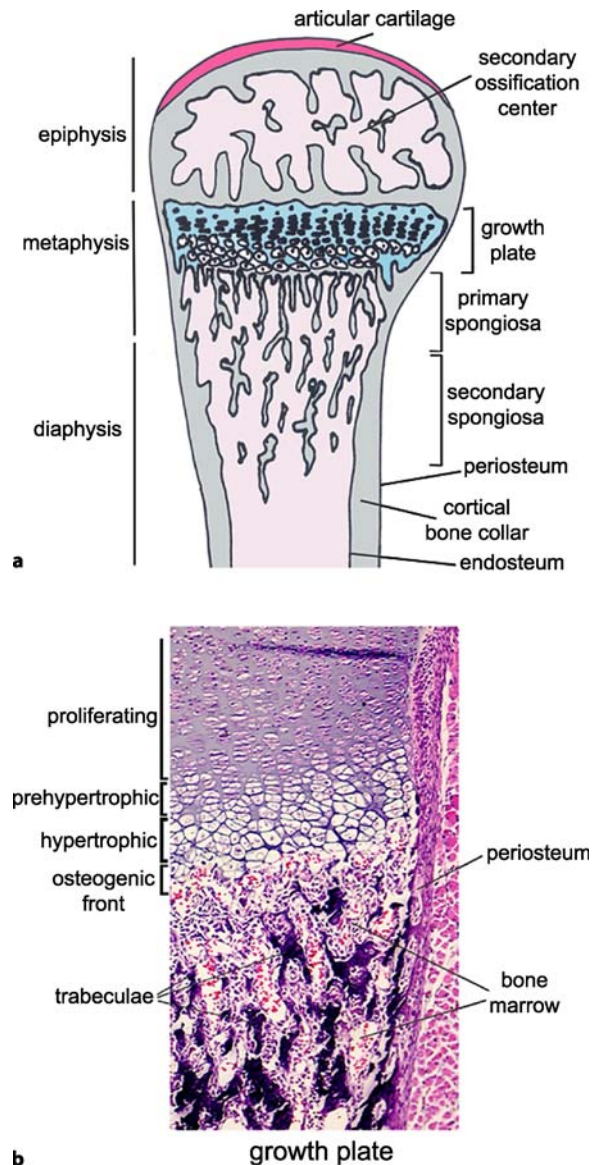
Bone is a highly specialized form of connective tissue, with a mineralized ►extracellular matrix (ECM), which confers rigidity and stability on the skeletal elements of higher vertebrates while still maintaining some degree of elasticity. In addition to its supportive function, bone is also the major source for inorganic ions, being actively involved in calcium ►homeostasis, and furthermore, the ►bone marrow is the main hematopoietic organ during ►postnatal life. Cartilage is also a specialized form of resilient, flexible connective tissue, which consists of one major cell type, the chondrocyte. Three different types of cartilage can be distinguished, hyaline-, fibro- and elastic-cartilages, based on their ECM and fiber arrangements. Hyaline cartilage is found in the trachea, nose, the growth plate and epiphysis of the long bones during postnatal growth and in the embryonic skeleton. Fibro-cartilage is found in the ►intervertebral discs and

joints, while elastic cartilage is found in the external ear and the ►[epiglottis](#).

Bones of the vertebrate skeleton are formed by two processes, membranous and endochondral ossification. The flat bones in the skull are formed by membranous ossification, whereby ►[neural crest](#)-derived, mesenchymal cells aggregate and directly differentiate into ►[osteoblasts](#). Most of bones within the vertebrate skeleton are formed by endochondral ossification, starting with a cartilaginous template, which ultimately will be replaced by bone. This process begins during embryonic development with the condensation of mesenchymal cells. Cells in the center of these aggregates will become ►[prechondrogenic](#) and will subsequently differentiate into chondrocytes, while cells at the periphery will go on to form the so-called ►[perichondrium](#). In order for the final replacement by bone to occur, chondrocytes have to undergo a well-controlled maturation program proceeding from proliferating chondrocytes, through prehypertrophic, to finally ►[hypertrophied](#) chondrocytes, which will gradually be removed and replaced by bone. The replacement of the hypertrophied, ►[calcified](#) region by bone is coupled with the entry of blood vessels into this region. Subsequently, this sequential process of chondrocyte proliferation, ►[hypertrophy](#) and replacement by osteoblasts becomes organized in the growth plates at the metaphyseal regions near each end of the growing bone, which regulate longitudinal growth of the skeleton until their disappearance at the end of puberty in humans (Fig. 1). The perichondrium adjacent to hypertrophic chondrocytes differentiates into a structure known as ►[periosteum](#). The innermost periosteal cells adjacent to the region of hypertrophic chondrocytes differentiate into osteoblasts, which secrete a bone matrix (►[osteoid](#)) that becomes progressively calcified, forming a membranous bone collar around the ►[diaphysis](#) (shaft region) of the skeletal element (1).

Characteristics

Bone is a complex, living tissue consisting of four different cell types, the ►[bone lining cells](#), osteoblasts, ►[osteocytes](#), and ►[osteoclasts](#). The ECM of bone is rich in type I collagen (Col1a1, Col1a2) and the tissue is heavily vascularized. Bone cells are of different origins; bone lining cells, osteoblasts and osteocytes are mesenchymal in origin, while the osteoclasts are derived from the hematopoietic lineage. Osteoblasts secrete the osteoid, which in its immature form, called woven bone, is characterized by irregular bundles and randomly oriented collagen fibers and is only weakly mineralized. Woven bone ultimately gets remodeled into lamellar bone, with a highly organized fiber structure and extensive mineralization. The shaft of



Bone and Cartilage. Figure 1 (a) Schematic drawing of the distal part of a long bone, indicating the different regions such as articular cartilage (red), secondary ossification center, growth plate (blue) and trabeculation (primary and secondary spongiosa). (b) Hematoxylin/eosin stained section of a growth plate region from a mouse humerus at birth, displaying the different zones of chondrocytes (proliferating, prehypertrophic, and hypertrophic).

endochondral-derived skeletal elements is surrounded by a thick layer of cortical bone, giving the bone its stability. Cortical bone is formed by osteoblasts originating from the periosteum through ►[appositional growth](#), while the trabecular bone, also known as

spongy bone or cancellous bone, found within the bone shaft is formed by a remodeling of calcified, hypertrophied cartilage, involving osteoblasts and osteoclasts.

The different skeletal disorders can be classified into metabolic (can be hereditary or acquired) and congenital and hereditary diseases and they can affect either bone or cartilage (2). Common bone diseases associated with a reduction in bone density are ►osteoporosis and ►osteopenia, which are due to an imbalance in bone remodeling, whereby osteoblasts produce less bone than the amount resorbed by osteoclasts. Osteoporosis is an age-related bone disease, associated with high fracture risk, due to the decreased bone mass. Metabolic bone diseases resulting from vitamin D deficiency are rickets in infants and growing children and ►osteomalacia in adults, due to insufficient mineralization of the bone. Von Recklinghausen's disease and osteitis fibrosa cystica are metabolic bone disorders due to hyperparathyroidism resulting in a diffuse or focal resorptive loss and fibrous replacement of bone, due to an excess of osteoclastic over osteoblastic activity caused by an over-production of ►parathyroid hormone (PTH).

Patients with ►osteopetrosis (hereditary disorder) have too high a bone mass, caused by an impairment in the differentiation of osteoclasts and/or their function. The inherited congenital bone disease, ►osteogenesis imperfecta (brittle-bone disease), is due to mutations in extracellular matrix proteins such as the type I collagen genes, resulting in structural or secretion abnormalities, affecting osteoid production and remodeling of woven into lamellar bone.

Cartilage on the other hand is ►avascular tissue and consists of chondrocytes secreting an ECM, which is rich in proteoglycans, sulfated mucopolysaccharides and type II ►collagens (Col2a1). When chondrocytes hypertrophy, they become vacuolated, produce type X collagen (Col10a1) and their matrix eventually becomes calcified. Chondrocytes, like osteoblasts, are of mesenchymal origin. Achondroplasia is an inherited ►congenital ►autosomal dominant disorder of the skeleton, characterized by a unique form of dwarfism and bone deformity resulting in a disproportionate shortness of the extremities relative to the trunk. Certain point ►mutations in one of the fibroblast growth factor receptor genes (FGFR3), leading to a constitutive activation of the receptor, cause achondroplasia. Disorders affecting the cartilage are generally referred to as chondrodysplasias, affect various segments of the bones, such as the metaphysis and ►epiphysis and can be associated with growth factors, their receptors or transcription factors. Furthermore, mutations in genes encoding components of the ECM often lead to skeletal disorders (3).

Molecular Interactions

Endochondral skeletogenesis starts with the aggregation of mesenchymal cells. The formation of these early aggregates is dependent on changes in cell adhesion properties and is under the control of positive and negative signals leading to the formation of the prechondrogenic condensations. Cells within this region express the transcription factor Sox9, which is absolutely required for the differentiation of condensed mesenchymal cells into chondrocytes. ►Haploinsufficiency of Sox9 in humans is associated with the severe skeletal dysmorphology syndrome, campomelic dysplasia. Sox9 controls the expression of chondrocyte-specific marker genes, such as aggrecan and members of the collagen family (Col2a1, Col9a2, Col11a2), as well as of two other Sox-gene family members, L-Sox5 and Sox6, which are also required for chondrocyte differentiation. Proliferation of non-hypertrophic chondrocytes within the cartilage elements is under the control of a number of signaling molecules, including members of the fibroblast growth factor family (FGFs), ►bone morphogenetic proteins (BMPs), Indian hedgehog (Ihh) and the Wntless/int-gene family (WNT). In addition, two other locally produced peptides have been shown to control chondrocyte proliferation, C-type natriuretic peptide (CNP) and parathyroid-hormone-related peptide (PTHrP). FGFs are negative regulators of chondrocyte proliferation, with probably Fgf18 as the preferred ligand of FGFR3 signaling through the ►transcription factor STAT1 leading to an up-regulation of the cell-cycle inhibitor p21. Ihh acts as a positive regulator of chondrocyte proliferation regulating CyclinD1 expression. BMPs have been shown to stimulate chondrocyte proliferation, while other members of the TGF- β superfamily such as TGF- β 1 can have a negative effect on proliferation; it is so far unclear how these distinct effects on proliferation are achieved. It is also unclear how CNP affects cell proliferation on the molecular level.

Regulation of chondrocyte proliferation is coupled to a control of differentiation rate. Experiments with ►transgenic animals have revealed that some WNTs (Wnt5a and Wnt5b) differentially control the expression of cell-cycle regulators such as the negative regulator p130/Rb2 and the positive regulator CyclinD1, thereby controlling the progression of cells from the proliferative to the non-proliferative state (4). Similarly PTHrP signaling also controls the progression of proliferating to non-proliferating chondrocytes. Signaling through its receptor PPR, a serpentine receptor interacting with heterotrimeric G-proteins, it positively affects proliferation of chondrocytes through adenylate cyclase-cAMP-mediated activation of protein ►kinase A (PKA), leading to a transcriptional up-regulation of positive regulators of the cell-cycle such

as CyclinA and CyclinD1, mediated through transcription factors of the CREB/ATF and AP-1 family. In contrast, activation of the phospholipase C (PLC) branch of PPR intracellular signaling is probably involved in dampening the cAMP/PKA effect on proliferation and promoting chondrocyte differentiation (5).

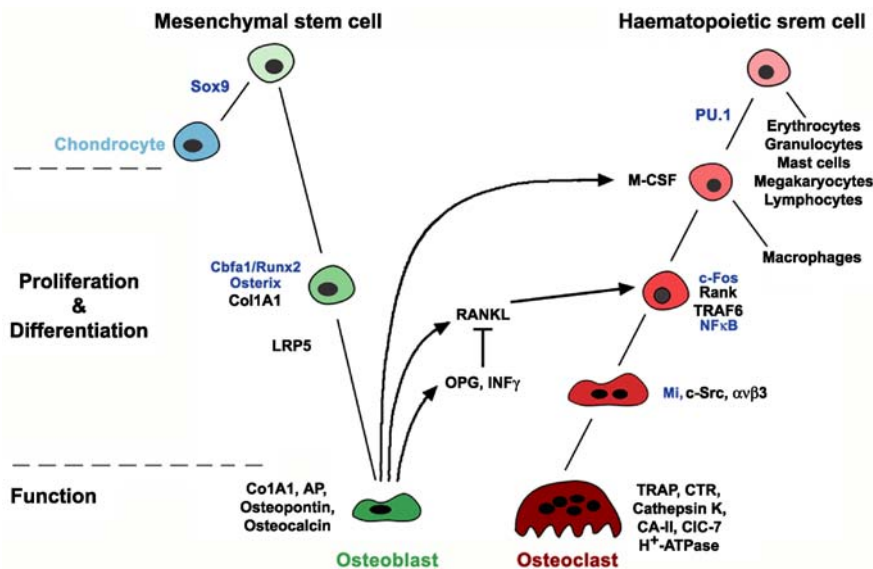
A number of transcription factors have been shown to be essential within the osteoblast lineage, such as Cbfa1/Runx2 which acts upstream of the ►zinc-finger transcription factor Osterix in osteoblast differentiation. Patients ►heterozygous for mutations in Cbfa1/Runx have cleidocranial dysplasia (CCD) syndrome, in which only bones formed through intramembranous ossification are affected. Cbfa1/Runx2 has also been shown to be required for osteoblast function. Loss-of-function studies in mice have further revealed that the homeobox-containing proteins Dlx5, Msx1, Msx2 and Bapx1/Nkx3.2 are involved in osteoblast differentiation. In addition, a signaling pathway has been identified which seems to only affect osteoblast proliferation, but not differentiation and might act independently of Cbfa1/Runx2. This signaling pathway requires the LDL-receptor-related-protein 5 (LRP5), which is ubiquitously expressed. Inactivating mutations in the human gene lead to the osteoporosis-pseudoglioma syndrome, while activating mutations result in a high-bone mass syndrome. Since LRP5 can also act as a co-receptor for Wnt proteins, these

findings suggest that Wnt-signaling is controlling peri- and post-natal osteoblast proliferation.

Osteoclasts differentiate from hematopoietic stem cells and share a common precursor with macrophages (monocyte-macrophage lineage). During development mononucleated precursors enter the mesenchyme surrounding the skeletal elements, divide and differentiate into tartrate-resistant acid phosphatase (TRAP-) positive cells that migrate together with endothelial cells through the nascent bone collar, invading the calcified cartilage where they fuse into multinucleated, fully functional osteoclasts. They require the transcription factor PU.1 for differentiation early within the macrophage lineage specification and c-Fos a member of the AP-1 transcription factor family for the specification of osteoclast precursors *versus* macrophages (Fig. 2).

Regulatory Mechanisms

In order to coordinate skeletal growth during development (embryonic and postnatal), a variety of factors are employed, which have been shown to regulate proliferation and differentiation. These include locally produced cytokines and growth factors during embryonic development and circulating hormones regulating post-natal growth. In addition, other factors such as nutritional intake and mechanical influences regulate growth and ►morphogenesis of the skeleton. Coordination between



Bone and Cartilage. Figure 2 Summary of the molecular mechanisms and regulation of osteoblastogenesis and osteoclastogenesis from stem cells to functional osteoblasts and osteoclasts. The particular molecules presumably acting at the different stages during precursor proliferation and differentiation are depicted: transcription factors involved in these steps are colored blue. Abbreviations (for factors not mentioned in the text): AP, alkaline phosphatase; CTR, calcitonin receptor; CA-II, carbonic anhydrase II; CIC-7, chloride channel-7; Mi, microphthalmia transcription factor; TRAF6, TNF receptor associated factor 6.

maturation of chondrocytes, ►osteoblastogenesis, and blood vessel invasion is particularly important during embryonic development for proper skeletogenesis. One of the central molecules coordinating these processes is *Ihh*. *Ihh* is expressed in a subpopulation of chondrocytes, the prehypertrophic chondrocytes. As mentioned above, it controls chondrocyte proliferation, but also their maturation, by regulating the progression from the proliferative to the non-proliferative, hypertrophic state. The latter is achieved *via* *Ihh*-dependent regulation of PTHrP expression levels in the articular region. *Ihh* signaling is further required for the induction of osteoblast progenitor differentiation from mesenchymal cells within the adjacent perichondrium/periosteum. The transcription factor *Cbfa1/Runx2*, which is absolutely required for osteoblastogenesis, is under the control of *Ihh* signaling in the periosteum. *Cbfa1/Runx2*, however, is also playing a role in chondrocyte maturation. In *Cbfa1/Runx2* mutant mice, hypertrophic chondrocytes are absent from several, but not all, skeletal elements and constitutive expression of *Cbfa1/Runx2* in non-hypertrophic chondrocytes induces their differentiation into hypertrophic chondrocytes; in addition, *Cbfa1* induces the expression of *Ihh*. Thus regulatory loops exist between *Cbfa1* and *Ihh* in chondrocytes and *Ihh*-signaling and *Cbfa1* in osteoblast precursors. *Cbfa1/Runx2* is further necessary for vascular invasion of the hypertrophic region, a function, which is independent of its role in regulating differentiation of hypertrophic chondrocytes. Attraction of blood vessels into the hypertrophic region is achieved *via* a *Cbfa1/Runx2* mediated up-regulation of the expression of vascular-endothelial growth factor (VEGF) within the hypertrophic chondrocytes. Thus, the *Ihh*/PTHrP regulatory loop controls the rate of chondrocyte progression at the level of proliferative to non-proliferative prehypertrophic chondrocytes, while *Cbfa1/Runx2* controls the level of *Ihh* expression and the progression from prehypertrophic to hypertrophic chondrocytes. In addition, *Cbfa1/Runx2* controls the process of blood vessel invasion *via* regulation of VEGF. *Ihh*-signaling, on the other hand, controls directly or indirectly the expression of *Cbfa1/Runx2* in osteoblast progenitors thereby coordinating the onset of bone formation with the status of chondrocyte maturation. However, additional modulators on post-transcriptional and post-translational levels are required, since *Ihh* as well as *Cbfa1/Runx2* expression precedes the appearance of functional osteoblasts. One of these modulators is probably Osterix, which is required downstream of *Cbfa1/Runx2* in osteoblastogenesis (1).

Regulatory mechanisms also exist between osteoblasts and osteoclasts ensuring bone homeostasis in healthy individuals. Osteoblasts secrete factors controlling the

differentiation of osteoclasts, such as macrophage colony-stimulating factor (M-CSF), required for the induction of osteoclast differentiation from the monocyte-macrophage lineage. Together with M-CSF, a second factor expressed by osteoblasts and stromal cells, the membrane-associated protein, receptor activator of nuclear factor κ B ligand (RANKL), a member of the tumor necrosis factor (TNF)-ligand family, is required for osteoclast differentiation. Osteoclasts express the counterpart RANK, a TNF receptor family member. Signaling through RANK leads to activation of nuclear factor κ B (NF- κ B) and c-Jun N-terminal protein kinase (JNK). In mature osteoclasts RANKL/RANK interaction is required for bone resorbing activity. The fine-tuning of the system is achieved by the fact that osteoblasts/stromal cells also express a soluble decoy receptor of RANKL, osteoprotegerin (OPG), which functions as an inhibitory factor of ►osteoclastogenesis. In summary osteoblasts/stromal cells produce stimulators and inhibitors of osteoclastogenesis (Fig. 2).

References

1. Karsenty G, Wagner EF (2002) Reaching a genetic and molecular understanding of skeletal development. *Dev Cell* 2:389–406
2. Zelzer E, Olsen BR (2003) The genetic basis for skeletal diseases. *Nature* 423:343–348
3. Aszodi A, Bateman JF, Gustafsson E et al (2000) Mammalian skeletogenesis and extracellular matrix: what can we learn from knockout mice? *Cell Struct Funct* 25:73–84
4. Yang Y, Topol L, Lee H et al (2003) Wnt5a and Wnt5b exhibit distinct activities in coordinating chondrocyte proliferation and differentiation. *Development* 13:1003–1015
5. Guo J, Chung UI, Kondo H et al (2002). The PTH/PTHrP receptor can delay chondrocyte hypertrophy *in vivo* without activating phospholipase C. *Dev Cell* 3:183–194

Bone Disease and Skeletal Disorders, Genetics

STEFAN MUNDLOS

Institute for Medical Genetics, Charité-University Medicine Berlin, Berlin, Germany
stefan.mundlos@charite.de

Definition

The skeleton is an extremely complex organ that fulfills multiple functions in the vertebrate body. It consists of

more than 230 individual elements, many of them connected by joints allowing movement through muscle contraction. Furthermore, the skeleton harbors the bone marrow, the primary source for hematopoietic progenitor cells and is the body's largest resource for calcium. Cells within the skeleton create the entire longitudinal growth generated during an individual's life. Even in adult life, the skeleton is an active organ that is constantly being remodeled. Giving the complexity of the system it is not surprising that a large number of genes are involved in the molecular control of growth and maintenance of the skeleton. Accordingly, genetic disorders of the skeleton are frequent and comprise a large group of genetically heterogeneous and clinically distinct conditions. The skeleton can be affected as part of a complex malformation syndrome or the genetic defect may affect solely the skeleton. In the latter case, the condition will be called a primary skeletal disease. Traditionally, these conditions have been subdivided into ►**dysostoses**, defined as malformations of individual bones or groups of bones or ►**osteochondrodysplasias** defined as disorders of chondro-osseous tissue. More recently, disorders of bone maintenance have been added that affect bone turn over and consequently result in too much or too little bone.

Characteristics

The clinical manifestations of skeletal disorders range from neonatal lethality to only mild growth retardation or even normal growth in late onset conditions. Their clinical diversity makes them often difficult to diagnose. Many attempts have thus been made to facilitate diagnosis by delineating single entities or groups of diseases. On the basis of clinical, radiological and molecular features, a group of experts created a taxonomy that has been called "Nosology" (1). Recent advances that were made in understanding the molecular pathology of these disorders were incorporated into this classification. However, the merging of clinical and molecular data to create a comprehensive picture of this disease group proved to be difficult. While the molecular data help to understand pathogenetic mechanisms and, in addition, have been extremely instrumental in grouping diseases into larger "families", they have been less helpful in the respective clinical aspects. Some conditions that have a common molecular origin have little in common clinically. Mutations in *FGFR3*, for example, can result in such diverse conditions as lethal ►**thanatophoric dysplasia**, ►**achondroplasia**, a condition characterized by moderate to severe growth retardation or ►**hypochondroplasia**, which is frequently only associated with borderline small stature. Thus, clinical and genetic data are often difficult to merge.

Based on clinical, embryological and molecular findings we have proposed classifying skeletal disorders into a) conditions that have defects in bone patterning, b) conditions with defects in early formation of cartilage/bone, c) conditions with abnormal growth and d) conditions with abnormal bone homeostasis (2).

Abnormal Patterning

Patterning genes control the overall body plan (ground plan) of the skeleton, i.e. the number, size and gestalt of individual skeletal elements. In spite of our increasing knowledge about developmental mechanisms *per se*, how skeletal pattern is established is still largely a mystery. Much of what we know comes from gene inactivation studies in the mouse or the study of rare human genetic conditions. Human skeletal malformations can generally be considered to be deficiencies of embryonic patterning that can be subdivided into defects involving the craniofacial, the axial or the limb skeletons.

Most of the craniofacial bones are of neural crest origin. Accordingly, many of the craniofacial dysostoses include abnormalities in migration and/or differentiation of cranial neural crest cells. Waardenburg syndrome, for example, is characterized by deafness and pigmentation anomalies in variable combination with skeletal anomalies and is caused by mutations in the *PAX3* gene coding for a transcription factor involved in cell migration of muscle and neural crest progenitors. A multitude of other gene defects are associated with craniofacial malformation. Principles and details of craniofacial development have been discussed in detail elsewhere (for review, see (3)).

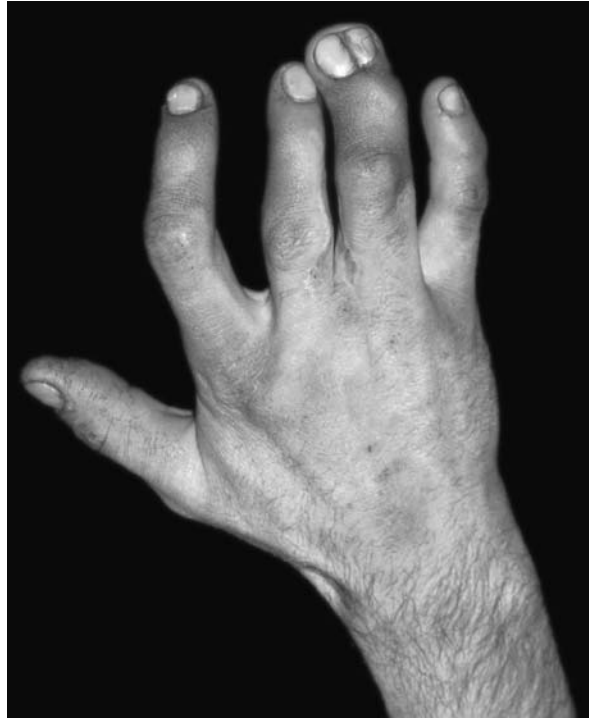
The axial skeleton, consisting of vertebrae and ribs, originates from the ►**somites**, transient organizational structures located at both sides of the embryonic body (4). Upon patterning signals from the ►**notochord**, cells from the somite differentiate and migrate to form the ►**sclerotome**. Four adjacent sclerotome compartments eventually fuse to form a single vertebra and the corresponding ribs. Defects of the axial skeleton occur frequently in association with other malformations, but those that primarily affect the axial skeleton are rare and are summarized under the term spondylocostal dysostoses. Such conditions can be caused by mutations in *DLL3*, which is part of the notch-delta pathway, involved in somite boundary formation.

The appendicular skeleton has its origin in mesenchymal cells located in the lateral plate mesoderm. Localization of limb bud outgrowth is determined by the axial Hox gene code. Outgrowth of the limb bud depends on expression of members of the fibroblast growth factor (FGF) family in the apical ectodermal ridge (►**AER**), an anatomical structure located at the

very tip of the bud. Once the bud is formed, the differentiation of each cell is controlled by a three-dimensional coordinate system consisting of signals that define the dorsoventral, proximodistal and anteroposterior axes. Subsequently, the various tissues of the limb are laid down in a proximodistal sequence, i.e. first the humerus, then the radius and ulna and finally the digits (5). Defects in the AER result in truncations or the ►split-hand-foot phenotype, as caused by mutations in the transcription factor p63 or by deletion on chromosome 7q encompassing the DLX5/DLX6 genes. Defects of the anteroposterior axis frequently involve the ►hedgehog pathway and result in polydactyly, as seen in Greig syndrome (mutations in GLI3), Townes Brooks syndrome (SALL1) or triphalangeal thumb-polysyndactyly syndrome (regulatory mutations of SHH). Defects in ►HOX genes result in digit hypoplasia (hand-foot-genital syndrome, mutations in HOXA13) or in central synpolydactyly (Fig. 1) (mutations in HOXD13). Other phenotypes involve deficiencies of the radial or the ulnar ray. Hypoplasia of the thumb and/or the radius can be observed in Holt-Oram syndrome, which is caused by mutations in TBX5. Ulnar ray deficiencies, on the other hand, are part of the ulna-mammary syndrome phenotype caused by mutations in TBX3. Tbx genes play important roles in the specification of the upper vs. the lower limb and in the initial limb outgrowth.

Abnormal Formation of Bone/Cartilage

After cells have received their positional information through patterning genes, they migrate and condense to form the skeletal ►anlage. Within the condensation, they differentiate into either ►chondrocytes that form cartilage or ►osteoblasts that form bone. Cartilage formation is controlled by the transcription factor Sox9 and two other members of the family, Sox5 and Sox6. Without Sox9 no cartilage forms. Mutations in SOX9, on the other hand, cause campomelic dysplasia, a commonly lethal condition characterized by severe osteochondrodysplasia, bowed femura and hypoplastic scapulae. The ►transforming growth factor β family (TGF β) consisting of bone morphogenetic proteins (►BMPs) and their receptors plays essential roles in the formation of the early condensations. Mutations in GDF5, a member of the BMP family and in its receptor, BMPRII, cause ►brachydactyly (shortening of the digits, absence of individual phalanges) type C and type A2, respectively. Although bone and cartilage may originate from a common progenitor cell, their further differentiation is regulated separately. The protein Runx2 (Runt-related transcription factor 2; also called Cbfa1) has been shown to be essential for osteoblast differentiation. Without Runx2, no bone is formed. Mutations in RUNX2 result in ►cleidocranial dysplasia, a condition characterized by absent/hypoplastic



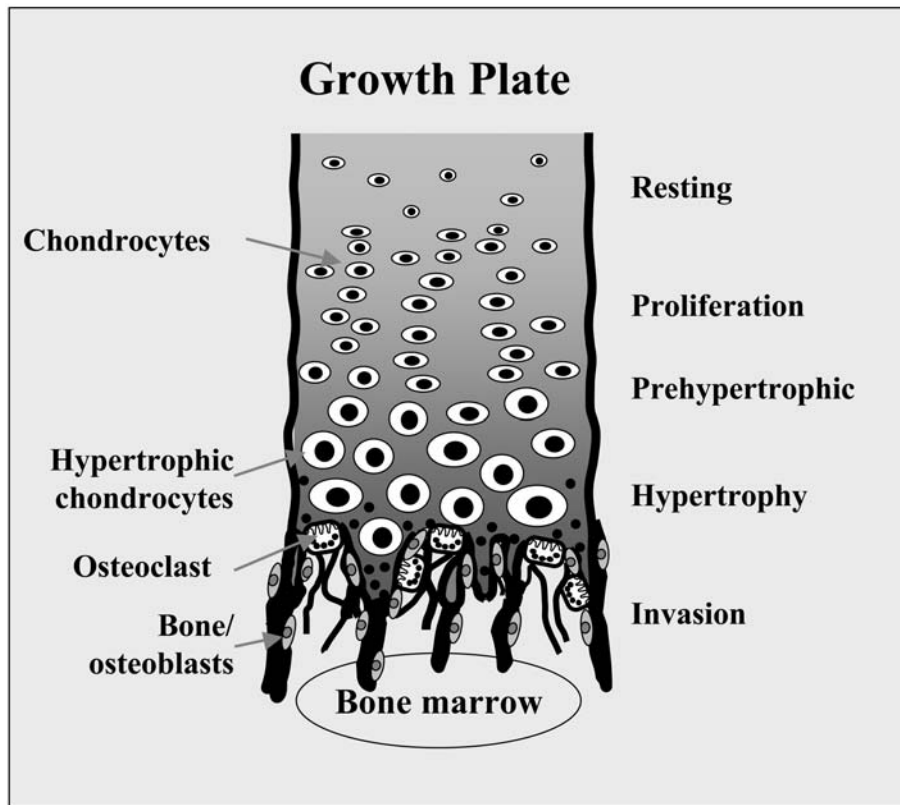
Bone Disease and Skeletal Disorders, Genetics.
Figure 1 Hand of individual with synpolydactyly caused by a mutation in HOXD13.

clavicles, ossification defects in the skull and abnormalities of tooth formation.

Abnormal Growth

Growth of the long bones is generated *via* a mechanism called ►endochondral bone formation. In contrast to the growth of the flat bones of the skull that grow by appositional growth (►desmal ossification), endochondral bone formation works *via* the transformation of a transient cartilaginous template into bone. Central to this process is the formation of a ►growth plate, a highly organized structure within the cartilaginous part of the growing bone that constantly produces new cartilage through cell proliferation, cell differentiation and matrix production. The newly produced cartilage calcifies and is then removed by ►osteoclasts and replaced by bone. A schematic of the growth plate architecture is given in Fig. 2.

The integrity of the extracellular matrix is essential to growth plate function. Accordingly, many defects in cartilaginous matrix constituents result in abnormal growth plate function and, consequently, in short stature. Collagen type II, probably the most important component of cartilaginous matrix, is mutated in spondyloepiphyseal dysplasia (SED) congenita, a condition characterized by severe dwarfism, flat

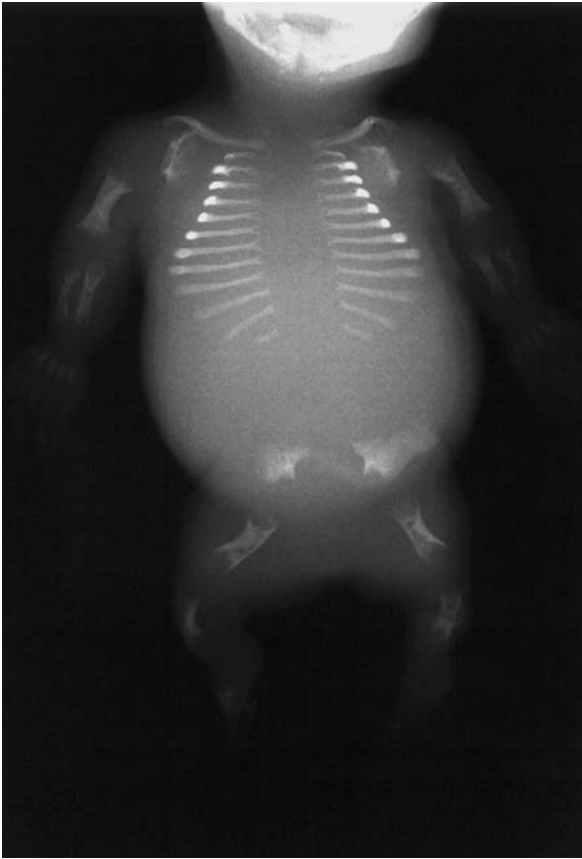


Bone Disease and Skeletal Disorders, Genetics. Figure 2 Schematic of growth plate architecture. The growth plate is a highly organized structure that continuously produces new cartilage through the proliferation, matrix production and differentiation of chondrocytes. During this process, chondrocytes differentiate from small resting cells with little proliferation to proliferating cells and finally become hypertrophic. Hypertrophic chondrocytes produce a specific extracellular matrix that subsequently is removed by osteoclasts and replaced by bone. The different zones of chondrocyte differentiation are shown on the right, the cell types on the left hand side.

vertebrae, precocious arthropathy and eye involvement (myopia). A more severe form of mutation in COL2A1 results in ▶**achondrogenesis** type II, a neonatally lethal chondrodysplasia with highly abnormal cartilage, severely delayed ossification and very short limbs (Fig. 3). Mutations in other collagens that connect type II collagen with the surrounding matrix, such as collagen types IX and XI give rise to milder phenotypes such as multiple epiphyseal dysplasia (MED) and ▶**Stickler syndrome**, respectively. Another form of MED and pseudoachondroplasia, a condition characterized by a short limb dwarfism with body proportions resembling those of achondroplasia but with a normal facies, are caused by mutations in cartilage oligomeric matrix protein (COMP). COMP interacts with collagens type I, II, III, and IX and connects them with the cartilaginous proteoglycans. Proteoglycans are highly sulfated, a process that involves several enzymatic steps as well as active transport of sulfur into the cell through the sulfate transporter DTDST. Mutations in

DTDST result in a spectrum of diseases including lethal achondrogenesis type 2, ▶**atelosteogenesis** type 2, ▶**diastrophic dysplasia** and a milder, recessive form of MED. Diastrophic dysplasia, the condition in which the transporter was first identified, is a severe dwarfing condition characterized by multiple joint contractures (club feet frequently), proximally set, hypermobile thumbs, cystic masses at the external ears, cleft palate and scoliosis.

The major matrix component of bone is collagen type I, encoded by two genes, COL1A1 and COL1A2. Mutations in this collagen result in a spectrum of conditions collectively termed ▶**osteogenesis imperfecta** (OI) or brittle bones disease. The hallmark of these conditions is an extreme bone fragility and ▶**osteoporosis**. Based on clinical and radiological findings, four major subtypes have been delineated. Type 2 OI is the most severe and lethal form. Multiple fractures already occur *in utero* and the bones are very thin (ribs), or extremely short and crushed (Fig. 4). OI



Bone Disease and Skeletal Disorders, Genetics.

Figure 3 X-ray of a fetus with achondrogenesis type II and a mutation in COL2A1. Note complete lack of ossification of vertebrae and very small long bones with cup shaped ends.



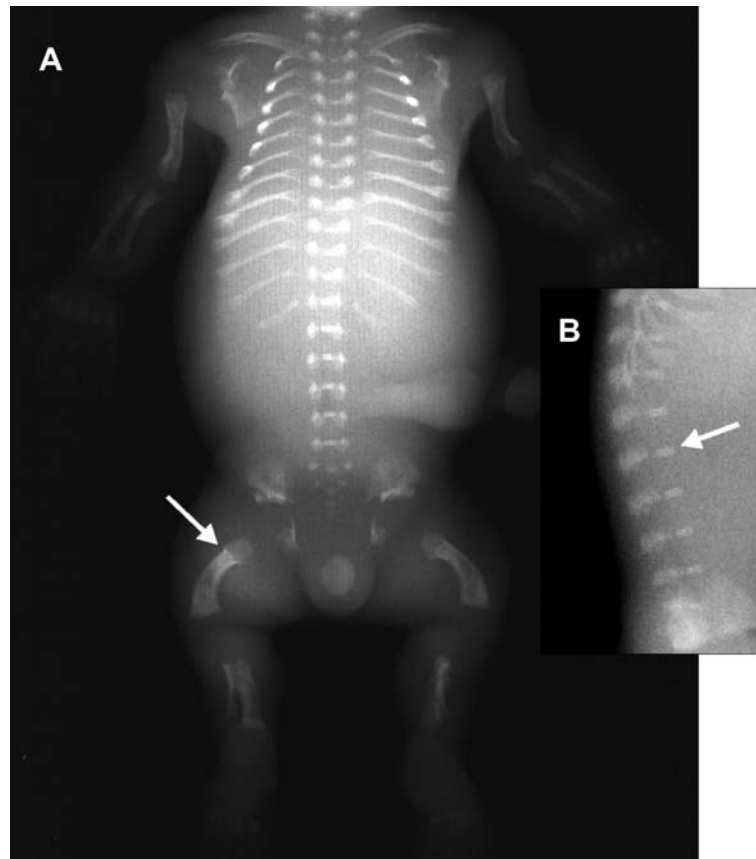
Bone Disease and Skeletal Disorders, Genetics.

Figure 4 X-ray of a fetus with osteogenesis imperfecta type II and a mutation in COL1A1. Note very thin ribs with multiple fractures, very short limbs with bowing and signs of multiple fractures and inefficient healing.

type 3 patients survive but frequently become wheelchair bound due to recurrent fractures and bone deformities. Type 4 OI shows moderate deformity, short stature and dentinogenesis imperfecta. OI type 1 is the mildest form characterized by blue sclerae, normal teeth, near normal stature and recurrent fractures that heal without deformities. OI types 2–4 are caused by point mutations in either of the collagen type I genes, usually affecting one of the repetitive glycines within the triple helical domain of the molecule. This disrupts the correct folding of the helix resulting in the degradation of normal chains as well, hence the severe phenotype. There is no clear phenotype/genotype correlation. OI type 1 is caused by nonsense mutations resulting in loss of one COL1 allele.

Growth of the cartilaginous template depends on proliferation and differentiation of chondrocytes. During their differentiation fate chondrocytes change from small round cells in the resting zone to flat stacked cells

in the columnar zone, and finally become hypertrophic before they are removed by osteoclasts. This differentiation process is highly regulated and depends on the interaction of multiple signaling pathways. One major pathway involves the hedgehog/parathyroid hormone (PTH)-related peptide (PTHrP) pathway. Indian hedgehog (Ihh), which is expressed by pre-[hypertrophic chondrocytes](#), controls cell proliferation and the differentiation into hypertrophic cells. Ihh regulates the expression of PTHrP *via* its receptor patched and the transcription factor Gli. Mutations in IHH result in a generalized skeletal dysplasia, acrocapitofemoral dysplasia and brachydactyly type A1, which can also be associated with short stature. Homozygous loss of function mutations of PPR, the receptor for PTHrP/PTH, result in Blomstrand chondrodysplasia, a severe, often lethal dysplasia with severely shortened tubular bones and an accelerated bone maturation. In contrast, dominant activating mutations of this receptor result in Jansen type



Bone Disease and Skeletal Disorders, Genetics. Figure 5 X-ray of a fetus with thanatophoric dysplasia type I and a mutation in the FGFR3 gene. Note narrow thorax, flattening of the vertebral bodies, also shown in the lateral view in (B), abnormal pelvic bones, short and bowed long bones (arrow shows bowing of femur).

metaphyseal dysplasia. Patients suffer from severe growth retardation, generalized demineralization, rachitiform changes and enlargement of the metaphyses. Another important pathway in the growth plate involves FGF signaling molecules and their receptors, the FGFRs. FGFs control multiple functions but most importantly appear to influence proliferation and differentiation. Mutations in the FGFR3 gene cause achondroplasia, a common form of dwarfism with disproportionately short limbs, a large head and a characteristic facies. The mutations lead to an activation of the receptor resulting in ligand independent signaling. Other mutations in FGFR3 that result in more or less activation result in more severe, lethal phenotypes (thanatophoric dysplasia) (Fig. 5), or milder phenotypes (hypochondrogenesis).

Defects in Homeostasis

Bone, although it appears rather static, is in fact a very active tissue. Our bones are constantly remodeled to adjust to the continually changing mechanical and

metabolic needs. It is important that the overall bone mass and bone structure is preserved in this process in order not to lose bone or to gain too much bone. This steady state is achieved through a concerted action of bone producing osteoblasts and bone removing osteoclasts. Any imbalance of this intricate system will result in bone loss and thus osteoporosis or too much bone as in ▶osteopetrosis. The RANKL/OPG system has recently been shown to be a major regulator of the osteoblast-osteoclast crosstalk. Recessive juvenile Paget disease is caused by mutations in the OPG encoding gene TNFSRF11B. This hereditary form resembles adult Paget disease showing hyperactive osteoclasts and excessive bone formation. Mutations in the RANK signaling peptide, which lead to intracellular accumulation and self-stimulation of the receptor, result in familial expansile osteolysis.

Inactivation of osteoclast function, on the other hand, leads to excessive bone formation and, in severe cases, to a solid bone without bone marrow. Mutations in the gene for cathepsin k, an acidic protease secreted by the

osteoclast, result in the sclerosing bone disorder pycnodysostosis. Strong acidification of the osteoblast absorption lacunae is needed to degrade the matrix components of bone with proteases such as cathepsin and to dissolve the hydroxyapatite of bone. Acidification is accomplished *via* proton pumps. Mutations in *TCIRG1*, which encodes a subunit of the osteoclast's proton pump, result in recessive, malignant osteopetrosis. A very similar phenotype is caused by mutations in a chloride channel, *CLCN7*, which transports chloride ions in parallel to the protons pumped into the resorption lacunae. Dominant missense mutation in the same gene may cause dominant osteopetrosis type 2, a late onset high bone density phenotype.

Cellular and Molecular Regulation

As outlined above, skeletal patterning and growth is determined by an extremely complex set of signaling pathways involving a multitude of genes. One emerging picture is that mutations within the same gene frequently result in clinically distinct conditions. However, a close look at the phenotype reveals overlapping, common features that are unique to these mutations. Such common features have previously been identified on purely clinical grounds and have been called “disease families”. The more recent molecular data support this concept and link the common phenotype to a specific molecular pathway. Such “disease families” frequently cover a whole spectrum of conditions. For example, thanatophoric dysplasia, achondroplasia and hypochondroplasia have, in spite of their completely different clinical outcome (see above), a common radiological pattern. This common feature results from activating mutations in *FGFR3*, which lead to an arrest of chondrocyte proliferation and differentiation. The difference in severity correlates with the degree of *FGFR3* activation, i.e. thanatophoric dysplasia has the strongest activation and hypochondroplasia the weakest. A similar spectrum of conditions results from mutations in the sulfate transporter *DTDST*. Complete inactivation results in the most severe, lethal phenotype (achondrogenesis type 2), whereas diastrophic dysplasia (severe dwarfism) and recessive MED correspond to partial function. Similar disease spectrums can be observed in the collagen type 1 (OI) and the collagen type 2 (SED congenita) related conditions.

In certain cases overlapping or even identical phenotypes can be observed even if the mutation occurs in a different gene. Such a phenotypic copy may be produced by pure chance, for example because the number of phenotypes that can be produced is limited or because the mutations affect genes that are closely linked in a common molecular pathway. Examples are the polydactyly phenotypes caused by perturbations of the hedgehog pathway.

Clinical Relevance

Although individually rare, disorders of the skeleton are of high clinical relevance because of their overall frequency. The clinical severity differs from minor handicaps to death in the neonatal period. In surviving patients, secondary complications of skeletal deformity and manifestations in extraskeletal organs add to the burden of disease. Genetic factors are important determinators of more common diseases such as osteoporosis and osteoarthritis. Osteoporotic fractures, for example, are a major healthcare problem in Europe. Understanding the molecular and genetic regulators of bone development and maintenance will help to understand these conditions better and to finally establish effective therapies.

References

1. Hall CM (2002) International nosology and classification of constitutional disorders of bone (2001) *Am J Med Genet* 113:65–77
2. Kornak U, Mundlos S (2003) Genetic Disorders of the Skeleton: A Developmental Approach. *Am J Hum Genet.* 73: 447–74
3. Wilkie AO, Morriss-Kay GM (2001) Genetics of craniofacial development and malformation. *Nat Rev Genet* 6:458–68
4. Saga Y, Takeda H (2001) The making of the somite: molecular events in vertebrate segmentation. *Nat Rev Genet* 11:835–45
5. Tickle C (2003) Patterning systems – from one end of the limb to the other. *Dev Cell* 4:449–58

Bone Lining Cell

Definition

Bone lining cells are flat cells containing very few organelles. They line the bone's surface and might be an inactive osteoprogenitor cell.

► Bone and Cartilage

Bone Marrow

Definition

Bone marrow is a cell and blood-rich material in the center of bone, which contains many stem cells and is involved in haematopoiesis.

► Bone and Cartilage

Bone Morphogenetic Proteins

► [BMP](#)

Bootstrapping

Definition

Bootstrapping refers to a simple test for the validity of an evolutionary tree to perform the tree building on input data with slight, randomly created changes. This step is repeated 500–1000 times, and the tree with the highest probability is selected.

► [Sequence Annotation in Evolution](#)

Boundary Conditions

Definition

Boundary conditions refer to algorithms that minimize edge effects in simulations of liquids, solutions, or solids.

► [Molecular Dynamics Simulation in Drug Design](#)

Bovine Spongiform Encephalopathy

► [BSE](#)

Brachydactyly

Definition

Brachydactyly is characterized by excessively shortened tubular bones in the hands and feet due to hypoplasia/aplasia of phalanges, and/or metacarpals of different genetic cause.

► [Bone Disease and Skeletal Disorders, Genetics](#)

► [Mendelian Forms of Human Hypertension and Mechanisms of Disease](#)

Bragg's Law

Definition

Bragg's law depicts the mathematical relationship ($n\lambda = 2d \sin \theta$) that relates the spacing d of a diffraction grating, the wavelength λ of diffracted light, the order n of the reflected wave, and the diffraction angle θ . In a three dimensional crystal, the grating spacing for first order ($n = 1$) reflections corresponds to the spacing between Miller planes.

► [X-ray Crystallography, Basic Principles](#)

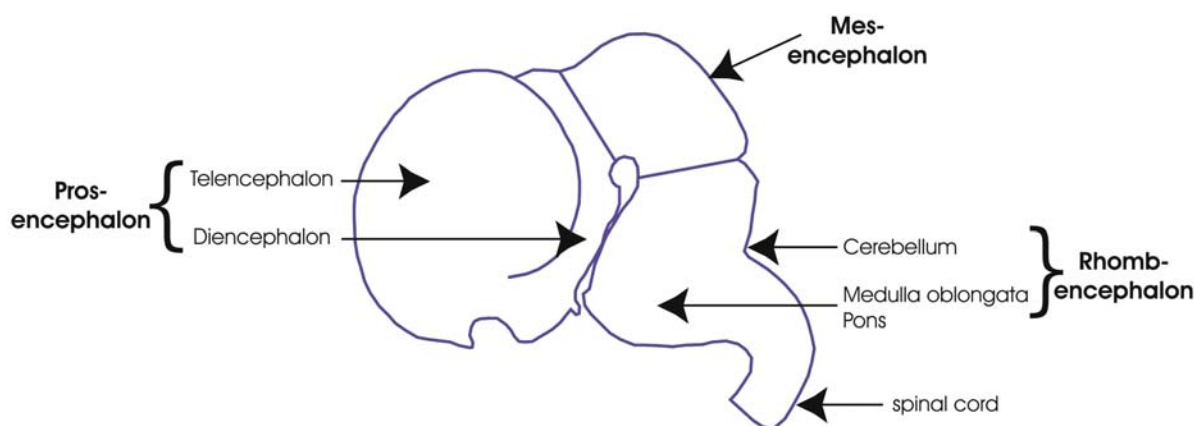
Brain

INGO BECHMANN

Experimental Neuroimmunology, Institute of Anatomy, Cell and Neurobiology, University Hospital Charité, Berlin, Germany
ingo.bechmann@charite.de

Definition

The brain develops from the rostral end of the ► [neural tube](#) and is thus continuous with the spinal cord, which evolves from its caudal portion. Although no sharp border exists between these two parts of the central nervous system (CNS), in vertebrates the brain may be defined as the part of the CNS enclosed in the cartilaginous or bony cranium. Here, the general organization of the spinal cord (gray matter ► [neurons](#) located in the center surrounded by ascending and descending fibers of white matter) is no longer maintained. Gray matter develops at the surface as well as in the deep regions of the brain and fiber tracts develop to interconnect different brain areas and transfer information to and from the periphery. The morphogenesis of the brain is initiated by the development of embryonic vesicles (Fig. 1, Table 1), which further differentiate into the rhombencephalon (hindbrain), mesencephalon (midbrain) and prosencephalon (forebrain). The last consists of the central diencephalon containing the hypothalamus, thalamus and pineal gland (epithalamus) and the telencephalon, i.e. the cerebral hemispheres covered by the cerebral cortex. As evolution continues, the cerebral cortex massively expands through the multidirectional growth of the telencephalic lobes (frontal, parietal, occipital and temporal, Fig. 2a, b, c) and gyrencephalisation, i.e. the development of sulci and gyri in monkeys and man.



Brain. Figure 1 Schematic drawing of the embryonic vesicles.

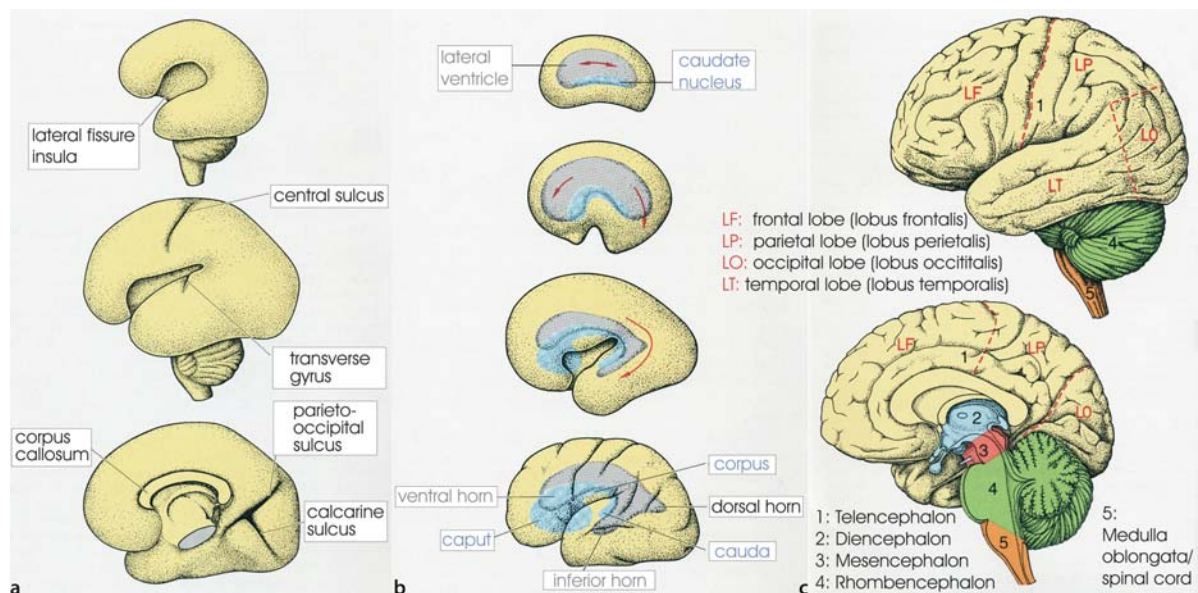
Brain. Table 1 Overview of the derivatives of the embryonic vesicles (compare Fig.1/2)

Embryonic vesicle	Derivative	Prominent structures	Derivatives of the neural tube's central canal
Paired prosencephalic vesicles (forebrain vesicles)	Telencephalon	Hemispheres/Cortex: Frontal, parietal, occipital, temporal lobe, hippocampus, amigdala subcortical nuclei (basal ganglia): Nucl. caudatus, Putamen, Pallidum	I.+II. Ventricle
	Diencephalon	Hypothalamus, Thalamus, Epiphysis (pineal gland)	III. Ventricle
Unpaired mesencephalic vesicle (midbrain vesicle)	Mesencephalon	Colliculi, substantia nigra, Central gray	Aqueduct
Unpaired rhombencephalic vesicle (hindbrain vesicle)	Metencephalon	Pons, Cerebellum	IV. Ventricle
	Myelencephalon	Medulla oblongata	Central canal of the cord

The direction in which the lobes grow is reflected in the form of the ventricles, which are the continuous derivatives of the neural tube's central canal (Fig. 2b). As a result of the massive expansion of the telencephalic lobes in man, the midbrain and upper hindbrain (i.e. the cerebellum) are completely covered by the hemispheres (Fig. 2c).

Functionally, the brain integrates sensory input from the spinal cord and afferent portions of the cranial nerves, controls motor activity and analyses visual, acoustic, gustatory and olfactory information. The integration of all the afferent information provides a picture of the world outside and the position of the body therein. Changes in the environment or sudden contact with

objects may be responded to by reflexive loops which do not require – or which occur in the absence of – consciousness. In addition, regulation of homeostasis (e.g. breathing, food intake and digestion, blood pressure), circadian rhythms and the ovarian cycle occur subconsciously *via* the autonomic nervous system and hormonal pathways controlled within the CNS. Upon conscious reflection on the situation in the world, purposeful acts can be planned and performed in the absence of actual environmental stimuli. Consciousness appears to require the thalamus of the diencephalon and the cortex of the telencephalon, yet such issues are difficult to address experimentally. The question as to whether an individual's responses to cognitive and



Brain. Figure 2 (a) Development of the telencephalon. The frontal lobe is distinguished from the temporal lobe by the lateral fissure, the parietal from the occipital lobe by the parieto-occipital sulcus and the occipital from the temporal lobe by the calcarine sulcus. The insula, a region involved in the processing of acoustic information, is completely covered by the temporal lobe due to its rostral expansion. The central sulcus distinguishes motor areas (precentral region) from sensory areas (postcentral region). (b) The growth of the lobes during the brain's morphogenesis is reflected in the form of the lateral ventricle, which follows their expansion. Its ventral horn is located within the frontal lobe, the corpus in the parietal lobe, the dorsal horn in the occipital lobe and the inferior horn in the temporal lobe. A subcortical nucleus, the caudate nucleus, is similarly formed by the U-turn of the hemisphere's expansion. (c) The final form of the brain; the lobes (upper panel) and parts of the brain are labeled with regard to their origin from the vesicles (lower panel). (From: Zilles&Rekämper, Funktionelle Neuroanatomie, Springer 1998)

social tasks occur autonomously or are controlled by a responsible “self” remains open.

Characteristics

Topography

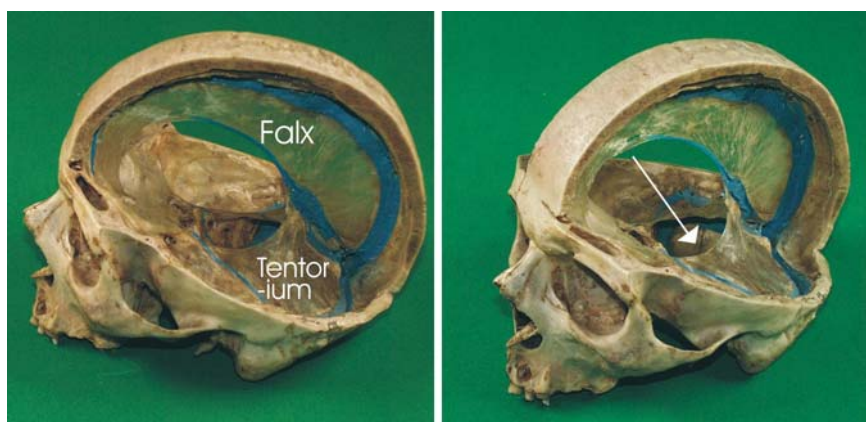
Meninges

The brain is enclosed by the cranium and enveloped by the meninges. Their outer layer, the dura mater (pachymeninx), consists of compact collagenous tissue covering the inner skull. From there, massive membranes, the falx cerebri and tentorium cerebri, separate the inner space of the skull into three compartments, one for each hemisphere and one for the cerebellum (Fig. 3). The leptomeninx consists of two layers of soft connective tissue, the arachnoid, which is attached to the dura and follows the bony surface of the inner skull and the pia mater, which is attached to the surface of the brain. The subarachnoid space between the two is filled with cerebrospinal fluid (CSF). This fluid is produced by the choroid plexus of the ventricles, which are connected to the subarachnoid space *via* three holes, the unpaired apertura mediana (Magendi, Fig. 4) and the paired aperturæ laterales (Luschke), all located at the caudal end of the cerebellum. One can thus picture the brain filled

with and “swimming” in the CSF, an important feature protecting this vital organ from damage induced by head trauma.

Vessels

The subarachnoid space also harbors the larger veins and arteries. Disruption of the latter thus leads to subarachnoid bleeding. Blood reaches the brain *via* four arteries, two internal carotid arteries and two vertebral arteries forming the basilar artery. These two systems are interconnected *via* small vessels building the arterial circle of Willis. Because of this interconnection, slow occlusion of one vessel, e.g. due to arteriosclerosis, can be compensated by the others. Blood drainage occurs *via* cervical and spinal veins and the venous sinus located within the dura mater (marked blue in Fig. 3). Disruption of the veins connecting the brain and these sinuses leads to subdural bleeding. A common reason for this kind of lesion is the shaking of infants (by their parents). Arterial pulsation leads to changes in the pressure within the CSF, which drives blood flow within the veins. Intravenous pressure depends mainly on hydrostatic pressure and, therefore, on the position of the head. In a reclining position this pressure is about 0 mmHg; in a standing



Brain. Figure 3 The dura separates three compartments, two for the hemispheres separated by the falx and one for the cerebellum separated by the tentorium. This keeps the brain in position, thereby avoiding damage during rapid movements. On the other hand, swelling of the brain can lead to injuries due to squeezing under the falx or tentorium. The brainstem passes through the small hole indicated by the arrow, where it is often damaged during edema or bleeding. The sinuses of the dura are labeled in blue.

position it is negative. Since cervical, facial and intracranial veins have no valves, blood can flow in both inside-out and outside-in directions. Thus, infectious agents can be transmitted from the face to the inside of the skull, where they may produce severe neurological problems. Increased intracranial pressure as occurs, for example, as a result of tumors, hinders venous drainage, thereby causing severe secondary damage. Moreover, parts of the brain may be squeezed under the falx and tentorium (Fig. 3) under these conditions. This points to a general problem in neurology/neuropathology; due to the limited space within the skull, even minor expansion of the tissue and thus swellings or tumors that are benign *per se*, easily produce life-threatening conditions.

Blood-Brain Barrier

The blood compartment and central nervous parenchyma are separated by a tissue barrier provided by the endothelial cells of CNS vessels. These cells are interconnected by ►tight junctions and – in contrast to those in most other regions of the body – are not fenestrated. Thus, diffusion of hydrophilic substances from blood to brain is inhibited in order to maintain homeostasis within the brain, a crucial prerequisite for proper neural transmission. Therefore, peripherally administered drugs (e.g. antibiotics) often do not reach the central nervous system, an important problem in neuropharmacology. In addition to this mechanical barrier, there is also a functional barrier for leukocytes, limiting their access to the CNS and their operations once inside the tissue. The highly differentiated networks apparently require special protection to keep inflammation-induced cell loss to a minimum, since post-mitotic neurons can hardly be replaced. As a result, many antigens, e.g. deriving from viruses, which are

readily eliminated elsewhere in the body, are tolerated within the CNS. From an evolutionary point of view, the persistence, for example, of the varicella-zoster virus illustrates that such tolerance is less detrimental for the individual than the elimination of all infected neurons.

Embryology/Morphogenesis

The brain differentiates from four embryonic vesicles. An overview of the general morphogenesis is provided in Figs. 1, 2. Table 1 gives an overview of the respective derived structures of the adult brain.

The Adult Brain of Man

An overview of the most important structures is provided in Figs. 2c, 4, 5.

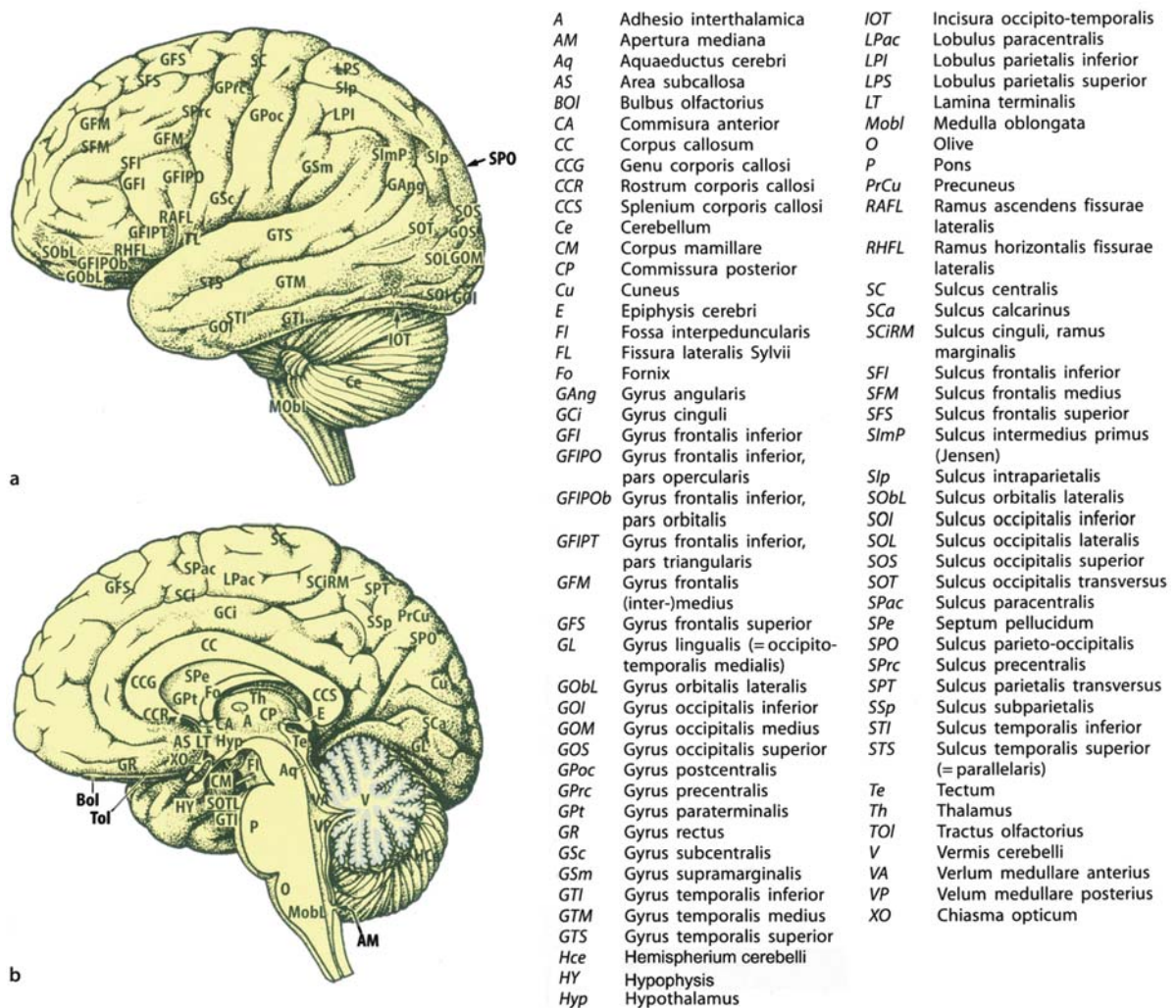
A simplified overview of structures and their function is provided in Table 2.

Functional Histology of the Brain

Central nervous tissue consists of neurons, ►glia, ependyma and a complex intercellular matrix.

Neurons

The morphology (shape and size) of neurons is extremely heterogeneous. Generally, they consist of a soma containing the ►nucleus, ►dendrites and an ►axon. Axons and dendrites are also referred to as neurites. Neurons are connected to each other at the ►synapse. The terminal of the axon forms the presynaptic part, while the postsynaptic part may be located at dendrites, the soma or the axon of the target neuron. Some neurons exhibit a pyramidal shape about 70 µm in size, others are round and no larger than 10 µm. In some regions, most



Brain. Figure 4 Overview of the adult brain from the lateral (upper panel) and medial (lower panel) perspectives. Some important areas, sulci and gyri are labeled. (From: Zilles&Rekämper, Funktionelle Neuroanatomie, Springer 1998)

prominently within the cortex of the telencephalon and cerebellum, neurons are located in separate layers of cells that are organized parallel to the cortical surface. In other regions, e.g. the brainstem of the mesencephalon and metencephalon, neurons are located in groups or columns forming specialized nuclei rather than layers. Lesion experiments and circumscribed pathologies have helped clarify the connectivity and function of many such nuclei and cortical areas.

The axons (synonym: fibers) of neurons may terminate at other neurons (their axon, dendrite or soma) or leave the CNS to innervate (smooth) muscle cells outside the skull. Axon endings often divide into collaterals and may therefore build synapses with many other structures. Target neurons may be located within the same region from which the axon derives or be located far away. Intrinsic fibers remain within one region (e.g. one particular gyrus), associational fibers connect areas

of the same hemisphere and commissural fibers terminate in the contralateral hemisphere. A prominent structure of commissural fibers is the corpus callosum (Fig. 2 a, c; Fig. 4). Particularly long fibers reach from the cortex down to the spinal cord, where they control motor functions. Bundles of many axons projecting from one region to another form circumscribed fiber tracts. In cross-sections of the brain, accumulations of neurons are the gray matter, while areas containing fiber tracts are visible as white matter to the naked eye. The white color of fiber tracts is due to the insulation of axons with myelin-containing fatty acids. Functionally, connections between certain neuronal populations form circuits related to particular operations. Knowledge of the step-by-step integration of information/response along such neuronal circuits allows localization of a lesion on the basis of given symptoms (topology).

Brain. Table 2 Overview of brain structures and their function

Structure	Function
Prosencephalon	
-Telencephalon	
Cortex	Sensory perception Motor control Consciousness
Basal ganglia: (Ncl. caudatus, Pallidum, Putamen, Claustrum)	Motor control Programmed movements
Hippocampus	Short-term memory Spatial orientation
Amygdala	Control of emotions Connection of memory with emotion Autonomous reactions
-Diencephalon	
Hypothalamus/Hypophysis	Regulation of hormonal pathways
Thalamus	Relay to the cortex, "Gate to consciousness"
Epithalamus (pineal gland)	"Biological clock"
Mesencephalon	
Superior colliculi of the tectum	Visual relay
Inferior colliculi of the tectum	Auditory relay
Substantia nigra	Motor control
Central gray	Perception of pain
Rhombencephalon	
Pons/Medulla oblongata	Relay for fiber tracts to and from the spinal cord; Regulation of vital functions including breathing
Cerebellum	Coordination and balance

given neuron is integrated in a binary way in that it generates an AP in neurons or not. There is, however, also information encoded in the frequency of firing. In addition, input from a given region may, although not providing enough stimuli to induce continuous APs, sensitize neuronal populations so that they easily generate APs in response to input from other regions. In that sense, EPSPs induced by a given system are not necessarily "wasted" if they alone do not reach the threshold for AP generation. Once generated at the initial segment of the axon, APs spread along the axon up to its terminal, where they may induce the release of vesicles filled with neurotransmitters into the synaptic cleft. These transmitters then bind to specific receptors at the postsynaptic membrane of the target neurons, where they induce an EPSP (excitatory transmitters) or an IPSP (inhibitory transmitters) by opening certain ion channels. Thus, within the synapse, electrical information (AP) is transformed to chemical information (transmitter

release), in turn inducing electrical changes (IPSP/EPSP) at the membrane of the next neuron in the network.

Some transmitters can be taken up by the presynaptic membrane and recycled or cleaved shortly after they have been released, while others exhibit long-term effects at the postsynaptic membrane. The former are called "classical" neurotransmitters (e.g. glutamate, GABA, serotonin, acetylcholine) while the latter have been termed co-transmitters (e.g. NPY, substance P). There is, however, no sharp borderline between the two. Interfering with cleavage, re-uptake and binding of neurotransmitters to their receptors are important targets in neuropharmacology.

A single neuron produces and secretes one classical neurotransmitter into the synaptic cleft, and is therefore termed glutamatergic, dopaminergic, cholinergic, etc., but receives input *via* many different transmitters for which it possesses the appropriate receptors at the postsynaptic membrane. A good example is the

α -motoneuron, the axon that innervates muscles in the periphery. This task is regulated by many cortical, subcortical and intrinsic (spinal) systems which all terminate, often *via* local interneurons, at the α -motoneuron, using different transmitters with opposite effects. Once this neuron is excited to generate APs, there is no physiological way to hold back contraction of those muscle cells innervated by its axon. The baseline generation of APs by motoneurons innervating a given muscle leads to a certain degree of tension, which, for example, may be important for withstanding gravity. Only the coordinated excitation of the neurons innervating a single muscle leads to visible contraction. The fact that the reader can move his little finger at his command demonstrates that it is possible to induce coordinated APs in the appropriate motoneurons at will and it is known that cortical activities induce this act *via* corticospinal projections. However, if the same reader puts the finger into boiling water, sensory input will induce a spinal reflex of coordinated movements withdrawing the finger before she/he can even think about what may be the proper response to rescue the finger from serious harm. Up to a certain point, however, it is possible to suppress this reflexive response if desired, but not beyond the point of AP generation in the motoneuron. This simple example shows that different, discernable systems terminate on the same groups of neurons, here the motoneuron. At the cognitive level, it is not only input from the eye that generates conscious pictures of the environment in the cortex. Similar pictures can be summoned with closed eyes using pure imagination. In like manner, many people are capable of imaging even complex pieces of music simply by inducing the respective activity within the cortex. Thus, while environmental stimuli still lead to uniform neuronal reactions, the human brain can also purposefully trigger some of these reactions in the absence of the corresponding stimuli.

References

1. Geyer S, Matelli M, Luppino G et al (2000) Functional neuroanatomy of the primate isocortical motor system. *Anat Embryol* 202:443–474
2. Martin R, Wallace BG, Fuchs PA et al (2001) *From Neuron to Brain: A Cellular and Molecular Approach to the Function of the Nervous System*, 1st edn. Sinauer Associates, Sunderland
3. Kandel ER, Schwartz JH, Jessell TM (2000). *Principles of Neural Science*. McGraw-Hill Medical, 4th edn. McGraw-Hill, New York
4. Duus P (2005) *Tropical Diagnosis In Neurology: Anatomy, Physiology, Signs, Symptoms*. Thieme Medical Publishers, Stuttgart New York

5. Kahle W, Frotscher M (2003). *Nervous System and Sensory Organs. Color Atlas and Textbook of Human Anatomy*. vol 3, 5th rev edn. Thieme Medical Publishers, Stuttgart New York
6. Blumenfeld H (2002). *Neuroanatomy Through Clinical Cases*. Sinauer Associates, Sunderland

Brain-Derived Neurotrophic Factor

► BDNF

Branchial Arches

Definition

Branchial arches are embryonic structures that contribute cells to various organs including skeletal and muscle elements of the face.

► Muscle Development

BRCA

Definition

BRCA is a gene for early-onset familial breast and ovarian cancer, which is a tumor suppressor gene located on chromosome 17q12-21. Mutations in that gene are responsible for breast cancer.

► Breast Cancer

BRCA1

Definition

Mutations in the BRCA1 (Breast Cancer Susceptibility Gene 1) gene are responsible for 45% of the cases of inherited breast cancer and over 80% of inherited breast and ovarian cancer cases.

► Chromosomal Instability Syndromes

BRCA2

Definition

Mutations in the BRCA2 (Breast Cancer Susceptibility Gene 2) gene are responsible for some forms of inherited breast cancer.

► [Chromosomal Instability Syndromes](#)

BRCT Domain

Definition

BRCT domain is an autonomously folding protein module of about 95 amino acids, first identified in the C-terminal region of the ► [BRCA1](#) tumor suppressor protein. However, it has also been found in a wide range of proteins involved in DNA replication, repair and checkpoint functions.

► [DNA Ligases](#)

BrdU

► [5-Chlorodeoxyuridine & 5-Bromodeoxyuridine](#)

BRE

Definition

BRE (TFIIB recognition element) is a sequence motif that is commonly found in core promoter elements.

► [Core Promoters](#)

Breast Cancer

ROLAND MOLL

Institute of Pathology, Philipp University, Marburg,
Germany
mollr@med.uni-marburg.de

Synonym

Breast Carcinoma

Definition

Breast cancer (breast carcinoma) is the most frequent cancer of women. It is a group of malignant epithelial tumors that develop in the mammary gland and subsequently metastasize to regional lymph nodes and to distant organs, ultimately leading to the death of the patients.

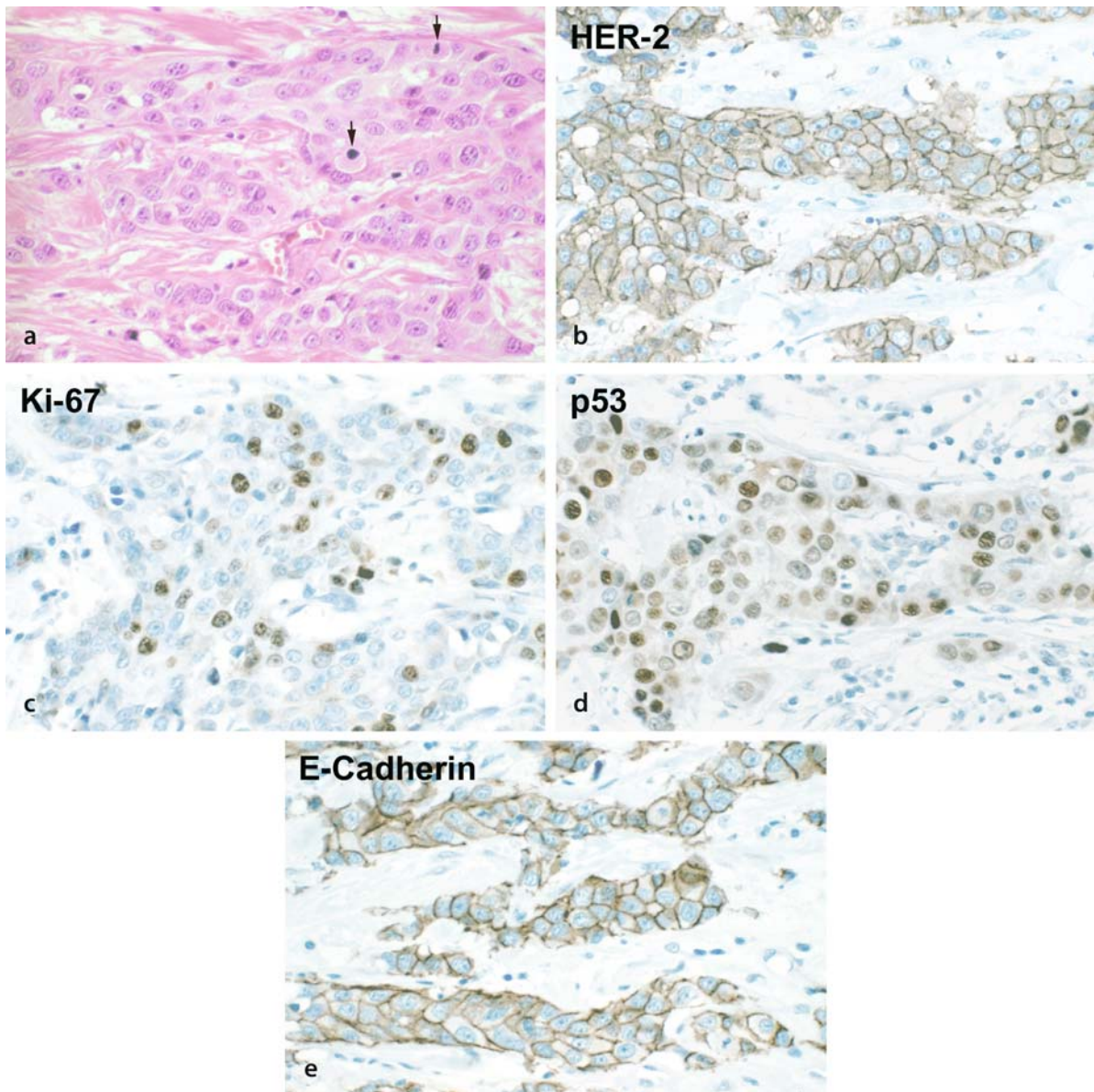
Characteristics

General Features

Human breast carcinomas arise, through malignant transformation, from the lining epithelium of the peripheral ducts (commonly the ► [terminal ductal-lobular unit](#)) of the female and, much more rarely, the male mammary gland (1). The malignant, indefinitely proliferating epithelial tumor cells first spread within ducts and/or acini (*in situ* stage) and subsequently invade the interstitial connective and fatty tissue of the breast (invasive stage). Macroscopically, this typically results in the appearance of a firm nodular, often stellate tumor mass. Later on, carcinoma cells may infiltrate adjacent tissues such as skin and pectoral (thorax wall) muscle. If invasion of local lymphatic vessels occurs, tumor cells drift into regional (mostly the axillary) lymph nodes, giving rise to lymph node metastases. The invasion of small blood vessels opens the way to hematogenous systemic metastases in distant organs, such as the lungs, liver, brain, adrenal glands and bones. Death of the patients is essentially due to the unlimited growth of these distant metastases.

Types of Breast Carcinoma

Breast cancer is no single disease; instead, a variety of histological types of breast carcinoma can be distinguished (1). The most frequent type (~75%) is invasive (infiltrating) ductal carcinoma, not otherwise specified (NOS) (Fig. 1). This type is itself heterogeneous, comprising a spectrum of morphological variants and different ► [grades of malignancy](#) [increasing from well (G1) through moderately (G2) to poorly differentiated (G3)]. Generally, the invasive tumor cells form coherent epithelial structures such as tubules, cords and sheets (Fig. 1a). The second most frequent type (~10%) is invasive lobular carcinoma (Fig. 2). Here, the tumor cells show little coherence and infiltrate in a markedly dispersed fashion in single-cell files and individual cells (Fig. 2a). Various rare types of breast cancer include, among others, tubular, mucinous, papillary and medullary breast carcinoma. These rare types behave less aggressively than the two common types.

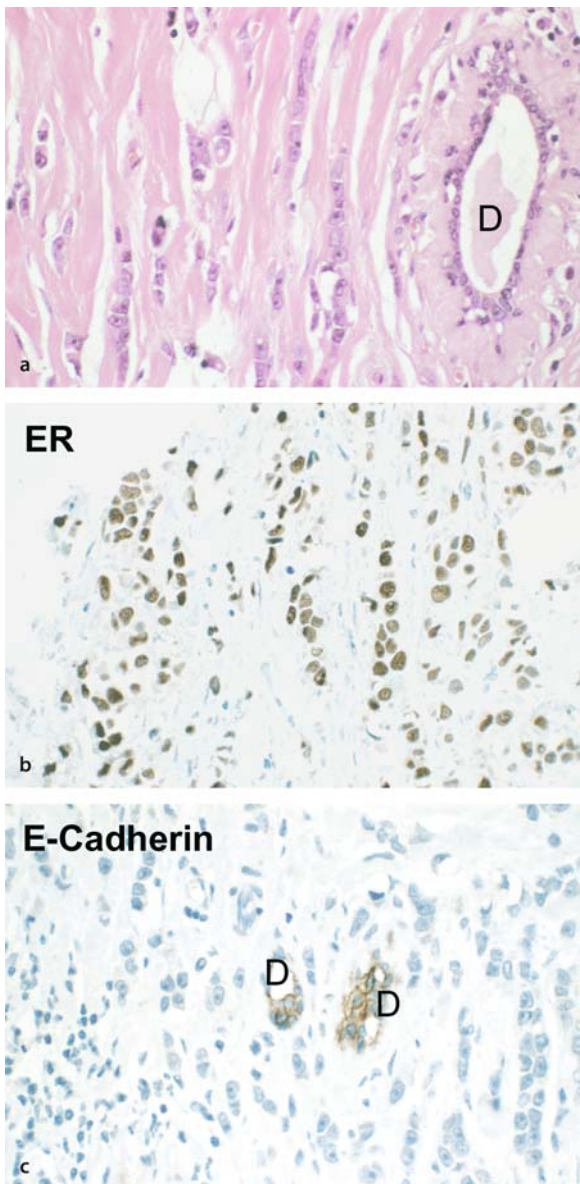


Breast Cancer. Figure 1 Invasive ductal carcinoma of the breast, poorly differentiated (G3): (a) Hematoxylin & eosin staining, showing solid cords of atypical tumor cells with pleiomorphic nuclei and mitoses (arrows). (b-e) Immunohistochemical stainings (paraffin sections; antigen retrieval by microwave oven heating), the detected antigen being visualized by brown color based on 3,3'-diaminobenzidine (DAB). (b) Strong complete membrane staining of the carcinoma cells for HER-2 protein, indicating HER-2 over-expression (the *HER-2* gene was amplified in this case). (c) Detection of proliferating tumor cells by nuclear staining for Ki-67, here revealing a proliferation index of 20%. (d) Clear-cut staining of tumor cell nuclei for p53, suggesting a *p53* gene mutation with resulting increased stability and nuclear accumulation of the p53 protein. (e) Preserved expression of E-cadherin with uniform staining of the intracellular borders. Magnification: x 220.

Histological Carcinogenesis

Concerning the development of invasive ductal carcinoma NOS, transformed cells first proliferate and expand for shorter or longer periods of time longitudinally within the intact ductal tree. This first stage is called ductal carcinoma *in situ* (DCIS) or intraductal

carcinoma (1). DCIS cell clones that have acquired invasive properties, may break through the ductal basement membrane and subsequently invade the surrounding tissue. Thereby, the second stage is reached, i.e. invasive ductal carcinoma.



Breast Cancer. Figure 2 Invasive lobular carcinoma of the breast, moderately differentiated (G2): (a) Hematoxylin & eosin staining showing the typical histomorphology with diffusely infiltrating single-cell files of tumor cells and dispersed tumor cells. Note a preserved normal mammary gland duct (D) included within the tumor tissue. (b,c) Immunohistochemical stainings (paraffin sections; antigen retrieval by microwave oven heating), the detected antigen being visualized by brown color based on DAB. (b) Strong nuclear staining for the estrogen receptor (ER) in most tumor cells, a typical finding in invasive lobular carcinoma. (c) Completely negative reaction of the tumor cells for E-cadherin, contrasting with the positive staining of two small residual normal ducts (D). Magnification: x 220.

As to the lobular type of carcinoma, early *in situ* growth of tumor cells involves the small ductules of mammary gland lobules, giving rise to lobular carcinoma *in situ* (LCIS) [also called lobular neoplasia or lobular intraepithelial neoplasia (LIN)] (1). Later, LCIS may proceed to invasive lobular carcinoma. LCIS thus represents a (nonobligatory) precursor lesion.

Invasive breast carcinomas in general are characterized by invasive and destructive growth, the induction of stromal connective tissue, often with pronounced desmoplasia, and the formation of new intra-tumoral blood vessels (►angiogenesis). As the clinically decisive steps of progression, invasive breast carcinomas give rise to lymphogenous and hematogenous metastases (see above). It has been recognized more recently that already in the early stages, single cells may detach from the primary tumor and disseminate in the body, notably in the bone marrow; however, most of these cells remain dormant.

Molecular Interactions and Regulatory Mechanisms

Molecular Carcinogenesis of Breast Cancer

Although there is ample evidence for the involvement of endocrinological and reproductive factors including sex hormones, a specific etiologic agent causing breast cancer has not yet been identified. This hampers the fight against this fatal disease, the incidence of which is still rising. Nevertheless, quite abundant knowledge on the molecular biology of breast cancer has accumulated. As holds for most other types of cancer, molecular carcinogenesis in the breast is based on genetic changes that breast epithelial cells undergo, involving the activation of ►oncogenes and the inactivation of ►tumor suppressor genes. The most important of them relevant to breast cancer will be summarized in the following section. First, a special biological feature of normal breast epithelium and many breast carcinomas, i.e. their sex hormone dependency, will be discussed.

Steroid Hormone Receptors

About two third of breast carcinomas express the ►estrogen receptor (ER) (1, 2). This is a transcription factor of the nuclear receptor superfamily, which is activated by its ligand, estrogen and then binds to certain genes containing estrogen responsive elements (EREs), activating their transcription. ER thus exerts various effects, notably the proliferation of mammary epithelial and carcinoma cells. The main ER is ER α encoded by *ESR1* (certain breast tumors express a second type of ER, ER β). Immunohistochemistry of ER-positive breast carcinomas reveals nuclear staining of variable proportions of tumor cells (Fig. 2b). ER-positive breast carcinomas behave less aggressively than ER-negative ones. Correspondingly, well/moderately differentiated invasive ductal carcinomas (grades

G1/G2) are more frequently ER-positive than poorly differentiated carcinomas (G3). Absence of ER expression in breast cancer cells has been associated with DNA methylation of the *ESR1* promoter. The great majority of invasive lobular carcinomas is ER-positive (Fig. 2b). Importantly, ER-positive tumors respond to treatment with the anti-estrogen substance, tamoxifen, making the determination of the ER status an obligatory diagnostic procedure for breast carcinomas. In the presence of a functional ER, breast carcinoma cells may also express the ER target gene encoding the ►progesterone receptor (PR), a transcription factor of the same superfamily (with two isoforms, PR-A and PR-B) stimulated by progestins.

Oncogenes

HER2

HER2 (also called ErbB2/neu; 185 kD), one of the four members of the EGFR family of transmembrane glycoprotein receptors possessing intracellular tyrosine kinase activity, is over-expressed in ~25% of invasive ductal breast carcinomas, usually due to amplification of the *HER2* gene on chromosome 17q12 (1, 3, 4). In these cases, it is an important oncogene. HER2 has no ligand of its own, but after binding of a ligand (such as EGF or TGF- α) to one of the other EGFR types, HER2 forms a heterodimer with the latter, upon which the intracellular tyrosine kinase domains are phosphorylated. This activates signaling cascades, in particular the ►PI3K pathway and the ►MAPK pathway, ultimately resulting in cell-cycle progression and increased proliferation.

Over-expression of HER2 shows some preference for poorly differentiated and ER-negative invasive ductal breast carcinomas. In tumor tissue specimens, HER2 protein over-expression can be analyzed by immunohistochemistry (Fig. 1b) and *HER2* gene amplification by fluorescence *in situ* hybridization (FISH). This is clinically important, since the group of patients with HER2-over-expressing breast carcinomas have reduced survival but respond to treatment with a humanized monoclonal antibody [trastuzumab (Herceptin)] targeted against the HER2 protein.

Epidermal Growth Factor Receptor (EGFR)

The prototype of the EGFR family of receptors, the ►epidermal growth factor receptor (EGFR) (ErbB1, ►HER1; 170 kD), is – as another oncogene – over-expressed in some breast carcinomas (associated with ER negativity), although its gene is not usually amplified (3, 4). This appears to be a factor conferring poor prognosis. There are also other mechanisms of EGFR deregulation in breast cancer, including activating gene mutation. After ligand-induced homo- or hetero-dimerization of EGFR, intracellular signaling

proceeds mainly through the MAPK pathway, resulting in increased cellular proliferation. Specific targeted therapies are in development.

Akt

►Akt (protein kinase B) (2, 4, 5) is part of the PI3K signaling pathway that may be activated by HER2 (see above). Akt is commonly activated in breast carcinomas, providing cell survival signals required for tumor progression thus suppressing apoptosis and is probably involved in mechanisms of resistance against therapeutic agents. Moreover, Akt may functionally inactivate the tumor suppressor gene product ►BRCA1.

c-Myc and Telomerase

►c-Myc is a nuclear protein (62 kD) and transcription factor involved in cell proliferation, apoptosis, and cellular transformation (1, 2, 5). The *MYC* gene on chromosome 8q24 is a mammary oncogene and amplified in ~20% of breast cancers, being correlated with ER negativity and worse prognosis. Among the target genes transactivated by c-Myc is human ►telomerase reverse transcriptase (hTERT), the catalytic subunit of telomerase (2, 6). This enzyme complex is essential for the unlimited proliferation and immortal state of cancer cells. Thus, c-Myc over-expression seems to be an important factor for increased telomerase activity in such cells. As regards breast cancer, telomerase is activated in most tumors and high levels of hTERT have been shown to be associated with worse prognosis in an independent manner.

Cyclin D1

►Cyclin D1 (1) plays an important role in progression of the cell cycle from the G1 phase to the S phase, contributing to phosphorylation of the ►retinoblastoma (Rb) protein. Its gene on chromosome 11q13, *CCND1*, is amplified in 15–20% of breast cancers, which is strongly associated with ER positivity. Over-expression of cyclin D1 is more frequent in invasive lobular carcinomas as compared to invasive ductal carcinomas. Data on the prognostic significance of cyclin D1 in breast cancer are still controversial.

Ki-67

Ki-67 (3, 7) is a unique, large nuclear protein (395 kD) whose expression is tightly connected to the proliferative status of cells: It is specifically expressed in the G₁, S, G₂, and M phase of the cell-cycle but absent in resting cells of the G₀ phase. Although, surprisingly, its physiologic function is not yet clear, it is widely used as a proliferation marker to determine the growth fraction of tissues and tumors. In breast cancer (Fig. 1c), Ki-67 has been extensively studied; a high Ki-67 fraction is an independent indicator of poor prognosis.

Tumor Suppressor Genes

p53

p53 is a nuclear transcription factor of 53 kD encoded by *TP53* on chromosome 17p13 and, upon DNA damage, may induce cell-cycle arrest or apoptosis (1, 2). *TP53* is extremely important in tumor biology and is the most commonly mutated tumor suppressor gene in sporadic breast cancer. *TP53* mutations are found in 20–25% of breast cancers and are mostly inactivating missense mutations accompanied by loss of the wild type allele. The mutated p53 protein is usually abnormally stable and thus can easily be detected in the tumor cell nuclei by immunohistochemistry (Fig. 1d). Mutations of *TP53* in breast carcinomas are associated with aggressive histopathological features and poor survival of the patients.

Rb 1-inducible coiled-coil 1 (*RB1CC1*)

For the ►retinoblastoma gene *RBI*, the classical tumor suppressor gene, no mutations have been observed in primary breast cancers. However, the *RB1CC1* gene, coding for retinoblastoma 1-inducible coiled-coil 1 (*RB1CC1*) (1), an important activator of *RBI*, has been described to be mutated in 20% of such tumors, in this manner still producing inactivation of the retinoblastoma pathway. The tumor suppressor gene *RB1CC1* is located on chromosome 8q11; its product represents a transcription factor localized in the nucleus.

PTEN

The *PTEN* tumor suppressor gene (on chromosome 10q23) (2) encodes for PTEN protein, which inhibits the activation of Akt (see Akt). Loss of PTEN due to gene mutation thus results in unopposed cell survival signals. *PTEN* is frequently mutated in sporadic breast cancer; its germline mutation causes Cowden's syndrome predisposing to breast cancer.

BRCA1

►*BRCA1* is a rather large protein of 220 kD with a predominantly nuclear localization and cell-cycle-dependent expression (5). It is encoded by *BRCA1* (on chromosome 17q12-21). The multiple functions of *BRCA1* include inhibition of cell proliferation, DNA damage-induced cell-cycle arrest at the G2/M checkpoint and DNA repair. It inhibits the activity of the mammary oncogene c-Myc. *BRCA1* is the most commonly mutated gene in the small group of hereditary (familial) breast cancers (comprising 5–10% of all breast cancer cases). *BRCA1* mutations are usually frame-shift mutations resulting in truncated *BRCA1* proteins. The resulting tumors are peculiar; they often have high nuclear grade, are exceptionally highly proliferative, are ER and PR negative, lack HER2 over-expression, show a high frequency of *TP53* mutations and often present histological features of the

medullary type of breast carcinoma, but the prognosis for these patients is probably no worse than that for women with sporadic breast cancer. Notably, this tumor suppressor gene also plays a role in sporadic breast cancer, since 30–40% of such tumors, in particular poorly differentiated ones, show decreased *BRCA1* expression caused by different mechanisms.

BRCA2

The ►*BRCA2* gene on chromosome 13q12 is the second most frequently mutated gene in hereditary breast cancer (5). The *BRCA2* protein is very large (380 kD) and localized in the nucleus. Its expression is cell-cycle-dependent and co-regulated with that of *BRCA1* and the two proteins have overlapping functions, e.g. in DNA repair. In particular, *BRCA2* is involved in the repair of DNA double strand breaks during the S phase; consequently, the loss of this tumor suppressor gene leads to chromosomal instability, promoting malignant transformation. Breast carcinomas arising on the basis of *BRCA2* mutations are usually ER positive.

Maspin

Maspin (42 kD) is a serine protease inhibitor (serpin) with various tumor-suppressing functions (3, 8). Its expression is down-regulated during breast cancer progression by gene silencing.

E-Cadherin

E-Cadherin is the prototype of the ►cadherin family of calcium-dependent homophilic cell-cell adhesion proteins and one hallmark of epithelial differentiation (9). It is a transmembrane glycoprotein of 120 kD localized in intercellular ►adherens junctions. In malignant epithelial tumors, its role as a potent invasion suppressor molecule has been demonstrated in various experimental systems.

As to E-cadherin's role in breast cancer (8, 9, 10), the situation differs between ductal and lobular carcinomas. Invasive ductal carcinomas show either preserved (Fig. 1e) or heterogeneously reduced E-cadherin expression. Reduced expression, which may be due to promoter hypermethylation of the E-cadherin gene (*CDH1*) and the presence of repressing transcription factors such as the zinc finger transcription factor Snail, is correlated with a higher degree of malignancy and poor prognosis. The down-regulation of E-cadherin expression in some invasive ductal carcinomas may be transient, since there may be strong re-expression in corresponding metastases.

In contrast, the majority of invasive lobular carcinomas characteristically exhibit inactivating mutations of one allele of the *CDH1* gene on chromosome 16q22.1, mostly truncating, together with loss of heterozygosity of the wild type allele (9). Immunohistochemically,

most of these tumors are completely E-cadherin negative (Fig. 2c) (9, 10). This is most probably correlated with the conspicuously diffuse and dissociated infiltrative growth pattern of invasive lobular carcinomas. Since E-cadherin is already absent in the precursor stage, LCIS (see above), its loss appears to be an early event during tumor development, with E-cadherin exerting a genuine tumor suppressor role. E-cadherin immunostaining may be used for diagnostic differentiation between ductal and lobular breast carcinomas. An E-cadherin-associated cytoplasmic plaque protein of adherens junctions, β -catenin, which in certain tumors has an oncogenic function, is usually also absent in invasive lobular breast cancers.

Molecular Mechanisms of Progression of Breast Carcinoma

After malignant transformation and acquisition of indefinite cell proliferation, further progression of breast carcinoma involves tissue invasion, induction of mesenchymal stroma, and metastasis. These processes are based on various molecular mechanisms (including e.g. E-cadherin as invasion suppressor; see E-Cadherin).

Invasion and Metastasis – Role of Proteases

An essential step for invasive growth and metastasis is proteolytic degradation of extracellular matrix components, including those of basement membranes of epithelia and blood vessels, thus opening up the way for invading tumor cells. Various proteases are involved in breast cancer invasion. Cathepsin D (2, 8) is a lysosomal aspartyl endoprotease belonging to the pepsin family. Its expression is regulated by estrogen. Its multiple functions include not only matrix degradation but also promotion of mitoses and tumor growth. The level of cathepsin D produced by breast cancers is correlated with poor prognosis. Another proteolytic system very important for breast cancer invasion and prognosis is the serine protease urokinase plasminogen activator (uPA) (2, 8), its receptor [uPAR (CD87)] and its inhibitor (PAI-1). uPA activates plasminogen to produce plasmin; it again has multiple functions. Associated with the cell surface, uPA-uPAR confers pericellular proteolytic activity and promotes the invasive and metastatic capacity of tumor cells. There is a strong correlation between high concentrations of uPA and PAI-1 in breast cancer tissue and poor prognoses for the patients. A third group of proteinases highly relevant in breast cancer are the **▶matrix metalloproteinases** (MMPs), a family of zinc-dependent endopeptidases (2, 8). MMPs degrade basement membranes and interstitial matrix but also have roles in tumor growth and angiogenesis. High levels of MMP-2 (collagenase type IV) and MMP-11 (stromelysin-3) have been found to correlate with poor prognosis.

There are specific tissue inhibitors of MMPs (TIMPs). Attempts are being made to develop targeted therapies with specific MMP inhibition but they have not yet succeeded.

Stroma Induction and Angiogenesis

The stromal compartment containing myofibroblasts and blood vessels also plays an important role in the biology of breast cancer. Macrophages stimulated by cytokines such as granulocyte-macrophage colony stimulating factor (GM-CSF) produce growth factors such as transforming growth factor- β 1 (TGF- β 1) which stimulates collagen-synthesizing myofibroblasts. This results in remodeling of the extracellular matrix and the formation of a reactive desmoplastic stroma, which, particularly in breast cancer, may be excessive. *Vice versa*, stromal fibroblasts promote tumor growth. In parallel, angiogenic factors, in particular vascular endothelial growth factor-A (**▶VEGF-A**), secreted by tumor and stroma cells induce the new formation of tumor-supplying blood vessels (3). VEGF-A stimulates endothelial cells by binding to the receptor tyrosine kinase VEGFR2/KDR. Effective anti-angiogenic therapies are not yet available.

Various Breast Cancer Antigens

Breast cancer cells, although rarely producing lactation-associated molecules, may express certain differentiation-related antigens with more or less tissue specificity for the mammary gland. Such antigens may serve as histological and/or serological tumor markers. Among these, gross cystic disease fluid protein-15 (GCDFP-15), a major glycoprotein of 15 kD present in breast cyst fluid, has been established as a useful histological marker for breast origin of unclear carcinoma metastases. Except for a few other rare tumor types expressing GCDFP-15, its expression is specific for breast carcinomas (about two thirds of cases being positive). Mammaglobin, a 10 kD glycoprotein of the uteroglobin family, is a relatively mammary-specific differentiation marker of still unknown function (3, 11). Its expression is associated with a less aggressive breast cancer phenotype. RT-PCR tests for mammaglobin mRNA may detect disseminated breast cancer cells in lymph node, bone marrow and blood specimens. MUC-1, a mucin-like transmembrane glycoprotein of high molecular weight recognized by monoclonal antibody CA15-3, is widely used as a serological marker for monitoring breast cancer progression; whether MUC-1 might serve as an immunotherapeutic target still remains to be clarified.

Cytokeratins are a large family of cytoskeletal (**▶intermediate filament**) proteins of epithelial cells, which are expressed in cell type and differentiation dependent patterns. Most breast carcinomas exhibit a simple (luminal) epithelial-type pattern consisting of

CK7, CK8, CK18 and CK19. A small proportion (15%–20%) of invasive ductal carcinomas reveals a bimodal pattern with additional, mostly focal expression of basal-cell markers CK5, CK14 and/or CK17 (12). This tumor group is characterized by aggressive behavior and poor clinical outcome and appears to be distinct, since it corresponds to the basal-like subgroup as identified by cDNA microarray profiling (see below).

Gene Expression Profiles

The novel technique of cDNA ►microarrays, allowing the simultaneous expression analysis of thousands of genes, has been applied to breast carcinomas (1, 2, 13). Although these studies are still limited, five tumor subtypes could be delineated on the basis of distinct mRNA patterns (13). Among these, a basal-like subgroup, with expression of basal cell-typical cytokeratins (12), showed a particularly poor prognosis. In contrast, a luminal-like subgroup was characterized by a good prognosis. It appears that gene expression profiling is a highly powerful method for predicting disease outcome and may be a potential approach to reach the important goal of individualizing breast cancer therapy. To what extent this expensive technique will become clinical routine remains, however, to be seen in the future.

Acknowledgements

I thank Dr. Melissa Conrad Stoppler and Dr. Hubert Stoppler for reading and help with the manuscript.

References

1. Tavassoli FA, Devilee P (eds) (2003) World Health Classification of Tumours. Pathology and Genetics of Tumours of the Breast and Female Genital Organs. Chapter 1: Tumours of the breast, and Chapter 8: Inherited tumour syndromes. IARC Press, Lyon, pp 9–112, pp 335–363, and (for references) pp 372–424
2. Ross JS, Linette GP, Stec J et al (2004) Breast cancer biomarkers and molecular medicine: part II. Expert Rev Mol Diagn 4:169–188
3. Ross JS, Linette GP, Stec J et al (2003) Breast cancer biomarkers and molecular medicine. Expert Rev Mol Diagn 3:573–585
4. Atalay G, Cardoso F, Awada A et al (2003) Novel therapeutic strategies targeting the epidermal growth factor receptor (EGFR) family and its downstream effectors in breast cancer. Ann Oncol 14:1346–1363
5. Rosen EM, Fan S, Pestell RG et al (2003) *BRCA1* gene in breast cancer. J Cell Physiol 196:19–41
6. Poremba C, Heine B, Diallo R et al (2002) Telomerase as a prognostic marker in breast cancer: high-throughput tissue microarray analysis of hTERT and hTR. J Pathol 198:181–189
7. Brown DC, Gatter KC (2002) Ki67 protein: the immaculate deception? Histopathology 40:2–11
8. Debies MT, Welch DR (2001) Genetic basis of human breast cancer metastasis. J Mammary Gland Biol Neoplasia 6:441–451
9. Berx G, Roy F van (2001) The E-cadherin/catenin complex: an important gatekeeper in breast cancer tumorigenesis and malignant progression. Breast Cancer Res 3:289–293
10. Moll R, Mitze M, Frixen UH et al (1993) Differential loss of E-cadherin expression in infiltrating ductal and lobular breast carcinomas. Am J Pathol 143:1731–1742
11. Zehentner BK, Carter D (2004) Mammaglobin: a candidate diagnostic marker for breast cancer. Clin Biochem 37:249–257
12. Rijn M van de, Perou CM, Tibshirani R et al (2002) Expression of cytokeratins 17 and 5 identifies a group of breast carcinomas with poor clinical outcome. Am J Pathol 161:1991–1996
13. Sørlie T, Tibshirani R, Parker J et al (2003) Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci USA 100:8418–8423

Breast Cancer Resistance Protein

Definition

Breast cancer resistance protein (BCRP), alias Mitoxantrone-resistance protein (MXR), alias ABC transporter in placenta (ABCP), alias ABCG2, belongs to the ABC transporter family, subfamily ABCG (4q22). It is an integral membrane protein of 655 amino acids acting dimerized as an energydependent drug efflux pump responsible for decreased drug accumulation in multi-drug resistant cells.

►Multi-Drug Resistance

Breast Cancer Susceptibility Gene 1

►BRCA1

Breast Cancer Susceptibility Gene 2

►BRCA2

Bromodomain

Definition

Bromodomains comprise of an extensive family of evolutionarily conserved protein modules, originally

found in proteins that are associated with chromatin, and in nearly all nuclear histone acetyltransferases. The structural motif binds with high affinity to acetylated histones.

- ▶ Nucleosomes
- ▶ Chromatin Acetylation

Bronchial Asthma

- ▶ COPD and Asthma, Genetics

Bronchospasm

Definition

Bronchospasm describes the contraction of smooth muscle in the walls of the bronchi and bronchioles.

- ▶ Hyper-and Hypoparathyroidism

Bronzed Diabetes

- ▶ Hemochromatosis

Brown Tumor

Definition

Brown tumor refers to a mass of fibrous tissue occurring in primary hyperparathyroidism containing macrophages and multinucleated giant cells, which replaces and expands part of a bone.

- ▶ Hyper-and Hypoparathyroidism

Brush Border

Definition

A brush border is a cellular structure specialised in absorption. It is formed by a large number of microvilli,

finger-like plasma membrane projections supported by an array of bundled actin filaments. Brush borders are mainly found on the apical surface of absorptive epithelial intestinal and kidney cells.

- ▶ Microvilli

BSE

Definition

BSE (bovine spongiform encephalopathy) is a prion disease of cattle and related to the ▶ Creutzfeldt-Jakob disease of humans and scrapie of sheep.

- ▶ Prion Diseases

BTB/Kelch

Definition

BTB/kelch is the short name for Bric-a-brac, Tram-track, Broad-complex or POZ (Poxvirus zinc finger) domain, a protein/protein interaction module, present in the *Drosophila melanogaster* Tramtrack (Ttk) and Broad Complex (BR-C) zinc finger proteins. *Drosophila* kelch binds and organizes actin filaments and is involved in oocyte development.

- ▶ Hereditary Neuropathies, Motor and/or Sensor

Bud Neck

Definition

Bud neck defines the site of cleavage between mother and daughter cells in budding yeast (*S.cerevisiae*).

- ▶ Sumoylation

Bullous (Blistering)

Definition

Bullous (blistering) is an adjective describing a dermatologic condition in which bullae are found. A bulla is a blister, or a fluid filled cavity that forms within or below the epidermis.

- ▶ Desmosomes

Bullous Impetigo

Definition

Bullous impetigo describes a skin disease that is characterized by epidermal blistering due to localized staphylococcal skin infection.

► [Desmosomes](#)

bZip

Definition

bZIP is a DNA binding domain of AP-1 subunits and other cellular proteins that is characterised by the “basic domain” and “leucine zipper” motifs.

► [Jun/Fos](#)

BX-C

► [Bithorax Complex](#)

bZIP Proteins

► [Leucine Zipper Transcription Factors: bZIP Proteins](#)

C Type Lectin Domain

Definition

C type lectin domains are involved in calcium dependent carbohydrate recognition.

- ▶ Autosomal Dominant (Inherited Disorder)
- ▶ Polycystic Kidney Disease, Autosomal Dominant

C. Elegans

- ▶ *Caenorhabditis elegans* as a Model Organism for Functional Genomics
- ▶ *C. elegans* Genome, Comparative Sequencing

C. elegans Genome, Comparative Sequencing

JAGAN SRINIVASAN, RALF J. SOMMER
Max Planck Institute for Developmental Biology,
Department of Evolutionary Biology, Tübingen,
Germany
ralf.sommer@tuebingen.mpg.de

Definition

Caenorhabditis elegans (*C. elegans*) is a small, free-living nematode found commonly in many parts of the world. It is 1 mm long and has a life cycle of 3–4 days at 20°C. *C. elegans* is a bacteriovore and feeds mainly on *Escherichia coli* under laboratory conditions. It exists as two sexes, as hermaphrodites and males. The hermaphrodite produces both sperm and oocytes and reproduces by self-fertilisation. Spontaneous males are very rare and occur at a frequency of 10^{-3} to 10^{-4} . During its reproductive life span, a hermaphrodite lays around 300 eggs. This number is controlled since

sperm is the limiting factor during self-fertilisation. Freshly hatched eggs undergo four larval stages (L1–L4), punctuated by moults. The adult arising after the fourth larval moult is reproductively mature for about four days and lives for an additional ten days.

Characteristics

The success of the worm as a model system in modern biology lies in its easy amenability under laboratory conditions and its simple anatomical architecture. *C. elegans* is made up of a relatively small number of cells and the cell lineage is invariant, that is, it is identical from individual to individual. Specifically, the adult hermaphrodite has 959 cells and the adult male has 1031 cells. The development of all of these cells is known and can be studied by differential interference contrast (DIC) microscopy. Pioneering work during early phases of *C. elegans* research led to the complete description of the cell lineage and the determination of the connectivity of the nervous system (1). To test functional aspects of the cell lineage and potential cell-cell interactions two approaches are used. First, by laser microbeam irradiation, individuals cells can be deleted from living animals and the consequences of these “cell ablations” be investigated during the further development of the individual. Second, mutants can be obtained following chemical mutagenesis or exposure to ionising radiation (2). *C. elegans* possesses six linkage groups (5 autosomes and 1 sex chromosome) with each linkage group having a genetic size of 50 cM. The physical size of the genome is 100 Mb. Together, the small genome size, the rapid generation time, the invariant cell lineage and the ability to generate mutants made *C. elegans* a popular model system for developmental genetics.

Further research in *C. elegans* focussed on getting a molecular understanding of the vast diversity of mutants isolated in various genetic screens. These studies were accelerated by the development of two major technologies; first the production of a ▶cosmid physical map by the *C. elegans* genome project and second the development of an effective transformation system (3). The physical map was constructed by ▶fingerprinting clones from a cosmid library. Gaps in the cosmid library were covered by the ▶yeast artificial chromosome (YAC) library. In addition, an ▶expressed sequence tag (EST) library was created to support the

identification of genes. Transgenic animals can easily be obtained by injecting all sorts of DNA into the syncytial gonad of gravid females. In a standard assay, 50–100 transgenic lines can be obtained by injecting 10–20 animals.

Genome Sequence

C. elegans was the first metazoan genome to be sequenced completely (4). The 100 Mb sequence comprised 2527 cosmids, 257 YACs and 113 fosmids. Genome sequencing revealed that at first sight the genome looks remarkably uniform. The GC content is 36% and remains essentially unchanged across all chromosomes. As of 1st July 2003, the genome contains the surprisingly high number of 20,621 predicted **▶open reading frames** (ORF). Gene predictions in *C. elegans* are done *ab initio* using the Gene Finder algorithm (<http://ftp.genome.washington.edu/cgi-bin/GeneFinder>). Sequence analyses indicated that centres of chromosomes have lower recombination frequencies and a higher gene density than chromosomal arms. The X chromosome has more genes than any of the autosomes. The chromosomal arms have a prevalence of **▶tandem repeats** and **▶pseudogenes** indicating that these regions are highly evolving.

Another unique feature of *C. elegans* and some other nematodes is that several genes are organised in polycistronic units called **▶operons**. Such operons are resolved by polyadenylation and **▶trans-splicing**, with the first gene in the operon receiving the spliced leader sequence 1 (SL1) and the downstream genes receiving either SL1 or a spliced leader of another family of leader sequences. Recent studies suggest the existence of around 1000 operons in *C. elegans*. It should be noted however, that genes within one operon are mostly unrelated in function and that the evolutionary origin of *C. elegans* operons and trans splicing remains a mystery. Information on all of the predicted ORFs and the complete annotated genome is stored in a centralized database known as WormBase (<http://www.wormbase.org/>).

Functional Genomics

Less than 5% of the 20,621 predicted ORFs have been characterised by conventional genetic or biochemical tools. To revisit the information that was generated by the genome sequence, high throughput functional genomic approaches are being carried out. These approaches are based upon the generation of loss of function phenotypes, analysis of expression profiles and protein-protein interactions.

Large-scale Generation of Loss-of-function Phenotypes by RNA Mediated Interference (RNAi)

The discovery that treatment of individuals with single or double stranded RNA of a gene could result in a

mutant phenocopy was a revolution in exploring gene function (5). The advantage that this technique provided over conventional mutagenesis screens was the speed with which one could relate sequence information to gene function. Gene knockdown by **▶RNAi** can be administered by several ways, injecting into the gonad of the worm, soaking the worm in the RNA solution or feeding the worm with dsRNA-expressing bacteria. The phenotype can be observed not just in the injected mother but also in the progeny of the treated worms.

With the completion of the genome sequence of *C. elegans*, a large-scale systematic RNAi analysis of 16,757 ORFs (87% of the total gene complement of *C. elegans*) has recently been carried out. 1721 (10.3%) of the genes analysed gave a detectable phenotype ranging from non-viable (lethals and steriles) to viable (growth and post-embryonic) phenotypes. Among the detectable phenotypes, the most common RNAi phenotype was embryonic lethality (5.5%). 92% of the observed phenotypes corresponded to the genetic mutant phenotype if known. An added benefit of this large-scale approach is the availability of a bacterial strain library of the entire genome complement.

Gene Expression Profiling

A way of complementing the genome wide RNAi analysis is to examine gene expression patterns on a global scale. A whole genome microarray of more than 17,000 ORFs has been developed. Different experiments utilising several different parameters such as growth conditions, developmental stages and mutants have been performed to obtain a global analysis of gene expression. Such a correlation reveals an **▶expression landscape** with several 'gene mountains' showing similar expression profiles (6). Each mountain contains sets of highly correlated genes with the mountain altitude representing the total number of genes present in the mountain and the mountain width denoting the overall level of correlation of all the genes present in that mountain. *In silico* microarray data are supported by *in situ* hybridization of a non-redundant EST dataset of *C. elegans* providing useful information on cells and tissues in which genes are expressed.

Protein – Protein Interaction Screens

Once an ORF is identified and characterized using genetic and/or RNAi approaches, one can study the nature of the protein encoded by the gene and its potential interactors. Large-scale **▶yeast two-hybrid analysis** allows determination of genome wide protein interactions that may provide valuable insight into protein function. To test the conceptual and functional feasibility of such a method, interactions of key proteins involved in vulva development were used as an example. 29 proteins were chosen and a matrix experiment was performed to identify the already known interactors.

50% (6/11 interactions) of the interactions already known in the literature were detected.

Comparative Nematode Genomics

To complement the *C. elegans* genome sequence, a draft genome sequence of *C. briggsae* has been released. This free-living soil nematode belongs to the same genus as *C. elegans* and has diverged from *C. elegans* ~50–150 million years ago. The *C. briggsae* genome has been estimated to be around 108 Mb, approximately, 8 Mb more than *C. elegans*. Comparison of segments from the two species reveals conservation of gene order making cloning by synteny easier. It also strengthens gene and exon predictions often providing candidate regulatory elements based on sequence similarities between *C. elegans* and *C. briggsae* in upstream and intronic regions. The coding regions between the two species are more similar than the non-coding regions. At a more global level, comparisons of genome sequence between *C. elegans* and *C. briggsae* suggest a rearrangement rate of 0.4–1.0 breakages/Mb/million years.

Nematode Diversity

Nematodes were treated as a minor phylum and originally classified as pseudocoelomates. Recent evidence indicates a close relationship between nematodes and arthropods in a newly suggested monophyletic taxon named Ecdysozoa. Nematodes inhabit nearly all ecosystems and are numerically abundant with over millions of animals per square meter. Although the number of described species is only ~20,000, estimates suggest that there are over 10 million different nematode species. They exist either, as free-living, parasitic or aquatic forms. Molecular data from several nematodes have been applied at a phylogenetic level, revealing a depth and diversity in many groups of these nematodes. The molecular phylogeny provides landmarks across the nematode phylum from a target gene of interest in *C. elegans* to socio economic aspects of a parasitic nematode.

Clinical Relevance

The annotated genome sequence was the dawn of a new era in the clinical relevance of the worm. The discovery of several genes in the worm having human homologues is beginning to make an impact on our understanding of human disease genes. It also proves that *C. elegans* can now be used as a model to study several human diseases, such as cancer, neurodegeneration, infectious diseases and physiological disorders. Sequence analysis of the worm genome suggested that 42% of human disease genes have orthologues in *C. elegans*. Identifying the human homologue in the worm and knocking it down using ►‘reverse genetics’ is a useful way of understanding the function of this

gene. In the long run, the functional analysis of orthologous genes in *C. elegans* makes it possible to understand the mechanisms of human disease with the added advantage of suggesting novel approaches in drug discovery. Given the several functional genomic approaches in *C. elegans* outlined above, one can identify candidate modifier genes and study their nature of action. Another aspect of nematode research is the biology of parasitic nematodes. Amongst the parasitic nematodes, many species affect humans or other vertebrates and agriculture worldwide. For example, *Onchocerca volvulus* a parasitic nematode, is the causative agent of river blindness, one of the most harmful human diseases. In order to ascertain the molecular nature of parasitism, numerous EST sequencing projects have been initiated. At present, 33 different nematode species are being sequenced (7). The partial genomic information from these projects reveals information that can lead to the development of novel control strategies and anti-nematode drugs.

►*Caenorhabditis elegans* as a Model Organism for Functional Genomics

References

1. Sulston JE, Horvitz HR (1977) Post-embryonic cell lineages of the nematode, *Caenorhabditis elegans*. Dev Biol 56:110–156
2. Brenner S (1974) The genetics of *Caenorhabditis elegans*. Genetics 77:71–94
3. Mello CC, Kramer JM, Stinchcomb D et al (1991) Efficient gene transfer in *C. elegans*: extrachromosomal maintenance and integration of transforming sequences. EMBO J 10:3959–3970
4. The *C. elegans* Sequencing Consortium (1998) Genome Sequence of the nematode *C. elegans*: a platform for investigating biology. Science 282:2012–2018
5. Fire A, Xu S, Montgomery MK et al (1998) Potent and specific genetic interference by double-stranded RNA in *Caenorhabditis elegans*. Nature 391:806–811
6. Kim SK, Lund J, Kiraly M et al (2001) A gene expression map for *Caenorhabditis elegans*. Science 293:2087–2092
7. McCarter JP, Clifton SW, Bird DM et al (2002) Nematode gene sequences, update for June 2002. J Nematol 34:71–74

C2H2 Zinc Finger Gene

Definition

C2H2 zinc finger genes comprise of a subfamily of one of the largest human gene families, the zinc finger gene family, which plays an important role in the regulation of transcription. They contain a C2H2 ►zinc finger, a small protein domain with a helical region that binds to the major groove of the DNA. The structure is held

together by a zinc atom bound to two cysteines and two histidines, from which the name is derived. In the (Cys)₂(His)₂ type of zinc finger genes, there is a highly conserved consensus sequence TGEKPYX (X representing any amino acid) between adjacent zinc finger motifs.

► [Gene Duplications](#)

an autosomal disorder caused by mutations in the notch3 gene. Patients with CADASIL show a variety of symptoms such as migraine, mood disorders, ischemic strokes, dementia and premature death.

► [Genetic Predisposition to Multiple Sclerosis](#)

► [Notch Pathway](#)

C57BL/6J

Definition

C57BL/6J is the inbred strain of mouse whose genome was sequenced by the Mouse Genome Sequencing Consortium (MGSC) in 2002. Bred by Clarence Little in 1921, this mouse strain has become one of the strains most widely used by mouse geneticists.

► [Mouse Genomics](#)

Cadherins

AKIRA NAGAFUCHI¹, MASATOSHI TAKEICHI²

¹Department of Cellular Interactions, Institute of Molecular Embryology and Genetics, Kumamoto University, Kumamoto, Japan

²RIKEN Center for Developmental Biology, Chuo-ku, Kobe, Japan

naga-san@kaiju.medic.kumamoto-u.ac.jp,

takeichi@cdb.riken.jp

CAAX-Box

Definition

The CAAX-box indicates the amino acid sequence cysteine – aliphatic amino acid – aliphatic amino acid – any amino acid at the C-terminus of proteins. Proteins with a CAAX-box are modified at the cysteine with isoprene-groups via a thioether linkage. Generally, a farnesyl-group (C 15) is transferred to the protein unless leucine is present at position X, which results in modification with the more hydrophobic geranyl-geranyl moiety (C 20). Two different enzymes, designated farnesyl- and geranylgeranyltransferase I, catalyze these reactions. The amino acids AAX are then cleaved by an endoproteinase, and the resulting carboxyl group of the cysteine residue is methylated. Rab-proteins, a large family of small G-proteins involved in vesicular transport, contain the C-terminal sequence CC or CXC. They are geranylgeranylated at both cysteine residues by yet another enzyme, geranylgeranyl-transferase II.

► [Fatty Acid Acylation of Proteins](#)

► [Protein Prenylation](#)

Definition

A cadherin was identified as a calcium-dependent cell-cell [adhesion molecule](#). A mature cadherin polypeptide is a type I transmembrane protein composed of three domains, an extracellular domain, a transmembrane domain and a cytoplasmic domain. The extracellular domain is composed of five repetitive sequences – the cadherin repeat or EC (extracellular) domain. Basically, each repeat has conserved sequences like PE, LDRE, DxDND and DxD, some of which have been implicated in calcium binding. Over 20 different cadherin subtypes have been found in a single vertebrate species. The cytoplasmic domain of cadherin is conserved among these subtypes and interacts with catenins through the carboxy terminal half domain as discussed below. This catenin-binding domain is required for the normal function of cadherins. The presence of the cadherin motifs in the extracellular domain and the catenin-binding site in the cytoplasmic domain serve as defining characteristics of cadherins (known as “classic” cadherins when using a more precise classification; see below). On the basis of specific amino acid sequences in the cytoplasmic domain, cadherins can be divided into two subfamilies, type I and type II cadherins. Type I cadherins comprise E-, N-, P- and R-cadherins, which were named according to the organ or tissue in which they were first identified (epithelial, neural, placental and retinal, respectively). Many type II cadherins were identified by the cloning of molecules with the cadherin repeats, and were named using a number system representing the chronological order of their discovery, in which the

CADASIL

Definition

CADASIL (cerebral autosomal dominant arteriopathy with subcortical infarcts and leukoencephalopathy) is

discovery of cadherin-5 predates that of cadherin-6, and so on. Type I cadherins are also represented as numbered cadherins; E-cadherin may be represented as cadherin-1, N-cadherin as cadherin-2, P-cadherin as cadherin-3 and R-cadherin as cadherin-4.

A large number of molecules have been reported that have cadherin repeats in an extracellular domain, but without a catenin-binding site in the cytoplasmic domain. These molecules are collectively called non-classic cadherins. In this respect, the cadherins discussed within this essay are classic cadherins. Non-classic cadherins include a variety of molecules. Major subfamilies are desmosomal cadherins, Fats, protocadherins and Flamingos/Celsrs. These non-classic cadherins display variation in the number of the cadherin repeats and cytoplasmic sequences. Desmosomal cadherins consist of desmogleins and desmocollins, whose cytoplasmic tails have some similarity with those of classic cadherins. These proteins are responsible for adhesion and anchorage to cytoskeletons in a specialized cell-cell junction: the ►desmosome (see below). Fat is a large cadherin with 34 cadherin repeats in the extracellular domain. This protein was first identified as a tumor-suppressor gene in *Drosophila*, and is involved in cell proliferation and maintenance of epithelial structure. Flamingos/Celsrs are also large cadherins that are characterized by a 7-pass transmembrane domain. These proteins are involved in the control of planar polarity and patterning of dendrites. The physiological functions of other non-classic cadherins, including whether these molecules function as cell adhesion molecules, have not been elucidated. The cadherins on which we will concentrate in this short essay are classic cadherins.

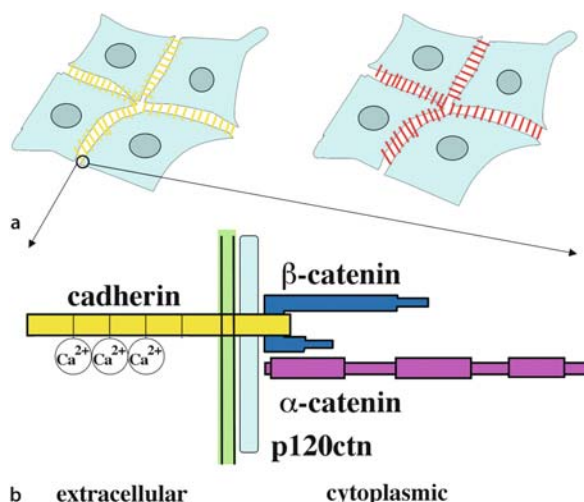
Characteristics

Major characteristics of cadherins include calcium dependency and the exhibition of strong cell adhesion activity, cytoskeletal interactions and cell adhesion specificity. In the absence of calcium, cadherins do not function as cell adhesion molecules and are sensitive to trypsin digestion. Cadherins mediate strong cell-cell adhesion when in the presence of calcium. As a consequence, cadherin-positive cells are seldom dissociated into single cells from cell aggregates, even under high shearing force. This strong cell adhesion activity is dependent on an interaction between the cadherin and the actin-based ►cytoskeleton at the proper temperature, and is thought to play a fundamental role in tissue formation and maintenance. Dysfunction of cadherins is believed to be one of the major causes facilitating ►cancer invasion and metastasis. Adhesion specificity is another characteristic property of cadherins. For example, E-cadherin-positive and P-cadherin-positive cells aggregate separately when mixed. Since different cadherin

subtypes are expressed in different tissues, cadherins are thought to play crucial roles in cell-type specific selective adhesions.

Molecular Interactions

A single cadherin repeat (EC domain) consists of about 110 amino acids and forms a barrel-like structure. It has been shown that 3 Ca^{2+} are bound at each interdomain boundary of the first four cadherin repeats. The tandem array of barrels is flexible when the concentration of Ca^{2+} is low, but rigid when the concentration of Ca^{2+} is high. It has also been suggested that two cadherin molecules form a homodimer in the presence of Ca^{2+} . The structural changes caused by the Ca^{2+} interaction seem to be necessary for cadherin to function as a cell adhesion molecule. The conformational change of cadherin in the presence of calcium also protects certain subtypes of cadherins from degradation by trypsin. Cadherin-mediated cell-cell adhesion is basically homophilic; the extracellular domain of cadherin molecules binds to that of other cadherin molecules on neighboring cell surfaces. Various analyses of crystal structure have so far yielded three models that attempt



Cadherins. Figure 1 Schematic drawing of cadherin-mediated cell-cell adhesion. (a) Cell-cell adhesion is mediated by the cell adhesion molecule cadherin (yellow or red bars). Different cadherins, yellow cadherins and red cadherins, may have different adhesion specificities. Cells expressing different cadherins preferentially form different colonies. (b) The cell adhesion molecule cadherin is a transmembrane protein. Ca^{2+} is required for the proper conformation of the extracellular domain, which is directly responsible for cell-cell adhesion. Cadherin forms a complex with several cytoplasmic components, collectively called catenins. β -catenin and p120ctn might regulate the adhesion activity of cadherins. α -catenin is involved mainly in cadherin-cytoskeleton interactions.

to explain the trans-interaction of cadherin extracellular domains. As these models vary in specific details, the molecular mechanisms of cadherin trans-interactions remain elusive.

Cadherin forms a relatively stable complex with α -catenin and **▶ β -catenin** (or plakoglobin) through its catenin-binding site in the cytoplasmic domain. β -catenin and plakoglobin are highly homologous proteins and belong to the armadillo family of molecules. β -catenin and plakoglobin bind directly to the catenin-binding site of cadherin and the amino terminal domain of α -catenin to establish a connection between these proteins. α -catenin has a weak homology with vinculin, and binds to various actin-binding proteins, such as vinculin, α -actinin, ZO-1 and actin itself. The interaction with actin-based cytoskeletons through α -catenin is essential for cadherins to function as strong cell adhesion molecules. Other armadillo family proteins interact directly with cadherins, including p120-catenin (p120ctn) and delta-catenin, but do so at the juxtamembrane domain of the cadherin that is situated at the N-terminal position of the catenin-binding site. These proteins are not required for the basic adhesive function of cadherins, but may be necessary for cadherin stability and other regulatory processes involved in cell adhesion.

Regulatory Mechanisms

The spatial and temporal expression of cadherin subtypes is strictly regulated during animal morphogenesis, during which adhesion specificity of cadherin subtypes might play an important role. It was reported recently that the transcription of E-cadherin is directly repressed by several transcription regulators, including snail, slug, ZEB-1/deltaEF1, ZEB-2/SIP-1 and E47. These E-cadherin repressors may be involved in **▶epithelial-mesenchymal transitions** during embryonic development and tumor progression.

Cadherin expression is also under post-transcriptional regulation. Cadherin molecules are synthesized as a precursor molecule with a pro sequence at the amino terminus. This sequence is removed prior to arrival at the plasma membrane, to form a mature cadherin. E-cadherin displaying a mutation that prevents the removal of this pro sequence fails to function as a cell adhesion molecule. It has also been reported that a specific cleavage, endocytosis and recycling mediated by various proteins, including presenilin-1, HAKAI, protein kinase C and Rac1, are involved in the regulation of mature cadherin expression. The physiological functions of such regulation have not been elucidated. On the other hand, the expression of β -catenin is post-transcriptionally regulated using the ubiquitin-proteasome system. This regulation is required, not for cadherin-mediated cell adhesion, but for the **▶Wnt signaling pathway**, which plays a

fundamental role in cell fate decisions and patterning of embryos. β -catenin that has not associated with cadherins is effectively degraded and absent from the cytoplasm and nucleus when a Wnt signal is absent. Wnt signals prevent this ubiquitination and degradation, causing the accumulation of β -catenin in the cytoplasm. Stabilized β -catenin finally translocates into the nucleus with the LEF/TCF transcription factor to activate LEF/TCF target genes.

Cadherin-mediated adhesion is regulated by intracellular signals in an 'inside out' direction. It has also been suggested that cadherins transduce signals in the reverse direction. Many **▶kinases**, including EGFR and src, and phosphatases, such as PTP μ and PTP κ , are localized at cadherin-based adhesion sites. Many reports indicate that major targets of kinases and phosphatases are the tyrosine residues of β -catenin and p120ctn. It is generally believed that phosphorylation of β -catenin and p120ctn affect cell adhesion activity. On the other hand, cadherin-mediated adhesion is involved in the activation of a signaling pathway through phosphorylation mechanisms. It has been reported that the AKT/MAPK pathway and the PI3-kinase/Akt cascade could be activated by cadherin-mediated cell adhesion. "The contact inhibition" of cell movement or cell growth is a well-accepted concept in the field of cell biology. Although the 'outside-in' signals mediated by cadherins are a good candidate for signals involved in this concept, little experimental evidence exists at the moment to associate cadherins with the contact inhibition of cell movement and growth.

Rho small GTPases are other strong candidates for regulators of cadherin-mediated cell adhesion. There is much evidence indicating that Rac1, cdc42 and RhoA types of GTPases affect cadherin function through the regulation of cadherin accumulation or cadherin-catenin complex formation. Furthermore, the activity of these small GTPases is modulated by cadherin-mediated cell adhesion or p120ctn. Since Rho small GTPases are involved in various dynamic cellular processes, such as cell motility, these proteins might regulate the balance between adhesive and motile cellular phenotypes.

Cadherin is essential for the formation of a highly specialized cell-cell junction, the **▶adherens junction** (AJ). The Ig-superfamily adhesion molecule nectin and its associated protein afadin have been shown to be components of the AJ. This adhesion system works cooperatively with the cadherin system to form AJs in certain cellular contexts. Three specialized cell-cell junctions, **▶tight junctions** (TJs), AJs and **▶desmosomes**, form junctional complexes at the apical edge of cell-cell boundaries in epithelial cells. The epithelial AJ is highly developed and is called the zonula adherens. Anti-E-cadherin antibodies not only disrupt AJs, but

also perturb the organization of TJs, indicating that cadherin-mediated cell adhesion regulates the formation of other junctions in epithelial cells. In fact, the dysfunction of the cadherin adhesion system, which occurs in many tumor cells, is correlated with the loss of junctional complexes and epithelial integrity. Recent studies have demonstrated that cadherins are also involved in regulation of synaptic junctions.

- Adhesion Molecules
- Adherens Functions
- Cell Adhesion
- Rho, Rac, Cdc42

References

1. Hirano S, Suzuki ST, Redices C (2003) The cadherin superfamily in neural development: Diversity, Function and interaction with other molecules. *Frontiers in Bioscience* 8:306–356
2. Nagafuchi A (2001) Molecular architecture of adherens junctions. *Curr Opin Cell Biol* 13:600–603
3. Anastasiadis PZ, Reynolds AB (2001) Regulation of Rho GTPases by p120-catenin. *Curr Opin Cell Biol* 13:604–610
4. Zhurinsky J, Shtutman Ben-Ze'ev A (2000) Plakoglobin and β -catenin: protein interactions, regulation and biological roles. *J Cell Sci* 113:3127–3139
5. Takeichi M (1977) Functional correlation between cell adhesive properties and some cell surface proteins. *J Cell Biol* 75:464–74

Caenorhabditis elegans

- Caenorhabditis elegans as a Model Organism for Functional Genomics
- C. elegans Genome, Comparative Sequencing

Caenorhabditis elegans as a Model Organism for Functional Genomics

ADAM ANTEBI
Baylor College of Medicine, Huffington Center on Aging, Houston, USA
aantebi@bcm.tmc.edu

Definition

► *C. elegans* is a small free-living soil nematode that has become an important model organism of intense

investigation. Complete descriptions of its development, nervous system and genome provide a unique platform to launch functional genomic studies of evolutionarily conserved genes, pathways and processes from worm to human.

Characteristics Features

Pioneered by Brenner as a genetic system in the 1960s, *C. elegans* was chosen in large part because of its small size (1 mm), rapid generation time (3.5 days) and ease of cultivation on an *E. coli* lawn. Moreover, stocks can be frozen in liquid nitrogen for years. Because of its clarity, cell constancy (959 somatic cells) and invariant development, the fate of every cell has been traced in entirety by Sulston, Horvitz and co-workers, forming a complete and precise framework for detailed biological exploration. Indeed, Sulston, Horvitz and Brenner were awarded the 2002 Nobel Prize in Medicine for their efforts.

Despite a simple anatomy, *C. elegans* has several major organ systems: a nervous system comprised of sensory neurons and motor circuits that control behavior, body muscle for movement, a pharynx for food ingestion and body fluid transport, an intestine for digestion and metabolic control, gonads and associated structures for reproduction, an excretory system and epidermis for fluid balance and barrier protection. Notably, the anatomy is described down to the ultrastructural level. In particular, with only 302 neurons, the entire hermaphrodite nervous system, including every synaptic connection and gap junction, has been reconstructed from serial electron micrographs, providing neural network models for behavioral studies.

Genetics

C. elegans has 5 pairs of autosomes and the sex (X) chromosome. Two X's specify the hermaphrodite sex, one X, the male. Because laboratory stocks are inbred, homozygous, and of single origin, powerful isogenic comparisons are possible. Self-fertilizing hermaphrodites produce broods of about 300 progeny. Self-fertilization greatly facilitates large-scale genetic screens, since a homozygous colony can be derived from a single worm, while mating with males permits markers to be crossed between strains. Basic forward genetic techniques such as chemical mutagenesis, complementation tests and gene mapping by morphological markers were established early on, followed by advanced techniques such as transgenesis, mosaic analysis, transposon insertion mutagenesis, balancer and deletion chromosomes, snp mapping, ►positional cloning and ►reverse genetics (below). Genes are described by a three letter name, usually an acronym for the mutant phenotype or function, followed by a

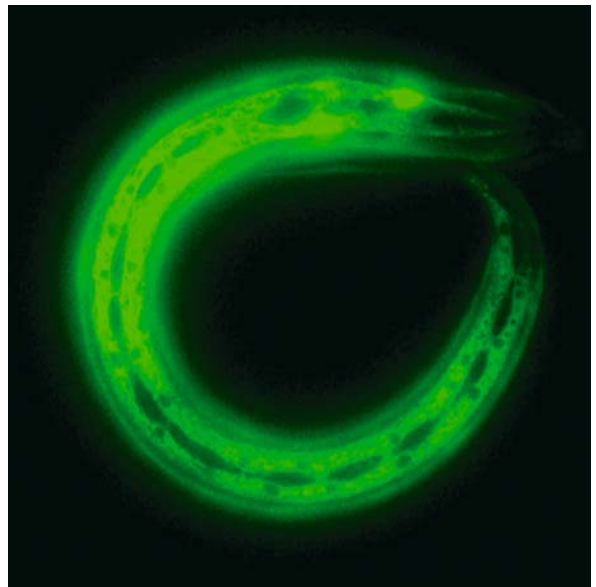
gene number and the allele in parenthesis (e.g. *unc-5* (*e53*), uncoordinated mutant gene 5, allele *e53*).

Genomics

C. elegans is a highly developed genomic organism. Extensive web-based bioinformatic development and central data archives are supported by Wormbase (www.wormbase.org) and nematode related links are found at <http://elegans.swmed.edu/>. The entire genome is cloned into overlapping cosmids and yeast artificial chromosomes (Sanger Center, UK) and ordered into a contiguous physical map. Because of its relatively small genome (97 megabases, 1/30 human), and dense physical and genetic map, *C. elegans* became the first metazoan whose genome was decoded (1), a collaborative effort between the Sanger Center (UK) and Washington University Genome Center (US). The catalog contains over 19427 predicted genes, about half that of human. Approximately, 36% of them are conserved with vertebrates.

cDNAs encoding most genes are available as expressed sequence tags (Yuji Kohara, National Institute of Genetics, Mishima) or full length cDNAs (Marc Vidal, Orfome project, Harvard). ▶**Expression profile** studies have been carried out using Affymetrix technology, cDNA or exon rich gene arrays, as well as serial analysis of gene expression. Such studies have included expression profiles describing temporal development and ▶**aging**, dauer formation, germline and sex-specific genes, muscle specific genes and pharyngeal organogenesis. One of the most promising approaches includes comparisons among the hundreds of array experiments, which suggest clusters of coexpressed genes visualized as “▶**expression mountains**” (2). ▶**Yeast two-hybrid** screens to identify protein-protein interactions were first employed systematically to study whole pathways such as vulval development, DNA damage and repair and germline expressed genes, an approach now being scaled up to build a comprehensive genome-wide ▶**interaction map** (3).

The ability to deduce gene function starting with a known gene, known as reverse genetics, has been achieved on a genome-wide scale by RNA mediated inhibition (RNAi, see below). In addition, the Gene Knockout Consortium systematically searches for deletions in a particular gene of interest by screening mutant pools of DNA by PCR. A library containing random insertion of transposons serves as another source of mutants as well as polymorphic markers for high-density mapping. Gene expression patterns are being documented systematically using *in situ* cDNA hybridization as well as promoter fusions to β -galactosidase and ▶**green fluorescent protein** (Fig. 1). Comparisons to the recently acquired sequence of *C. briggsae*, a related species, is proving



***Caenorhabditis elegans* as a Model Organism for Functional Genomics. Figure 1** A worm expressing green fluorescent protein (GFP) fused to DAF-9/cytochrome P450, which regulates *C. elegans* life span.

important for understanding evolution of genes and their regulatory regions.

Clinical Relevance

C. elegans is a mainstay model organism for fundamental pathway discovery, as well as a ▶**disease model** for analysis of clinically relevant gene function. Some areas of inquiry include organogenesis, muscle structure and physiology, neural development, ▶**synaptogenesis**, behavior, body patterning, ▶**developmental timing**, cell signaling, endocrine regulation, innate immunity, pathogenesis, sex determination and aging (4). Fundamental problems in developmental cell biology, such as cell-cycle, cell death, migration and fusion, cell polarity, spindle and chromosome dynamics, secretion, endo/exocytosis, extracellular matrix and cell-cell junctions have also been studied in *C. elegans*. Here we highlight a few prominent examples where *C. elegans* has led the way.

Programmed Cell Death

Elucidation of the *C. elegans* cell lineage shows that 131 cells are destined to undergo ▶**programmed cell death** at stereotyped times and positions. Notably, the first cell death mutants in any metazoan were isolated in the worm and genetic ▶**epistasis** experiments revealed a pathway that selects this fate. Included in this pathway are transcription factors that specify cell death (CES-1, CES-2, TRA-1), proteases that execute

cell death (e.g. ►CED-3/caspase) and their regulators (CED-4/Apaf1, EGL-1/BH3 and CED-9/Bcl2), as well as genes involved in cell corpse engulfment (CED-1/SREC scavenger receptor, CED-2/CrkII, CED-5/DOCK 180, CED-6/GULP adaptor, CED-7/ABCA1 transporter, CED-10/Rac1 GTPase, CED-12/Elmo1). Furthermore, cell death in the germline was shown to be sensitive to radiation damage and linked to CEP-1/p53 mediated pathways. Importantly, work in flies and mammalian systems demonstrated that the apoptotic machinery is evolutionarily conserved. Thus, mechanistic insights gleaned from work in the nematode have broadened our understanding of some cancers in which aberrant cells fail to undergo ►apoptosis, of autoimmunity where removal of autoreactive clones has gone awry and of neurodegenerative diseases and stroke where massive cell death leads to organismal demise. Finally such studies have revealed how programmed cell death plays an intimate part in development.

Neurobiology

Brenner identified numerous behavioral mutants that affected the nervous system in early genetic screens. Among them were a set of genes that guide the migrations of neuroblasts and growth cones, identifying the first true ►axon guidance cues in metazoans. ►UNC-6 is a homolog of the vertebrate netrin, which both attracts and repels axonal outgrowth in the floorplate. This secreted protein is proposed to polymerize in the extracellular matrix, providing graded cues for its receptors, UNC-40 and UNC-5. UNC-40 transmembrane receptor is a DCC (deleted in colorectal cancer) homolog, which alone mediates attraction to netrin sources, but together with coreceptor UNC-5 mediates repulsion. Other molecules involved in axon guidance and cell migration conserved across taxa include ephrins (EFN-1-4) and the eph receptor (VAB-1), semaphorins (SMP-1, SMP-2, MAB-20) and their receptors plexin (PLX-1) and neuropilin, as well as *slit* (SLT-1) and its cognate receptor round about (SAX-3). Molecules such as these are important for understanding neural development and regeneration.

Finally, numerous molecules involved in synaptogenesis (e.g. SYD-2/alpha liprin, UNC-104/KIF1 kinesin) and synaptic function (e.g. UNC-13/MUNC13, DGK-1/diacylglycerol kinase, SNT-1/synaptotagmin, SNB-1/synaptobrevin), have been intensely studied in the worm and have vertebrate counterparts performing similar functions. The use of green fluorescent protein (GFP) reporters to monitor neural structure and function, actually first pioneered in the worm, has facilitated the discovery of genes central to these pathways. Conceivably, mental retardation in humans

could be associated with mutational loss of such molecules.

RNAi and MicroRNAs

One of the most surprising findings first discovered in *C. elegans* is that double stranded RNA of a particular gene acts as a potent and specific inhibitor of gene function, known as RNA mediated inhibition (RNAi). In the worm, double stranded RNA is injected, soaked in or ingested and spreads systemically throughout the organism. Double stranded RNAs are cleaved into 21-mer fragments, called siRNAs, by the enzyme Dicer. Once incorporated into the RNA induced silencing complex, the antisense RNAi then serves as a template for mRNA degradation leading to loss of gene function. Remarkably, this technique works in mammalian cell culture and even in live mice. Surely, RNAi will prove to be an important therapeutic innovation and revolutionize high throughput analysis of gene function in higher organisms.

Indeed, RNAi has greatly accelerated functional genomics in the worm. Worms fed RNAi from 16757 genes gave rise to basic embryonic and larval phenotypes in 10.3% of the cases (5). Interestingly, such studies divulge unexpected aspects of chromosomal architecture, such as the relative dearth of essential genes on the X chromosome. Genome-wide screens for genes influencing ►fat metabolism, spindle dynamics, genome stability, ►cell migration, ►cell division, ►axon guidance, life span and other processes promise to illuminate numerous areas of biology. A particularly powerful approach combines ►microarray or yeast two-hybrid data with functional RNAi screens. In this way, many germline specific transcripts have been assigned functional roles in meiosis and chromosome dynamics. Similarly, yeast two-hybrid interactors of known DNA repair components have led to the discovery of new genes that influence the damage response.

Interestingly, the endogenous RNAi machinery in the worm is employed to translationally regulate a number of genes in the heterochronic pathway, which controls developmental timing. Initially discovered in the worm, ►lin-4 and let-7 ►microRNAs are 21–22 nucleotide long molecules that base pair imperfectly with the 3'UTRs of several genes, down-regulating their translation. Subsequently, it was discovered that scores of such genes exist not only in the worm but also in mammalian genomes. Remarkably, clear human orthologs to *lin-4* and *let-7* are among them. The function of these tiny RNAs is just beginning to unfold but already they are suggested to be involved in several diseases and syndromes. In addition, the role of the endogenous RNAi machinery in metazoan gene silencing is just beginning to emerge.

Aging and Life Span Regulation

A fascinating area of research spawned by findings in *C. elegans* is the biology of aging and longevity. Single gene mutations were found in *age-1*/PI3 kinase and

daf-2/insulin/IGF receptor mutants, that double or triple nematode life span, with little compromise in the fertility, vitality or metabolism of the organism. The molecular identity of these genes and similar mutants,

***Caenorhabditis elegans* as a Model Organism for Functional Genomics. Table 1** A handful of *C. elegans* orthologs of human disease loci

Human Syndrome	Human Gene	<i>C. elegans</i> predicted gene	<i>C. elegans</i> locus
Adenomatous polyposis of the colon	APC	K04G2.8	<i>apr-1</i>
Acute promyelocytic leukemia	PLZF	R11E3.6	<i>eor-1</i>
Alzheimer's	APP	C42D8.8	<i>apl-1</i>
Alzheimer Type 3, 4	Presenilin	F35H12.3	<i>sel-12</i>
Ataxia telangiectasia	ATM	T06E4.3	<i>atl-1</i>
Bardet-Biedl	BBS7	Y75B8A.12	<i>osm-12</i>
Cockayne's syndrome	CKN1	K07A1.12	<i>rba-2</i>
Darier-White Disease	SERCA	K11D9.2	<i>sca-1</i>
Duchenne's Muscular Dystrophy	Dystrophin	F15D3.1	<i>dys-1</i>
Friedrich ataxia	FRDA	F59G1.7	<i>frh-1</i>
	Frataxin		
Gaucher disease	Prosaposin	C28C12.7	<i>spp-10</i>
Hailey-Hailey	Ca ²⁺ -ATPase	ZK256.1	<i>pmr-1</i>
Juvenile Neuronal Ceroid Lipofuscinosis	CLN3	F07B10.1	<i>cln-3.1</i>
Kallmann	KAL1	K03D10.1	<i>kal-1</i>
Limb-girdle muscular dystrophy type 2B	Dysferlin	F43G9.6	<i>fer-1</i>
Miller Dieker Lissencephaly	LIS1	T03F6.5	<i>lis-1</i>
Mucopolipidosis 1	MCOLN1	R13A5.1	<i>cup-5</i>
	Mucolipin-1		
Niemann-Pick Type C1	NPC1	F02E8.6	<i>ncr-1</i>
		F09G8.4	<i>ncr-2</i>
Parkinson's disease juvenile form 2	Parkin	K08E3.7	<i>pdr-1</i>
Phosphatase and tensin homolog tumor suppressor	PTEN	T07A9.6	<i>daf-18</i>
Polycystic Kidney Disease	PKD1 polycystin 1	ZK945.9	<i>lov-1</i>
Polycystic Kidney Disease	PKD2 polycystin 2	Y73F8A.1	<i>pkd-2</i>
Rubinstein-Taybi	CBP	R10E11.1	<i>cbp-1</i>
Spinocerebellar ataxia 2	SCA2	D2045.1	<i>atx-2</i>
Usher syndrome 2a	USH2A	K08C7.3	<i>epi-1</i>
Werner's	WRN	F18C5.2	<i>wrn-1</i>
Zellweger syndrome 3	PXMP3	ZK809.7	<i>prx-2</i>

reveal insulin/IGF-signaling components as central regulators of longevity. While up-regulated insulin signaling is pro-aging, modest inhibition of the pathway leads to animals that are stress resistant and long lived. Longevity depends entirely on ►DAF-16, a FOXO forkhead transcription factor. Remarkably, recent work in fly and mouse indicate that reduced insulin/IGF signaling also extends life span. In addition, molecular dissection of this pathway in the worm has identified pivotal molecules, such as DAF-18/PTEN and DAF-16/FOXO, previously unknown to be involved in insulin signal transduction, illuminating disease states such as ►diabetes. Another endocrine pathway that influences life span is nuclear hormone receptor signaling. In particular, DAF-9/CYP450 produces a lipophilic hormone that regulates the life extending properties of nuclear receptor DAF-12 (Fig. 1). Currently over 50 genes modulate *C. elegans* life span and include components involved in mitochondrial function, eating behavior, thermotolerance and heat shock, checkpoint control, germline maturation, gene silencing and endocrine systems. Moreover, expression profiling and bioinformatic approaches to identify DAF-16 targets have uncovered genes that incrementally shorten or extend life span. Genome-wide screens using RNAi have also yielded a plethora of molecules, which by unknown mechanisms modestly influence life span. It will be interesting to understand which amongst them play a similar role in humans.

Prospectus

Future work will utilize high throughput functional analysis of genes by RNAi or other systematic methods, integrated with expression profiles and proteomic approaches, to elucidate gene interaction networks (3). From a clinical perspective, as new disease genes are unveiled, *C. elegans* will help elucidate their function as a simple genetic disease model. Already orthologs of several important disease loci have been identified and are being studied (Table 1). The primary challenge will be to clarify the relationship from gene, pathway and mutant phenotype to the relevant physiology in human.

►*Caenorhabditis elegans* Genome, Comparative Sequencing

►Molecular Aging Research

►Polycystic Kidney Disease, Autosomal Dominant

References

1. The *C. elegans* Sequencing, Consortium (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282:2012–2218
2. Stuart JM, Segal E, Koller D et al (2003) A gene-coexpression network for global discovery of conserved genetic modules. *Science* 302:249–255
3. Li S, Armstrong CM, Bertin N et al (2004) A map of the interactome network of the metazoan *C. elegans*. *Science* 303:540–543
4. Riddle D, Meyer B, Priess J et al (1997) *C. elegans* II. Cold Spring Harbor Laboratory Press
5. Kamath RS, Fraser AG, Dong Y et al (2003) Systematic functional analysis of the *C. elegans* genome using RNAi. *Nature* 421:231–237

Caged Compound

Definition

Caged compounds characterizes a photolabile inactive derivative of a biologically active molecule, from which the biomolecule is released by means of a flash of light.

►Uncaging and Photoconversion/Activation

►Functional Imaging

Caging Group

Definition

Caging group refers to a photolabile protecting group for the key pharmacophoric functionality of a biomolecule, rendering it inactive. Photorelease of the protecting group regenerates the active biomolecule.

►Uncaging and Photoconversion/Activation

Calcified

Definition

Calcified describes the property of an organ, or part of an organ, that is composed of, or contains, calcium carbonate.

►Bone and Cartilage

Calcium Release Channel

Definition

Release of Ca^{2+} ions from intracellular stores (sarcoplasmic reticulum and endoplasmic reticulum) can occur via

two classes of intracellular tetrameric calcium release channel (CRC) proteins, the inositol 1,4, 5 trisphosphate receptors (InsP3Rs) and the ryanodine receptors (RyRs). Multiple isoforms and subtypes of each CRC class display distinct but overlapping distributions within mammalian tissues. The activity of the CRC is modulated by interaction of a series of accessory proteins.

► [Peptidyl Prolyl Cis/Trans Isomerases](#)

Calnexin

Definition

Calnexin is a calcium-binding transmembrane protein in the endoplasmic reticulum, which binds to Glc1Man9GlcNAc2–Asn modifications within newly synthesized glycoproteins. It is involved in retention and quality control of glycoprotein biosynthesis.

► [Glycosylation of Proteins](#)

Calpain 10

Definition

Calpain 10 refers to the gene for a calcium activated protease-like cysteine protease family member, which has been implicated in type 2 diabetes in Hispanic Americans.

► [Diabetes Mellitus, Genetics](#)

Calreticulin

Definition

Calreticulin is an abundant, 46 kDa calcium-binding protein with a KDEL-(Lys-Asp-Glu-Leu) retention signal in the endoplasmic reticulum, which binds to Glc1Man9GlcNAc2–Asn modifications within newly synthesized glycoproteins. Calreticulin participates in Ca^{2+} storage and intracellular Ca^{2+} signaling, cell adhesion, gene expression, as well as the retention and quality control of glycoprotein biosynthesis.

► [Endoplasmic Reticulum](#)

► [Glycosylation of Proteins](#)

► [Signal Transduction: Integrin-Mediated Pathways](#)

Camel as a Model for Functional Genomics

DIRK SAERENS, MIREILLE PELLIS, CÉCILE VINCKE, ELS PARDON, KATJA CONRATH, SERGE MUYLDERMANS
Cellular and Molecular Immunology, Vlaams Inter-University Institute of Biotechnology, Free University of Brussels, Brussels, Belgium
svmuyld@vub.ac.be

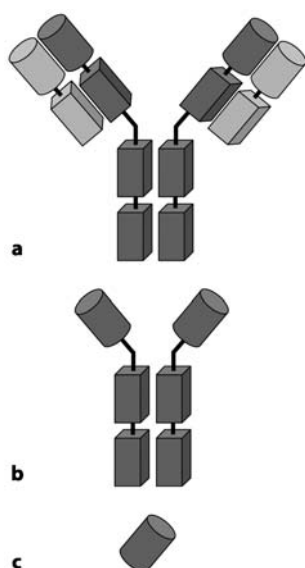
Definition

► [Immunoglobulins](#) raised against non-self molecules are key-players in the vertebrates' adaptive immune response. An elicited antibody associates with its cognate ► [antigen](#) with high ► [affinity](#) and specificity. The IgG are the most abundant antibodies in the peripheral blood and it has been well established that they are invariably composed of two heavy chains and two light chains (Fig. 1). The heavy and light chains contain four and two domains, respectively. The first domain of both the light and heavy chain is more variable in sequence (referred to as ► [VH](#) and ► [VL](#)) than the remaining domains and the paired VH and VL domains form the antigen-binding site. The discovery in *Camelidae* (dromedaries, camels and llamas) of naturally occurring IgG molecules consisting of heavy chains only and lacking light chains was totally unexpected (Fig. 1). These so-called ► [heavy-chain antibodies](#) (HCAbs) are functional in antigen binding (4). However since the homo-dimeric HCAbs lack a light chain and thus a VL, the antigen is bound by only one single domain, the ► [VHH](#) (VHH is the abbreviation of variable domain of the heavy chain of HCAbs). A VHH can easily be cloned and expressed in bacteria or yeasts to yield high amounts of soluble protein. Such a single-domain antibody format (VHH) has a MW of only 15 kD, which is at most half the size of the intact antigen-binding site of a conventional antibody, the ► [VH-VL pair](#). The VHHs possess a number of biophysical and biochemical properties that offer instrumental advantages over the antigen-binding fragments of conventional antibodies (i.e. the paired VH and VL) for their versatile usage in functional genomics.

Characteristics

General Characteristics and Properties of VHH

The VHHs possess a number of 'hallmark' amino acid substitutions that distinguish them from the conventional variable domain of the heavy chain of conventional antibodies (VH). These hallmarks are all located at the side of the domain that corresponds to the



Camel as a Model for Functional Genomics.

Figure 1 Schematic representation of a conventional antibody, a heavy-chain antibody of camelids and the VHH antibody fragment derived from such a heavy-chain antibody (from top to bottom). The heavy chains are in dark grey, the light chains are in light grey. The variable domains of the chains are represented as cylinders, the constant domains are represented by cubes.

VL-binding site of a VH (4). The nature of the VHH hallmark substitutions renders this area much more hydrophilic, probably leading to an enhanced solubility of the isolated domain.

The HCAb presence in serum of ►**camelids**, in conjunction with the possibility of immunizing these animals, allows for a straightforward cloning and selection procedure and for the production of an antigen-binding unit consisting of a single-domain only, the VHH (4). Hence, a technology is available to obtain *in vivo* matured antigen binders in a single domain format. To this end, a dromedary (or a llama) is first immunized with low amounts of antigen (50 µg), or a mixture of purified antigens. The peripheral blood ►**lymphocytes** are isolated to extract mRNA, which is used in an RT-PCR with VHH-specific primers to amplify the *VHH* genes. The *VHH* gene fragments are then ligated in a phage display vector and an 'immune' library is created from which potent antigen binders can be selected.

With this technology at hand, we have already selected over 100 binders against some 20–25 different antigens. The chosen *VHHs* are recloned in an expression vector behind a signal sequence to direct the VHH into the ►**periplasm** of bacteria and with a histidine tag used for purification by immobilised metal

affinity chromatography. After a further gel filtration we routinely obtain pure VHH at yields of 5–10 mg/l of culture in shaking flasks. This yield is on average 10× higher than most VH-VL constructs, and even higher yields have been reported when VHH expression was carried out in yeast under fermentation conditions.

The VHH domains are highly soluble, they can be brought to concentrations of 10–20 mg/ml without any sign of aggregation. The VHH tested so far are remarkably stable, e.g. they resist a one week incubation at 37°C and some of them can be heated to 90°C without loss of antigen binding activity (4). This feature might be correlated with the frequent occurrence of an extra disulfide bond between the antigen binding loops.

The VHH isolated from the immunised libraries are highly specific for the target antigen and do not cross-react with other, non-related antigens. The interaction between the VHH and its cognate antigen occurs with affinities in the nanomolar range, an affinity that is similar to that obtained for the best VH-VL pairs. The crystal structure of more than a dozen have been determined, creating a large structural database on how these binders interact structurally with their cognate antigen.

Surprisingly, many of the camelid VHHs directed against enzymes act as competitive inhibitors (4). This will definitely be an asset for provoking an immunomodulation (see further). This feature is probably related to the VHH's small size and especially to its small footprint on the associated antigen.

Although not of high priority in functional proteomics, the VHH possess two additional properties that might become important for subsequent applications. First, the low complexity of the antigen-binding site of VHHs consisting of only three loops (instead of 6 loops in VH-VL pairs) offers particular advantages over classical antibodies in designing novel inhibitors or cell receptor blockers. Secondly, the large sequence identity shared by VHH and human VH predicts a negligible ►**immunogenic** reaction against the VHH when administered to humans.

VHHs Against the Proteome

During the immunisation of a camelid, the B-cells producing antigen-specific HCAbs are enriched in the pool of antibody producing cells, and might even reach 1% of the population. Moreover, it is possible to immunise with a complex mixture of antigens, possibly even the whole proteome and to raise an immune response to each antigen or at least to the majority of the antigens. The cloned VHH repertoire of such an immunised camelid might be an ideal source to retrieve *in vivo* matured, high-affinity binders against any antigen of the proteome. This approach is quite feasible since the cloning of 10⁸ individual *VHHs* amplified from 10⁸–10⁹

B-cells should yield 10^8 different VHH sequences, each of which might be directed against a different epitope. It is unlikely that this might be achieved after immunising a mouse or rabbit because the *VH* and *VL* gene fragments of the 10^8 conventional antibodies will be amplified separately and assembled randomly in the subsequent step to arrive at 10^{16} possible *VH-VL* constructs. Only a fraction of this can be transformed in bacteria as a result of transformation efficiency limitations. In addition, starting from a large naïve library might turn out to be inefficient as well, because binders retrieved from such libraries are frequently of too low an affinity for subsequent applications.

Camelid VHH as Probe in Protein ► Biochips

Antibody-based assays are among the most commonly used diagnostic tools. In addition, the use of antibodies is ever expanding to detect and purify bio-molecules in complex samples. The introduction of phage-displayed antibodies and antibody engineering resulted in greater flexibility in assay and probe design. The generation of a panel of highly stable, antigen-specific binders that can be directionally immobilised at a high density on the sensor layer opens great opportunities in novel diagnostic tools and proteome research (1).

The incorporation of the camel single domain antibody fragments, in conjunction with their small size, high specificity and high k_{on} rates for the antigen allows the detection of low analyte concentrations (in the range of ng/ml). The robust behaviour of the camel VHH probe increases the lifetime of the biosensor.

Coupling the various antigen binders, obtained from our proteome immunisation technology, side by side in a micro-array should in principle allow the simultaneous detection of multiple antigens from the proteome present in various cell-types. The outcome of this protein profiling provides the basic information to identify new targets for therapy to be developed in the near future.

Camelid VHH for Immuno-targeting and Immuno-modulation

After unveiling a new gene it will be necessary to find out the function of its encoded product. Determination of the cell-types that express this gene and the localisation of the gene product and identifying its interacting partners are crucial steps in the target validation process.

The nucleotide-based micro-array is a perfect tool to investigate the gene expression profile, however, it might be substituted by antibody-based micro-arrays (see above).

For years, labelled antibodies have been used to determine the localisation of the antigen. The strictly monomeric nature of the VHH, and its free C-end located at the opposite side of the paratope within

the domain make it easy to tether tags (e.g. green fluorescent protein, hexahistidine, or another VHH to generate bispecific constructs...) by genetic fusion. These immuno-conjugates are well expressed, soluble and retain full antigen binding activity rendering them suitable for immuno-targeting. The VHH coupled by chemical means to paramagnetic beads and mixed with a protein cell extract should enable the capturing of the antigen together with its interacting partners. With the help of a magnet, the interacting protein complex can then be isolated for further identification.

In addition, the unique preference of a VHH to interact with cavities on the surface of the antigen and the likelihood that the catalytic sites of enzymes are within such cavities often make VHHs competitive enzyme inhibitors. Therefore, besides targeting the antigen, the introduction of a VHH inside cells might decrease and even knockout the function of the target. Such an immuno-modulation by the ectopic expression of *VHH* genes might become a powerful technique to manipulate cellular metabolism, signal transduction or pathogen infectivity. Proof of the principle has been given by the intracellular expression and proper folding of VHHs in eukaryotic cells (intracellular immunization) leading to interference with biological processes that take place inside such cells (3).

Camelid VHH as Crystallisation Aids

Now that most genes of the genome have been identified, there will be a desire to determine the structure of all gene products by crystallography or NMR. Both of these techniques require highly concentrated samples of pure protein. Keeping these proteins soluble at high concentration and the crystallisation of flexible proteins or membrane proteins for high-resolution structural studies remain challenging because of the non-uniformity in folding and the amphipathic surface of the molecules, respectively. Embedded in the membrane bilayer, the surfaces of the protein in contact with the acyl chains of the phospholipids are hydrophobic, whereas the polar surfaces of the protein are exposed to the polar head groups of the lipids and to the aqueous environment. Because many membrane proteins contain exposed hydrophobic areas, a strategy to increase the probability of getting well-ordered crystals consists in covering the hydrophobic areas by specifically bound antibodies. This association might increase the solubility of the antigen, and fix the flexible proteins in a suitable conformation for crystallisation (2).

Monovalent recombinant antibody fragments such as *VH-VL* pairs and especially the rigid VHH should be suitable crystallisation aids. The VHHs can be first used to purify the antigens by affinity chromatography and can later be added in a 1:1 stoichiometry to

stabilize the antigen by forming a tight complex, in which hydrophobic patches of the antigen are shielded or flexible protein subdomains are stabilized. Several VHH-antigen complexes, including unstable antigens, have already been crystallised and their structure solved by molecular replacements.

Clinical Relevance

Microarrays or biochips provide an efficient way to obtain a global molecular perspective of human disease. DNA micro-arrays have been used extensively to generate transcriptional fingerprints of cancer. More recently, high-throughput techniques to study proteins have also been developed. Protein-based micro-arrays containing labelled antigens or antibodies offer the ability to study protein-protein interactions, identify biomarkers and study the humoral response to cancer.

However, the role of a VHH could go well beyond the target validation step, since many *in vivo* diagnostic and therapeutic molecules are derived from antibodies. The high degree of sequence identity between a VHH and a human VH and the minimal immunogenicity of a VHH suggest that the selected VHH used for target validation could easily be readapted into a diagnostic or therapeutic compound. Moreover, the VHH might create special opportunities for fast and reliable *in vivo* diagnosis at an early stage of the disease. The ideal cancer-imaging agent should deliver an amount of label sufficient to detect the smallest metastases against a low level of non-specific background signal. The minimal sized VHH based imaging agents would represent a valuable imaging tool. The rapid distribution of VHH over the whole organism, the superior potential in tissue penetration of minimal-sized VHH, combined with high affinity target binding and fast clearance of excess VHH from the circulation, represent the ideal basis for imaging purposes.

The employment of single domain antibodies as crystallisation aids will boost the process of crystallisation and structure determination of flexible and membrane proteins. The structural information about these antigens will be a valuable resource in drug design.

References

1. Borrebaeck CAK (2000) Antibodies in diagnostics – from immunoassay to protein chips. *Immunology Today* 21 (8):379–382
2. Hunte C, Michel H (2002) Crystallisation of membrane proteins mediated by antibody fragments. *Current Opinion in Structural Biology* 12:503–508
3. Jobling SA, Jarman C, Teh MM et al (2003) Immunomodulation of enzyme function in plants by single-domain antibody fragments. *Nature Biotechnology* 21:77–80

4. Nguyen VK, Desmyter A, Muyldermans S (2001) Functional Heavy-Chain antibodies in Camelidae. *Advances in Immunology* 79:261–296

Camelid

Definition

Camelid refers to species of the family of *Camelidae*, suborder of *Tylopoda*, order of *Artiodactyla*. The existing species are dromedary, camel, and llamas.

► [Camel as a Model for Functional Genomics](#)

cAMP

Definition

Cyclic adenosine monophosphate (cAMP) is a cellular second messenger that is involved in diverse signal transduction pathways, typically originating from activation of G protein coupled receptors.

► [Polycystic Kidney Disease, Autosomal Dominant](#)

Cancer

Definition

Cancer is the name given to a disease characterized by the accumulation of cells, causing the formation of a tumor, which is always fatal when untreated and is therefore called malignant. Cancer may originate from almost every tissue in the body; it grows into neighboring tissues and organs and in half of the cases forms distant metastases. Benign tumors differ from cancers because they remain confined to the tissue of origin and are easily removed. The development of tumors involves frequent mutations in genes termed "oncogenes and tumorsuppressor genes". Cells in highly malignant human tumors often bear five to seven mutations. Many mutations afflict genes that regulate the cell cycle directly or have activities upon cell division. Thus, tumorigenic cells become insensitive to growth factors, replicate their DNA content more than once per cell cycle, and typically have more

than one centrosome per cell. The cancer cells have the ability to escape from destruction by the immune system.

► **Cancer Invasion and Metastasis: Cellular, Molecular and Clinical Aspects**

► **Cell Division**

► **Protein Interaction-Phage Display**

► **Tumor Suppressor Genes**

Cancer Invasion and Metastasis: Cellular, Molecular and Clinical Aspects

MARIA JOSÉ OLIVEIRA^{1,2}, MARC MAREEL¹, ANCY LEROY¹

¹Laboratory of Experimental Cancerology, Ghent University Hospital, Ghent, Belgium

²IPATIMUP, Institute of Molecular Pathology and Immunology, University of Porto, Porto, Portugal
ancy.leroy@ugent.be

Definition

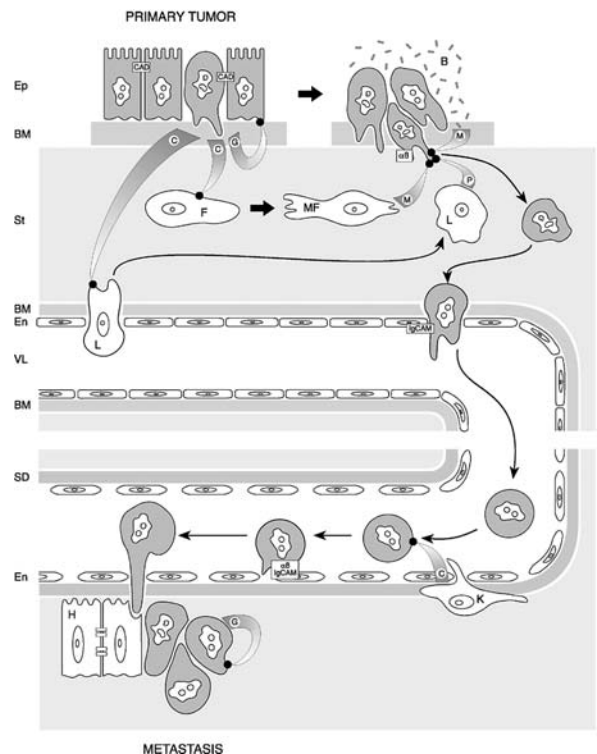
Invasion and Metastasis as Markers of ► **Cancer Malignancy**

Cancer is a disease of those cellular activities that are essential for the development and the maintenance of the organism, namely growth, ► **differentiation**, tissue integrity and ectopic ► **apoptosis** (Fig. 1) (3). Loss of growth control leads to accumulation of cells producing a tumour. Loss of tissue integrity and ectopic survival marks the difference between benign non-invasive tumours and malignant invasive ones. Loss of differentiation is found in invasive tumours and in their non-invasive precursors. Invasion is characterized by the presence, i.e. survival and eventual growth, of cells in tissues different from their tissue of origin, implying their penetration through the ► **basement membrane** tissue barrier, e.g. epithelial cells in the stroma. Metastasis is characterized by the growth of cells in organs distant from the organ of origin (secondary site), e.g. breast cancer cells in the bone and colon cancer cells in the liver (Fig. 1). To form metastases, cancer cells need to be transported by the lymph and blood circulation.

Characteristics

Bioassays for Invasion and Metastasis

Various *in vivo* and *in vitro* assays have been used to analyse cellular and molecular mechanisms of invasion and metastasis. Some assays aim at a close mimic of the



Cancer Invasion and Metastasis: Cellular, Molecular and Clinical Aspects.

Figure 1 The multi-step invasion process of metastasis. Consecutive steps are: invasion from the primary tumour through the basement membrane into the surrounding tissue; intravasation; transport through the blood circulation; extravasation; survival and growth at a secondary site. Molecular crosstalks take place between cancer cells, host cells, extracellular matrix elements and environmental factors. Cancer cells are represented in grey and host cells in white. Thick arrows indicate transformation. Thin arrows indicate direction of movement. Gradients of chemokines (C), growth factors (G), pro-invasive and motility factors (M), proteases (P) point at their respective receptors at the cell surface (black dots). $\alpha\beta$, integrins; B, bacteria; BM, basement membrane; CAD, Ca^{2+} -dependent cadherins; En, endothelial cells; Ep, epithelium; F, fibroblast; H, hepatocyte; IgCAM, Ca^{2+} -independent immunoglobulin cell adhesion molecules; K, Kupffer cell; L, leukocyte; MF, myofibroblast; SD, space of Disse; St, stroma; VL, vascular lumen.

natural situation; others are limited to one invasion- or metastasis-associated cellular ► **phenotype**. Invasion is frequently accompanied by the transition from an epithelial differentiation towards a mesenchymal one, a phenomenon called ► **EMT** (4). In brief, when cells pass through a bare filter in a two-compartment chamber *in vitro*, their activity is called ► **migration**; when the filter is covered by a layer of ► **extracellular matrix**, it is called invasion. When cells produce

colonies in the lungs, liver or bone after subcutaneous or intravenous injection into syngeneic or immunosuppressed mice, they are called metastatic. ▶ **Transgenic** animals, comprising mice, fish, flies and worms are valuable tools in the analysis of invasion and metastasis.

Invasion and Metastasis is not Unique to Cancer

Invasion and metastasis, as defined above, are also observed in diseases caused by microorganisms, during embryonic development and in adults when leukocytes leave the bone marrow and home to various organs (3). Microorganisms invade the cells lining the alimentary and the respiratory tracts, where they cause enteritis and bronchitis; they eventually metastasise to distant organs such as the heart and the brain, where they cause endocarditis and meningitis.

Molecular Interactions

▶ Tumour Promoter and ▶ Suppressor Genes

Abnormal activation and inactivation of promoter and suppressor genes respectively are the cause of cancer (3). Loss of genes that are implicated in DNA repair will increase the risk of alteration in all genes that are sensitive to mutation and so influence the acquisition of invasion and metastasis in a non specific manner. Examples of invasion- and metastasis-specific genes altered in human cancer, as well as the criteria by which they were recognized are shown in Table 1A and 1B. Only for a minority of these genes do we understand the mechanisms by which the proteins they encode determine the invasive and metastatic phenotype. For example, the kinase *MKK4* is a metastasis suppressor because it restricts growth of cancer cells at the secondary site. Sometimes the abnormal expressions of promoter and suppressor gene products make the cells responsive to pro-invasive factors present in the tumour micro-environment (2). For example, a background of Ras activation makes the cells invasive upon loss of regulators of cell polarity or upon contact with cytokines such as TGF- β . The same pro-invasive molecules may emerge through various genomic alterations. For example, enhanced transcription of the pro-invasive matrix metalloprotease MMP7 may result from oncogenic mutation of ▶ **β -catenin** or of the ▶ **adenomatosis polyposis coli (APC)** gene. Several genomic changes are usually needed for the development of an individual tumour and its acquisition of invasiveness and metastasis. DNA micro array-based gene expression profiling of cancers as compared to their precursors reveals hundreds of differences in mRNAs and proteins. This multiplicity may be explained by the fact that the product of the oncogene controls the expression of many other genes. For example, the enhancer of zeste homolog 2 (EZH2) stimulates invasion and metastasis because it changes

the balance between suppressors and promoters through modulation of multiple genes. It is a matter of debate whether metastasis genes are already activated or inactivated in earlier stages of tumour development or only in later stages.

Regulatory Mechanisms Proteomes of Invasive Cells

Several ▶ **protein families** play a role in invasion and metastasis. Some of these proteins are the products of promoter or suppressor genes (Table 1A, B), others are targets of the proteins encoded by these genes. For example, when the tumour suppressor ▶ **Von Hippel Lindau (VHL)** protein is mutated, the DNA-binding hypoxia inducing transcription factor (HIF) fails to be degraded; this leads to the constitutive activation of the CXCR4 gene encoding a ▶ **receptor** for a motility factor that attracts metastatic cells. The members of these protein families emerged as key players in distinct activities of invasive cells. Cadherins and Ig-CAMs are crucial for homotypic and heterotypic ▶ **cell-cell adhesion**, respectively counteracting primary invasion and stimulating metastasis (1); their activity should not, however, be solely considered as mechanistic, since they participate extensively in ▶ **signal transduction** through their linkage with cytoplasmic components. Moreover, cadherins may undergo ▶ **ectodomain shedding** and their soluble fragments may regulate invasion. Integrin receptors and their extracellular matrix protein ligands regulate cell-matrix adhesion and de-adhesion mechanistically as well as inside-out and outside-in signalling, i.e. from the cell to the matrix and vice versa (5). This interaction has a dual function, arresting cells in the matrix and assisting migration. Motility factors and their receptors stimulate the locomotory machinery of the cancer cells and guide them through the invasion pathways (5). Proteases break down the extracellular matrix, paving the way for invaders. They may also release pro-invasive factors from the cell surface as well as from the extracellular matrix. Proteases act in cascades, with inactive precursors being activated by other proteases and active proteases being neutralized by specific inhibitors. Most normal cells are ▶ **anchorage dependent** for survival; when they are released from their substrate or when they arrive in an alien environment they go into ▶ **apoptosis**. Invasive cancer cells escape from apoptosis through activation of growth and survival pathways, as well as inactivation of death pathways. The role of the abovementioned proteins is not limited to single categories of cellular activities as they participate in integrated invasion programs. They all form multi-protein complexes, most of which establish networks together with other complexes. Remarkably, the same protein, e.g. β -catenin, may serve as an invasion-suppressor in one complex, e.g. the E-cadherin/catenin complex, and as

Cancer Invasion and Metastasis: Cellular, Molecular and Clinical Aspects. Table 1 Promoter Genes Implicated in Invasion and Metastasis

Gene	Locus	Protein/Function	Postulates					I/M
Acronym			(A1)	(A2)	(B)	(C)	(D)	
A. Promoter Genes Implicated in Invasion and Metastasis								
CTNNB1	3p22–p21.3	Regulator of transcription factors	yes	no	no	yes	yes	I
ERBB2 (HER-2;Neu)	17q21.1	Receptor tyrosine kinase	yes	no	yes	yes	yes	M
EZH2	7q35–q36	Regulator of transcription factors	yes	yes	no	yes	no	I & M
FGF3	11q13	Heparin-binding growth factor	yes	no	yes	yes	no	I & M
KRAS2 HRAS	12p12.1–11p15.5	Small GTPases	yes	yes	no	yes	yes	I
MADH2	18q21	Transcription factor	yes	no	no	yes	no	M
MET (HGFR)	7q31	Receptor tyrosine kinase	yes	yes	yes	yes	yes	I & M
MTA1	14q32.3	Regulator of transcription	yes	no	yes	yes	no	I & M
MYC	8q24.12–q24.13	HLH transcription factor	yes	no	no	yes	no	I
PI3Kγ	7q22.3	Protein and lipid tyrosine kinase	no	no	no	yes	yes	I
S100A4 (mts1)	1q21	Ca ²⁺ -binding	yes	no	yes	yes	yes	M
SNAI1	20q13.1–q13.2	Zinc-finger transcription factor	yes	no	yes	yes	no	I
SRC	20q12–q13	Non-receptor tyrosine kinase	yes	yes	yes	yes	yes	I
TIAM1	21q22.1	Rho GEF	no	no	no	yes	yes	I
B. Suppressor Genes Implicated in Invasion and Metastasis								
BRMS1	11q13.1–q13.2	Regulator of transcription	no	no	yes	yes	no	M
CDH1	16q22.1	Ca ²⁺ -dependent cadherins	yes	yes	yes	yes	yes	I & M
CRSP3	6q22.33–q24.1	Regulator of transcription	no	no	yes	yes	no	M
KAI1	11p11.2	TCR/CD3 Tetraspanin coreceptor	yes	no	yes	yes	no	M
KISS1	1q32	Cytoskeletal reorganiser	yes	no	yes	yes	no	M
MADH4	18q21.1	Transcription factor	yes	no	no	yes	yes	I
MAP2K4 (MKK4)	17p11.2	Serine/threonine kinase	no	no	no	yes	no	M
NME1 (NM23)	17q21.3	Nucleoside diphosphate kinase	yes	no	yes	yes	no	M
PTEN	10q23.3	Protein and lipid phosphatase	yes	yes	yes	yes	yes	I
PBP (RKIP)	12q24.23	Raf kinase inhibitor and negative regulator of the MAPK pathway	yes	no	yes	yes	no	M
Rb1	13q14.2	Regulator of transcription	yes	yes	no	yes	yes	I
TIMP2	17q25	Inhibitor of proteases	yes	no	yes	yes	no	I

Cancer Invasion and Metastasis: Cellular, Molecular and Clinical Aspects. Table 1 Promoter Genes Implicated in Invasion and Metastasis (Continued)

Gene Acronym	Locus	Protein/Function	Postulates					I/M
			(A1)	(A2)	(B)	(C)	(D)	
<i>TP53</i>	17p13.1	Regulator of transcription; growth arrest and apoptosis	yes	yes	yes	yes	yes	I
<i>TXNIP (VDUP)</i>	1q21.2	Thioredoxin-binding protein	no	no	yes	yes	no	M
<i>VHL</i>	3p26-p25	Proteasomal degradation; regulator of transcription	yes	yes	yes	yes	no	I

(A1) The genomic change is present in invasive (metastatic) lesions or their precursors and not in benign lesions or normal precursor tissues of a given tumour type; (A2) more rigorously, invasive and benign areas are obtained by micro-dissection and quantitative PCR from the same tumour. (B) The genomic change marks the difference between invasive and noninvasive variants of a cell line derived from the same tissue (same genetic background as evidenced by DNA profiling). (C) Transfection of cells with sense or anti-sense cDNA or siRNA causes conversion of the invasive (metastatic) phenotype. (D) Transgenic animals expressing or silencing the gene of interest in a specific tissue develop tumours with different invasive (metastatic) behaviour. To distinguish between invasion genes and metastasis genes, models are needed showing conversion from non-invasive towards invasive, non metastatic and from invasive, non metastatic to metastatic. I & M, the invasive and metastatic phenotypes conveyed by the genomic alteration have not been distinguished from one another; I, invasion but not metastasis; M, metastasis is conveyed upon an invasion-positive tumour. Gene names are according to UCL/HGNC/HUGO Human gene nomenclature databases.

BRMS-1, breast cancer metastasis-suppressor 1; *CDH1*, cadherin 1; *CRSP3*, Cofactor required for Sp1 transcriptional activation subunit 3; *CTNNB1*, beta-catenin; *ERBB2*, avian erythroblastic leukaemia viral homologue 2; *EZH2*, enhancer of zeste homologue protein 2; *FGF3*, fibroblast growth factor 3; GEF, Rho guanyl-nucleotide exchange factor; *HER-2*, human epidermal growth factor receptor 2; *HGFR*, hepatocyte growth factor receptor; HLH, helix-loop-helix; *HRAS*, v-HA-RAS Harvey rat sarcoma viral oncogene homologue; *KRAS2*, v-KI-RAS2 Kirsten rat sarcoma 2 viral oncogene homologue; *KAI1*, "kang ai" 1 (Chinese word for anticancer); *MADH2*, mothers against decapentaplegic homologue 2; *MADH4*, mothers against decapentaplegic homologue 4; *MAP2K4* or *MKK4*, mitogen-activated protein kinase kinase 4; *MTA1*, metastasis-associated protein 1; *mts1*, metastasin 1; *MYC*, Myelocytomatosis virus protein; *neu*, transformed in ethylnitrosourea (ENU)-induced rat neuroblastomas; *NM23*, nonmetastatic protein 23; *NME1*, protein expressed in nonmetastatic cells 1; *PBP*, prostatic binding protein; *PI3K γ* , Phosphatidylinositol-4,5-bisphosphate 3-kinase gamma isoform; *PTEN*, phosphatase and tensin homologue deleted on chromosome 10; *RKIP*, Raf kinase inhibitor protein; *S100A4*, S100 calcium-binding protein A4; *SNAI1*, snail homologue 1; *SRC*, Rous sarcoma virus protein; *TIAM1*, T-lymphoma invasion and metastasis inducing protein 1; *TIMP2*, Tissue inhibitor of metalloproteinase 2; *TP53*, tumour protein p53; *TXNIP*, thioredoxin interacting protein; *VDUP*, Vitamin D3-upregulated gene 1; *VHL*, Von Hippel-Lindau.

an invasion-promoter in another complex, e.g. the APC/GSK-3 β complex. The networks constitute the positive and negative invasion signalling pathways (3). Such pathways are understood in terms of invasion factors binding to specific receptors that mediate transduction of the signal to stimulate or inhibit the cellular activities implicated in invasion and metastasis. Pivotal in the signal transduction are [small GTPases](#) of the Rho family that modulate assembly/disassembly of the [actin cytoskeleton](#) and as a consequence cell migration. Note that different types of tumours may use different invasion pathways and that even the same tumour may change invasion pathways following alterations of the context. This is realized in a versatile manner as the participant proteins can be modulated at multiple levels affecting their structure and their localization; examples are transcriptional modulation, proteolysis, [phosphorylation](#) and dephosphorylation and glycosylation. Taken together, this profusion of

regulatory systems permits cancer cells to invade following various strategies.

The Tumour Ecosystem

In the first issue of the Lancet (1889), Paget launched the "seed" and "soil" hypothesis, which is still valid today: the development of metastases results from the interaction between the cancer cells (the "seed") and the organ environment (the "soil"). Malignant tumours are not just a collection of cancer cells. In most of them, about half of the population consists of host cells that moved into the tumour from immediate or distant sites and all of which participate in invasion (2). Remarkably, these host cells show characteristics of invasive cells. When cancer cells lack oxygen, they produce vascular endothelial growth factors (VEFG) that attract blood- and lymph vessels to the tumour, a process called angiogenesis. Cancer cells produce cytokines, e.g. TGF- β , that cause the transition from stromal

fibroblasts into myofibroblasts, producing pro-invasive factors such as SF/HGF and tenascin C. Cancer cells release ►[chemokines](#) that attract white blood cells that release proteases, so paving the way for cancer cell invasion through a passive countercurrent mechanism. Considering this mutual cross talk between cancer cells and host cells, one may rightly ask the question who is invading whom (Fig. 1).

It is still not yet clear whether cancer cells spread to most organs and selectively grow as a consequence of the molecular conversation with the host or spread selectively to organs as a consequence of molecular attractants recognizing markers on the cancer cell and grow there. It has recently been found that not only host cells but also bacteria may participate in the production of pro-invasive factors. Taken together the recent observations confirm Paget's hypothesis and provide a molecular basis for it.

Host Defence Against Invasion and Metastasis

Metastasis is an inefficient process, as only a few of the cells that enter the circulation will ultimately form metastases. Such cells may remain dormant for several years and then form "late" metastases, well documented in the bone for breast cancer and in the liver for ocular melanoma. It is not clear whether inflammatory or immunological reactions target specifically metastatic cells and in this way are responsible for tumour ►[dormancy](#). Escape of cancer cells expressing tumour specific antigens from the defence mechanism has been shown through loss of ►[HLA](#) class I molecules that are needed for recognition by the immune effector cells or through a cancer cell-mediated anergic state of immune effector cells.

Medical Applications

Invasion and metastasis are the major determinants of tumour ►[prognosis](#). For most cancers, prognosis is based on staging according to depth of invasion and on grading according to degree of differentiation. It is the aim of adjuvant chemotherapy to destroy small nests of metastatic cells that remain after primary surgery and radiotherapy. The actual staging and grading is, however, not refined enough to spot those patients that bear such nests and are, therefore, good candidates for adjuvant therapy. The consequence is that many patients are treated with toxic agents without any benefit, for the simple reason that there are no cancer cells left. One of the expectations of molecular oncology is that the presence of metastasis will be predicted with high probability from the characterization of individual primary cancers, so that toxic treatment can be avoided. Moreover, a better molecular understanding of the mechanisms of invasion and metastasis might lead to new, more efficient and less

toxic treatment strategies. Here, the major obstacle to be overcome is the cancer cells' capability to reprogram their mechanisms of invasion and metastasis.

Acknowledgements

The authors acknowledge G. De Bruyne for preparing the figure. M. J. Oliveira is supported by the Portuguese Foundation for Science and Technology (FCT) BD15980, A. Leroy is a postdoctoral fellow from the FWO, Vlaanderen, Belgium.

►Genomic Analysis of Single Disseminated Cancer Cells

References

1. Birchmeier W, Behrens J (1994) Cadherin expression in carcinomas: role in the formation of cell junctions and the prevention of invasiveness. *Biochim Biophys Acta* 1198:11–26
2. Liotta LA, Kohn EC (2001) The microenvironment of the tumour-host interface. *Nature* 411:375–379
3. Mareel M, Leroy A (2003) Clinical, cellular, and molecular aspects of cancer invasion. *Physiol Rev* 83:337–376
4. Thiery JP (2002) Epithelial-mesenchymal transitions in tumour progression. *Nat Rev* 2:442–454
5. Trusolino L, Bertotti A, Comoglio PM (2001) A signalling adapter function for $\alpha 6 \beta 4$ integrin in the control of HGF-dependent invasive growth. *Cell* 107:643–654

Candidate Gene

Definition

A candidate gene is a gene that is believed to be a disease relevant gene, and/or a gene responsible for the susceptibility for the disease.

- [Atopy Genetics](#)
- [COPD and Asthma Genetics](#)
- [Diabetes Mellitus, Genetics](#)
- [Genetic Predisposition to Multiple Sclerosis](#)
- [Manic Depression](#)
- [Mendelian Forms of Human Hypertension and Mechanisms of Disease](#)

Candidate Gene Approach

Definition

In the candidate-gene approach, genes with known or proposed function, with the potential to influence the disease phenotype, are investigated for a direct role in disease.

- ▶ Common (Multifactorial) Diseases
- ▶ Large-Scale ENU Mutagenesis in Mice

5' Cap

Definition

5' cap refers to a 7-methyl guanosine modification at the 5' end of mRNAs.

- ▶ RNA Stability

Cap/Cap Structure

Definition

Cap/Cap structure refers to the sequence of methylated terminal oligonucleotides. The modification is found at the 5' end of all eukaryotic mRNAs. It is added post-transcriptionally, and consists of a 7-methyl guanosine linked by a 5'-5' triphosphate bridge, with the first nucleotide encoded in the DNA.

- ▶ Polyadenylation
- ▶ RNA Capping

Capacitation

Definition

Capacitation describes the final maturation of sperm in the female reproductive tract that enables sperm to bind to eggs and undergo the acrosome reaction.

- ▶ Mammalian Fertilization

Cap-Dependent /Cap-Independent Translation

Definition

Cap-dependent/Cap-independent translation refers to protein synthesis using only capped mRNAs or uncapped as templates, respectively.

- ▶ RNA Capping

Capillary Electrophoresis

Definition

Capillary electrophoresis is a method for electrophoretic separation of biomolecules, or other large molecules, in (fused silica) capillaries with an inner diameter of 50–100 μm and 30–80 cm in length. The capillaries are filled with gel-material and electrophoresis buffer. Molecules such as DNA are separated by high voltage (5–30 kV). The use of a high voltage gradient (200–1000 V/cm) results in significantly faster electrophoresis than conventional gel electrophoresis in flat gels (about five times shorter). This is possible due to rapid heat dissipation of the capillary.

- ▶ SNP Detection and Mass Spectrometry

Cap-Independent Translational Control

MARTIN HOLČÍK

Apoptosis Research Center, Children's Hospital of Eastern Ontario Research Institute, Ottawa, Canada
martin@mgcheo.med.uottawa.ca

Synonyms

Control of protein synthesis

Definition

Translational control is defined as a change in the rate or efficiency of translation of one or more [mRNAs](#) per unit of time (1). The rate of protein synthesis is determined by (i) the amount of mRNAs present, (ii) the activity of the protein synthesis machinery, (iii) the abundance of [ribosomes](#) and (iv) the rate of peptide chain elongation. Therefore, changes in any of these factors may limit the rate of protein synthesis and thus regulate overall protein synthesis.

Characteristics

Translational control is the final step in a complex network of regulatory processes involved in the control of gene expression. In the process of translation the genetic information carried in the molecule of mRNA is converted into protein. This process takes place in three stages – initiation, elongation and termination. Although all three stages are targets of regulatory mechanisms, in most cases it is the initiation of

translation that is the rate-limiting step. This is probably due to the fact that it is more efficient to control the onset of a biological process rather than to interrupt it later and deal with the consequences of aberrant protein synthesis (1). Since translation is the last step in the flow of the genetic information, regulation at this level also allows for an immediate and rapid response to changes in physiological conditions without the involvement of mRNA synthesis and transport.

Cap-Dependent Initiation of Translation

The initiation of translation in a eukaryotic cell involves the recognition of mRNA by a specific subset of eukaryotic initiation factors (eIF) followed by recruitment of the ribosome, recognition of the initiation codon AUG and initiation of protein synthesis (Fig. 1). This mechanism is known as the scanning model (2). Because of the compartmentalization of the eukaryotic cell, the translation is spatially and temporally separated from mRNA synthesis (►transcription). Following mRNA synthesis in the ►nucleus, an m⁷G-cap structure is added on the 5' terminus and a poly(A) tail on the 3' terminus of the mRNA. The mRNA is further processed by ►splicing to remove non-coding introns before it is exported into the cytoplasm as a messenger ribonucleoprotein (►mRNP) complex (3). Once in the cytoplasm the m⁷G-cap structure is specifically recognized by cap-binding protein eIF4E, which is part of a larger multiprotein complex called eIF4F. eIF4F consists of eIF4E (cap binding protein), eIF4G (a scaffold protein) and eIF4A (an RNA helicase). The binding of eIF4F to mRNA commits the translation machinery to the translation of that mRNA by facilitating the binding of the 40S ribosomal subunit carrying the initiator methionine-►tRNA to the mRNA. Once the 40S subunit of the ribosome is bound to the mRNA it "scans" in the 5'–3' direction until an initiation codon is encountered. The 40S ribosomal subunit then stably binds at the initiation codon and is joined by a 60S subunit to form an 80S ribosome that commences protein synthesis.

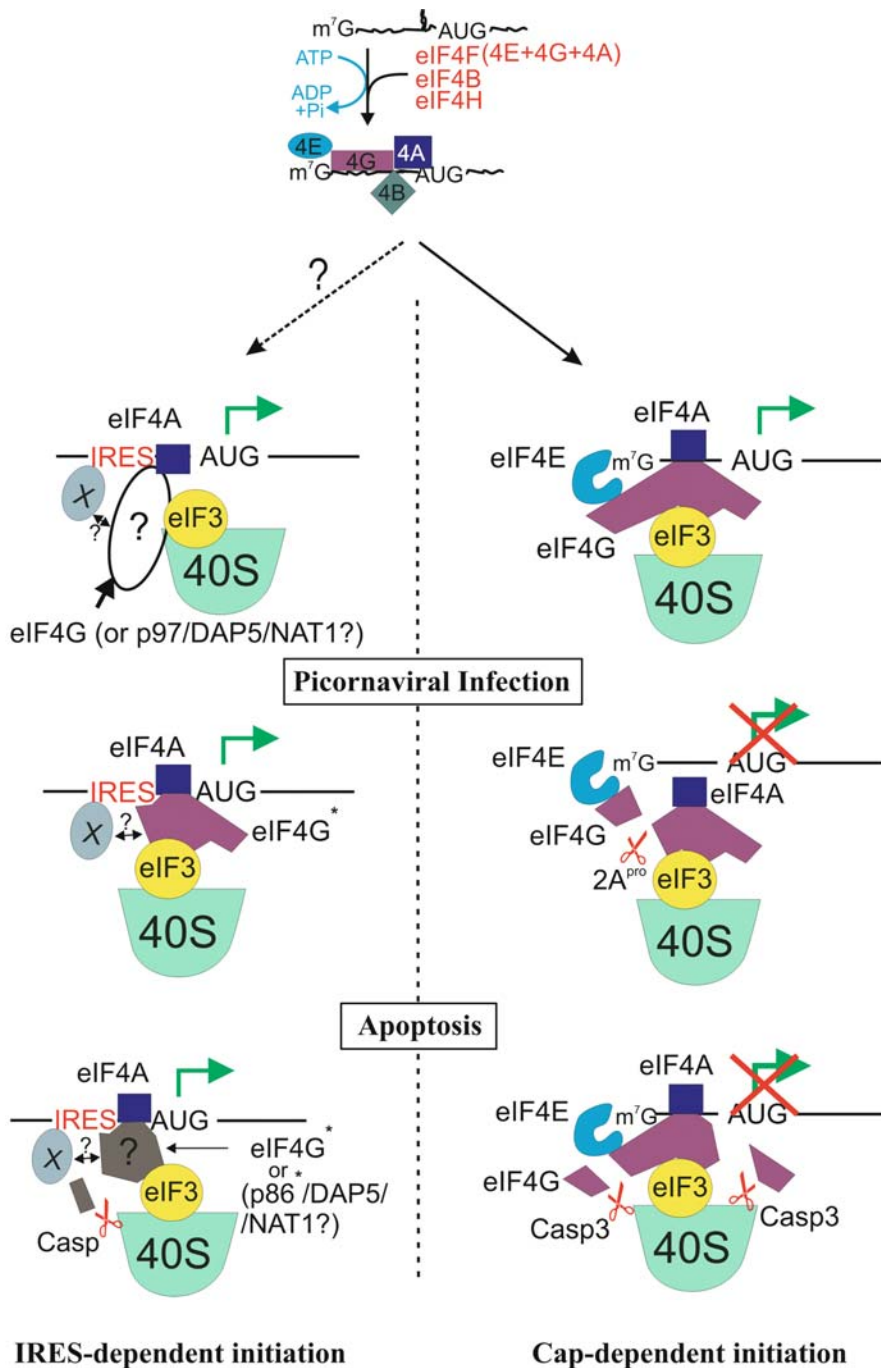
The rate of translation for a given mRNA is controlled by both RNA structural elements and protein factors. eIF4F is a key factor in selecting mRNA for translation and its activity is regulated by several mechanisms. (i) The affinity of the cap-binding complex is regulated by the phosphorylation of eIF4E and eIF4G, (ii) a group of proteins called 4E-BPs bind to eIF4E and prevent its association with eIF4G (the binding of eIF4E and 4E-BPs is also regulated by phosphorylation) and (iii) proteins that are structurally related to eIF4G such as PAIP and p97/DAP5/NAT1 can compete with eIF4G and thus regulate the affinity of eIF4F for mRNA. In addition, global protein synthesis can be controlled by regulating the formation of the ternary complex by changing the phosphorylation status of

eIF2 by four distinct ►protein kinases HRI, PERK, PKR and GCN2 (3). In addition to proteins, the binding efficiency of a given mRNA is affected by the length and secondary structure of the 5' ►UTR and placement of the initiation codon. The 5' UTR is a region between the cap structure and the initiating codon and is typically 50–70 nucleotides long with low degree of secondary structure. Extensive secondary structure within the 5' UTR is thought to be refractory to efficient translation. Indeed, the RNA helicase activity of the eIF4A subunit of eIF4F aided by eIF4B and eIF4H serves to melt the secondary structures in the 5' UTR and to allow binding of the 40S ribosome to RNA. The placement and type of the initiation codon also determine the efficiency of translation initiation. Most initiation in mammalian cells occurs at the AUG initiation codon, although other codons such as GUG, UUG or CUG are sometimes used. In addition, the initiation codon is often found in the favorable sequence context (Kozak sequence) with purines at positions –3 and +4. Deviation from this surrounding sequence may result in the ribosome scanning through and initiating at a different AUG in a more favorable context.

Cap-independent Initiation of Translation

Since eukaryotic translation machinery utilizes cap structure, the binding of the ribosome necessarily occurs at the 5' terminus of the mRNA. However, there exists an exemption to this rule. Some cellular mRNA transcripts and viral RNAs are translated by an internal ribosome entry (IRES) mechanism that is independent of the cap structure (4). The genomic RNA of ►RNA viruses is naturally uncapped and yet translated efficiently. In addition, during viral infection of cells by some members of the *Picornaviridae*, *Flaviridae* and retrovirus families, host protein synthesis is shut off to compromise the host anti-viral response. These viruses produce a protease that cleaves the eIF4E-binding domain from eIF4G so that it is unable to bind to the mRNA cap binding protein eIF4E (Fig. 1). The cleavage of eIF4G disables cellular translation but still allows the viral mRNA to be translated through a cap-independent mechanism using an IRES for translation initiation. Interestingly, the truncated form of eIF4G can stimulate translation of some of the viral RNAs containing an IRES. By this effective and clever mechanism the virus hijacks the cellular translation machinery, allowing the virus to maintain high levels of viral protein synthesis.

The IRES element is a specific regulatory sequence that is found in the 5' UTR of RNA and mediates translational initiation by directly recruiting and binding ribosomes, thus allowing for cap-independent translation. In addition to viral RNA, IRES elements are found in a limited but growing number of cellular



Cap-Independent Translational Control. Figure 1 Mechanisms of cap-dependent (right) and IRES-dependent (left) control of translation initiation. Conditions and initiation factors thought to favor one mode of translation initiation over the other under normal conditions (top), picornaviral infection (middle) and apoptosis (bottom) are shown. For clarity only eukaryotic initiation factors (eIF) pertinent to this essay are indicated. The cap-binding protein eIF4E (blue) recognizes and binds to the 5' m⁷GpppX cap structure. The capped end of the mRNA is bridged with the 40S ribosomal subunit (light blue) by an adapter molecule eIF4G (purple). The binding of eIF4G to the 40S subunit is facilitated by eIF3 (yellow). eIF4A (dark blue) is an RNA-dependent ATPase and RNA helicase involved in the unwinding of the secondary structure of the 5' UTR. Individual components of the translation machinery are not drawn to scale. Factor X represents a putative IRES-binding protein(s) that has been suggested by several laboratories to be involved in internal initiation. Asterisks denote initiation factors proteolytically modified (indicated by scissors) by viral protease (2A^{pro}), caspase 3 (Casp3) or other caspases (Casp).

Cap-Independent Translational Control. Table 1
Cellular mRNAs with Internal Ribosome Entry Sites

Gene product	Regulation (if known)
Transcription Factors	
Antennapedia	-
Ultrabithorax	-
MYT2	-
NRF	-
AML1/RUNX1	Development
Gtx homeodomain protein	-
Hypoxia Inducible Factor 1 α	Hypoxia
Mnt	-
Proto-oncogenes	
c-Myc	Genotoxic Stress, Apoptosis
N-Myc	-
Pim1	-
Protein kinase p58 ^{PITSLRE}	Cell-Cycle
c-Jun	-
ornithine decarboxylase	Cell-Cycle
Blk	-
Tumor Suppressors	
APC	-
Proteins involved in apoptosis	
Apaf-1	Apoptosis, anoxia
XIAP	Apoptosis, anoxia, radiation
HIAP2	Apoptosis, Endoplasmic reticulum stress
Bcl2	-
Bag 1	-
Growth Factors	
FGF2	Heat Shock, Oxidative Stress
PDGF/c-sis	Differentiation
VEGF	Hypoxia
Cyr61	-
IGFII	-
Translation Factors	
eIF4G	-

Cap-Independent Translational Control. Table 1
Cellular mRNAs with Internal Ribosome Entry Sites
(Continued)

Gene product	Regulation (if known)
p97/DAP5/NAT1	Apoptosis
Transporters/Receptors	
CAT-1	Amino acid starvation, ER stress
Notch 2	-
Estrogen Receptor α	-
IFG-1 Receptor	-
Dendritically localized proteins	
ARC	-
MAP2	-
RC3	-
α -subunit of calcium	
calmodulin-dependent kinase II dendrin	-
Others	
BiP	ER stress
La autoantigen	-
β subunit of mitochondrial H ⁺ - ATP synthase	-
p27 ^{Kip1}	-
Hairless	Cell-Cycle, Development
Protein Kinase C delta	Apoptosis
Connexin 43	-
Connexin 32	-
FMR1	-

mRNAs (Table 1). Interestingly, these mRNAs include growth factors, [▶oncogenes](#) and genes that are critically involved in the regulation of [▶apoptosis](#). There is no structural or sequence homology among cellular IRES elements so their discovery remains largely experimental. In addition, the mechanism by which IRES elements recruit ribosomes is not clear. The experimental evidence indicates that the IRES translation requires the canonical initiation factors as well as auxiliary cellular proteins. However, the IRES-mediated translation escapes many of the control mechanisms that regulate cap-dependent translation and was therefore proposed to be critical

for selective translation of mRNAs under physiological conditions, such as apoptosis, hypoxia, amino acid starvation or cell-cycle or during viral infection, when translation of most cellular mRNAs is compromised (5, 6). The regulation of cellular IRES translation is only poorly understood. The best-characterized mechanism so far is the regulation of IRES translation during apoptosis, where it was shown that the ►caspases cleave eIF4G and p97/DAP5/NAT1, resulting in the inhibition of cellular cap-dependent translation. The caspase-generated fragments of eIF4G and p97/DAP5/NAT1 however, selectively drive the translation of Apaf-1 and p97/DAP5/NAT1, presumably to enforce the commitment of cells to die.

Clinical Relevance

Since translation plays such a critical role in the control of gene expression it is not surprising that aberrant translational control correlates with many human pathologies. Increased levels of eIF4E and eIF4G were found in a broad spectrum of transformed cell lines and primary cancers. Furthermore, over-expression of eIF4E, eIF4G and some other initiation factors causes ►malignant transformation, presumably by allowing indiscriminate translation of cellular mRNAs. The accumulated data suggests that over-expression of eIF4E could be used as a valuable prognostic marker in breast cancers. Abnormal IRES translation also correlates with several human disorders. A point mutation in the IRES element of connexin 32 leads to a remarkable inhibition of connexin 32 translation and is found in some patients with ►Charcot-Marie-Tooth disease, a neurodegenerative disorder. Similarly, a point mutation in the IRES of an oncogene, c-Myc, found in patients with multiple myeloma resulted in a significant increase in c-Myc protein levels. Finally, many human viral pathogens depend on IRES function for translation of their proteins. For example, mutations in the IRES element of hepatitis C virus were found to correlate with increased virulence; conversely the attenuated neurovirulence of the Sabin strain of poliovirus used for vaccination is caused by a single nucleotide change in its IRES.

►RNA capping

References

1. Mathews MB, Sonenberg N, Hershey JWB (2000) Origins and Principles of Translational Control. In: Sonenberg N, Hershey JWB, Mathews MB (eds) Translational control of gene expression. Cold Spring Harbor Laboratory Press, Cold Spring Harbor Laboratory, pp 1–32
2. Kozak M, Shatkin AJ (1978) Identification of features in 5' terminal fragments from reovirus mRNA which are important for ribosome binding. Cell 13:201–212

3. Hershey JWB (2000) Pathway and mechanism of initiation of protein synthesis: In: Sonenberg N, Hershey JWB, Mathews MB (eds) Translation control of gene expression. Cold Spring Harbor Laboratory Press, Cold Spring Harbor Laboratory, pp 33–88
4. Hellen CU, Sarnow P (2001) Internal ribosome entry sites in eukaryotic mRNA molecules. Genes Dev 15:1593–612
5. Holcik M, Sonenberg N, Korneluk RG (2000) Internal ribosome initiation of translation and the control of cell death. Trends in Genetics 16:469–473
6. Holcik M, Sonenberg N (2005) Translational control in stress and apoptosis. Nat Rev Mol Cell Biol 6:318–327

Capping Site

Definition

Capping site denotes the initiation site of mRNA synthesis where capping occurs.

►RNA Capping

Cap-Snatching

Definition

Cap-snatching refers to a replication strategy used by influenza and related viruses, to generate pre-formed 5' capped sequences from host cell mRNA as primers for viral mRNA synthesis.

►RNA Capping

CAR

►Cyclic AMP Receptor

Carbohydrate Sulfotransferase

Definition

Carbohydrate sulfotransferase displays one of a family of Type II transmembrane Golgi enzymes, which

catalyze salvation of various monosaccharide units in Asn-linked and Ser/Thr-linked glycols in glycoproteins and proteoglycans.

► [Methylation of Proteins](#)

Carboxymethylated Dextran

Definition

Dextran is a linear polymer of glucose units. In order to use it on sensor chip surfaces for chemical modification, it has to be modified by carboxymethyl groups.

► [Surface Plasmon Resonance](#)

Carboxy-Terminal Transactivation Domain

► [C-TAD](#)

Carcinoma

Definition

Carcinoma refers to a malignant neoplasm derived from epithelial cells (commonly glandular) with a tendency to invade adjacent tissues and to spread to adjacent sites by metastasis.

► [Cancer](#)

► [Cancer Invasion and Metastasis: Cellular, Molecular and Clinical Aspects](#)

► [Hyper- and Hypoparathyroidism](#)

CARD15

Definition

CARD15 is a member of a family of genes (Apaf-1/Ced-4/CARD-4/NOD1) that is implicated in activation of NFκB proinflammatory cytokine induction and apoptosis.

► [Crohn's Disease](#)

Cardiac Connexins

Definition

Cardiac connexins are protein components of ► [gap junctions](#). Connexins 43, 40 and 45 link cardiac myocytes and provide electrical continuity between the cells, which is the basis of synchronised contractions.

► [Intermediate Filaments](#)

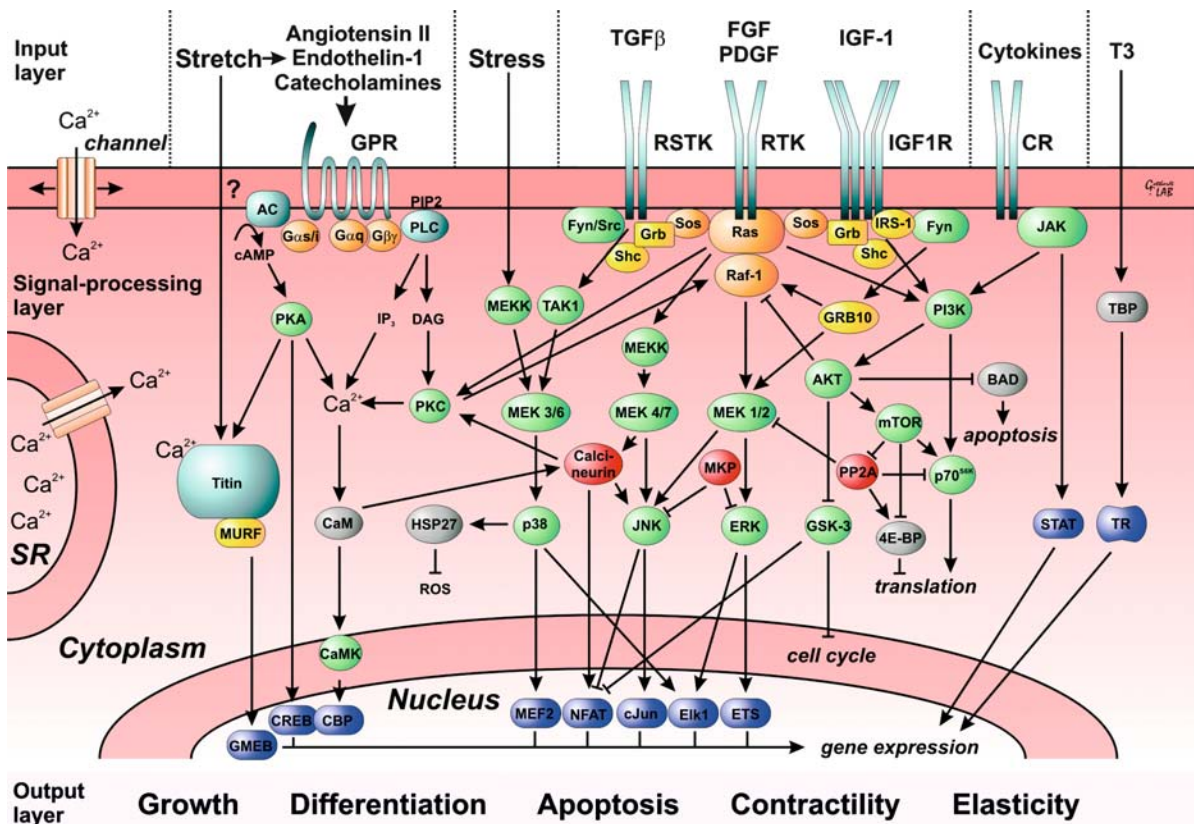
Cardiac Signaling: Cellular, Molecular and Clinical Aspects

MICHAEL GOTTHARDT, KATY RADDATZ
MDC, Berlin, Germany
gotthardt@mdc-berlin.de

Definition

The cardiovascular system distributes oxygen and nutrients in the organism according to the demands of the environment. The required short- and long-term adjustments depend on a complex system of signal transduction pathways that integrate biochemical and biomechanical cues to regulate cardiac growth and function (Fig. 1). The same signaling pathways that are essential for normal cardiac function mediate changes that accompany heart disease e.g. the hypertrophic gene response in ► [cardiomyopathy](#) or ► [apoptosis](#) in ► [myocardial infarction](#). Considerable effort has been put into elucidating these pathways as a basis for developing novel therapeutic strategies.

Overall the responses of the ► [cardiomyocyte](#) to diverse stimuli such as exercise, hormonal or pharmacological stimulation and hemodynamic overload are surprisingly uniform. A growing number of cardiac signaling pathways have been characterized, which include ► [G-protein](#) signaling pathways, the mitogen activated protein kinase (► [MAPK](#)) pathways, calcium signaling, protein kinase C, calcineurin, ► [AKT](#) and many others. Most of these pathways are interconnected and operate in concert to regulate growth, differentiation, cell death and the generation of active and passive forces in the ► [sarcomere](#). A summary of proteins involved in the various pathways is provided in Table 1.



Cardiac Signaling: Cellular, Molecular and Clinical Aspects. Figure 1 Signal transduction in the cardiomyocyte. Mechanical and biochemical signals (input layer) are communicated through the action on transmembrane or intracellular receptors. In the cytoplasm multiple signal transduction pathways, such as G-protein coupled receptor signaling, Akt/PKB signaling, mitogen activated protein kinase (MAPK) cascades and the Jak/Stat pathway integrate these signals to affect gene expression. Ultimately changes in gene expression help adjust the number of cardiomyocytes (growth and apoptosis) and their function (differentiation, contractility, and elasticity) to the needs of the organism. Heart disease is accompanied, and in part caused by, deregulation of cardiac signaling cascades that leads to excessive hypertrophy, dilation, and impaired ventricular function. Kinases are depicted in green, phosphatases in red, adaptor proteins in yellow, G-proteins and their regulators in orange, and transcription factors in blue. Arrows indicate stimulation (pointed) or inhibition (blunted). For abbreviations used also refer to table 1. SR, sarcoplasmic reticulum; ROS, reactive oxygen species; GPR, G-protein coupled receptor; RSTK, receptor serine/threonine kinase; RTK, receptor tyrosine kinase; CR, cytokine receptor; PIP, phospho-inositol-phosphate; IP₃, inositol phosphate; DAG, di-acylglycerol; Ca, calcium.

Characteristics

Cardiac Function

The cardiomyocyte has to perform the daunting task of producing sufficient energy for muscle contraction and ion homeostasis, turn over its protein content in less than three weeks and relay signals to adjust to environmental cues, while contracting at more than 60 beats per minute. These functions coexist and are mediated and regulated through signaling pathways that crosstalk, utilizing shared components, as exemplified for the calcium ion: Calcium is not only essential to develop the action potential, but also to develop active force (contraction of the heart in

►systole), to modulate the elastic properties (filling of the heart in ►diastole), to mediate the ►hypertrophy signal and to regulate glucose metabolism (through activation of phosphorylase kinase). These functions can only coexist through compartmentalization and rapid movement of calcium within the cell.

Not only can a single molecule cause a plethora of changes in the cardiomyocyte – conversely a number of pathways can converge on a common response such as the hypertrophic gene response, which can be activated through beta-adrenergic signaling, growth factors, cytokines, hormones and stretch. If these biochemical or mechanical stimuli persist for extended periods of

Cardiac Signaling: Cellular, Molecular and Clinical Aspects. Table 1 Proteins involved in cardiac signaling

Hormones, vasoactive peptides and catecholamines	Growth factor signaling
AT-II, angiotensin-II	4E-BP, eukaryotic initiation factor-binding protein
ET-1, endothelin-1	Akt, oncogene =PKB, protein kinase B
T3, 3,5,3'-triiodo-L-thyronine	ATF2, Activating transcription factor2
T4, L-thyroxine	Cdc42, homologous to yeast cell division cycle gene 42
NO, nitrogen oxide	c-Jun, oncogene "ju-nana" = 17 from avian sarcoma virus 17
epinephrine	ELK1, oncogene, ETS-domain protein
norepinephrine	ERK, extracellular signal regulated kinase;
acetylcholine	ERK1= p44; ERK2 =p42
Cytokines	ETS, E twenty-six
CT-1, cardiotrophin-1	Fyn, Src-family tyrosine kinase
IL, interleukin	GAP, GTPase activating protein
TNF, tumor necrosis factor	GEF, guanine nucleotide exchange factor
Growth factors and receptors	Grb, growth factor receptor bound protein
EGF, epidermal growth factor	GSK3, glycogen synthase kinase 3
FGF, fibroblast growth factor	HSP27, heat shock protein, 27 kDa
FGFR, FGF receptor	IRS, insulin receptor substrate
IGF, insulin-like growth factor	JNK, c-Jun N-terminal kinase
IGFR, IGF receptor	MAP, microtubule associated protein
PDGF, platelet-derived growth factor	MAPK, mitogen activated protein kinase
TGF β , transforming growth factor β	MAPKAPK, MAPK activated protein kinase
G-protein signaling	MAPKK, MAPK kinase
AC, adenylate cyclase	MAPKKK, MAPKK kinase
CaM, calmodulin	MEF2, mads box transcription enhancer factor 2
CaMK, calcium/calmodulin-dependent protein kinase	MEK, Map/Erk kinase
cAMP, cyclic adenosine monophosphate	MEKK, MEK kinase
CREB, cAMP responsive element-binding protein	MKP, MAP kinase phosphatase
CBP, CREB binding protein	mTOR, mammalian target of rapamycin
DAG, 1,2-diacylglycerol	NFAT, nuclear factor of activated T cells
G-protein, heterotrimeric guanine nucleotide-binding protein	p38=RK, reactivating kinase
GPR, G-protein-coupled receptor	p70S6K, p70 ribosomal protein S6 kinase
IP3, inositol-1,4,5-trisphosphate	PI3K, phosphoinositide 3-kinase
PA, phosphatidic acid	PLD, phospholipase-D
PI, phosphatidylinositol	PP2A, protein phosphatase 2A
PIP2, phosphatidylinositol-4,5-bisphosphate	Rac, a member of the Ras superfamily
PKA, protein kinase-A	Raf, a MAPKKK

Cardiac Signaling: Cellular, Molecular and Clinical Aspects. Table 1 Proteins involved in cardiac signaling (Continued)

Hormones, vasoactive peptides and catecholamines	Growth factor signaling
PKC, protein kinase-C	Ral, Ras-related protein
PLC, phospholipase-C	Ras, a monomeric GTPase (p21)
	Rho, Ras homologous
	RSTK, receptor serine/threonine kinase
	RTK, receptor tyrosine kinase
	SAPK, stress-activated protein kinase (JNK)
Cytokine signaling	SH domain, Src homology domain
CR, cytokine receptor	Shc, SH2-domain containing adaptor protein
gp130, membrane glycoprotein 130	Sos, mammalian homologues of Son-of-sevenless (a GEF)
gp130R, membrane glycoprotein receptor	Src, derived from the Rous sarcoma virus oncoprotein
JAK, Janus kinase	TAK1, TGF β activated kinase 1
LIF, leukemia inhibitory factor	
STAT, signal transducer and activator of transcription	
Apoptosis	
BAD, Bcl2-antagonist of cell death	
Stretch signaling	
GMEB, glucocorticoid modulatory element binding protein	
MLP, muscle LIM protein	
MURF, muscle specific ring finger protein	
Hormone signaling	
TBP, T3 binding protein	
TR, T3 receptor	

time, signaling pathways required for the adaptation of cardiac function have the potential to produce cardiac decompensation.

Signal Transduction

Biochemical and biomechanical signals are processed through transmembrane receptors and ion-channels (input layer, Fig. 1). In the cytoplasm and nucleus, the incoming signals are amplified and relayed through an intricate network of signaling molecules (signal-processing layer). Through changes in gene expression, the fate (differentiation, apoptosis) and function (growth, contractility, elasticity) of the cardiomyocyte are altered (output layer).

Signal transduction can lead to effects that are short-term (posttranslational modifications such as

phosphorylation), intermediate (changes in gene expression, alternative splicing), or long-term (adaptation through structural changes and changes in cell number). Regulation of signal transduction itself usually involves short-term changes such as translocation of proteins to the nucleus, redistribution of calcium between cytoplasm and **sarcoplasmic reticulum**, and changes in activity through phosphorylation or exchange of cofactors such as GDP for GTP.

Molecular Interactions

Biochemical Signals

G-Protein Coupled Receptors

►Catecholamines and ►vasoactive peptides exert their signaling functions through binding to the G-protein coupled seven-transmembrane-spanning domain receptors.

In the cardiovascular system there are three classes of **►G-protein coupled receptors**. The β -adrenergic receptor (β AR) is stimulated by epinephrine and norepinephrine and coupled to $G_{\alpha s}$. Beta-adrenergic signaling leads to increased heart rate and myocardial contractility. The second class comprises the cholinergic receptors, which are activated by acetylcholine and coupled to $G_{\alpha i}$, thus counteracting the stimulatory signals relayed through $G_{\alpha s}$. The third class includes angiotensin II, endothelin and α -adrenergic receptors, which are coupled primarily to $G_{\alpha q}$. They are less important in cardiac function but play a major role in the cardiac hypertrophic gene responses to pathological stimuli.

Growth Factor Signaling

Peptide growth factors such as TGF β (transforming growth factor β), FGF (fibroblast growth factor), and IGF (insulin-like growth factors) have been shown to play an important role in the hypertrophic gene response of cardiomyocytes. A common feature of these growth factor signaling pathways is ligand induced dimerization and cross-phosphorylation of their receptors, stimulation of MAP kinase and PI $_3$ kinase pathways to regulate cell growth, cell division, differentiation and apoptosis.

TGF β binds to three different receptors. Type-I and -II are receptor serine/threonine kinases (RSTK). Ligand binding to the type-II receptor leads to heterodimerization with the type-I receptor. Subsequently type-II receptor phosphorylates the type-I receptor, which transmits the signal to TAK1 (TGF β -activated kinase). TAK1 directly activates MEKK and leads to JNK and/or p38 activation (Fig. 1). For Type-III TGF-receptors no signaling function has been described so far.

FGF binds to a transmembrane receptor with intracellular tyrosine kinase activity (RTK). After ligand induced dimerization and cross-phosphorylation, the phospho-tyrosine residues serve as high-affinity binding sites for a number of intracellular signaling proteins which contain SH2 and **►SH3 domains**. Insulin-like growth factors (IGF) I and II promote cellular proliferation and/or differentiation through binding to a heterotetrameric receptor with tyrosine kinase activity. Ligand binding leads to phosphorylation of the catalytic receptor domains, which in turn phosphorylate insulin receptor substrates (IRS 1 and 2) on multiple tyrosines. IRS serves as a docking site for **►SH2 domain**-containing proteins such as PIK, Fyn, and Shc, resulting in Grb2, Sos and Ras activation.

Cytokine Signaling

►Cytokines are signaling molecules that act as local mediators in cell-cell communication, such as colony-stimulating factors, **►interferons**, and **►interleukins**. Ligands such as cardiotrophin, a member of the **►IL-6**

family and leukemia inhibitory factor (LIF) can induce cardiac hypertrophy or regulate energy metabolism.

Cardiotrophin 1 interacts with the gp130 receptor and the low affinity leukemia inhibitor factor receptor (LIFR). Ligand binding to the gp130-LIFR complex results in phosphorylation of associated Janus kinase signaling factor, which in turn phosphorylates the gp130 receptor, generating a docking site for SH2 domain-containing proteins (Shc, STAT). Thus the cytokine receptors are able to signal *via* several downstream pathways. The family of signal transducers and activators of transcription (STAT) are SH2 domain-containing factors that are phosphorylated by gp130. Activated STATs form homo- and heterodimers, translocate to the nucleus and bind response elements to induce transcription. Gp130 and/or LIFR activation has also been shown to induce MAPK signaling through activated Ras. In addition gp130 activation leads to PIK activation by a mechanism involving Jak1-mediated phosphorylation.

Hormone Signaling

The thyroid hormone is associated with several cardiovascular disorders. In **►hyperthyroidism**, heart rate, contractility, ejection fraction and coronary blood flow are increased, while peripheral vascular resistance is reduced. The active thyroid hormone is 3,5,3'-triiodo-L-thyronine (T3) which is formed by deiodination of L-thyroxine (T4) in the periphery. About 0.3% of serum T3 is in a free form, which can be taken up by the cells by an energy dependent process. The translocation of T3 into the nucleus is still unclear although different cytosolic T3 binding proteins (TBP) have been identified. The nuclear T3 receptor (TR) belongs to the superfamily of hormone responsive nuclear transcription factors. There are two TR genes with two alternative splice products per gene (TR α_1 , TR α_2 , TR β_1 , TR β_2). Cardiomyocytes express TR α_1 , TR α_2 (down-regulated by T3) and TR β_1 (upregulated by T3). T3 furthermore regulates expression of several proteins involved in muscle contraction (e.g. actin and myosin heavy chain) and Ca $^{2+}$ -handling (α_2 , α_3 and β subunit of the **►sarcolemmal** Na-K-ATPase) as well as markers of hypertrophy such as ANF. In addition it has been shown that T3 represses PKC α and PKC ϵ activity in **►neonatal** and PKC ϵ in the adult rat.

Biomechanical Signals

The Stretch Signal

Stretch leaves a mark on the subsequent contraction of cardiomyocytes, which enables the heart to adjust the ejection of blood in systole to the filling of the ventricle in diastole.

So far, it is largely unknown how the stretch signal is sensed and transmitted, but the giant elastic protein titin has been proposed to play a critical role in the process:

Not only do titin's elastic subdomains unfold according to the mechanical forces applied, but furthermore a plethora of titin binding proteins are related to signal transduction through their domain structure and/or translocation to the nucleus upon the application of stretch. Titin forms a continuous filament overlapping at the z-disc and m-line of the sarcomere. It contains a kinase domain, with a binding site for the ring-finger protein MURF-1 in close proximity. MURF-1 can translocate to the nucleus and bind the transcription factor GMEB-1. Titin's N-terminus has been implicated in sensing stretch through the titin binding protein MLP, although the mechanism remains elusive.

Stress

Deformation of the cardiomyocyte during contraction represents a mechanical stress for the cellular structures and induces several responses that modify cardiac rate and rhythm. Acute mechanical stress depolarizes the cell membrane and shortens the duration of the action potential by activation of mechano-sensitive ion channels including K^+ , Cl^- , Na^+ and Ca^{2+} channels. So far, it is largely unknown how the signal is relayed, but it has been shown that stretch activates PKC and AT-II and that both lead to an activation of Raf-1 and MAPK-signaling pathways.

Environmental stress can also be transmitted through proteins expressed in response to injury, such as the cytokines TNF or IL-1 and the cytoprotective factors bradykinin or NO. Stress in response to increased mechanical load is signaled in part through TGF β and angiotensin II. Overall these pathways work to either delimit injury through up-regulation of cytoprotective factors, or to induce tissue repair, when cytoprotection is insufficient to prevent apoptosis. The nature and coordination of these signaling pathways is only poorly understood.

Regulatory Mechanisms

G-Protein Signaling Pathways

All heterotrimeric **G-proteins** consist of $G\alpha$ and $G\beta\gamma$ subunits. Agonist occupation of a membrane-bound receptor leads to an exchange of GDP for GTP on the $G\alpha$ subunit and dissociation of $G\alpha$ from $G\beta\gamma$. The ratio of GTP bound active to GDP bound inactive state determines signal strength. Accordingly the G-protein signal is modulated by proteins that facilitate the exchange of GDP to GTP (such as the guanine nucleotide exchange factors (rhoGEF) or that increase hydrolysis of GTP to GDP (**GTPase** activating proteins – GAPs).

The $G\alpha$ and $G\beta\gamma$ subunits modulate the activity of different downstream signaling effectors – $G\alpha$ works through the adenylate cyclase and PKA to modulate the stretch signal, while the $G\beta\gamma$ subunit activates PLC and PKC linking to the growth factor signal transduction pathways (Fig. 1). Both pathways converge on

adjusting calcium levels. PKA phosphorylates a number of intracellular protein targets at serine and threonine residues e.g. **L-type calcium channels**, **phospholamban** (modulator of the sarcoplasmic reticulum associated ATP-dependent Ca^{2+} -pump), titin's elastic N2B domain and modulatory proteins associated with the contractile apparatus such as troponin I. Activation of PKA targets increases the intracellular Ca^{2+} -level, myocardial contractility and rate of relaxation. Increased Ca^{2+} concentrations in turn affect several Ca^{2+} -dependent proteins, such as calmodulin and calcineurin. In addition PKA phosphorylates G-protein coupled receptors (e.g. the β -adrenergic receptor), resulting in partial uncoupling and desensitization of receptors for further agonist stimulation.

Ras/MAPK Signaling

The oncogene ras is a low molecular weight GTPase that is activated by **G-protein-coupled receptors** (GPCR), receptor tyrosine kinases (RTK), Janus kinase 1 (**JAK**) or an increase in intracellular Ca^{2+} . Ras activation results in GDP to GTP exchange and activation of numerous effector proteins, Raf-1, PI $_3$ K, small GTPase Ral proteins, PKC and all three MAPK signaling branches. Ral promotes activation of PLD which catalyses the hydrolysis of phosphatidylcholine to phosphatidic acid (PA) and choline. PA is then further metabolized to 1,2-diacylglycerol (DAG), an activator of PKC. In addition Ral activates **cdc42** and **Rac-1**, two members of the **Rho family** of small GTPases. Both have been implicated in the regulation of the actin **cytoskeleton** and have been shown to stimulate two MAP kinases, p38 and **JNK/SAPK**.

Mitogen-activated protein kinase signaling pathways (**MAPK**) are activated in cardiomyocytes by G-protein-coupled receptors (GPRs), receptor tyrosine kinases (RTKs), transforming growth factor beta-receptor (TGFR), protein kinase C (PKC), calcium or stress stimuli. The upstream events result in the activation of Map/Erk kinase kinase (MEKK), which leads to the activation of Map/Erk kinase (MEK) factors and in turn to activation of the three terminal MAPK effectors, c-Jun NH2-terminal kinases (JNK1/2/3), extracellular signal-regulated kinases 1 and 2 (ERK1 and 2) and p38. The three mammalian MAPK families are defined by their distinct phosphorylation motifs, *Thr-Glu-Tyr* for **ERKs**, *Thr-Phe-Tyr* for the JNKs and *Thr-Gly-Tyr* for the p38. They activate various nuclear and extranuclear substrates by phosphorylation at serine and threonine residues. Extranuclear substrates of ERK comprise microtubule-associated proteins (MAP1, 2, 4 and **tau-protein**), which mediate **microtubule** rearrangements and changes in cell morphology.

Nuclear substrates of ERK and JNK include the ternary complex factor Elk-1 and the transcription factor

ANF2. ANF2 can heterodimerize with c-Jun, an additional substrate of JNK, to activate gene expression through the cAMP response element (Cre).

Downstream targets of p38 are the transcription factors MEF2, ATF-2 and ELK-1 as well as the cytoplasmic protein MAPKAPK2, which phosphorylates HSP27 and might protect cells against ►oxidative stress and apoptosis.

PI3K/Akt/GSK-3-Signaling

The serine threonine kinase ►Akt/PKB, plays a central role in diverse cellular processes, including glucose metabolism, proliferation and apoptosis. It mediates many events downstream of receptor tyrosine kinases (RTK), such as the IGF-1 receptor, ►Janus kinase (JAK-1), Ras and ►G-protein-coupled receptors (GPR) that induce the accumulation of phosphatidylinositol-triphosphates (PI3P) by the phosphoinositide 3-kinase (PI3K). Akt controls cell growth through mTor (mammalian target of rapamycin), which activates translation through p70S6 kinase and the translation initiation factor 4EBP1/eIF4E. In addition Akt also directly phosphorylates the proapoptotic protein BAD and glycogen synthase kinase 3 (GSK-3), thereby inhibiting their activity.

Ca²⁺-Signaling

Ca²⁺ is essential in cardiac electrical activity and is a direct activator of the myofilaments that cause contraction. Local calcium levels in microdomains within the cell are crucial to enable signaling function, while the electrical demands on the cardiomyocyte cause rapid changes in total cytoplasmic Ca²⁺ levels. Ca²⁺ influx and release from the sarcoplasmic reticulum (SR) is modulated after adrenergic stimulation through PKA. Accelerated relaxation is mediated by phosphorylation of phospholamban and troponin, which leads to increased Ca²⁺ reuptake and dissociation from the myofilament. Ca²⁺ reuptake is mediated by four pathways, sarcoplasmic Ca²⁺-ATPase, sarcolemmal Na⁺/Ca²⁺ exchange, sarcolemmal Ca²⁺-ATPase or the mitochondrial Ca²⁺ uniport. Elevated sarcoplasmic Ca²⁺ stores increase the developed contraction.

Ca²⁺ channels are regulated by stretch, receptors, membrane potential and internal Ca²⁺ stores. After activation, Ca²⁺ signals can act locally or recruit other channels to activate processes through a global increase in Ca²⁺. Since high levels of Ca²⁺ are toxic, the signal is modulated in amplitude or frequency. Different pathways can respond according to the duration or frequency of the Ca²⁺ signal, such as calmodulin dependent kinase II (CaMK) or induction of the serine-threonine phosphatase calcineurin, which in turn dephosphorylates members of the NFAT-transcription factor family or calmodulin dependent kinase II

(CaMK). Multiple Ras exchange factors and GTPase activating proteins are regulated by Ca²⁺ through CaMKII. Thus Ca²⁺ microdomains can influence growth through crosstalk with the ERK/MAPK pathway.

Medical Applications

►Heart disease is the main cause of death in industrialized nations. Patients with impaired cardiac function from various causes have benefited from treatment with drugs that interfere with signal transduction to counteract hypertrophy and improve remodeling of the heart muscle.

►Beta-blockers inhibit beta-adrenergic signaling and help improve cardiac performance by decreasing heart rate, thus improving ventricular filling. Additionally they reduce myocardial oxygen demand. Similar effects are obtained with Ca²⁺-channel blockers, such as verapamil, which do not impair responsiveness to sympathetic stimulation and help maintain blood pressure.

Angiotensin converting enzyme ►(ACE) inhibitors are used to treat congestive cardiac failure. Their anti-hypertrophic and anti-apoptotic action is mediated by increased bradykinin signaling and inhibition of myocardial renin-angiotensin-aldosterone system. The underlying mechanisms are still under investigation and include attenuation of PKC translocation to prevent down-regulation of Ca²⁺ cycling and activation of ERK signaling.

Cardiac signal transduction can also mediate the side effects of drug therapy such as the cardiotoxicity of the erbB2 receptor antibody (Trastuzumab/Herceptin). Understanding cardiac signaling, such as the ErbB2 signaling pathway in adult cardiomyocytes might help develop strategies to alleviate the cardiotoxic effects of Trastuzumab treatment, allowing for safer treatment of cancers over-expressing erbB2.

Acknowledgements

K. Raddatz is supported by the Alexander von Humboldt Foundation. M. Gotthardt is a Humboldt Fellow and supported by the DFG and the NIH (RO-1 HL69008).

Further Reading

1. Bers DM (2002) Cardiac excitation-contraction coupling. *Nature* 415:198–205
2. Frey N, Olson EN (2003) Cardiac hypertrophy: the good, the bad, and the ugly. *Annu Rev Physiol* 65:45–79
3. Mann DL (2003) Stress-activated cytokines and the heart: from adaptation to maladaptation. *Annu. Rev. Physiol.* 65:81–101
4. Molkentin JD, Dorn Jr (2001) Cytoplasmic signaling pathways that regulate cardiac hypertrophy. *Annu. Rev. Physiol.* 63:391–426
5. Post RP, Hammond HK, Insel PA (1999) Beta-adrenergic receptors and receptor signaling in heart failure. *Annu. Rev. Pharmacol. Toxicol.* 39:343–360

Cardiochip

CHOONG-CHIN LIEW

Harvard Medical School, Boston, MA, USA

cliew@rics.bwh.harvard.edu

Synonyms

Cardiovascular microarray; Cardiovascular DNA chip

Definition

The Cardiochip is a customized human cardiovascular gene-specific DNA-based [▶microarray](#), constructed in-house at our institution from genes derived from clones of our heart [▶cDNA](#) library. The Cardiochip enables us to review gene expression profiles for the study of cardiac growth and development, as well as study the progression of cardiovascular diseases. Such information may be utilized to develop specific molecular portraits of individual diseases and disease subtypes, to monitor disease progression at the molecular level, for diagnostic and prognostic purposes, to assess potential drug candidates or to investigate drug responses.

Rationale for the Cardiochip

Each living organ and organ system is associated with a set of genes that are involved in maintaining and regulating its function. The heart cells, for example, include cardiomyocytes, smooth muscle cells, endothelial cells, conduction cells for maintaining cardiac rhythm, neuroendocrine cells and other cell types; specific genes are associated with each of these cell types. Thus, transcripts derived from the cells of the heart represent the genes of the entire cardiovascular system. Our human heart [▶cDNA libraries](#) are representative of gene transcription in the heart tissues and thus of mRNA populations of all the genes of the cardiovascular system. Since the Cardiochip is based on our heart tissue specific library and the cDNA library, the chip provides a window for the study of the function, development and diseases of the cardiovascular system.

Cardiac specific genes are first expressed during embryonic development and function to control and regulate that development in a cell appropriate fashion. Embryonic heart cell genes signal stem cells to differentiate and to develop mature cells (i.e., cardiac myocytes) with a cardiac-specific structure. Cardiac myocytes divide during embryogenesis but lose their ability to divide soon after birth. Subsequent cardiac cell growth is achieved by an increase in cell size (hypertrophy), during maturation and later by accommodating fibroblasts when the heart is damaged during disease processes.

Microarray technology such as the Cardiochip tracks tissue specific gene expression at all stages of development and disease; thus, the technology can be used to characterize gene expression at various stages of disease. Microarray-based portraits of cardiovascular gene expression enable cardiovascular diseases to be compared at the gene level, using “snapshots” of different diseases or of the same disease over time.

Characteristics

We constructed our flexible, tissue-specific cDNA based “Cardiochip” microarray in-house at the Cardiovascular Genome Unit of Brigham and Women’s Hospital, Harvard Medical School. (1) We used PCR products generated from our comprehensive collection of cardiovascular [▶expressed sequence tags](#) (EST) derived from human fetal and adult heart, familial hypertrophic cardiomyopathic tissue and vascular cDNA libraries. After [▶BLAST](#) sequence similarity searching, we compiled the non-redundant Cardiochip which comprises to date, 10,848 spots. Of these spots, 2761 (25%) are known genes, 3406 (31.4%) are matched ESTs, 4489 (41.4%) are unmatched novel ESTs, and 192 (1.8%) are bacterial clones (for quality control purposes).

We have conducted a number of [▶hybridization](#) studies using the Cardiochip. This technique involves extracting RNA from the tissue of interest. The RNA probe is then color-coded (or “labeled”) using fluorescence technology and subsequently hybridized onto the chip. Hybridization capitalizes on the phenomenon that DNA of genes of interest recognize and bind complementarily to corresponding DNA on the microarray chip. After labeling and hybridization is complete, color images of the bound genes can be derived from each chip. These images allow convenient visualization of differential gene expression; they are also used to generate a quantitative measure for each gene’s expression level that can be used for statistical data analysis. These microarray results may be verified, either one gene at a time or in a high-throughput capacity, using real-time RT-PCR. Identifying up- or down-regulated genes may also provide insight into gene “pathways”, where certain genes trigger the expression of other genes in order to attain a disease state.

Our preliminary Cardiochip study compared samples of failing heart tissue (obtained at the end stage of hypertrophic cardiomyopathy) with normal heart tissue. (1) We identified 38 key transcripts among which were genes previously suspected to be involved in heart failure as well as several novel, previously uncharacterized genes. In a later study, we utilized the Cardiochip to explore end-stage dilated cardiomyopathy (2). We found more than 100 genes consistently differentially expressed in dilated cardiomyopathy as

compared to normal samples. Atrial natriuretic peptide was highly up-regulated in dilated cardiomyopathy samples, confirming its pivotal role as a marker of cardiac stress. In addition, numerous sarcomeric and cytoskeletal proteins, stress response proteins and transcription/translation regulators were up-regulated in dilated cardiomyopathy. Cell signaling channel and mediator proteins were prominently down-regulated, especially those involved in calcium signaling. Intriguingly, we also found several novel expressed sequence tags identified as differentially regulated. Such tags may represent key genes in the pathophysiology of heart failure. We have also developed gene expression profiles comparing dilated and hypertrophic cardiomyopathy, and identified several genes differentially expressed between these two diseases. (3) Most recently, we used the Cardiochip to determine that cardiovascular related genes form a distinctive cluster in the chromosome structure (4). A current study in progress is the characterization of the genetic profile of Chagas disease, a form of cardiomyopathy prevalent in Latin America; preliminary results indicate an up-regulation in chemokine- and cytokine-related immune genes, consistent with the established knowledge regarding the disease.

Clinical Relevance

Understanding the genetic basis of highly complex, multifactorial cardiovascular diseases is a major challenge. Many genes and gene pathways remain to be elucidated before we are able to convincingly map out the molecular mechanisms driving such diseases as hypertension, heart failure, atherosclerosis and other conditions.

Microarray technology is facilitating our exploration of complex cardiovascular diseases. The use of the Cardiochip and similar microarrays in cardiovascular medicine has been favorably reviewed (5, 6). Applied to the cardiovascular system, microarray technology has made it possible to carry out large-scale global gene expression profiles. Thousands of genes can be compared simultaneously, rather than one by one, and microarray technology distinguishes genes that are expressed and genes that are not. Such gene snapshots of disease samples can be compared with normal samples to indicate which genes are specifically up- or down-regulated; similarly, different cardiovascular diseases can be compared with one another to discover disease-specific differential gene expression. This type of information is essential for characterizing diseases at a molecular (as opposed to a histological or clinical) level. With the development of the Cardiochip and related microarray technologies, we now have genomic tools wherewith we can examine the genetic changes that

distinguish different stages of heart failure or any other heart condition (7). Further use of microarray technology should lead to clues about gene function, gene pathways and the identification of novel genes. Differing molecular disease portraits allow for the diagnosis of disease subtypes; using the Cardiochip, this information may be used to identify the molecular changes that occur in the cardiovascular system at different clinical stages of disease or drug response. Recent research has focused on further refinement to Cardiochip and other microarray systems (8). For instance, microarray studies have relied on tissue samples in the past. However, disease tissue is not always readily obtainable, especially in the case of heart disease. Thus, our laboratory and other groups are beginning to establish blood RNA-based microarray systems as alternatives to tissue-based systems for disease diagnosis and prognosis (9-11).

►Cardiac Signaling: Cellular, Molecular and Clinical Aspects

►Common (Multifactorial) Diseases

References

1. Barrans JD, Stamatou D, Liew C et al (2001) Construction of a human cardiovascular cDNA microarray: portrait of the failing heart. *Biochem Biophys Res Commun* 280:964-969
2. Barrans JD, Allen PD, Stamatou D et al (2002) Global gene expression profiling of end-stage dilated cardiomyopathy using a human cardiovascular based cDNA microarray. *Am J Pathol* 160:2035-2043
3. Hwang JJ, Allen PD, Tseng GC et al (2002) Microarray gene expression profiles in dilated and hypertrophic cardiomyopathic end-stage heart failure. *Physiol Genomics* 10:31-44
4. Barrans JD, Ip J, Lam CW et al (2003) Chromosomal distribution of the human cardiovascular transcriptome. *Genomics* 81:519-24
5. Cook S, Rosenzweig A (2002) DNA microarrays: implications for cardiovascular medicine. *Circ Res* 91:559-564
6. Napoli C, Lerman LO, Sica V et al (2003) Microarray analysis: a novel research tool for cardiovascular scientists and physicians. *Heart* 89:597-604
7. Liew CC, Dzau VJ (2004) Molecular genetics and genomics of heart failure. *Nat Rev Genet* 5:811-25
8. Liew CC (2005) Expressed Genome Molecular genetics and genomics of heart failure. *Clin Chem Lab Med*, in press
9. Ma J, Liew CC (2003) Gene profiling identifies secreted protein transcripts from peripheral blood cells in coronary artery disease. *J Mol Cell Cardiol* 35:993-8
10. Tsuang MT, Nossova N, Yager T (2004) Assessing the validity of blood-based gene expression profiles for the classification of schizophrenia and bipolar disorder: a preliminary report. *Am J Med Genet B Neuropsychiatr Genet* 133:1-5

11. Morello F, de Bruin TW, Rotter JI (2004) Differential gene expression of blood-derived cell lines in familial combined hyperlipidemia. *Arterioscler Thromb Vasc Biol* 24:2149–54

Cardiogenesis

Definition

Cardiogenesis defines the process of heart formation.

- ▶ *Drosophila* Model of Cardiac Disease
- ▶ Heart

Cardiomyocyte

Definition

Cardiomyocytes are the cross-striated muscle cells that form the myocardium.

- ▶ Cardiac Signaling: Cellular, Molecular and Clinical Aspects
- ▶ Heart

Cardiomyopathy

Definition

Cardiomyopathy is a disease of the heart, specifically affecting the middle layer of tissue in the heart. The heart muscle is weakened, which usually causes inadequate heart pumping.

- ▶ Cardiac Signaling: Cellular, Molecular and Clinical Aspects
- ▶ *Drosophila* Model of Cardiac Disease
- ▶ Duchenne Muscular Dystrophy
- ▶ Familial Dilated Cardiomyopathy
- ▶ Limb Girdle Muscular Dystrophies

Cardiovascular Disease

Definition

Cardiovascular disease describes diseases of the heart and vascular system. Examples are arteriosclerosis and deep vein thrombosis. Cardiovascular disease may ultimately lead to myocardial infarction (heart attack) or stroke.

- ▶ Diabetes Mellitus, Genetics

Cardiovascular DNA Chip

- ▶ Cardiochip

Cardiovascular Microarray

- ▶ Cardiochip

Caretaker Gene

Definition

A gene is designated a caretaker gene whose function is to keep the mutation rate low, for example, a DNA repair gene. The inactivation of the caretaker gene leads to genomic instabilities resulting in increased mutation of all genes, including gate keepers.

- ▶ DNA-Repair Mechanisms
- ▶ Tumor Suppressor Genes

Cargo

Definition

RNA molecules and RNA-protein complexes, which are exported from the nucleus to the cytoplasm by export receptors, are referred to as cargo.

- ▶ Nuclear Import and Export
- ▶ RNA Export

Carpopedal Spasm

Definition

Carpopedal spasm describes a sudden involuntary contraction of one or more muscles of the hands and feet.

- ▶ Hyper- and Hypoparathyroidism

Carrier Ethics

Definition

Carrier ethics reviews moral responsibilities and obligations of carriers of a genetic disorder or disease towards themselves and members of the family, who might or might not have the same genetic heritage.

► [Ethical Issues in Medical Genetics](#)

Carrier Frequency

Definition

Carrier frequency means the frequency of individuals in the general population who are heterozygous for a (recessive) trait.

► [Heritable Skin Disorders](#)

Carrier(s)

Definition

Carrier(s) are subjects bearing only one allelic copy of an altered gene responsible for an autosomal recessively transmitted disease; Carriers are also referred to as heterozygotes and are frequently clinically asymptomatic.

► [Epistasis in Cystic Fibrosis](#)

Cartilage

► [Bone and Cartilage](#)

Cas

Definition

Cas stands for Crk-associated substrate. It is a docking protein which plays a central coordinating role

for tyrosine-kinase-based signaling related to cell adhesion.

► [Crk](#)

► [Signal Transduction: Integrin-Mediated Pathways](#)

Casein Kinase I

Definition

Casein kinase I (CKI) represents a large family of ► [serine-threonine kinases](#) that are implicated in a variety of cellular functions and processes. CKI exists in multiple forms in mammalian tissues and is present in the nucleus, cytosol, plasma membrane and microsomes. CKIs prime β -catenin for phosphorylation by glycogen synthase kinase-3 (GSK3). It also binds and phosphorylates multiple other pathway components.

► [Wnt/Beta-Catenin Signaling Pathway](#)

Casein Kinase II

Definition

Casein Kinase II (CKII) are serine-threonine kinases that bind and phosphorylate ► [Dvl \(Dishevelled\)](#) in *Drosophila*. In vertebrates it has multiple roles in the Wnt pathway and phosphorylates multiple pathway components.

► [Wnt/Beta-Catenin Signaling Pathway](#)

Caspase

Definition

Caspases are cysteine proteases that facilitate ► [apoptosis](#) by cleaving a growing number of cellular targets. Caspases are synthesized as inactive zymogens that are proteolytically processed to produce mature, active proteases. In addition, caspases are arranged into a cascade, such that the upstream caspases (initiator caspases, such as caspase-9) activate downstream caspases

(effector caspases, such as caspase-3). Caspases also have apoptosis independent functions, e.g. in T-cell proliferation or IL1 maturation.

- ▶ Apoptosis, Regulation and Clinical Implications
- ▶ Cap-Independent Translational Control
- ▶ TNF Receptor/Fas Signaling Pathways

CAT

Definition

Computed axial tomography (CAT) is a widely employed medical imaging method for macroscopic organ diagnosis. The principle of image reconstruction is similar for all tomographic imaging methods, including electron tomography.

- ▶ Electron Tomography

Catalysis

Definition

Catalysis describes the increase in the rate of a chemical reaction by the introduction of a catalyst such as an enzyme.

- ▶ Proteases and Inhibitors

Catalytic Antibodies

Definition

Catalytic antibodies do not only bind an antigen, but are able to convert the bound molecule in the same way as is known for enzymes. Most catalytic antibodies produced so far have cleaved a chemical substance. They are produced by immunizing animals with stable transition state analogues, or by selecting antibodies by ▶ phage display with the help of transition state analogues.

- ▶ Monoclonal Antibodies

Catalytic RNAs

MASAYUKI SANO¹, KAZUNARI TAIRA²

¹Gene Function Research Center, National Institute of Advanced Industrial Science and Technology (AIST), Tsukuba Science City, Japan

²Department of Chemistry and Biotechnology, School of Engineering, University of Tokyo, Tokyo, Japan
taira@chembio.t.u-tokyo.ac.jp

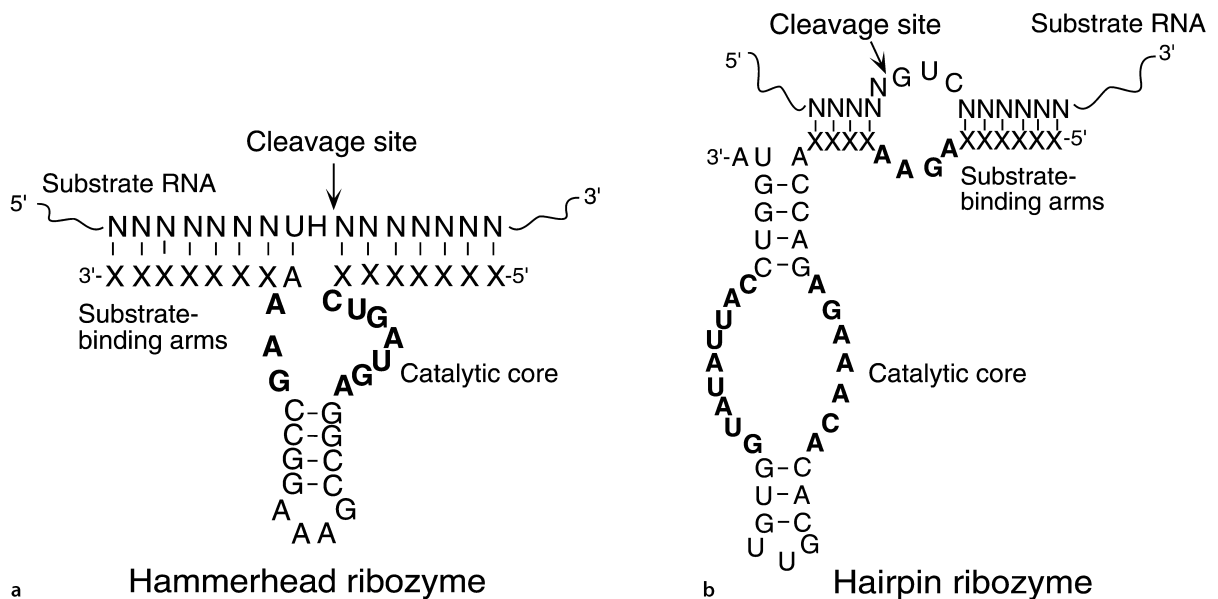
Definition

Rapid developing technologies that involve several types of functional RNA have received increasing attention as possible approaches to gene therapy. The functional RNAs, such as ▶ antisense RNAs, ▶ ribozymes and ▶ small interfering RNAs (siRNAs), inactivate genes in a sequence-specific manner at the mRNA level. ▶ RNA aptamers, namely, RNAs that can bind to proteins, have also been used to modulate gene function at the protein level. Selective modulation of gene expression is very important not only for the elucidation of the biological functions of individual genes but also for the phenotypic analysis of genes in living cells. In addition, functional RNAs have been tested for their ability to target specific disease-related genes in efforts to combat human diseases. With the accumulation of vast amounts of data from genome sequencing, potential targets for functional RNAs are rapidly being identified and validated and RNA-based systems that attack such targets are being evaluated in clinical trials. This essay focuses on recent advances that have been made with two types of functional RNA, ribozymes and siRNAs, as efforts continue to exploit them as therapeutic agents that will be useful in a clinical setting.

Characteristics

Application of Ribozymes

Ribozymes have been successfully employed to suppress the expression of disease-related genes in cultured cells and tissues and in living organisms. Ribozymes that cleave their substrates “*in trans*”, such as ▶ hammerhead and ▶ hairpin ribozymes, can cleave target RNAs specifically after they have bound to their substrate RNAs *via* base-pairing interactions (Fig. 1). Such ribozymes can target and cleave their respective substrates repeatedly, acting as authentic catalysts. Synthetic nuclease-resistant ribozymes have been evaluated in clinical trials. One such ribozyme is ANGIOZYME, which is directed against the transcript of the *Flt-1* gene, which encodes the high-affinity



Catalytic RNAs. Figure 1 Secondary structures of a hammerhead ribozyme (a) and a hairpin ribozyme (b). The polyX corresponds to the sequence that is complementary to the target RNA. These ribozymes can cleave any RNA substrate that contains an NUH triple, where N is any nucleotide and H is A, C or U in the case of the hammerhead ribozyme, or GUC in the case of the hairpin ribozyme.

receptor (VEGF-R1) for vascular endothelial growth factor (VEGF) (1). Since VEGF is involved in angiogenesis, namely, the formation of new blood vessels, it is required for the sustained growth of tumor. VEGF appears to be a promising target for therapeutic antagonists such as ribozymes. In mice, ANGIOZYME efficiently and dose-dependently inhibits the metastasis of colorectal cancer to the liver. This ribozyme was tested in patients with advanced malignancy.

Although synthetic ribozymes can inhibit the expression of specific genes, their application over long periods of time is severely restricted in living cells or organisms. To overcome this limitation, numerous viral and non-viral vectors have been developed for the stable expression of ribozymes. In such cases, strong and constitutively active **▶promoters** are employed to control the expression of each ribozyme. The **▶RNA polymerase III** (pol III) system has proved to be useful in this context since (i) large numbers of extra sequences are not required for appropriate transcription and termination and (ii) levels of expression are higher than those obtained with the **▶RNA polymerase II** (pol II) system. One of the specific pol III systems involves the promoter of genes for tRNAs. Hairpin ribozymes produced under the control of the promoter of the human gene for tRNA^{Val} efficiently inhibit the replication of various strains of HIV-1 in cells in culture (2). In addition, **▶hammerhead ribozymes** transcribed from the human tRNA^{Val} promoter have been used to confirm the roles of **▶p300** and **▶CBP**, which are

transcriptional co-activators, in retinoic acid-induced differentiation, cell-cycle arrest and apoptosis of embryonal carcinoma cells (3).

The promoter of the human gene for tRNA^{Val} has been also used for the production of a dimeric ribozyme known as the **▶maxizyme**. The maxizyme is an allosterically controllable ribozyme that has cleavage activity only in the presence of its specific target sequence. Thus, a maxizyme can be designed such that its activity is evident only, for example, in the presence of an abnormal RNA sequence, such as the junction sequence of the chimeric mRNA generated by the *BCR/ABL* rearrangement. This rearrangement is the molecular equivalent of the Philadelphia chromosome that arises from the reciprocal chromosomal translocation t(9;22)(q34;q11). Chronic myelogenous leukemia (CML) is a clonal myeloproliferative disorder of hematopoietic stem cells that is associated with the Philadelphia chromosome. Appropriate targeting of the chimeric *BCR/ABL* fusion would kill tumor cells specifically, while normal cells would remain unaffected. A maxizyme directed against the junction sequence of *BCR/ABL* did, indeed, disrupt the chimeric transcript specifically without affecting the normal *BCR* and *ABL* transcripts in murine cells. Furthermore, the maxizyme efficiently killed BV173 cells, which were derived from a patient with leukemia who carried the Philadelphia chromosome. The potential utility of the maxizyme in the treatment of CML has also been assessed in a murine model of

CML. All mice, without exception, that had been injected with cells from a patient with CML died of diffuse leukemia within 13 weeks. By contrast, all mice remained disease-free when the CML cells that were injected had been treated with the maxizyme prior to injection. These results suggest that maxizyme technology might provide a useful approach to the treatment of CML (4).

In addition to the potential utility of ribozymes in a clinical setting, combinatorial ribozyme libraries are useful for the identification of genes that are associated with specific phenotypes in mammalian cells. Several reports have described systems for screening for novel functional or disease-related genes using ribozymes with randomized binding arms. For example, a library of hammerhead ribozymes with randomized binding arms allowed the identification of certain genes that are involved in the migration of cells that is an essential feature of ►metastasis, one of most serious problems in cancer therapy (5). A line of highly invasive cancer cells, HT1080 fibrosarcoma cells, normally migrates toward a chemoattractant. However, after treatment with the library of randomized ribozymes, some of the cells did not migrate toward the chemoattractant. Sequencing of the randomized regions of the ribozymes that were then isolated from the nonmigrating cells and a search for the sequences in databases allowed the rapid identification of genes that are likely to play a role in metastasis. Two ribozymes directed against the transcript of *ROCK1*, one of the genes that regulate the ►actin cytoskeleton, were found among 924 positive ribozyme clones obtained in such an experiment. These two ribozymes inhibited the migration and invasion by invasive HT1080 cells almost completely without any effect on the proliferation of the cells. Similar experiments with a library of ribozymes with randomized binding arms have also allowed identification of genes involved in ►Fas- and ►TNF α -mediated apoptosis and cell invasion. Such ribozyme technologies are powerful tools for the development of gene-inactivating reagents of therapeutic importance and for the rapid identification of disease-related genes.

RNAi – a Brief Description

RNA interference (RNAi) has emerged as another novel tool for the silencing of gene expression in animals and plants. In this phenomenon, double-stranded RNA (dsRNA) induces the sequence-specific degradation of a cognate mRNA. In mammalian cells, RNAi with long dsRNAs was first described in mouse embryonal carcinoma cells and embryonic stem cells. However, long dsRNAs (of more than 30 bp) also induce non-specific responses as a result of inhibitory effects on protein synthesis and the degradation of mRNA in mammalian somatic cells.

Tuschl and coworkers demonstrated that 21-nucleotide (21-nt) synthetic RNA duplexes with 2-nt 3' overhangs, defined as small interfering RNAs (siRNAs), specifically suppressed the expression of endogenous genes in mammalian cells (6). It became apparent subsequently that siRNAs are associated with a multicomponent nuclease, the ►RNAi-induced silencing complex (RISC) and that they guide this enzyme so that it can catalyze the sequence-specific degradation of the cognate mRNA (Fig. 2).

Successful experiments with synthetic siRNAs were followed by the development of several vector-based systems for the expression of siRNA that could also induce RNAi in mammalian cells. This technical advance not only allowed the long-term expression of silencing effects *in vivo* but also opened the door to the therapeutic application using vector-mediated RNAi.

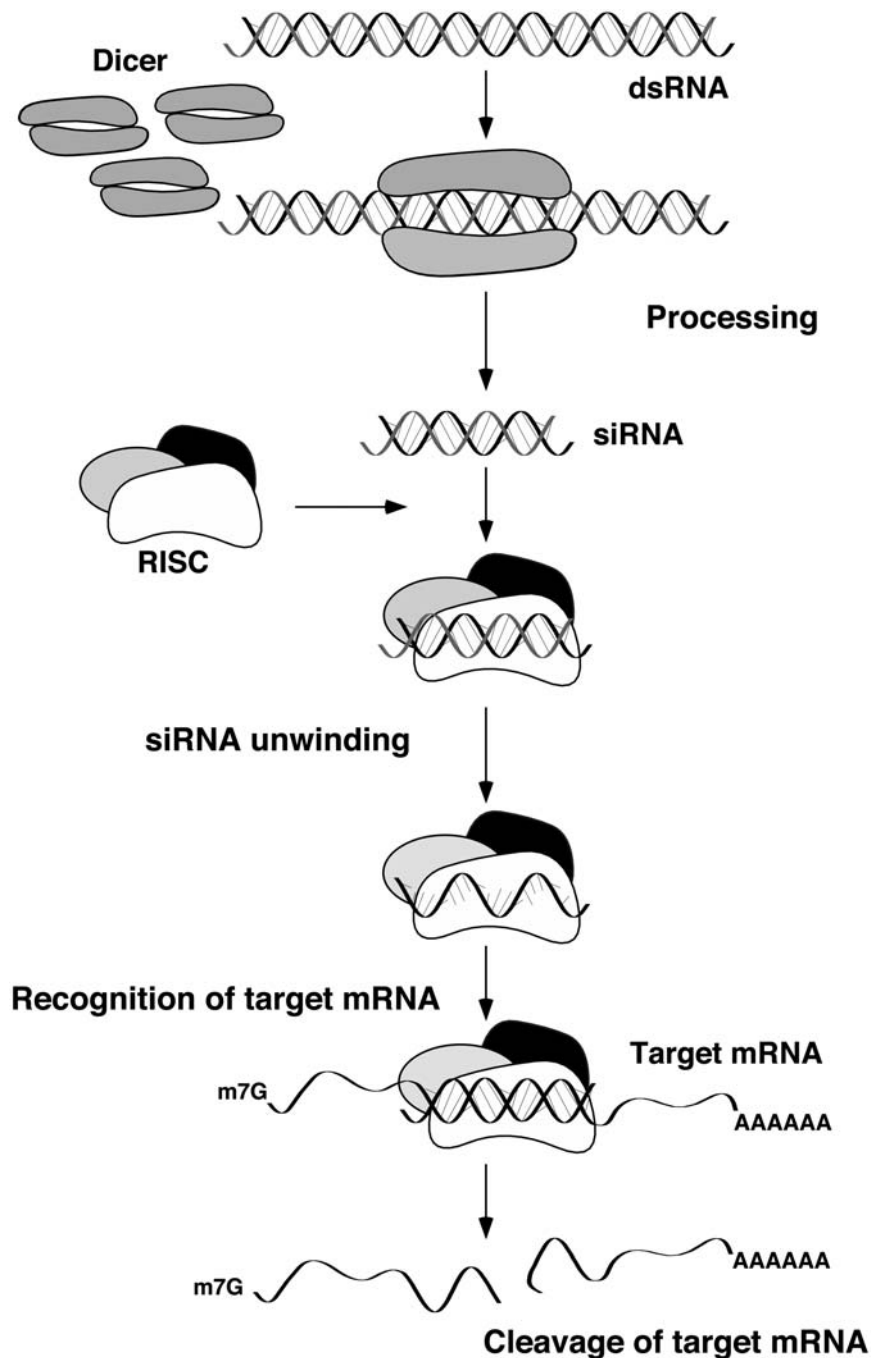
Clinical Relevance

Clinical Application of RNAi

The phenomenon of siRNA-mediated gene silencing *in vivo* is extremely powerful and sequence-specific. Thus, RNAi technology clearly has significant potential for application in a clinical setting. Indeed, considerable progress has been made in assessing the potential utility of RNAi that is directed against specific pathogenic genes (Table 1).

Studies of the potential utility of siRNAs as anti-viral agents have already shown that synthetic and vector-based siRNAs can inhibit the replication of HIV-1 at several stages of infection, including the very early stage. Moreover, infection by HIV-1 was blocked by targeting either viral genes or host genes that are involved in viral infection or proliferation. RNAi technology has also been applied to the evaluation of drugs against hepatitis C virus (HCV) using an HCV-replication system that supports the replication of the virus, but not the production of infectious virions. siRNAs directed against HCV RNA efficiently eliminated replicating HCV RNA from human hepatoma cells. Furthermore, siRNAs targeted to the 5'-untranslated region (5'UTR) of HCV suppressed HCV replication by 80% at a relatively low dose of siRNAs. Vector-based siRNAs also significantly suppressed the replication of HCV. Nonetheless, most RNA viruses can become resistant to single-drug therapies because they can mutate rapidly. Thus, for example, siRNAs directed against the poliovirus genome failed to prevent production of viruses completely because of the rapid mutation of this virus. Multidrug and multitarget strategies will have to be developed as effective methods for preventing viral proliferation.

The potential utility of RNAi in cancer therapy is also being examined. Most human tumors have multiple mutations that include dominant mutations in oncogenes or tumor suppressor genes. For example, tumor



Catalytic RNAs. Figure 2 A model for RNAi. The RNase III-related endonuclease, known as Dicer, cleaves dsRNA into siRNAs. Subsequently, siRNAs associate with the RISC, and then this enzyme catalyzes the sequence-specific degradation of target mRNA.

cells with mutated genes for p53 can proliferate at abnormally high rates. p53 is a transcription factor that plays a crucial role in the maintenance of cell-cycle arrest at the G1 phase after DNA damage. The *RAS* genes (*K-RAS*, *H-RAS* and *N-RAS*) encode guanine nucleotide-binding proteins that interact with the inner

plasma membrane and transduce external signals to the interior of the cell. *RAS* genes are frequently mutated in some human cancers, in particular, pancreatic and colon carcinomas. Single base-pair mutations that affect the functions of the *p53* or *RAS* genes occur frequently during oncogenesis. RNAi can specifically

Catalytic RNAs. Table 1 Examples of the application of siRNA, targeted against transcripts of genes related to infectious diseases or cancer

Virus or disease	Target gene or protein	Cell line	Source of cells	Type of siRNA	Phenotype
Human immunodeficiency virus type 1 (HIV-1)	rev-EGFP	293T	Human embryonic kidney	Non-viral vector	Suppression of viral proliferation
	Tumor susceptibility gene 101 (Tsg 101)	293T	Human embryonic kidney	Synthetic siRNA	Suppression of budding
	<i>tat, rev</i>	293T	Human embryonic kidney	Synthetic siRNA	Suppression of viral proliferation
	<i>tat, rev</i>	Jurkat	Acute T-cell leukemia	Synthetic siRNA	Suppression of viral proliferation
	Reverse transcriptase, NFκB p65	Jurkat	Acute T-cell leukemia	Synthetic siRNA	Suppression of viral proliferation
	CD4	Magi-CD4	Cervical carcinoma	Synthetic siRNA	Suppression of viral infection
	Long terminal repeat (LTR), <i>nef, vif</i>	Magi-CD4	Cervical carcinoma	Synthetic siRNA	Suppression of viral proliferation
	p24	HeLa-CD4	Cervical carcinoma	Synthetic siRNA	Suppression of viral proliferation
	p24	ACH2	T-cell lymphocytes	Synthetic siRNA	Suppression of viral proliferation
	GFP	H9	T-cell lymphocytes	Synthetic siRNA	Suppression of viral proliferation
	CCR5	CD4+T	T-cell lymphocytes	Lentiviral vector	Suppression of viral infection
	<i>gag, pol</i>	HOS.T4. CXCR4	Osteosarcoma	Synthetic siRNA	Suppression of viral proliferation
Hepatitis C virus (HCV)	Nonstructural protein 5B (NS5B)-Fluc	Mouse liver	Mouse liver	Non-viral vector	Suppression of viral proliferation
	NS3, NS5B	Huh-7	Human hepatoma	Synthetic siRNA	Suppression of viral proliferation
	NS5B	Huh-7	Human hepatoma	Non-viral vector	Suppression of viral proliferation
	5'UTR	Huh-7.5	Human hepatoma	Synthetic siRNA	Elimination of virus
Hepatitis B virus (HBV)	Core or X open reading frame (ORFs)	Huh-7	Human hepatoma	Non-viral vector	Suppression of viral replication
	Core ORF	HepAD38,79	Human hepatoma	Synthetic siRNA	Suppression of viral replication
Poliovirus	Capsid, polymerase	HeLaS3	Epitheliod carcinoma	Synthetic siRNA	Suppression of viral proliferation
	Firefly luciferase (Fluc)	MEF	Mouse embryo fibroblast	Synthetic siRNA	Suppression of viral proliferation
Human papillomavirus	E6, E7	Caski, SiHa	Cervical carcinoma	Synthetic siRNA	Apoptosis
	E6	Caski, SiHa	Cervical carcinoma	Synthetic siRNA	Increase in level of p53

Catalytic RNAs. Table 1 Examples of the application of siRNA, targeted against transcripts of genes related to infectious diseases or cancer (Continued)

Virus or disease	Target gene or protein	Cell line	Source of cells	Type of siRNA	Phenotype
Breast cancer	<i>p53</i>	MCF-7	Mammary adenocarcinoma	Non-viral vector	Suppression of G1 arrest
Pancreatic cancer	K-RAS V12	CAPAN-1	Human pancreatic carcinoma	Retroviral vector	Suppression of tumorigenicity
Colon cancer	K-ras	SW480	Human colon carcinoma	Non-viral vector	Suppression of proliferation
Prostate cancer	Zeste homolog2 (EZH2)	RWPE	Prostate carcinoma	Synthetic siRNA	G2/M arrest
Prostate cancer	Fatty acid synthase	LNCap	Prostate carcinoma	Synthetic siRNA	Suppression of cell proliferation
Liver cancer	<i>cyclin E</i>	Hep3B, HepG2	Hepatocellular carcinoma	Synthetic siRNA	Suppression of cell proliferation
Angiogenesis	VEGF	cJ4	Rat fibrosarcoma	Synthetic siRNA	Suppression of TSP1 resistance
Colon cancer	hTERT, hTR	HCT-15	Colon carcinoma	Synthetic siRNA	Suppression of telomerase activity
CML	<i>bcr/abl</i>	K562	Chronic myeloid leukemia	Synthetic siRNA	Suppression of bcr/abl

inhibit the expression of *p53* or *RAS* genes with single point mutations, without affecting the wild type genes. In attempts to combat leukemia, siRNAs directed against *BCR/ABL* mRNA specifically disrupted the expression of the chimeric gene but did not affect the expression of normal *BCR* and *ABL* genes in cultured cells. The development of other siRNAs might lead to effective treatments for other cancers. For example, it was reported that siRNAs can inhibit telomerase activity in human cancer cells. ▶ **Telomerase** activity protects cells from senescence and might be an important target for anti-cancer drugs. siRNAs, targeted to either human telomerase reverse transcriptase (hTERT) or to human telomerase RNA (hTR), have been shown to reduce telomerase activity in carcinomas and sarcomas.

Successful RNAi has been demonstrated not only in cultured mammalian cells but also in mice. For example, when siRNA that was designed to target a gene for luciferase was introduced into mice that expressed luciferase, by the rapid injection in a large volume of physiological solution into the tail vein, levels of expression of luciferase were suppressed by 80–90% in the liver, spleen, lung, kidney and pancreas. Delivered by the same technique, siRNAs designed to target the NS5B region of HCV suppressed the expression of luciferase in mice that harbored a gene

for luciferase fused to the NS5B of HCV. Viral and non-viral vectors should also be useful in a clinical context.

Future Applications

Recent advances have shown that RNA-based technologies have obvious and significant potential in the treatment of patients with malignant disorders. Ribozymes are now being developed not only as potential therapeutic agents but also as tools for the identification of genes involved in certain diseases, such as certain cancers. In addition, RNAi technology now allows researchers to modulate gene expression more easily and, thus, to clarify and identify the role of known and unknown genes. The recent discovery of yet another new class of RNAs, micro-RNAs (miRNAs), suggests that they too hold great promise as future therapeutic agents. miRNAs are small non-coding RNAs of 18–25 nt in length. Numerous miRNAs have been found in animals and plants. Some act as repressors of their target mRNAs at the translational level. Such repression requires the presence of partially complementary sequences in the 3'-untranslated regions (UTRs) of the target mRNAs. It was reported that miRNA-23 participates in the retinoic acid-induced neuronal differentiation of NT2 cells, which are human embryonal carcinoma (EC) cells (7). miRNA-23 is expressed in differentiated NT2 cells

and promotes neuronal differentiation *via* regulation of the expression of Hes1, which is a negative regulator of the differentiation of NT2 cells. These results suggest that it might even be possible to regulate the differentiation of stem cells exogenously by introducing appropriate miRNAs. This technology suggests that, in the future, it might be feasible to engineer the phenotypes of certain cells and tissues, in addition to preventing the unchecked proliferation of malignant cells.

References

1. Pavco PA, Bouhana KS, Gallegos AM et al (2000) Antitumor and antimetastatic activity of ribozymes targeting the messenger RNA of vascular endothelial growth factor receptors. *Clin Cancer Res* 6:2094–2103
2. Yu M, Ojwang J, Yamada O et al (1993) A hairpin ribozyme inhibits expression of diverse strains of human immunodeficiency virus type 1. *Proc Natl Acad Sci USA* 90:6340–6344
3. Kawasaki H, Eckner R, Yao TP et al (1998) Distinct roles of the co-activators p300 and CBP in retinoic-acid-induced F9-cell differentiation. *Nature* 393:284–289
4. Tanabe T, Kuwabara T, Warashina M et al (2000) Oncogene inactivation in a mouse model: tissue invasion by leukaemic cells is stalled by loading them with a designer ribozyme. *Nature* 406:473–474
5. Suyama E, Kawasaki H, Kasaoka T et al (2003) Identification of genes responsible for cell migration by a library of randomized ribozymes. *Cancer Res* 63:119–124
6. Elbashir SM, Harborth J, Lendeckel W et al (2001) Duplexes of 21-nucleotide RNAs mediated RNA interference in mammalian cell culture. *Nature* 411:494–498
7. Kawasaki H, Wadhwa R, Taira K (2004) World of small RNAs: from ribozymes to siRNA and miRNA. *Differentiation* 72:58–64

Cataplexy

Definition

Cataplexy is a sudden and short-lasting loss of muscle tone triggered by emotions. Cataplexy is the pathognomonic symptom of narcolepsy.

► [Narcolepsy](#)

Catecholamines

Definition

Catecholamines comprise of a group of amines that are produced in the medulla of the adrenal gland. They can

act as neurotransmitters and hormones. Examples are epinephrine, norepinephrine, and dopamine.

► [Cardiac Signaling: Cellular, Molecular and Clinical Aspects](#)

Catenin

► [Adhesion Molecules](#)

► [Beta-Catenin](#)

► [Cadherins](#)

Caveolin-1

Definition

Caveolin-1 is a 21 kDa integral membrane protein, which is a major structural component of caveolae, small invagination of the plasma membrane characteristic of many mammalian cells and associated with ► [endocytosis](#). Caveolin is inserted in the plasma membrane as a hair pin structure with both the N and the C terminal facing the cytosol.

► [Signal Transduction: Integrin-Mediated Pathways](#)

CBF

► [Core Binding Factor](#)

CBP

Definition

CBP describes a Cap-binding protein complex that binds to the first nucleotides of mRNA, to protect the RNA molecule from degradation.

► [Fragile X Syndrome](#)

► [RNA Polymerase II Transcription](#)

CBP (CREB-Binding Protein)/p300

Definition

CBP stands for CREB-binding protein, a transcription co-activator with intrinsic histone acetyltransferase activity. P300 is another histone acetyltransferase closely related to CBP.

► [Methylation of Proteins](#)

► [Polyglutamine Disease, the Emerging Role of Transcription Interference](#)

CCALD

► [Adrenoleukodystrophy](#)

► [Childhood Cerebral ALD](#)

CCD

► [Charge Coupled Device cDNA](#)

CCFDN

Definition

CCFDN refers to a disorder that is characterized by Congenital Cataracts, Facial Dysmorphism and Neuropathy Syndrome.

► [Hereditary Neuropathies, Motor and/or Sensor](#)

CD

► [Cluster of Differentiation](#)

CD8

Definition

The CD8 cell surface protein is expressed by T lymphocytes, and provides a costimulatory signal in the specific peptides bound to MHC class I molecules, by interacting with the $\alpha 3$ domain of the latter.

► [DNA-based Vaccination](#)

CD147

Definition

CD147 is a broadly expressed plasma membrane glycoprotein of the immunoglobulin superfamily, which stimulates the production of ► [matrix metalloproteinases](#) (MMPs).

► [Peptidyl Prolyl Cis/Trans Isomerases](#)

CDC42

► [Rho, Rac, Cdc42](#)

► [Tangier Disease](#)

CDGs

► [Congenital Disorders in Glycosylation](#)

CDK

CDK defines cyclin-dependent kinase. They are closely related catalytic subunits of serine/threonine kinases, which require association with the positive regulatory cyclin for activation to control progression through the cell cycle.

► [Cell Cycle – Overview](#)

► [Cyclin-Dependent Kinase](#)

► [Growth Factors](#)

► [Receptor Serine/Threonine Kinase](#)

► [Rho, Rac, Cdc42](#)

► [Senescence](#)

cDNA

Definition

cDNA (complementary DNA) is a DNA that is copied from an RNA molecule with the help of the enzyme “reverse transcriptase”, and therefore does not contain any introns that are present in genomic DNA. In contrast to the unstable and single stranded RNA, the cDNA is chemically more stable and can be easily manipulated (e.g. cloning, sequencing).

- ▶Automated High-Throughput Functional Characterization of Human Proteins
- ▶Full Length cDNA Sequencing
- ▶Multifactorial or Common Diseases
- ▶Protein Microarrays as Tool for Protein Profiling and Identification of Novel Protein-Protein Interactions
- ▶Recombinant Protein Production in Mammalian Cell Culture
- ▶Repetitive DNA

reading frame being present, especially at the 5 prime end. In contrast to a cDNA library, the clones could originate from various sources.

- ▶Automated High-Throughput Functional Characterization of Human Proteins

cDNA Microarray

Definition

cDNA microarray describes a cDNA library that is arranged on a glass slide in a defined pattern by which each cDNA clone is defined by its position on the slide.

- ▶Medaka as a Model Organism for Functional Genomics

cDNA (Expression) Library

Definition

CDNA (expression) library represents a collection of all mRNA molecules present in a cell, tissue or organism, which have been reversely transcribed containing the complete protein coding part of the respective genes into cDNA molecules, and inserted into expression vectors. This type of library contains only coding DNA (genes and 5'and 3' untranslated regions) and does not include any non-coding DNA.

- ▶Automated High-Throughput Functional Characterization of Human Proteins
- ▶DNA Microarrays/DNA Arrays
- ▶Full Length cDNA Sequencing
- ▶Protein Microarrays as Tool for Protein Profiling and Identification of Novel Protein-Protein Interactions

Cdx Genes

Definition

Cdx genes encode ▶homeobox transcription factors. Cdx–1 and Cdx–2 genes are the mammalian homologues of the *Drosophila caudal* gene. They are expressed in extraembryonic areas, and in later developmental and adult stages are restricted to the intestinal epithelium, where they are involved in the definition of intestinal epithelial cell identity and homeostasis between proliferation and differentiation.

- ▶Gut Epithelium

cDNA Collection (Full Length)

Definition

A collection of cDNA clones based upon accurate sequencing determination to ensure the complete

ced

Definition

Caenorhabditis elegans programmed cell death mutant.

- ▶Caenorhabditis elegans as a Model Organism for Functional Genomics

Cell Adhesion

Definition

Cell adhesion describes the binding of cells to other cells (cell-cell adhesion) or to extra cellular matrix (cell-matrix adhesion) through proteins embedded in the plasma membrane. Major families of cell adhesion molecules comprise of the Ca^{2+} -dependent cadherins, Ca^{2+} -independent immunoglobulin CAMS, selectins and integrins. Cell-cell adhesion occurs between cells of the same (homotypic), or of different (heterotypic), nature through the same (homophilic) or through different (heterophilic) adhesion molecules.

- ▶ Adhesion Molecules
- ▶ Cadherins
- ▶ Cancer Invasion and Metastasis: Cellular, Molecular and Clinical Aspects
- ▶ Cell Polarity
- ▶ Desmosomes

Cell Based Assay

Definition

Cell based assays comprise of a variety of assays that measure cell adhesion, proliferation, toxicity, motility, production of a measurable product, morphology or staining-patterns of subcellular structures. Cell-based assays offer a more accurate representation of the real-life model, since living cells are used in contrast to biochemical experiments.

- ▶ Automated High-Throughput Functional Characterization of Human Proteins

Cell Cycle Checkpoints

Definition

Cell cycle checkpoints refer to mechanisms that monitor cell cycle progression and prevent premature entry into the next phase of the cell cycle. In general, activation of cell cycle checkpoints leads to the arrest of cell cycle progression in the presence of damaged DNA. If repair is successful, cell cycle proceeds. In cases where the repair fails, the cell undergoes apoptosis or becomes tumorigenic (cancer).

- ▶ Cell Cycle – Overview
- ▶ Cell Division

- ▶ Centromeres
- ▶ Double-Strand Break Repair

Cell Cycle – Overview

LUDGER HENGST¹, ERICH A. NIGG²

¹Biocenter-Innsbruck Medical University, Innsbruck, Austria

²Max-Planck-Institute for Biochemistry, Martinsried, Germany

ludger.hengst@uibk.ac.at,
nigg@biochem.mpg.de

Definition

The cell division cycle describes the ordered sequence of events that cells undergo in order to duplicate. Cell divisions are not only required for producing the approximately 10^{14} cells that make up an adult human being, starting from a single fertilized egg, but also during adulthood, when many cell types need to be continuously replaced. Appropriate control over the cell-cycle is critical in order to produce the appropriate number of cells in every tissue and to secure the error-free propagation of the genome. Major tasks to be accomplished during the cell cycle are the precise duplication of the genome (DNA) and its accurate segregation (in the form of chromosomes) into the two daughter cells. In eukaryotic cells, the duplication and segregation of the genome usually occur as two discrete and non-overlapping processes; DNA replication occurs during ▶S phase (S for synthesis) and chromosome segregation during ▶M phase (M for mitosis). M and S phases are separated by two gap phases, termed G1 and G2. This basic outline of the cell-cycle and the molecular wiring of its controls have been highly conserved during eukaryotic evolution, but individual cell types and species may exhibit significant variations in the temporal and spatial choreography of cell-cycle events. Cell-cycle control relies not only on transcriptional and translational regulation but also, to a large extent, on posttranslational mechanisms, primarily phosphorylation and the proteolytic degradation of key regulatory proteins. Surveillance mechanisms, so called checkpoint controls, assure the proper order and completion of cell-cycle events. Internal or external signals also control the transitions between the cell-cycle and a state of quiescence. Deregulation of the cell-cycle may result in aberrant responses to mitogenic stimulation and/or the destabilization of the genome and thus give rise to cancer and other hyperproliferative disorders.

Characteristics

Most somatic cells in an adult organism exist in a non-dividing, resting state (G0). However, stem cells enter the cell-cycle as required for tissue homeostasis and some resting cells can also be re-activated to replace cells that have been lost due to injury or cell death. Most cells exit the cell-cycle from early ►G1 phase, but exit from ►G2 phase also occurs (e.g. with certain epidermal cells). Arguably the most spectacular stage of the cell-cycle is M phase, when chromosome segregation (mitosis) and cell division (cytokinesis) occur. Proliferating human cells in culture usually undergo M phase in about 1 h, whereas ►interphase occupies about 17–24 h. Interphase, the time between one M phase and the next, thus consists of G1 phase, S phase and G2 phase.

G1 Phase

During G1 phase, cells grow in size. Moreover, they choose during early G1 between continued division or withdrawal from the cell-cycle. Cells may exit into ►G0 phase if unfavorable conditions prevail, or if they receive anti-mitogenic signals from other cells. Cells in G0 remain biochemically active but undergo no DNA synthesis or cell divisions. They may be prompted to re-enter the cell-cycle (G1 phase) by appropriate mitogenic stimulation. In the presence of favorable conditions and mitogenic stimuli, cells proceed through the cycle. A point of no return exists in G1 phase where cells become committed to undergo a complete round of cell division, even if unfavorable signals subsequently prevail. This checkpoint in G1 of animal cells was termed the restriction point (R-point; equivalent to Start in yeast) and divides the G1 phase into early G1 (when cells are responsive to growth or differentiation signals) and late G1 (after commitment to cell division, when cells prepare for DNA replication).

S Phase

Both the genome and the ►centrosome (the microtubule organizing center of animal cells) need to be duplicated exactly once per cell-cycle. In mammalian cells, DNA replication and duplication of the centrioles (short cylinders of microtubules that lie at the center of the centrosome), take place during S phase. Both events are coordinated by the cell-cycle control system. This system regulates the expression of enzymes required for DNA synthesis; it activates origins of replication at the beginning of S phase and prevents the reinitiation of DNA synthesis (and centriole re-duplication) at later stages of the cell-cycle.

G2 Phase

The G2 phase separates S phase from M phase. It can provide additional time for growth before cells enter mitosis. Checkpoint controls during G2 phase can

arrest the cell-cycle if DNA has been damaged or if not all DNA has been replicated.

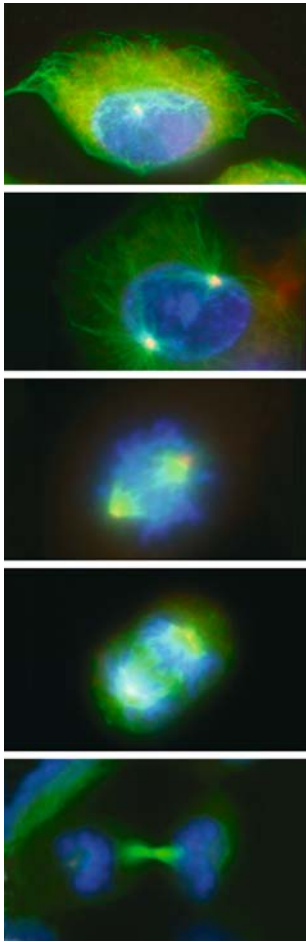
M Phase

During M phase, sister ►chromatids, representing duplicated and connected strands of DNA that were originally generated during S phase, are separated and segregated. Moreover, in tight temporal and spatial coordination with the segregation of the duplicated genome, cytokinesis, the actual process of cell division takes place. Each daughter cell thus receives an exact copy of the genome, one centrosome and similar shares of other cytoplasmic organelles and macromolecules. Mitosis in vertebrate cells is usually subdivided into five stages. In the first stage, prophase, replicated chromosomes undergo condensation, the duplicated centrosomes migrate apart and spindle assembly begins (Fig. 1). Breakdown of the nuclear envelope then marks the beginning of prometaphase and sister chromatids can thus be captured by spindle microtubules. (The ►mitotic spindle is a bipolar microtubule array that serves to segregate attached sister chromatids to opposite poles. This requires that two microtubule bundles, emanating from opposite poles of the spindle, attach to ►kinetochores, proteinaceous complexes assembled at the so-called centromeres of the sister chromatids). Prometaphase ends when all sister chromatids are properly bound to the spindle and have migrated to its center, the metaphase plate. During metaphase the sister chromatids remain aligned on the spindle. A signal to separate sister chromatids then initiates anaphase and mitotic exit. Importantly, this signal is delayed by the spindle checkpoint until all chromosomes have undergone bipolar attachment. In anaphase the sister chromosomes are pulled to opposite poles of the spindle (anaphase A) and the spindle poles themselves move further apart (anaphase B). Finally, in telophase, the spindle disassembles and the nuclear envelope reforms around the decondensing chromosomes.

Temporally overlapping with exit from mitosis, cells proceed through cytokinesis. Animal cells form a cleavage furrow with the help of a contractile ring (composed of actin bundles and myosin II) that gradually pulls their membranes inwards. Membrane depositions then allow the cell surface to grow and the final separation (abscission) of two newborn daughter cells occurs.

Specialized Cell-Cycles

Even though the basic outline of the cell-cycle described above is used by most eukaryotic cells, there are important variations from this scheme. Most notable is ►meiosis, the process by which diploid parental cells give rise to haploid germ cells. In this case, a single round of DNA replication is followed by



Cell Cycle – Overview. Figure 1 Cell division seen under the fluorescence microscope: Dramatic changes in cell architecture occur as cells (here human osteosarcoma U2OS cells) enter mitosis. DNA (blue, stained with Hoechst) condenses, the centrosomes (orange, highlighted with an antibody against γ -tubulin) move apart and microtubules (green, detected by an antibody against α -tubulin) reorganize to form the mitotic spindle. From left to right: interphase; prophase; prometaphase; anaphase and cytokinesis. (Courtesy of Elena Nigg)

two sequential divisions. During meiosis I, the homologous chromosomes pair and segregate from each other. Sister chromatids then separate during meiosis II, which resembles mitosis in somatic cells. Another deviation of the canonical cell-cycle is seen during endoreduplication (e.g. in salivary glands of insects) when multiple rounds of DNA replication occur without intervening M phases. Finally, in the embryonic cell-cycles of some species, G1 and G2 phases are largely suppressed, so that the fast alternation of S and M phases results in extremely rapid divisions.

Regulatory Mechanisms

To ensure that cell-cycle events are triggered in the correct order and only once per cycle, a complex regulatory network governs all major cell-cycle transitions. Moreover, surveillance mechanisms monitor the completion and accuracy of cell-cycle events. When irregularities are detected (e.g. DNA damage, incomplete DNA replication, incorrect chromosome attachment to the spindle), these surveillance mechanisms block cell-cycle progression at specific times and are thus referred to as checkpoints.

Cell-cycle control is to a large extent based on a combination of regulated protein synthesis, phosphorylation and proteolysis. These mechanisms are interdependent in that phosphorylation controls the machinery regulating synthesis and degradation, and conversely, many enzymes involved in phosphorylation are controlled by synthesis and degradation. Together, these mechanisms assure the unidirectional progression through the cell-cycle. Irreversibility is brought about by positive feedback loops and protein destruction.

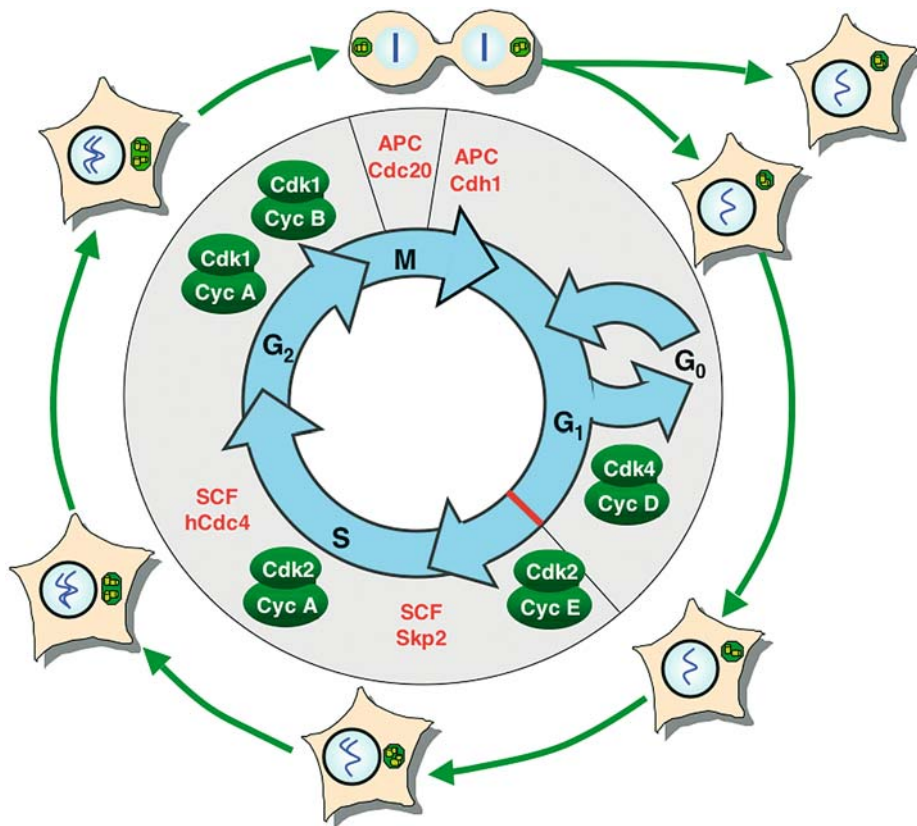
Principles

Kinases and Phosphatases

At the center of cell-cycle control are the protein kinases of the **cyclin-dependent kinase (CDK)** family. The most prominent members of this family function in cell-cycle control (CDK1, 2, 4 and 6) and CDK activation (CDK7), others control basal transcription by RNA polymerase II (e.g. CDK7,8 and 9) as well as migration and differentiation of post mitotic neurons (CDK5). As indicated by their name, activation of the inactive CDK requires binding of a positive regulatory subunit called cyclin. Several cyclins are typically expressed at specific stages of the cell cycle (Fig. 2). These cyclins are usually classified as G1, G1/S, S and M phase cyclins.

The catalytic activity of CDKs is controlled by distinct mechanisms. In addition to cyclin binding, these include activating and inhibitory phosphorylations and binding of CDK inhibitory proteins. Inhibitory phosphorylation of CDK1 by specific kinases (Wee1, Myt1) and activating dephosphorylation by the Cdc25C phosphatase are especially important at the onset of mitosis. During interphase, the binding of members of two families of CDK inhibitors (**CKIs**) plays a key regulatory role especially during the G1/S transition.

In their function as master regulators of the cell cycle, CDKs are supported and controlled by additional protein kinases. Some of these kinases are implicated in the regulation (both positive and negative) of CDK activity. In turn, some of these non-CDK kinases (as well as phosphatases) are under the control of CDKs. For example, Cdc7 cooperates with Cdk2 during replication and Polo-like kinases (Plks)



Cell Cycle – Overview. Figure 2 The cell cycle. Key regulatory proteins are indicated next to the position of the cell cycle where they are active. Cyclin dependent kinase complexes are shown in green and E3 ubiquitin ligases are shown in red. The red bar in G₁ phase represents the restriction point.

and Aurora kinases play central roles, along with Cdk1, in regulating the progression through mitosis.

Synthesis and Proteolysis

Periodic expression of most cell-cycle regulatory proteins relies on a balance between their synthesis and their destruction. In most cases, regulation of synthesis depends on transcriptional mechanisms. This frequently involves activation of transcription factor complexes or inactivation of repressors, processes that can in turn be regulated by phosphorylation.

Some cell-cycle regulatory proteins are relatively stable during most phases of the cell-cycle but become unstable at other specific times. Destruction of these proteins frequently involves attachment of poly-ubiquitin chains and subsequent degradation by the proteasome. Key determinants conferring specificity for degradation are ubiquitin ligases. Of this family of enzymes, two sub-families play a central role in the proteolysis of cell-cycle regulators. At the G₁/S transition, ubiquitin ligases belonging to the ►SCF sub-family target CKIs, G₁ cyclins and other substrates for proteolysis. SCF complexes (SCF stands for Skp1,

Cullin and F-box protein) contain three core subunits (Skp1, Cul1 and Rbx1) and one interchangeable subunit known as F-box protein. Each F-box protein targets a subset of substrates for destruction. Substrate recognition frequently requires prior modification by phosphorylation.

The second class of ubiquitin ligases important for cell cycle control is the ►APC/C complex (APC/C stands for anaphase-promoting complex/cyclosome). Proteolysis initiated by APC is required for the metaphase to anaphase transition. In contrast to SCF, APC function does not require phosphorylation of the substrate but activation of the APC core enzyme. At least two different activating subunits, termed Cdc20 and Cdh1, associate consecutively with APC and control its activity from the metaphase/anaphase transition (Cdc20) until late G₁ phase (Cdh1).

Subcellular Localization

The function of several cell-cycle regulatory proteins is also controlled at the level of their localization. Some cyclins, for instance, shuttle between the nucleus and the cytoplasm. Import into and export from the nucleus

usually depend on specific signal sequences, termed NLS (nuclear localization sequence) and NES (nuclear export signal), respectively. The functionality of these signals during the cell cycle can be subject to regulation by phosphorylation/dephosphorylation and/or interactions with other proteins.

Cell-Cycle Concepts

The G₁/S Transition and the Retinoblastoma Protein (pRb) Pathway

The decision as to whether to continue proliferation or exit from the cell-cycle is made during G₁ phase. Upon passage through a specific point in G₁, called the restriction point (R-point), cells become insensitive to lack of growth factors or differentiation signals (Fig. 3). In contrast to normal cells, most cancer cells appear to have lost R-point control. At the molecular level, R-point passage was proposed to depend on a short-lived protein, whose synthesis would be sensitive to growth factors. D-type cyclins meet this criterion. Through their association with CDKs, and in concert with cyclin E, they indeed control the G₁/S transition. The major substrate of the CDKs that are activated by D-type cyclins is the **retinoblastoma protein** (pRb). This tumor suppressor gene product becomes increasingly phosphorylated at multiple sites during G₁ progression. When hypophosphorylated, pRb binds transcription factors, notably members of the E2F family, and recruits them into a repressor complex. Mitogenic stimulation leads to activation of cyclin D/CDK4 or 6. Cyclin D dependent initial

phosphorylation of pRb then leads to release of transcriptional repression and activation of the cyclin E promoter. The cyclin E/CDK2 kinase is initially kept inactive by CDK inhibitor proteins of the Cip/Kip family. Destruction of these inhibitors allows the kinase to further phosphorylate pRb, leading to further activation of E2F-mediated transcription. As a result, genes are expressed that encode proteins required for DNA synthesis, inactivation of APC/C, and further progression through the cell-cycle. These genes include cyclin A and CDK1.

DNA damage signaling activates transcription by the p53 tumor suppressor. Among the proteins induced by p53 is the CDK inhibitor p21. This inhibitor inactivates cyclin E/CDK2 and prevents the hyperphosphorylation of pRb, thereby leading to cell-cycle arrest. Thus, DNA damage inhibits the G₁/S transition, in part through the pRb pathway. Components of the p53 and the pRb pathways are among the most frequently mutated proteins in human cancer.

A Single Round of DNA Replication

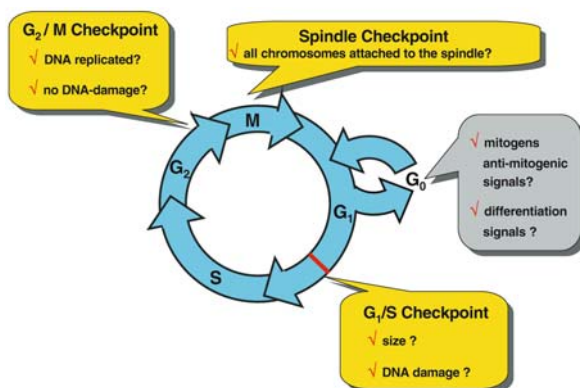
DNA replication initiates at origins of replication. The assembly and activity of the replication machinery at origins is controlled by several kinases, including CDKs. CDK activity is first required for the activation of origins and later prevents reinitiation of replication during late S, G₂ and M phases. During these latter stages, active CDKs inhibit the formation of pre-replication complexes and thus prevent “licensing” of DNA for replication.

The G₂/M Transition

The G₂/M checkpoint arrests cells prior to entry into mitosis if DNA is unreplicated or damaged. Damage signaling involves several kinases that up-regulate Wee1 and Myt1 and inhibit Cdc25C. This prevents the activation CDK1, which is required for entry into mitosis. DNA damage signaling also activates p53, which induces transcription of proteins such as p21, GADD45 and 14-3-3 and, moreover, inhibits the mitotic Plk1 kinase. In the absence of checkpoint signaling, both CDK1 and Plk1 become activated and a positive feedback loop supports maximal activation of CDK1/cyclin B.

The Spindle Assembly Checkpoint and Exit from Mitosis

A key checkpoint prevents sister chromatid separation until all chromosomes are properly attached to the mitotic spindle. Bipolar attachment of microtubules to kinetochores generates tension. Both attachment and/or tension can be monitored. Even a single unattached kinetochore activates the checkpoint and prevents the metaphase to anaphase transition. After all chromosomes are properly aligned on the spindle, the APC/C is activated by Cdc20 and triggers proteolysis of a



Cell Cycle – Overview. Figure 3 Cell cycle checkpoints. Negative signals from incomplete earlier cell cycle events or damage signals can provoke cell cycle arrest at specific points called checkpoints (yellow). External signals (grey) like lack of growth factors or differentiation signals may lead to withdrawal from the cell cycle into the G₀ phase. After transition through the restriction point, indicated by a red bar, cells become insensitive to growth factors and differentiation signals and committed to one round of cell division.

subset of proteins, notably cyclins and the anaphase-inhibitor securin. Destruction of securin activates separase, a protease that cleaves cohesion proteins, allowing the separation of sister chromatids. In parallel, the progressive inactivation of Cdk1, due to cyclin destruction, initiates mitotic exit. Finally, chromosome segregation has to be coordinated in time and space with the formation of a cleavage furrow (Fig. 1). The error-free segregation of complete sets of chromosomes in every cell division is essential to prevent the onset of chromosomal instability, a hallmark of human cancers.

- [Cell Division](#)
- [Double-Strand Break Repair](#)
- [Growth Factors](#)
- [Methylation of Proteins](#)
- [Muscle Development](#)
- [RNA Polymerase III](#)
- [Tumor Suppressor Genes](#)

References

1. Nurse P (2000) A long twentieth century of the cell cycle and beyond. *Cell* 100:71–78
2. Hartwell LH, Kastan MB (1994) Cell cycle control and cancer. *Science* 266:1821–1828
3. Morgan DO (1997) Cyclin-dependent kinases: engines, clocks, and microprocessors. *Annu Rev Cell Dev Biol* 13:261–291
4. Nigg EA (2001) Mitotic kinases as regulators of cell division and its checkpoints. *Nat Rev Mol Cell Biol* 2:21–32
5. Bell SP, Dutta A (2001) DNA replication in eukaryotic cells. *Annu Rev Biochem* 71:333–374
6. Sherr CJ, McCormick F (2002) The RB and p53 pathways in cancer. *Cancer Cell* 2:103–112

Cell Death

Definition

Cell death refers to the complete degeneration of cells by ► [necrosis](#) or ► [apoptosis](#).

- [Repeat Expansion Diseases](#)

Cell Differentiation

Definition

Cell differentiation designates the development of specialised structures and functions from unspecialised precursor cells.

- [RNA Polymerase III](#)

Cell Division

CAROLINE BODEN, ULRIKE ZIEBOLD

Max Delbrück Centre for Molecular Medicine, Berlin, Germany

uziebold@mdc-berlin.de

Definition

Cell division is the process by which a cell separates itself into two parts (infrequently more parts) to generate two genetically equal daughter cells.

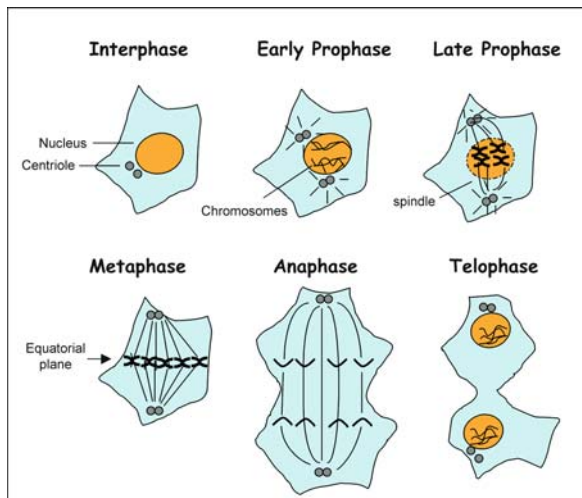
Characteristics

The physical division of cells is a requirement of natural growth in unicellular organisms as much as in complex organisms. In order to multiply a cell needs to duplicate its own composition, cytoplasmic organelles and genetic material. Only then will cell division proceed by separation of the nucleus (► [mitosis](#)) and cytoplasm (► [cytokinesis](#)). This process is restricted to somatic cells, since germ cells form four haploid cells out of one diploid cell (► [meiosis](#)). In normal somatic cells, division occurs as part of a complete cell-cycle (“cell division cycle”). Moreover, the decision for cytokinesis and mitosis are linked to the cell-cycle at several stages. This limits ► [centrosome](#) amplification and DNA replication to once per cell-cycle. In all organisms cell division is highly regulated. Molecules required for the regulation of this process are often highly conserved in evolution. During development of metazoans cell division may occur symmetrically or asymmetrically. The orchestrated events leading to cell division are important contributors to human diseases, specifically ► [cancer](#), if de-regulated.

Plant cells differ significantly from animal cells. Plant cells lack certain structures such as centrosomes. On the other hand, plant cells have specialized structures such as specific cytoskeletal structures, phragmoplasts and unique microtubule associated proteins. These differences lead to subtle differences in cell division. To describe cell division and its dependence on proper execution of mitosis and cytokinesis, we will focus on details in animal cells.

Mitosis and Cytokinesis

Mitosis has long been recognized to occur in stages. These stages are inter-, pro-, meta-, ana- and telo-phase and are microscopically distinguishable (Fig. 1). Directly after chromosomal DNA replication (in the S phase of the cell-cycle) interphase marks the beginning of mitosis. ► [Chromosomes](#), with their sister ► [chromatids](#), are discrete but not visible. Centrioles are now being replicated into two small daughter centrioles.



Cell Division. Figure 1 Mitosis and its stages.

Early prophase is marked by centrosome movements towards opposite sides of the cell. Now, the chromosomes are visible and the nuclear membrane breakdown begins. In late prophase condensation of chromosomes is finished. Structurally each chromosome contains two chromatids attached by a centromere. Specific spindle fibers (►mitotic spindle) are built. These consist of microtubules and are made inside the cell in a region close to the centrosomes. While the centrosomes move towards the opposite poles, some of the fibers contact the other pole. In addition the fibers contact chromatids and attach to the kinetochores. In metaphase the chromosomes orient themselves in one plane, which is sometimes called the equator of the cell. It is not until anaphase that the two sister chromatids are separated. Now two separated chromosomes are present, each containing a centromere. The spindle fibers connect each centromere to one pole. The separated chromosomes begin to move towards the poles. Telophase is characterized by the formation of new membranes. These form to build new nuclei around the chromosomes, which now uncoil again and thus become less visible. The spindle fibers disappear and microtubules are depolymerized. Cytokinesis is nearly complete. Throughout this process, both centrioles at the poles have grown. Finally, both descendants have a full-sized centriole. Only after the end of cytokinesis can both new cells enter another cell-cycle.

Regulatory Mechanisms

The molecules required for cell division, are believed to be under the control of a complex network of “cyclins”, kinases and inhibitors. These periodically expressed molecules are active at discrete stages of the cell division cycle but mainly act in the ►checkpoints of the cell-cycle. At any of these points cell division may be

halted, potentially corrected or proceed. The core of the molecules acting in the cell-cycle and checkpoints are dimers consisting of a catalytic and a kinase molecule, a cyclin and its “cyclin dependent kinase” (CDK). These complexes are themselves under the control of secondary modifications (like phosphorylation and ubiquitination). As part of the cell-cycle these events may lead to an increase in gene expression, intracellular translocation, protein degradation, nuclear envelope breakdown, DNA replication, chromosome separation and lastly to cell division.

Sometimes cytokinesis is viewed as the last stage of mitosis. Mitotic events regulate the location of the future cleavage plane through an unknown mechanism. It is known, that a contractile ring of actin and myosin constricts the cell during cytokinesis. This contractile ring and future cleavage will develop where the chromosomes line up during metaphase. It is still unclear which components of the mitotic machinery predict where the contractile ring will assemble. Analyses in sea urchin embryos suggest that the two ►asters, not the spindle itself, are critical for the prospective cleavage plane. These astral microtubules may signal to cortical regions, activating the polymerization of actin and myosin, which form the contractile ring where the cleavage furrow later starts. Since the signal has yet to be identified, it has been hypothesized that the cyclin dependent kinase CDC2 may be responsible. This kinase is not only activated throughout the relevant phases of the cell-cycle but is known to have a number of important substrates including myosin. CDC2 induced phosphorylation can promote the assembly of myosin filaments. Therefore, down-regulation of CDC2 is required for mitotic exit. This example serves to illustrate how cytokinesis may be linked to the cell-cycle.

Cell Division Concepts

Cell Division in Development

As exemplified in the early development of *Drosophila melanogaster*, cell division may be altered to fit the needs of speedy growth; for a short time nuclear growth is uncoupled from cell division. During early development of this fruit fly, nuclear division occurs through 13 rapid and synchronized rounds in the fertilized egg. Within this shared cytoplasm all nuclear divisions occur in short cycles. DNA synthesis alternates with mitosis without distinguishable interphases. This unusual stage of development is termed the “syncytial blastoderm”. After another round of duplication, the bulk of the nuclei migrate to the cortex of the embryo. Then, cellular walls emerge that divide the embryo into individual cells. At this stage, the “cellular blastoderm”, cell division gradually lengthens including a gap-phase. There is strong evidence to suggest that sequential destruction of mitotic cyclins by the

► **anaphase-promoting complex (APC/C)** may be important. In post-blastoderm embryos, cyclins A and B are degraded prior to the metaphase/anaphase, at the transition and during anaphase. It seems that, cyclin A degradation is required to disjoin sister chromatids. In addition, cyclin B degradation is required for cytokinesis and stable kinetochore-spindle attachments during anaphase. Stabilized cyclin B is able to block cytokinesis completely during early divisions. Conversely, homozygous *cyclin B* mutant embryos have an advanced onset of cytokinesis during early development. Cyclin B seems to act partly through a Rho-GEF, which is also required for initiation of cytokinesis. Lastly, cytokinesis in early and late *Drosophila* embryos depends on regulation by mitotic cyclin degradation. This is an example of tight-knit interaction between regulators of cell division with an active role in development.

Aberrant Cell Division Leads to Cancer

Molecular analysis of human cancer revealed that the genome of tumor cells usually acquires a number of mutations. In many cases these mutations are harbored within specific classes of genes. These genes that were later shown to be responsible for the initiation of the tumors were called “oncogenes and tumor suppressors” (among these genes some have viral counterparts, viral oncogenesis). Most mutations are inflicted upon genes that regulate the cell-cycle directly or have an activity upon cell division. As a consequence of these mutations, tumorigenic cells are insensitive to growth factors, replicate their DNA content more than once per cell-cycle and have typical aberrant ► **mitotic figures** or more than one centrosome per cell. Cells in highly malignant tumors often bear five to seven mutations. Not surprisingly, cells in such tumors are able to execute cell division in an uncontrollable fashion. Recent findings suggest a new group of molecules that link cytokinesis, chromosome segregation and tumorigenesis. The ► **serine/threonine kinases** Aurora are a family with three members in mammals. All Aurora genes map to loci frequently altered in human tumors. The proteins differ in their activities, timings and localizations. Aurora A is implicated in centrosome maturation and spindle assembly, whereas Aurora B has been proposed to regulate chromosome condensation, cohesion and segregation as well as kinetochore assembly and control of ► **spindle check-points**. In the case of Aurora A direct evidence suggests it to be a cancer susceptibility gene. Ectopic expression of Aurora A can transform rodent cells. More importantly, over-expression yields tetraploid cells. These cells display numerous extra centrosomes. Usually, additional numbers of centrosomes are recognized at the spindle checkpoint. Thus a cell division failure leads here to centrosome over-duplication, which finally produces

cellular aneuploidy. The resistance of some tumors to chemotherapeutic agents may partly be explained by similar failure in their mitotic spindle checkpoint.

Specialized Cell Division: Asymmetry

In specific cases cell division establishes two unequal descendants. Often this asymmetry is based on the polarity within the dividing cell. The establishment of polarity is considered to be the basis of differentiation and cellular diversity in higher organisms. Examples are embryonic stem cells (ES-cells) and specialized sheets of cells (including epithelial and neuronal cells) in nematodes.

In embryonic stem cells (ES-cells) there is a fine balance between symmetric and ► **asymmetric cell division**. In ES-cells symmetrical division occurs as part of their unlimited proliferative capacity. Self-renewal is a hallmark of stem cells. Only upon differentiation does unequal distribution of specific factors create two unequal descendants, one committed cell with a restricted fate and another progenitor cell with unlimited self-renewal potential. This asymmetrical division of an isotropic cell (in all directions completely symmetrical) is a rather unusual behavior. A number of mechanisms or specific factors may regulate this specific cell division in ES-cells. These potentially involve the generation of cell polarity within the cell, the orientation of the mitotic spindle and the segregation of differentiation or stem cell determinants into the two daughter cells. All these mechanisms have been likened to the establishment of polarity and epithelial-mesenchymal transition in the embryo. Molecules like APC, dynein-dynactin, PKC ζ , actin and myosins are currently under scrutiny for potential roles in this process.

In a genetic screen of the nematode ► *Caenorhabditis elegans*, defects in the axis determination of the early embryo were observed. The underlying defects were attributed to mutation of the PAR molecules. This family of proteins is now reckoned to be important for the planar polarity of a variety of cells in many organisms. In *C. elegans* in response to fertilization, PAR3, PAR6 and PKC become localized to the posterior pole of the zygote. Thus, PAR proteins polarize the embryo. The function of the PAR protein is also required to position and rotate the spindles along the anterior-posterior axis of the embryo. Again PAR3, PAR6 and PKC are important to tether the posterior centrosome close to the cortex (thereby forcing its change in shape). This action is believed to originate from asymmetric forces applied by the astral microtubules. Cortical forces applied to these microtubules in turn are controlled by PAR action. Apart from its interaction with PKC, PAR is also downstream of and dependent on CDC42, another link to cell polarity and asymmetrical cell division.

References

1. Lee AL, Orr-Weaver TL (2003) Regulation of cell cycles in *Drosophila* development: intrinsic and extrinsic Cues. *Ann Rev Genet* 37:545–578
2. Macara IG (2004) Parsing the polarity code. *Nat Rev Mol Cell Biol* 5:220–231
3. Murray AW (2004) Recycling the cell cycle: cyclins revisited. *Cell* 116:221–234
4. Fuchs E, Tumber T, Guasch G (2004). Socializing with the Neighbors: Stem Cells and Their Niche. *Cell* 116:769–778
5. Meraldi P, Honday R, Nigg EA (2004). Aurora kinases link chromosome segregation and cell division to cancer susceptibility. *Curr Opin Genet Dev* 14:29–36
6. Carmena M, Earnshaw WC (2003) The cellular geography of aurora kinases. *Nat Rev Mol Cell Biol* 4:842–854
7. Blagden SP, Glover DM (2003) Polar expeditions-provisioning the centrosome for mitosis. *Nat Cell Biol* 5:505–511

Cell Division Cycle 42 Protein

► CDC42

Cell Lineage Specification

Definition

Cell lineage specification refers to the process by which signals direct a multipotent cell (capable of entering a number of differentiation programmes) towards a particular tissue fate.

► Somitogenesis

Cell Membrane

► Biological Membranes

► Cell Polarity

Cell Migration

Synonyms

Cell Motility

Definition

Cell migration is a highly integrated multistep cellular process involving the active movement of individual cells or cell sheets. Migration (motility) is important throughout whole life of an organism. During gastrulation, large groups of cells migrate collectively as sheets to form the resulting three-layer embryo. Later on in embryonic development, cells (for example the neural crest cells) migrate from various epithelial layers to target locations, where they then differentiate to form the specialized cells that make up different tissues and organs. In the adult organism, cell migration governs tissue repair and regeneration (during the renewal of skin and intestine, new epithelial cells migrate up from the basal layer and the crypts, respectively). Migration is also crucial for immune surveillance, in which leukocytes from the circulation migrate into the surrounding tissue to destroy invading microorganisms and infected cells and to clear debris. Migration contributes to several important pathological processes, including vascular disease, osteoporosis, chronic inflammatory diseases such as rheumatoid arthritis and multiple sclerosis, cancer metastasis, and mental retardation.

Cell migration is initiated by chemical signals (► chemokines, ► growth factors, signals from the ► extracellular matrix), that via cell surface receptors initiate a complex cascade of events, including the activation of ► G proteins or ► tyrosine kinases, the production of PI (3,4)P₂/PIP₃, and the stimulation of Rho family proteins.

In general, cell migration can be described as a cyclic process involving cell polarization, adhesion and traction of the cell, and detachment of the cell rear. In response to a migration-promoting agent (chemotactic agents like chemokines) the cell starts to polarize and extend protrusions in the direction of migration. These protrusions can be large, broad lamellipodia or spike-like filopodia. Their formation is usually driven by actin polymerization. Rho family proteins are central regulators of actin and adhesion organization and thereby control the formation of lamellipodia and filopodia. ► Cdc42 also contributes to cell polarization by mediating reorientation of the microtubule organizing center (MTOC) toward the cell front, inducing the growth of microtubules and release of vesicles into this region. Adhesion molecules, like integrins, are activated by PI3Ks, PKCs, and/or Rap via talin, and they stabilize the protrusion via structural connections to the actin filaments. These adhesion sites then serve as traction sites. While the cell moves forward over them, these contacts are broken up at the cell rear and the cell can detach. The disassembly of adhesions is controlled by pathways that include FAK, ERK, Src, and the protease calpain, as well as microtubule dynamics.

Retraction at the rear requires Rho kinase and is a myosin-dependent process.

- ▶ Actin Cytoskeleton
- ▶ Cytoskeleton: Microtubules and Intermediate Filaments
- ▶ Neural Development
- ▶ Neural Tube
- ▶ Focal Adhesions

Cell Motility

- ▶ Cell Migration

Cell Polarity

ELENA CIBRIAN-UHALTE,
SALIM ABDELILAH-SEYFRIED
Max Delbrück Center, Berlin, Germany
salim@mdc-berlin.de

Definition

Cell polarity is a fundamental property of eukaryotic cells and refers to the polarized organization of the ▶ cell membrane with associated proteins, as well as to the polarized arrangement of the ▶ cytoskeleton and organelles within the ▶ cytoplasm and to the vectorial transport of ▶ secretory vesicles. Some types of neurons are highly polarized cells harboring elaborate dendritic extensions (▶ dendrites) at the postsynaptic side and ▶ axonal outgrowths at the presynaptic side. Similarly, epithelial cells are organized along a cellular axis that extends from the apical side facing an external lumen to the basal side facing either the ▶ extracellular matrix or adjacent cells. In addition to this apical-basal axis of polarity, epithelial cells are often oriented within the plane of the tissue along a proximal-distal axis, which is referred to as “tissue polarity” or “planar polarity”.

Genetic studies (performed mainly in the genetic model organisms *Caenorhabditis elegans* and *Drosophila melanogaster*) along with biochemical and tissue culture experiments have identified a set of genes that are essential for establishing apical-basal and “planar polarity” within epithelial cells. Some of these genes are also required for neuronal cell polarity. Cell polarity

genes regulate diverse biological processes including cell and tissue morphogenesis, ▶ cell migration, and organ morphogenesis. Moreover, the polarized state of neuronal and epithelial cells is a prerequisite for the physiological functions of mature tissues. Examples of these functions include the directional transport of nutrients through the gut ▶ epithelium, the secretion or reabsorption of ions and solutes in epithelia of glands and kidney ducts and the correct establishment of neuronal connectivity. The failure to establish or maintain cell polarity causes abnormal cellular functions and can result in disease. This chapter reviews the molecular mechanisms that lead to the establishment of apical-basal cell polarity in vertebrate epithelial cells and gives an overview of related human disorders.

Characteristics

The correct establishment and maintenance of cell polarity begins with the correct formation of an ▶ apical junctional complex that initiates ▶ cell adhesion and controls the creation of different membrane compartments. Next, apical and basolateral membrane domains are supported by sorting and targeting mechanisms for membrane material and proteins. The latter process requires that the cytoskeletal tracks used for vesicle trafficking are oriented towards cellular cues that propagate the maintenance of cell polarity.

The Formation of the Apical Junctional Complex

Evidence from tissue culture systems indicates that the initial spatial cues for the establishment of epithelial cell polarity are generated at the cell surface by cell-cell contacts or cell-extracellular matrix (ECM) contacts. Cell adhesion is mainly mediated by E-cadherin, a member of the Ca^{2+} -dependent ▶ cadherin superfamily of ▶ adhesion molecules, and nectins. On the future basal side, cell-ECM adhesion is mediated by the integrin superfamily of adhesion receptors. Cell adhesion initiates the restricted localization of basolateral membrane proteins whereas cell-ECM adhesion is required for organizing the apical-basal axis of polarity (reviewed in 1).

The apical junctional complex comprises several protein complexes, each assembled around a membrane protein linked to several submembraneous scaffolding proteins (2, 3). Many of the proteins involved contain the PSD-95/Discs large (Dlg)/ZO-1 (PDZ) protein-binding domain and establish links with the ▶ actin or ▶ microtubule cytoskeleton. The principal players within the apical junctional complex are the cadherin/catenins, the Par6/atypical protein kinase C (aPKC) and the Crumbs (Crb)/ Stardust (Sdt)/Pals-1 protein complexes. The latter two complexes localize to the subapical region of the membrane and are required to define the apical character of the cell membrane. A fourth protein complex formed by the

proteins Scribble (Scrib), Dlg and Lethal giant larvae (Lgl) appears to define basolateral membrane character. Genetic and biochemical studies have demonstrated that these proteins act in a regulatory pathway that is essential for the establishment of apical and basolateral membrane domains. However, the biochemical mechanisms responsible for the ►genetic interactions are currently not known.

On an ultrastructural level, the formation of the apical junctional complex results in the establishment of circumferential apical-lateral ►adherens junction (AJ) belts. The AJ is comprised of the E-cadherin/catenin cell adhesion system and is associated with an apical ►microfilament ring that regulates cell shape. In addition to mediating cell adhesion, the AJ is essential for the coordinated movement of epithelial tissues. Moreover, ►tight junctions (TJ) form in close proximity just apical to the AJ. In vertebrates, the TJ presents an important morphological difference between epithelial and non-epithelial cells. The TJ provides the most apical cell-cell junction and acts as a selective permeability barrier restricting paracellular diffusion across the epithelium. In addition, the TJ forms an intracellular fence against lateral diffusion of membrane proteins and lipids, thereby contributing to the maintenance of apical-basal polarity. The mechanisms by which non-epithelial cells that lack the TJ maintain a highly polarized surface are currently unknown. TJ function is dependent on transmembrane proteins including claudins, occludin and the junctional adhesion molecule (JAM). Currently, there are no ►animal models for these TJ genes.

Several human diseases have been linked to genes involved in the establishment of the apical junctional complex. These diseases fall into two classes.

- The first class includes several malignant tumors of an epithelial origin. Some of these tumors are associated with a down-regulation of E-cadherin. Loss or down-regulation of E-cadherin causes the breakdown of tissue integrity and epithelial barrier function and can result in inappropriate cell motility. Similarly, β -catenin has been implicated in tumorigenesis, possibly due to a role as a component of the ►Wnt signaling pathway. ►Peutz-Jeghers Syndrome (PJS) is an autosomal dominant disorder characterized by the development of multiple gastrointestinal hamartomatous polyps (benign epithelial and stromal overgrowth). Subsequently, a high percentage of PJS patients will develop malignant tumors at an early age. PJS is caused by mutations in the *Lkb1/Par4* gene that encodes a serine/threonine ►kinase required for the establishment of cell polarity (4, 5).
- The second class of diseases is comprised of several diseases that affect tissue integrity and result in

epithelial degenerations or malformations. This class includes retinitis pigmentosa (RP) type 12 that is caused by loss-of-function mutations in *crbl1* (6). RP is characterized by the progressive degeneration of the ►photoreceptor cells associated with disrupted AJ and results in severe visual impairment. Since a wide range of phenotypes can be observed in patients with mutations in *crbl1*, the ►genetic background may modulate the ►expressivity of the symptoms. Mutations in the *nectin-1* gene resulting in truncated forms of the protein are responsible for the cleft lip/-palate-ectodermal dysplasia syndrome. Whether the nectin adhesion complex is essential for the establishment and maintenance of adherens junctions is unclear.

Polarized Secretion and Protein Targeting

The best evidence for the involvement of vesicle targeting during cell polarization comes from studies in cultured epithelial cells. According to these studies, both endocytic and transcytotic pathways are utilized for membrane protein sorting. In addition, membrane proteins synthesized *de novo* go through the ►Golgi into the ►trans-Golgi network (TGN), where they are directly sorted as basolateral or apical membrane proteins. Apical targeting of proteins appears to require the assembly of membrane proteins onto ►lipid rafts. The glycosylphosphatidylinositol (GPI) domain present in several apical membrane proteins is sufficient for direct apical sorting. Similarly, basolateral sorting signals are present in the cytoplasmic tails of several proteins. PDZ domain-containing proteins have been shown to play a key role in protein targeting. The PDZ domain facilitates binding of the carboxy-terminus of transmembrane proteins with cellular signaling complexes.

The correct delivery of transport vesicles into different membrane domains requires ►SNAREs (soluble *N*-ethylmaleimide-sensitive factor attachment protein receptors) on the target membrane (►t-SNARE) and on the vesicular membrane (►v-SNARE). In epithelial cells, apical membranes contain the t-SNARE syntaxin-3 and basolateral membranes the t-SNARE syntaxin-4. Blocking their function disrupts vesicle delivery. Little is known about the mechanisms that lead to the polarized distribution of the t-SNAREs. The proper maintenance of apical syntaxin-3 is dependent on microtubules, as their disruption causes syntaxin-3 to mislocalize to the lateral membrane. Consequently, apical proteins are mistargeted into the lateral membrane.

An essential prerequisite for polarized sorting is the organization of the ►cytoskeleton. Both actin and microtubule cytoskeletons have been implicated in vesicle transport and polarized sorting mechanisms. The actin cytoskeleton may be involved in the short-range delivery of vesicles to the membrane whereas microtubules may function in long-range transport of

Cell Polarity. Table 1

Human gene	Protein structure (domains)	<i>Drosophila</i> homolog mutant phenotype	Mouse homolog mutant phenotype	Molecular function	Disease
occludin	Transmembrane (TM)	No homolog reported	occludin: Development	Cell adhesion, paracellular barrier	Unknown
claudin family: claudin (cldn)-1-18	TM	sinuous (sinu): Paracellular barrier Megatrachea: Cell adhesion	claudin-1: Cell adhesion claudin-5: Cell adhesion claudin-11: Development claudin-14: Paracellular barrier	Cell adhesion, paracellular barrier	Autosomal recessive deafness (claudin-14) Primary hypomagnesemia (claudin-16)
junctional adhesion molecule-1 (JAM-1)	Extracellular IgV loops	No homolog reported	No mutant reported	Cell adhesion	Unknown
crumbs1 (crb1) crb2 crb3	TM, Epidermal growth Factor (EGF), Laminin G (LamG)	crb: Cell polarity	Retinal degeneration 8 (rd 8)/crb1: Cell polarity; proliferation	Scaffolding protein	Retinitis pigmentosa type 12 (crb1)
zonula occludens-1 (ZO-1) ZO-2 ZO-3	PDZ, SH3, GUK, Proline-rich domains	tamou: Development	No mutant reported	Scaffolding protein	Familial hypercholesterolemia (ZO-2)
membrane protein, palmitoylated 5 (mpp 5)	PDZ, SH3, Guanylate kinase (GUK)	stardust (sdt): Cell polarity	No mutant reported	Scaffolding protein	Unknown
pals1-associated tight junction protein (patj)	PDZ, metal response element (MRE)	Dmpatj: No defects	No mutant reported	Scaffolding protein	Unknown
partitioning-defective protein 3 (par-3)	PDZ	bazooka (baz): Cell polarity	No mutant reported	Scaffolding protein	Unknown
par-6	Cdc42/Rac-interactive-binding (CRIB), PDZ, octicos peptide repeat (OPR)	DmPar-6: Cell polarity	No mutant reported	Scaffolding protein	Unknown
protein kinase C ζ (PKC ζ) PKC ι	Kinase, OPR, protein kinase C conserved region 1 (C1)	DaPKC: Cell polarity	No mutant reported	Cell signaling	Unknown

Cell Polarity. Table 1 (Continued)

Human gene	Protein structure (domains)	<i>Drosophila</i> homolog mutant phenotype	Mouse homolog mutant phenotype	Molecular function	Disease
cdc42	GTPase	cdc42: Cell polarity; development	cdc42: Development	Signal transduction	Unknown
cingulin	Coiled-coil	No homolog reported	No mutant reported	Cytoskeletal adaptor	Unknown
4.1R	Spectrin-actin-binding (SAB)	coracle(cor): Development; cell adhesion	4.1R: Development	Cytoskeletal adaptor	Hereditary elliptocytosis
Epithelial-cadherin (E-cadherin)	Cadherin repeats, TM	shotgun (shg). Cell adhesion	E-cadherin: Development	Cell adhesion	Implicated in initiation of several epithelial malignant tumors
nectin-1 nectin-2 nectin-3 nectin-4	Ig-like domains, TM	No homolog reported	nectin-2: Development	Cell adhesion	Cleft lip/palate-ectodermal dysplasia syndrome (CLPED-1) (nectin-1)
β -catenin	Armadillo/ β -catenin like domain (ARM)	armadillo (arm): Development; proliferation; cell adhesion	β -catenin: Proliferation; development	Cell adhesion, cytoskeletal adaptor, signal transduction	Implicated in initiation of several malignant tumors
α -catenin 1	ZO-1 binding domain, vinculin/ α actinin binding domain, adhesion-modulation domain	α -catenin: Development	α -catenin: Cell polarity; development	Cytoskeletal adaptor	Implicated in initiation of several epithelial malignant tumors
afadin	PDZ, Proline rich domains.	No mutant reported	afadin: Development	Cell adhesion	Implicated in epithelial ovarian cancer and acute myeloid leukemias
ponsin/c-Cbl associated protein (CAP)	Sorbin homology (SoHo), SH3	DCAP: No defects	No mutant reported	Signal transduction, cytoskeletal adaptor	Unknown
profilin-1 (pfn-1) pfn-2		chikadee (chick): Development	pfn-1: Cell proliferation	Cytoskeletal organization	May contribute to Miller-Dieker syndrome (pfn-1)
Wiskott-Aldrich syndrome protein (WASP)	P21-Rho-binding domain (PBD), WASP homology region 1 (WH1), WH2	DmWASP: Development	WASP: Proliferation	Cytoskeletal organization, signal transduction	Wiskott-Aldrich syndrome X-linked thrombocytopenia X-linked severe congenital neutropenia

Cell Polarity. Table 1 (Continued)

Human gene	Protein structure (domains)	<i>Drosophila</i> homolog mutant phenotype	Mouse homolog mutant phenotype	Molecular function	Disease
adenomatous polypsis of the colon (apc)	ARM	apc2: Cell adhesion	apc: Cell proliferation	Cell adhesion, signal transduction, cytoskeletal adaptor	Familial adenomatous polypsis (FAP)
rac-1 rac-2	GTPase	rac-1, rac-2: Development	rac1: Development	Cytoskeletal organization	Unknown
microtubule affinity-regulating kinase 1 (MARK1) MARK2 MARK3 MARK4	Kinase, TM	par-1: Cell polarity	ELKL motif kinase (emk): Development	Cytoskeleton organization; signal transduction; cell communication	Unknown
lkb1	Kinase	lkb1: Cell polarity; proliferation	lkb1: Development; proliferation; cell adhesion	Signal transduction	Peutz-Jeghers syndrome
α -spectrin β -spectrin	Pleckstrin homology (PH), SH3, spectrin repeats	α - and β -spectrin: Cell polarity	elf (embryonic live fodrin), β -spectrin: Development	Cytoskeletal organization, scaffolding protein	Unknown
human scribble (hScrib)	Leucine-rich repeats (LRR), PDZ	scrib: Cell polarity; cell adhesion; proliferation	circletail/ scrib1: Planar cell polarity; development	Scaffolding protein	Unknown
discs large1 (dlg1) dlg2 dlg3 dlg4	PDZ, SH3, GUK	dlg: Cell polarity; proliferation	dlg1: No defects	Scaffolding protein, cytoskeletal adaptor	Unknown
lethal giant larvae (lgl)	WD-40	l(2)gl: Cell polarity; cell adhesion	No mutant reported	Unknown	Unknown
integrin- β 4 (ITGB4)	Fibrinectin type 3 (FN3), von Willebrand factor type A (VWA), TM	No homolog reported	itgb4: Cell adhesion; proliferation	Cell adhesion, cell communication, signal transduction	Junctional epidermolysis bullosa with pyloric atresia
desmocollin-1(dsc-1) dsc-2 dsc-3	Cadherin	No homolog reported	No mutant reported	Cell adhesion	Unknown

Cell Polarity. Table 1 (Continued)

Human gene	Protein structure (domains)	<i>Drosophila</i> homolog mutant phenotype	Mouse homolog mutant phenotype	Molecular function	Disease
desmoglein-1 (dsg-1)	Cadherin	No homolog reported	dsg2: Proliferation; development; adhesion	Cell adhesion	Keratosis Palmoplantaris strata I (dsg1)
dsg-2			dsg3: Cell adhesion; proliferation		Autoimmune disease
dsg-3			dsg4: Cell adhesion; development		
dsg-4					
plakoglobin (PKGB)	ARM, β -catenin like	No homolog reported	pkgb: Cell adhesion.	Cell adhesion	Naxos disease
plakophilin-1 (pkp-1)	ARM	No homolog reported	No mutant reported	Cell adhesion	Ectodermal dysplasia/skin fragility syndrome (pkp1)
pkp-2					
pkp-3					
pkp-4					
desmoplakin (DP)	Plakin, coiled-coil rod, plakin repeat domain (PRD), glycine-serine-arginine (GSR)	No homolog reported	DP: Cell adhesion; development	Cell adhesion	Keratosis palmoplataris striata II Dilated cardiomyopathy with wooly hair and keratoderma Arrhythmogenic right ventricular dysplasia 8 Skin fragility-woolly hair syndrome
erbin (ERBB2IP)	LRR, PDZ	No homolog reported	No mutant reported	Signal transduction	Unknown

(Status: 2003)

vesicles to distinct membrane domains. In epithelial cells, a cortical meshwork of actin filaments is associated with each membrane domain and there is a tight apical actin belt at the AJ. The dynamics of actin reorganization/polymerization (during cellular morphogenesis and cell adhesion) are regulated by the Rho-family of small GTPases ([►Cdc42](#), Rho and Rac1), the Arp2/3 complex and several regulators of actin dynamics including the [►Wiskott-Aldrich syndrome protein](#) (WASP). The distribution of lateral membrane proteins is controlled by membrane scaffolds containing ankyrin and spectrins cross-linked with actin. Microtubules are organized along the apical-basal axis with their plus ends at the basal pole. Minus-end

(dyneins) and plus-end (kinesins) directed [►motor proteins](#) appear to be involved in vesicle transport. The small GTPase Cdc42 has some role in basolateral targeting of transport vesicles. Similarly, the basolateral determinant Lgl has been found to bind the t-SNARE syntaxin-4. Although, the function of Lgl in basolateral targeting has not been directly tested, it is possible that Lgl may be involved in regulating the delivery of transport vesicles containing basolateral proteins. The incorrect sorting of membrane proteins can cause loss of the polarized surface distribution and can affect cell polarity with pathological consequences. Most of the defects currently known in humans are due to mutations in sorting signals or defective [►endocytosis/](#)

►**transcytosis** that result in inappropriate protein localization. Examples include the mistargeting of the low-density lipoprotein receptor (LDLR) in ►**familial hypercholesterolemia** and the inhibition of endocytosis of the amiloride-sensitive epithelial Na⁺-channel in a hereditary form of arterial hypertension.

Regulatory Mechanisms

Many pathologies display defective cell polarity. However, the loss of cell polarity in most diseases may only be secondary to physiological changes or cell/tissue damage. Only a few examples of monogenetic diseases (►**Monogenic Disorder/Monogenic**) are currently known in which ►**apical junctional complex** proteins, components of the polarized ►**sorting** machinery or regulators of ►**cytoskeletal** assembly have been implicated. More insight into the potential involvement of these cell biological processes in disease formation has been gained through work in various animal models (Table 1).

The protein complexes assembled at the ►**apical junctional complex** have primarily two functions: 1) as mechanical components that link neighboring cells with each other and control the connection between membrane proteins and the ►**cytoskeleton** and 2) as signaling components for the regulation of multiple cellular functions. Therefore, most pathologies caused by defects in ►**apical junction complex** proteins may be caused by divergent molecular mechanisms as exemplified during tumor formation. Many ►**epithelial** malignant tumors are characterized by the inactivation of the E-cadherin/catenins ►**cell adhesion** system, breakdown of the ►**basement membrane** (which facilitates the inappropriate motility of cells) and misregulation of proliferation. Experiments carried out in *Drosophila melanogaster* have shown that the invasive behavior of epithelial malignant tumor cells requires the loss of apical junctional complex proteins (which may cause a failure in the assembly or maintenance of basement membrane components) and misregulated expression of ►**oncogenes** (leading to increased proliferation). In addition, malignant cells may increase the turnover of basement membrane components through the release of ►**proteases** and may therefore be more sensitive to already reduced basement membrane assembly. Tumor progression may also be associated with defective signaling as seen with the ►**Wnt-signaling pathway** component β -catenin that binds to the transcription factor LEF-1/TCF and regulates gene expression. The stability of β -catenin is regulated via another binding partner, Adenomatous polyposis coli (Apc), which is mutated in ►**familial adenomatous polyposis**.

Clinical Relevance

Currently, only a small number of human diseases have been directly linked to defects in components of the cell

polarity machinery. However, a large number of genes required for cell polarity have been identified through studies in animal models whose ►**mutant phenotypes** suggest a possible involvement in human diseases. Therefore, increased ►**genotyping** and mapping projects of candidate genes within human patient populations will continue to uncover additional disease genes. Further studies in model organisms are required to provide better insight into the relationship between cell polarity and disease and, ultimately, to allow for the development of new therapies.

References

1. Nelson WJ (2002) Adaptation of core mechanisms to generate cell polarity. *Nature* 422:766–774
2. Bilder D, Schober M, Perrimon N (2003) Integrated activity of PDZ protein complexes regulates epithelial polarity. *Nat Cell Biol* 5:53–58
3. Hurd TW, Gao L, Roh MH et al (2003) Direct interaction of two polarity complexes implicated in epithelial tight junction assembly. *Nat Cell Biol* 5:137–142
4. Hemminki A, Markie D, Tomlinson I et al (1998) A serine/threonine kinase gene defective in Peutz-Jeghers syndrome. *Nature* 391:184–187
5. Jenne DE, Reimann H, Nezu J et al (1998) Peutz-Jeghers syndrome is caused by mutations in a novel serine threonine kinase. *Nat Genet* 18:38–43
6. den Hollander AI, ten Brink JB, de Kok YJ et al (1999) Mutations in a human homologue of *Drosophila crumbs* cause retinitis pigmentosa (RP12). *Nat Genet* 23:217–221

Cell Transformation

Definition

Cell transformation refers to a complex series of events, which leads cells to lose growth control and become immortalized, thus resembling a tumorigenic condition.

►**Splicing**

Cellular Disease Model

Definition

Cellular disease model refers to a cell culture where cells display a phenotype that is acknowledged to be involved in a disease process.

►**Automated High-Throughput Functional Characterization of Human Proteins**

Cellular Immune Response

Definition

Cellular immune response describes antigen-specific immune responses of T cells.

►DNA-based Vaccination

Cellular Localization

►Immunochemical Methods, Localization

Cellular Proteomics

►Automated High Throughput Functional Characterization of Human Proteins

Cellular Senescence

IGNACIO PALMERO

Institute of Biomedical Research, CSIC-UAM,
Madrid, Spain
ipalmero@iib.uam.es

Definition and Characteristics

The concept of cellular senescence was first established by Leonard Hayflick in the 1960s from the seminal observations he made on the growth behaviour of ►fibroblasts obtained from human embryos (1). Hayflick observed that once explanted and maintained under standard laboratory tissue culture conditions, the fibroblast cultures could be expanded for only a limited number of passages, after which the cells ceased dividing and there was no net gain in cell number. Hayflick reasoned that this terminal non-proliferative state resulted from an intrinsic feature of the cells and named it senescence. Although it was first observed and remains best characterised in fibroblasts, senescence has been described in many different cell types, including T lymphocytes, epithelial cells, macrophages

and glial cells among others. Senescence occurs after a fixed number of cell divisions for each species and cell type, this number being around 70–80 for human fibroblasts. Senescent fibroblasts display a number of morphological and molecular features and a specific gene expression profile that distinguishes senescence from other non-proliferative states such as ►quiescence (2, 3). One of the most distinctive features of senescent fibroblasts is their change in cell shape from thin, elongated cells to a characteristic flat, enlarged morphology, frequently with multiple nuclei per cell. Senescent cells also display a characteristic beta-galactosidase enzymatic activity at neutral pH, named senescent-associated beta Gal activity or SA-beta Gal, which is absent from actively dividing pre-senescent cells or quiescent cells. Although viable and metabolically active, senescent cells have lost their ability to respond to mitogens, as evidenced by the lack of induction of early genes by serum. Entry into senescence is typically accompanied by changes in chromatin organisation and a defined gene expression program (2). Among the changes in gene expression associated with senescence, the up-regulation of several negative cell-cycle regulator proteins, such as the ►cdk inhibitors p16INK4a and p21CIP1 or the ►p53 regulator ARF, together with the down-regulation of cyclins should be mentioned. The expression of gene products involved in cell-extracellular matrix interactions, such as metalloproteases or collagen, and proteins related to inflammation, such as interleukin-1, is also altered in senescent cells.

Regulatory Mechanisms

Senescence can be triggered by several types of stimuli (2, 3, 4, 5, 6, 7). The primary mechanism responsible for the onset of senescence in primary human fibroblasts is ►telomere dysfunction. Telomeres are the DNA-protein structures located at the ends of eukaryotic chromosomes and consist of tandem repeats of short sequences (TTAGGG in mammals). The number of repeats and therefore the telomeric length is variable between species, being around 10 kb in human germ cells. Telomeres act as protective structures preserving chromosome integrity. As a consequence of the semi-replicative nature of eukaryotic DNA replication, each round of DNA synthesis results in the loss of telomeric sequences at the end of the chromosomes. In the absence of a telomere maintenance mechanism, telomeres suffer progressive shortening at each cell division until a critically short telomeric length is reached. The protective action of telomeres also requires the formation of an appropriate structure, dictated by the length of the telomeric DNA and the participation of specific telomeric proteins. The lack of telomere integrity appears to be the signal that triggers the onset of senescence in human fibroblasts.

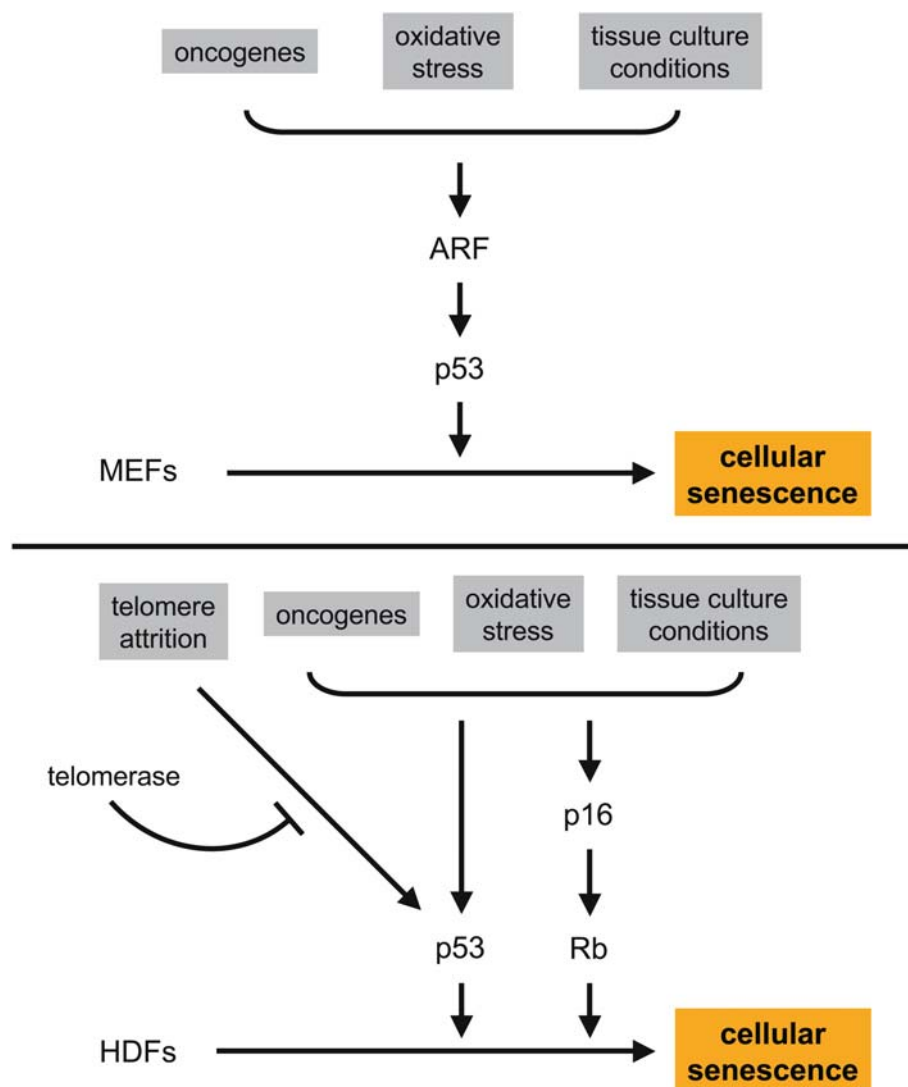
The importance of telomere erosion in the onset of senescence is revealed by studies using the enzyme telomerase. ▶**Telomerase** is a ribonucleoprotein that elongates telomeres using the telomeric repeats as templates and thus maintains telomere length. Human primary fibroblasts do not normally show telomerase activity. When telomerase is ectopically expressed in these cells, telomere erosion is prevented and this results in a significant extension of their normal life span.

Telomere exhaustion is not the only signal capable of triggering senescence in primary fibroblasts. Several other different stimuli are capable of causing a non-proliferative state indistinguishable from senescence before telomere exhaustion (5, 6). These stimuli include oxidative stress, exposure to DNA damaging agents, chronic expression of activated oncogenes and growth under suboptimal conditions. Senescence caused by these stimuli has been variedly referred to as stress-induced premature senescence (SIPS), oncogene-induced premature senescence (for oncogenes) or stress or aberrant signalling-induced senescence (Stasis), whereas the term replicative senescence tends to be used to describe exclusively telomere-dependent senescence (7). In an attempt to unify the concept of senescence, the term cellular senescence is normally used to include all types of senescence, irrespective of the initiating signal. The current view is that all the above-described stimuli, ranging from growth on plastic dishes to telomere shortening are interpreted by the cell as “stress” signals that ultimately activate an anti-stress response mechanism that leads to a cessation of active division (Fig. 1) (5, 6). The importance of telomere-dependent and independent pathways to senescence varies between cell types and species (7). For instance, mouse embryonic fibroblasts (▶**MEFs**) have very long telomeres and express telomerase. However, they stop dividing and enter senescence after a short number of divisions under normal tissue culture conditions. This form of telomere-independent senescence is an example of stress-induced senescence, triggered by the abnormal environment conditions caused by standard tissue culture practice and can be bypassed by changing culture conditions including oxygen level or mitogen supply. Similarly, some human epithelial cells suffer a senescent arrest well before telomere exhaustion, which again can be avoided by growing these cells in more favourable conditions. The stringency of the senescence program determines the different rates at which a particular cell type can be immortalised in culture. MEFs, where senescence is controlled by a relatively simple genetic pathway, undergo ▶**immortalisation** spontaneously in culture at a low but measurable rate. In contrast, immortalisation of human fibroblasts, which requires several concurrent genetic alterations, is an extremely rare event under standard culture conditions.

Interestingly, all forms of senescence ultimately activate a similar molecular machinery (2, 7). Different forms of stress-induced senescence impinge in the cellular pathways controlled by the tumour suppressors ▶**Rb** and ▶**p53**, most frequently through the action of the products of the ▶**INK4a/ARF** locus, the cdk-inhibitor p16INK4a and the p53 regulator ARF (p14ARF in humans and p19ARF in the mouse). In MEFs, senescence triggered by accumulation of cell divisions in culture is mainly mediated by the ARF protein, which in turn activates p53. Similarly, oncogene-induced senescence triggered by the activated form of the Ras oncogene is also mediated by ARF and p53 in this cell type. In agreement with this, MEFs genetically deficient for ARF or p53 do not undergo senescence upon continued culture or after exposure to the Ras oncogene. In other cell types, p16 plays the main role as sensor of stress. This is the case in some human epithelial cells or mouse macrophages, where inactivation of p16 is sufficient to extend their life span. Telomere-dependent senescence also feeds into similar molecular pathways. Telomeres act as protective structures of chromosome ends, and critically short or unprotected telomeres are probably interpreted by the cell as a DNA damage signal that ultimately activates the p53 pathway.

Cellular Senescence, Cancer, and Ageing

Cellular senescence has been extensively characterised as a phenomenon in ▶**primary cells** grown under laboratory conditions. How this phenomenon can be extrapolated to the behaviour of normal cells in living organisms and in particular its possible implication in cancer or ageing has been a matter of intense investigation. It has been assumed that cellular senescence in culture is a reflection of the limited proliferative potential of normal cells in the organism. Escape from senescence is a prerequisite for oncogenic ▶**transformation** of primary cells in the laboratory. It has been proposed that senescence acts as a tumour suppressor mechanism in living organisms (2, 3). Although, in theory, a cell could give rise to a tumour before exhausting its normal lifespan, the existence of high levels of ▶**apoptosis** and the selection of malignant clones in the early stages of tumorigenesis mean that bypassing the lifespan limit established by senescence is necessary for tumour formation. In support of this view, the genes that play important roles in the control of senescence are frequently mutated in tumours. This is clearly the case for the p53 tumour suppressor and the two products of the INK4a/ARF locus. It can be speculated that in the progression from normal to tumoral, cells experience stress situations similar to the ones studied *in vitro*, such as excessive mitogenic signalling, abnormal oxygen levels or DNA damage. These stress situations



Cellular Senescence. Figure 1 A schematic representation of the regulatory pathways that contribute to the onset of cellular senescence in human and murine primary cells. *MEFs*: Mouse Embryo Fibroblasts, *HDFs*: Human Diploid Fibroblasts.

would activate the senescence checkpoint and impede the formation of tumours by cells carrying potentially oncogenic alterations. Disabling of the senescence response by mutation of its regulators would then facilitate tumour formation. This model is supported by the observation that chemotherapeutic treatments that result in DNA damage can induce in living organisms a cell-cycle arrest reminiscent of senescence with the appearance of senescence-specific markers, such as SA-beta Gal staining.

Maintenance of telomere length is required for the continued growth of tumour cells. Although most normal tissues in the adult do not express telomerase, most tumour cells activate telomere-elongating mechanisms, implicating either *de novo* expression of

telomerase or poorly defined alternative mechanisms. Inhibition of telomerase in tumour cells has been proposed as a possible antitumour therapy. However, the potential use of this approach to restrain uncontrolled growth of tumour cells might have undesirable side effects. Inhibition of telomerase can result in critically short telomeres that in turn can provoke chromosomal instability in the target cells, which might have deleterious effects outweighing the beneficial antiproliferative effects.

The possible contribution of the senescence program to organismal ageing is more controversial. Several hypotheses have been put forward but a definitive proof for a link between senescence and ageing *in vivo* is still needed (2). Firstly, the Bmi protein, a known

regulator of senescence, has been shown to have an important role in the control of proliferation of stem cells in some tissues, mediated in part by the regulation of senescence effectors such as p16 or ARF. Based on this observation, it has been proposed that senescence plays an important role in the control of self renewal of stem cells *in vivo* and accordingly the depletion of the pool of stem cells that normally occurs during ageing could be due to the onset of senescence in these cells. Secondly, as mentioned above, senescent cells typically display alterations in gene products involved in the interaction between cells and the extracellular matrix and it has been speculated that these senescence-associated changes could contribute to the disruption of tissue organisation in ageing organisms. This must be weighed against the lack of conclusive evidence showing that senescent cells are more abundant in ageing organisms, although this observation is made difficult by the reduced number of senescent-specific markers available. Finally, based on the premature ageing phenotype described for some mouse models with over-expression of p53, it has been argued that premature ageing could be a consequence of reinforced tumour suppression and senescence. However, this view is not supported by other models and the possible connection between cellular senescence and organismal ageing remains to be firmly established.

References

1. Hayflick L, Moorhead P (1961) The serial cultivation of human diploid cell strains. *Exp Cell Res* 25:585–621
2. Campisi J (2003) Cancer and ageing: rival demons? *Nat Rev Cancer* 3:339–349
3. Mathon NF, and Lloyd AC (2001) Cell senescence and cancer. *Nat Rev Cancer* 1:203–213
4. Wright WE, Shay JW (2002) Historical claims and current interpretations of replicative aging. *Nat Biotechnol* 20:682–688
5. Sherr CJ, De Pinho RA (2000) Cellular senescence: mitotic clock or culture shock? *Cell* 102:407–410
6. Serrano M, Blasco MA (2001) Putting the stress on senescence. *Curr Opin Cell Biol* 13:748–753
7. Drayton S, Peters G (2002). Immortalisation and transformation revisited. *Curr Opin Genet Dev* 12:98–104

CENPs

Definition

This acronym identifies a group of six centromere proteins that are required for various functions, for example, for the attachment of microtubules to chromosomes.

► **Centromeres**

Central Element

Definition

Central element refers to a thin proteinaceous fibril that is present in the central region of the ► **synaptonemal complex**. This structure unites homologous chromosomes during the prophase of meiosis.

► **Meiosis and Meiotic Recombination**

Centromere

WILLIAM R. BRINKLEY, SCOTT SLATTERY
Department of Molecular and Cellular Biology, Baylor College of Medicine, Houston, Texas, USA
brinkley@bcm.tmc.edu

Synonyms

The older literature contains up to 27 different terms for the centromere, including the use of “kinetochore” as a synonym. Most of these alternate names have been abandoned and current studies in molecular genetics and cell biology clearly identify centromeres and kinetochores as structurally and functionally distinct components.

Definition

The centromere is a specialized site on a chromosome that directs the segregation of chromosomes to daughter cells during ► **cell division** (mitosis). The goal of mitosis is to divide and partition the genome equally by distributing one copy of each chromosome to the nucleus of each daughter cell and centromeres play a key role in this process.

The centromere appears only once on each chromosome. It can be identified microscopically within a prominent constriction (► **primary constriction**) along the chromosome's length. The centromere can assume one of three possible positions, near the tip of the chromosomes (► **telocentric chromosomes**), in the middle of the chromosome, with arms of equal length flanking the centromere (► **metacentric chromosomes**) or near one end of the chromosome, with arms of unequal length flanking the centromere (► **acrocentric chromosomes**).

The centromere consists of a discrete DNA-protein complex that assembles the ► **kinetochore** (1), a plate-like structure located on the outer surface of the centromere that connects the chromosome to ► **spindle**

microtubules. The kinetochore is composed of multiple centromere proteins (►CENPs) that associate with centromere DNA on one surface and attach to microtubules on the other. Microtubules, in conjunction with certain ►motor proteins of the kinetochore, provide the force needed to split duplicated chromosomes apart and move them to opposing ►mitotic spindle poles, where they are enclosed within the nucleus of each new daughter cell at the end of mitosis.

Characteristics

The centromere functions to organize and direct the assembly of the kinetochore at the primary constriction of the chromosome. Each centromere assembles two ►sister kinetochores, one on each half of the duplicated chromosome (►chromatid), that face opposite poles of the mitotic spindle. The kinetochore is a sophisticated motile device that couples chromosomes to spindle microtubules enabling one copy in a set of duplicated chromosomes to be pulled to each spindle pole and partitioned into the nucleus of daughter cells. In addition to supplying force and motility, some kinetochore proteins function in cell signaling as ►checkpoints to regulate mitosis and chromosome movements (2).

Centromere DNA

Each chromosome is composed of a single linear DNA duplex, associated with nuclear proteins forming a basic 30-nanometer ►chromatin fiber, which is folded and compacted into the chromosome arms. DNA in the centromere (cenDNA) is continuous with the DNA of the chromosome arms. CenDNA differs from the DNA in the chromosome arms, however, because it has a unique nucleotide sequence composition, it remains condensed (►heterochromatin) and it is genetically silent. The size and sequence composition of cenDNA varies in different organisms, ranging in complexity from 124 ►base pairs (bp) in budding yeast, to 40–100 kilobase (kb) pairs in fission yeast, to 0.5–5 megabase (mb) pairs in humans. Human cenDNA belongs to a family of DNA known as alpha satellite or ►alphoid DNA. Alphoid DNA consists of repeating units of 171 bp arranged in tandem arrays of 1,500 to 30,000 copies. The sequence of nucleotides within the cenDNA of higher plants and animals, unlike that of budding yeasts, does not appear to specify the location, organization or function of the centromeres. Instead, their centromere location and function is determined largely, if not exclusively, by ►epigenetic determinants thought to be specific CENPs (2, 3).

Centromere/Kinetochore Proteins

The kinetochore is a multi-protein complex that is bound to cenDNA. The table in Fig. 1 lists known

centromere/kinetochore proteins along with symbols depicting their approximate localization within five centromere sub-domains, as diagrammed in Fig. 1. The sub-domains represent five specialized zones of the centromere, the corona, a layer of fibrous loops that extend out from the outer plate, an inner plate adjacent to the underlying heterochromatin and the pairing domain that connects the two sister kinetochores as illustrated in Fig. 1. Of the seven groups of proteins in the centromere, some, such as CENPA, are constitutive and remain at the centromere throughout the life cycle of the cell. Others, such as motors and mitotic checkpoint proteins, are facultative and appear transiently at the kinetochore to regulate specific functions required for movement and progression of mitosis (4).

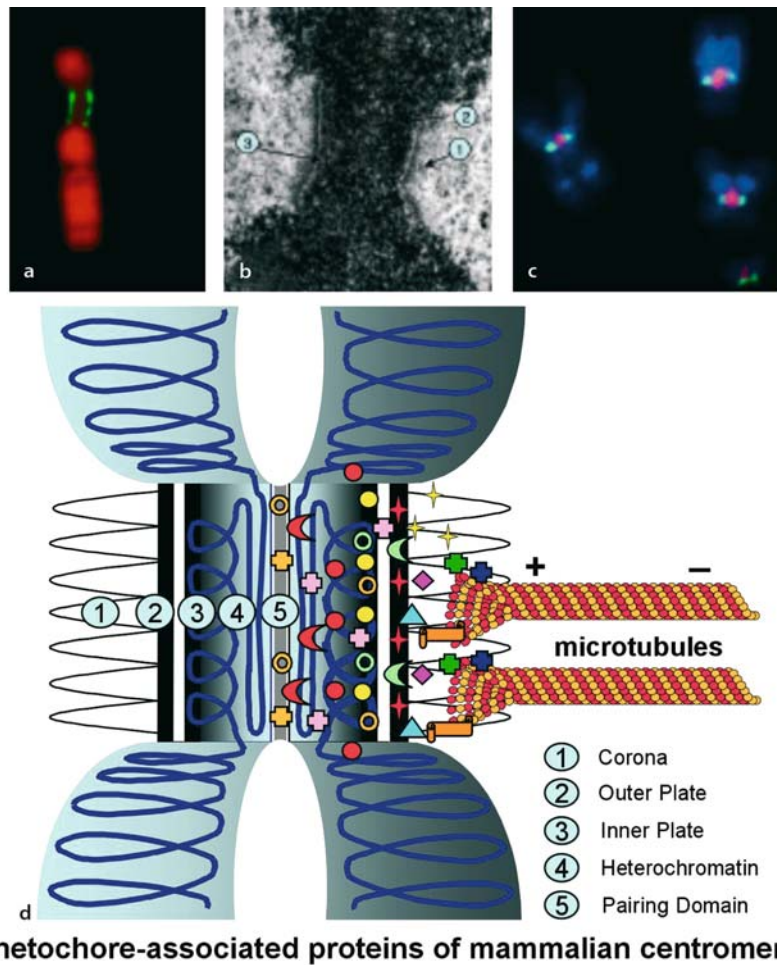
Chromosome Movements

The kinetochore must be physically tethered to spindle microtubules to enable the complex movements necessary to align chromosomes on a mitotic spindle and segregate them to the nuclei of daughter cells. Spindle microtubules are long, hollow tubules with a slow growing (minus) end, attached to poles at opposite ends of the spindle and a fast growing (plus) end, attached to the kinetochores facing each pole. In the budding yeast, a single spindle microtubule attaches to a discrete point at the centromere. In human chromosomes, bundles of microtubules attach to the trilaminar kinetochore plates located on opposite sides of the centromere (Fig. 1).

Cells in mitosis progress through five stages, including ►prophase, ►prometaphase, ►metaphase, ►anaphase and ►telophase. At the end of prophase, the ►nuclear envelope breaks down, releasing chromosomes into the cytoplasm. As chromosomes maneuver to the cell equator during prometaphase (►congressional movement) and align on the metaphase plate, sister kinetochores on each side of the centromere attach to spindle microtubules extending from opposite poles, thereby subjecting chromatids to equal force in opposing directions. Subsequently, the chromatids split apart and the kinetochores pull the chromosomes to opposite poles (►anaphase movement) (5).

Regulatory Functions

For many years, the centromere and its associated kinetochores were thought to function as passive “handles” that spindle microtubules grasp to pull microtubules along the mitotic spindle. More recent studies have discovered a remarkable set of kinetochore proteins that form a complex signaling circuitry for regulating chromosome movement (2, 3) and progression through mitosis. As shown in the diagram in Fig. 1, a group of proteins known as checkpoint proteins are strategically located along the outer face of



Kinetochore-associated proteins of mammalian centromeres

Constitutive	Microtubule Associated	Motors	Motor Associated	Passenger	Mitotic Checkpoint	Cohesion	Microtubule-Kinetochore Attachment	Misc. Kinetochore Proteins
CENP-A CENP-C CENP-G CENP-H CENP-I Mis12 DC8 c20orf172 CENP-B HP1	CLASP1 CLIP-170 MCAK APC EB1 ICIS	Dynein Dynactin CENP-E	ZW10 ROD Zwilch Zwint-1 Lis1	CENP-F Aurora B Kinase Aurora C Kinase INCENP Survivin Borealin	Bub1 Bub3 Mad1 Mad2 BubR1 Mps1	SMC1 SMC3 Sgo	Hec1 Nuf2 Spc24 Spc25	Nup107 Nup133 Nup358 RanGAP1 PP1 γ

Centromere. Figure 1 (a) Acrocentric chromosome from the muntjac deer with the centromere (green) located in the primary constriction between the long arm and short arm (red). The centromere proteins are stained with a specific antibody directed against CENPs A, B, and C (green). (b) Electron micrograph of muntjac deer chromosome showing prominent inner and outer kinetochore plates and corona. (c) Triple-stained chromosome showing proteins (green) on sister kinetochores. DNA in the chromosome arms is stained blue. (d) Diagram of the centromere/kinetochore showing the 5 sub-domains of the centromere. The colored symbols depict the approximate localization of various known kinetochore proteins. Two microtubules have attached to one sister kinetochore with the plus ends, associated with the corona and various anchoring and motor proteins, found at the splayed ends. The minus ends of the microtubules are directed toward the spindle poles (not shown). The table below lists currently known centromere-associated proteins of human chromosomes. Symbols correspond to the sub-domain localization of groups of proteins within the centromere/kinetochore. (Fig. 1a–c are reproduced from Van Hooser, AA. 2001. Specification of kinetochore-forming chromatin by unique nucleosome structure [Ph.D. dissertation]. Houston (TX): Graduate School of Biomedical Sciences, Baylor College of Medicine.)

the kinetochore, in sub-domains 1–3. ►**Checkpoint** proteins are enzyme complexes that modify protein substrates by inserting phosphate groups at specific sites to regulate their functions. As chromosomes align on the metaphase plate, a “wait anaphase” checkpoint prevents further progression, assuring that all chromosomes are properly aligned and fully attached to spindle microtubules extending from opposite poles. When alignment is complete and equal pole-directed forces are exerted on opposing sets of sister kinetochores, a second checkpoint “switch” signals the chromatids to split apart and move to opposite poles. This regulation achieves proper segregation and corrects any improper orientation of chromosomes, thus assuring equal partitioning of the genome. When checkpoints are abrogated or fail, however, errors in chromosome segregation occur that may lead to serious genetic disorders.

Clinical Relevance

Unequal distribution of chromosomes to the nuclei of daughter cells can produce altered genomes in cells characteristic of a number of serious human diseases. Normal mitosis leads to the perpetuation of a balanced ►**diploid genome** in which each nucleus contains 46 chromosomes. Errors in mitosis can lead to an imbalanced genome – a condition known as ►**aneuploidy**. Aneuploidy has numerous causes. Defective centromeres can interfere with normal chromosome splitting and chromosome segregation in mitosis. Premature chromosome separation can also lead to genomic imbalance. A centromere defect can cause chromosomes to lag behind and be excluded from the nucleus. Mutations in checkpoint proteins in the centromere can prevent chromosomes from splitting apart at anaphase, a process known as ►**nondisjunction**. Chromosome breakage and translocations can produce a condition in which chromosomes contain more than a single copy of the centromere, producing dicentric or tricentric chromosomes that fail to divide and segregate properly. Defects in chromosome structure and distribution can also arise when ►**neo-centromeres** form on chromosomes. Numerical disorders involving human sex chromosomes include the 47XXY condition in Klinefelter's syndrome and 45X in Turner's syndrome. ►**Downs syndrome** results from duplication of ►**chromosome 21**. Aneuploidy is also associated with most forms of human cancer.

►**Meiosis and Meiotic Recombination**

►**Mitotic Spindle**

►**Transposons**

References

1. Brinkley BR, Valdivia MM, Tousson A et al (1989) The kinetochore: structure and molecular organization. In: Hyams J, Brinkley BR (eds) *Mitosis: molecules and mechanisms*. Academic Press, New York, pp 77–118
2. Cleveland DW, Mao Y, Sullivan KF (2003) Centromeres and kinetochores: from epigenetics to checkpoint signaling. *Cell* 112:407–421
3. Karpen GH, Allshire R (1997) The case for epigenetic effects on centromere identity and function. *Trends Genet* 13:489–496
4. Sullivan KF (2001) A solid foundation: functional specialization of centromeric chromatin. *Curr Opin Genet Dev* 11:182–188
5. Reider CL, Salmon ED (1998) The vertebrate cell kinetochore and its role during mitosis. *Trends Cell Biol* 8:310–317

Centrosome

Definition

The vertebrate centrosome, also called the “microtubule-organizing center”, consists of two barrel-shaped centrioles. These are made up of nine triplets of short microtubules embedded within a so-called pericentriolar matrix. The two centrioles are arranged such that one is perpendicular to the other. Centrosomes influence a variety of processes: organelle transport, cell shape, polarity and motility. Centrosomes are vital for proper execution of mitosis and control spindle bipolarity, spindle positioning and cytokinesis. They serve as nucleation points for microtubular polymerisation and constrain the lattice structure of a microtubule to 13 protofilaments. Thus, in animal cells, the centrosome is the major microtubule organizing center (MTOC).

►**Cell Cycle – Overview**

►**Cell Division**

►**Cytoskeleton**

►**Mitotic Spindle**

Centrosome Cycle

Definition

Throughout the cell cycle centrosomes undergo a series of morphological changes. These discrete events are called centrosome cycle (also termed growth cycle of the centrosome). In higher organisms it consists of a duplication, maturation, separation and centrosome disorientation phase. Centrosome duplication and segregation is strictly coordinated with the duplication and segregation of the genome.

►**Cell Cycle – Overview**

►**Cell Division**

Cephalochordates

Definition

Cephalochordates comprise of a subphylum of chordates that includes the amphioxus *Branchiostoma* which has a notochord, dorsal nervous system and segmented trunk, and in this respect are the precursors of vertebrates.

► Somitogenesis

abnormalities; later the name was changed to Zellweger syndrome.

► Peroxisomal Disorders

CF

► Cystic Fibrosis (CF)

Cer

► Cerberus

CF I/CF II

► Cleavage Factor I (CF I)/Cleavage Factor II (CF II)

Cerberus

Definition

Cerberus (Cer) is a multifunctional secreted inhibitor of ► Nodal, ► BMP, and ► Wnt.

► Wnt/Beta-Catenin Signaling Pathway

CFTR

Definition

► Cystic Fibrosis Transmembrane Conductance Regulator

Cerebellum

Definition

Cerebellum refers to a brain structure that is involved in the learning of motor skills, in modulating force and range of movement.

► Brain

► Neutrophilic Factors

CGH

► Comparative Genomic Hybridization

Cerebro-Hepato-Renal Syndrome

Definition

Cerebro-hepato-renal syndrome describes a syndrome characterized by multiple congenital

Definition

Channel assembly designates the oligomerisation of ► connexin subunits into hexameric hemichannels (connexons) and their docking with hemichannels on the surface of neighbouring cells. In cells expressing more than one type of connexin, connexons are homomeric or heteromeric and gap junctions homotypic or heterotypic.

► Intermediate Filaments

Channelopathy

Definition

Channelopathy denotes a disease (e.g., certain forms of epilepsy, deafness, migraine or heart arrhythmia) that is associated with ion channel malfunction, they are often linked to mutations of the respective channel gene.

► Ion Channels/Excitable Membranes

in ► [connexin](#) trafficking and assembly in cells. CMT affects 1 in 2,500 people.

► [Glial Cells and Myelination](#)

► [Intermediate Filaments](#)

► [Hereditary Neuropathies, Motor and/or Sensor](#)

► [Hereditary Neuropathies, Motor and/or Sensory](#)

Chaotrope

Definition

A chaotrope or chaotropic agent is a compound that modifies the solubility of, e.g. a protein, by altering the structure of the surrounding water.

► [Mass Spectrometry: MALDI](#)

Chaperone (Proteins)

Definition

Chaperone (proteins) are a family of cellular proteins responsible for regulation of the folding of newly synthesized polypeptide chains, and in some cases, their assembly into oligomeric structures. Chaperone proteins are not components of those final structures.

► [Defective Protein Folding Disorders](#)

► [Hereditary Spastic Paraplegias](#)

► [NFκB Pathway](#)

► [Protein Folding](#)

Charcot-Marie-Tooth Disease

Definition

Charcot-Marie-Tooth disease (CMT) is the most common hereditary sensorimotoric neuropathy. It is characterized by weakness and atrophy, primarily in leg muscles. CMT results from progressive damage to nerves and involves abnormal myelination of the peripheral nerves. The X-chromosome linked disease is associated with connexin 32 mutations, with over 200 mutations detected and many resulting in problems

Charge Coupled Device cDNA

► [DNA Microarrays/DNA Arrays](#)

Charged Coupled Device (CCD) Camera

Definition

Charged coupled device (CCD) camera designates a video camera, in which a charged coupled device (CCD) is used as a two-dimensional light detector. A CCD is a light-sensitive silicon chip containing a two-dimensional array of light sensitive elements. The single elements represent photodiodes, where the produced photoelectrons are collected in potential wells, which are the picture elements ("pixels") of the CCD. CCDs show an extremely linear response to the light irradiance, and can collect the charges for long exposure times. Scientific grade CCDs feature extremely low dark currents and high quantum efficiencies for the photon-to-charge conversion process. The main source of noise in CCD images is produced by the charge readout. The noise level is related to the readout speed, i.e. the digitization rate. Therefore, scientific grade CCDs are often read out with speeds below the video standard, and are then designated as slow-scan CCDs.

► [Electron Tomography](#)

► [Fluorescence Microscopy: Single Particle Tracking](#)
[Single Particle Tracking](#)

Charge-Tagging

Definition

Charge-tagging refers to tagging nucleic acids (and other biomolecules) with a single fix charge.

► [SNP Detection and Mass Spectrometry](#)

CHD

- Familial Hypercholesterinemia

Checkpoint

- Cell Cycle Checkpoints

Checkpoint Kinase 2

- CHK2

Chemical Cross-Linking

Definition

Chemical cross-linking describes the process of covalent coupling of molecules with multi-functional reagents.

- Protein Interaction Analysis: Chemical Cross-Linking

Chemical Exchange

Definition

Chemical exchange defines an exchange of a molecule between different environments due to chemical reactions, binding events, or conformational transitions.

- Protein-Ligand-Interaction by NMR

Chemical Shift

Definition

The relative difference of resonance frequencies (in Hertz) between a probe ►nucleus and a reference

nucleus multiplied by 10^6 : $\delta(X) = 10^6 \Delta\nu/\nu$ (where $\Delta\nu = \nu(X) - \nu(\text{Reference})$).

- Protein-Ligand-Interaction by NMR

Chemiluminescence

Definition

The emission of light during a chemical reaction.

- Protein Microarrays as Tool for Protein Profiling and Identification of Novel Protein-Protein Interactions

Chemokine Receptors

RICHARD HORUK

Department of Immunology, Berlex Biosciences,
Richmond, CA, USA
horuk@pacbell.net

Definition

Chemokine receptors are members of the GPCR superfamily (1). Synonyms are chemoattractant receptors, chemokine receptors, ►G-protein coupled receptors.

At the latest count well over 850 members of this G-protein coupled receptor (GPCR) superfamily have been identified and classified into families (2). Six CXC, ten CC and one CX3C and XC chemokine receptors have been cloned so far (3). Receptor binding initiates a cascade of intracellular events mediated by the receptor associated heterotrimeric G proteins. These G-protein subunits trigger various effector enzymes, which leads to the activation not only of chemotaxis but also to a wide range of functions in different leukocytes, such as an increase in the respiratory burst, degranulation, phagocytosis and lipid mediator synthesis (4).

Characteristics

Chemokines belong to a large family of small, chemotactic cytokines characterized by a distinctive pattern of four conserved cysteine residues (5). They are divided into two major (CXC and CC) and two minor (C and CX3C) groups dependent on the number and spacing of the first two conserved cysteine residues. Although originally identified on the basis of their ability to regulate the trafficking of immune cells, the biological role of chemokines goes well beyond this simple description of their function as chemoattractants, and they have been shown to be

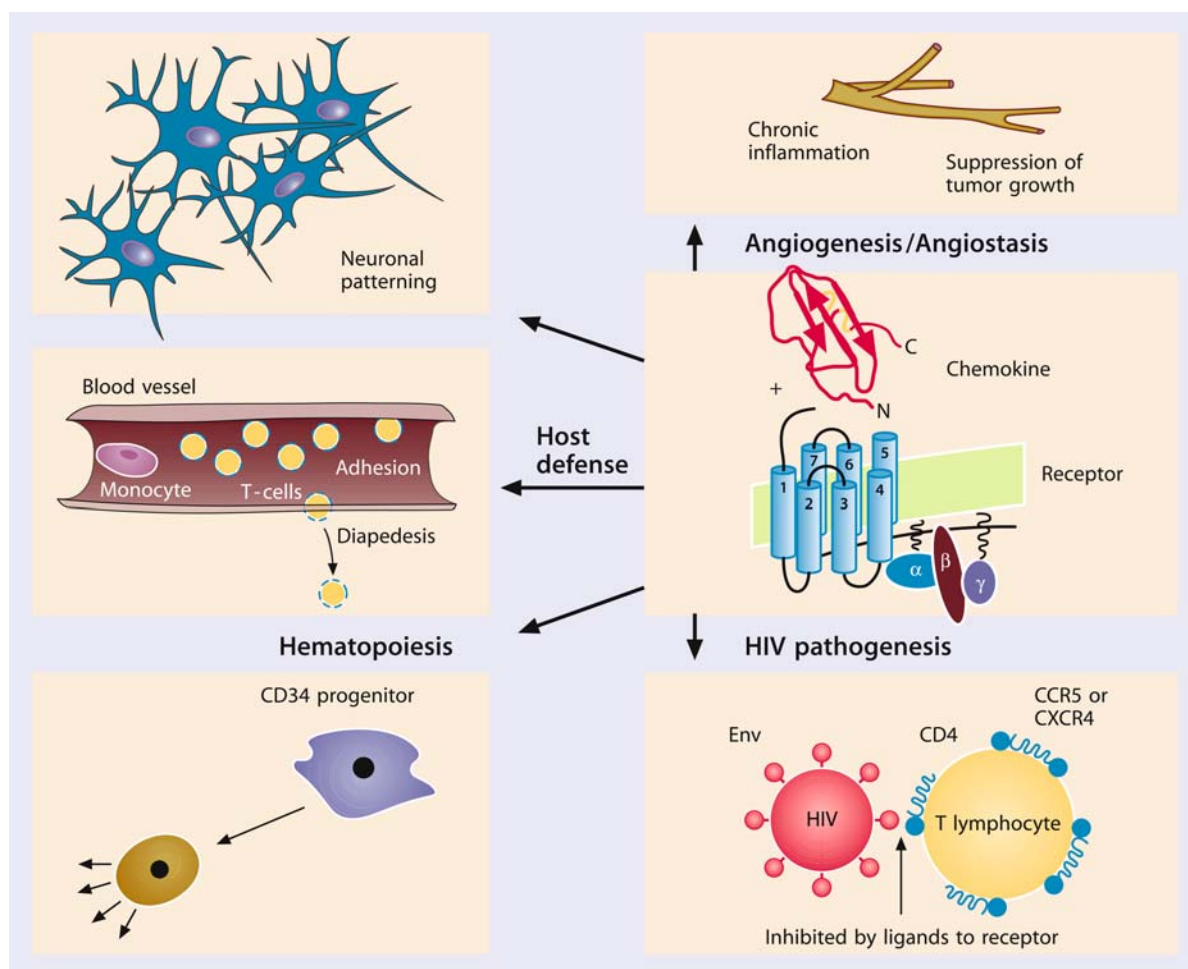
involved in a number of biological processes, including growth regulation, hematopoiesis, embryologic development, angiogenesis and HIV-1 infection (6) (Fig. 1). Chemokines have been shown to be associated with a number of autoimmune diseases including multiple sclerosis, rheumatoid arthritis, atherosclerosis, dermatitis, organ transplant rejection, etc. (7). Evidence, reviewed below, is mounting that chemokines may play a major role in the pathophysiology of these diseases and thus chemokine receptor antagonists could prove to be useful therapeutics in treating these and other proinflammatory diseases.

Chemokine Receptors

Although leukocytes continue to be the major site of expression of chemokine receptors, several studies have recently demonstrated chemokine receptor expression on neurons in the CNS. A number of

chemokine receptors including CXCR2, CXCR4, CCR1, CCR5 and DARC have been demonstrated in either adult or in fetal brain. Not only were these receptors present on the cell surface but they were also functional. Clearly the role of these receptors on CNS neurons must be very different from their role on immune cells. Given that human astrocytes can be stimulated with cytokines to tend to up-regulate the expression of chemokines, it is tempting to speculate that perhaps *in vivo* during CNS development chemokines secreted by astrocytes can engage specific receptors expressed on neurons and may play a role in the directed migration of specific subsets of neurons to distinctive regions of the brain.

Genetic mutations of receptors both natural and induced (by targeted gene disruption) can help to unravel their biological roles. Nature has been generous in this regard by providing us with two naturally



Chemokine Receptors. Figure 1 Biological functions of chemokine receptors as exemplified by the chemokine receptor CXCR4.

occurring examples of gene inactivation for chemokine receptors. Humans homozygous for inherited inactivating mutations of the Duffy (DARC) gene and the CCR5 gene have been identified and appear to be phenotypically normal and healthy. Indeed these gene inactivations appear to be beneficial to their hosts, rendering them resistant to certain infectious diseases. For example, the Duffy negative individuals are resistant to malaria induced by *Plasmodium vivax*, which utilizes DARC to attach to and enter erythrocytes and the CCR5 negative individuals are resistant to HIV-1, which utilizes this chemokine receptor as a coreceptor for invasion (see section role of receptors in HIV infection).

Analysis of receptor-inactivated individuals can also be useful in clarifying their role in disease. For example, it is known that MIP-1a (CCL3) appears to play an important role in multiple sclerosis. This chemokine is a potent agonist for both CCR1 and CCR5 receptors opening the possibility that either of these receptors could be involved in mediating the development of the pathophysiological changes seen in this disease. However, analysis of a large group of individuals, comprising both normal subjects and those suffering from relapsing/remitting multiple sclerosis, showed that there was no significant difference in the allele frequency of the CCR5 mutation between the groups. These studies indicate that CCR5 is not an essential component in the expression of multiple sclerosis and further implicate CCR1 in the disease.

Role in Immune Response

Chemokines are potent chemoattractants that provide directional cues to summon leukocytes. Leukocyte recruitment is a three-step process that involves the formation of solid phase chemokine gradients generated by chemokine binding to extracellular matrix proteins like glycosaminoglycans that decorate the cell surface of endothelial cells. These gradients then attract immune cells which first undergo selectin mediated rolling along the endothelial cells. Chemokine mediated up-regulation of CD11/18 complexes then results in a much firmer adherence of immune cells to the endothelium and this culminates in diapedesis of leukocytes across the endothelial space into the tissues. By regulating the movement of different subsets of leukocytes from the peripheral blood to extravascular sites such as organs, skin or connective tissue, chemokines play a critical role in the maintenance of host defense as well as in the development of the immune response. However, sometimes these molecules can inappropriately target immune cells to attack their own tissues and organs leading to inflammation and cellular destruction. Indeed there is now strong evidence in support of the idea that chemokines play an

important role in the pathogenesis of a number of autoimmune diseases such as ►multiple sclerosis and ►rheumatoid arthritis.

Rheumatoid arthritis is a chronic inflammatory disease characterized in part by a memory T lymphocyte and monocyte infiltrate. The interaction of the same cell types also plays a major role in the demyelinating processes that culminate in multiple sclerosis. Recent studies using neutralizing antibodies have provided strong *in vivo* concept validation for a role of chemokines in animal models of both diseases. For example, in an adjuvant-induced arthritis (AIA) model in the rat, antibodies to CCL5 were able to abrogate the development of the disease in the animals by greatly reducing the infiltration of mononuclear cells into tissue joints. Similarly antibodies to CCL3 prevented the development of both initial and relapsing paralytic disease as well as infiltration of mononuclear cells into the central nervous system of a mouse experimental autoimmune encephalomyelitis (EAE) model of multiple sclerosis. These results strongly suggest that these chemokines play important roles in these T-cell mediated autoimmune diseases.

In contrast to the role of T cells and monocytes in chronic inflammation, the primary hallmark of acute inflammatory diseases such as empyema, acute lung injury, acute respiratory distress syndrome (ARDS) and dermatitis is tissue infiltration by neutrophils. Thus, neutrophil activating CXC chemokines like CXCL8 are most often associated with these diseases. Clear evidence for a role of CXCL8 in acute lung injury and pleurisy has been provided by the finding that antibodies to CXCL8 dramatically increased the survival time of rabbits in models of disease. Not only did the antibodies increase the alveolar-arterial oxygen difference of the animals, increasing the oxygenation of the blood and therefore decreasing breathing difficulties, but they also affected a significant reduction in the infiltration of neutrophils into the lung.

Based on the demonstrated role of chemokines in disease, the generation of small molecule chemokine receptor antagonists has received great interest from pharmaceutical companies, since they are seen as attractive therapeutic approaches. GPCRs like chemokine receptors have in the past been an extremely fertile source of biological targets in the pharmaceutical industry and compound library screening has proven successful in the discovery of antagonists for a number of these receptors, i.e. CCK and neurotensin antagonists. Using similar approaches, several drug companies have now identified potent small molecule antagonists of a number of chemokine receptors, which should find broad utility in a variety of acute and chronic inflammatory diseases.

Clinical Relevance

Chemokine Receptor Inhibitors

Insight into the physiological and pathophysiological roles of chemokine receptors have been provided by studies with potent receptor antagonists for CCR1, CCR2, CCR3, CCR5, CXCR2, CXCR3 and CXCR4, reviewed in Ribeiro and Horuk (8). A number of these antagonists, CCR1, CCR3, CCR5, CXCR3 and CXCR4, are in human clinical trials. For example, the CCR5 antagonists are in phase III clinical trials for the treatment of ►AIDS. In addition, the CCR1 antagonist is in several phase II clinical trials.

The utility of chemokine receptor antagonists as therapeutic agents in disease can be exemplified by the CCR1 antagonist BX 471. BX 471 is a potent functional inhibitor specific for CCR1 based on its ability to inhibit a number of CCR1-mediated effects including CCL3 and CCL5 mediated increases in Ca^{2+} mobilization and extracellular acidification rate, CD11b expression and leukocyte migration. In addition, BX 471 demonstrates a greater than 10,000 fold selectivity for CCR1 compared with 28 different GPCR's. Pharmacokinetic studies demonstrated that BX 471 was orally active with a bioavailability of 60% in dogs. A recent model of renal fibrosis in the mouse showed that inhibition of CCR1 with BX 471 reduced leukocyte infiltration and renal fibrosis. BX 471-treated mice (day 0–10 and day 6–10) revealed a 40–60% reduction in interstitial macrophage and lymphocyte infiltrate compared with controls (v). Treated mice also showed a marked reduction of CCR1 and CCR5 mRNA levels and FACS analysis showed a comparable reduction of CD8+/CCR5+ T cells. Markers of renal fibrosis, such as interstitial fibroblasts, interstitial volume, mRNA and protein expression for collagen I, were all significantly reduced by BX 471- treatment compared with vehicle controls. In summary, blockade of CCR1 substantially reduced cell accumulation and renal fibrosis after UUO. Most interestingly, late onset of treatment was also effective and this is the first published example showing that a CCR1 antagonist is effective when given therapeutically as compared to prophylactically in an animal model of disease.

The CCR1 receptor antagonist BX 471 is also efficacious in a rat heterotopic heart transplant rejection model. Animals treated with BX 471 and a sub-therapeutic dose of cyclosporin, 2.5 mg/kg, which is by itself ineffective in prolonging transplant rejection, showed much more prolonged transplantation rejection than animals treated with either cyclosporin or BX 471 alone. Immunohistology of the rat hearts for infiltrating monocytes confirmed these data. 3 days after transplantation the extent of monocytic graft infiltration was significantly reduced by the combined therapy of BX 471 and cyclosporin. Thus, BX 471 given in combination

with cyclosporin resulted in a clear increase in efficacy in heart transplantation compared to cyclosporin alone. These data were in line with the observed effects of BX 471 in dose-responsively blocking the firm adhesion of monocytes triggered by CCL5 on inflamed endothelium. Together, these data demonstrate a significant role for CCR1 in allograft rejection. The promise of highly specific therapies for a number of devastating diseases is on the horizon, thanks to the identification of chemokine receptor antagonists and we can look forward with anticipation to the day when these drugs are finally marketed as potent therapeutics.

Chemokine Receptors in Pathogen Infection

A number of viruses including those in the *Herpes* and *Pox* families express chemokine-like or chemokine receptor-like molecules that presumably help them to survive immune attack and proliferate. In addition to these strategies, chemoattractant receptors have themselves been targeted as vehicles of cellular invasion by a wide variety of microbes. These range from the Duffy blood group antigen, a promiscuous chemokine receptor on human erythrocytes, which serves as a binding protein for the malarial parasite *Plasmodium vivax*, to the fractalkine receptor, CX3CR1, which is a portal of entry for the respiratory syncytial virus and the HIV-1 virus, which utilizes the chemokine receptors CXCR4 and CCR5 as coreceptors for cellular entry. CCR5 is an entry cofactor for M-tropic isolates of HIV-1 and is important in the early proliferative part of the disease, while CXCR4 is a coreceptor for T-tropic isolates of HIV-1 whose emergence in infected individuals usually correlates with accelerated disease progression.

HIV-1 resistance exhibited by some exposed but uninfected individuals is due, in part, to a 32 bp deletion in the CCR5 gene (CCR5D32), which results in a truncated protein that is not expressed on the cell surface. About 1% of Caucasians are homozygous for the CCR5D32 allele and appear to be healthy with no untoward signs of disease. In fact, recent findings suggest that homozygosity for the CCR5D32 alleles confers other selective advantages to these individuals, rendering them less susceptible to rheumatoid arthritis and asthma and prolonging survival of transplanted solid organs.

References

1. Baggiolini M (2001) Chemokines in pathology and medicine. *J Intern Med* 250:91–104
2. Venter JC, Adams MD, Myers EW et al (2001) The sequence of the human genome. *Science* 291:1304–1351
3. Horuk R (2003) Development and evaluation of pharmacological agents targeting chemokine receptors. *Methods* 29:369–375

4. Thelen M (2001) Dancing to the tune of chemokines. *Nat Immunol* 2:129–134
5. Mackay CR (2001) Chemokines: immunology's high impact factors. *Nat Immunol* 2:95–101
6. Horuk R (1998) Chemokines beyond inflammation. *Nature* 393:524–525
7. Gerard C, Rollins BJ (2001) Chemokines and disease. *Nat Immunol* 2:108–115
8. Ribeiro S, Horuk R (2005) Chemokine receptor antagonists from the bench to the clinic. Wiley-VCH, Weinheim, pp 371–402
9. Anders HJ, Vielhauer V, Frink M et al (2002) A chemokine receptor CCR-1 antagonist reduces renal fibrosis after unilateral ureter ligation. *J Clin Invest* 109:251–259

Chemokines

Definition

Chemokines (term fused from “chemotactic cytokines”) are protein factors that have initially been defined as attractants for different types of blood leucocytes to sites of infection and inflammation. They are produced locally in the affected tissue and act on leucocytes through selective chemokine receptors. Chemokines help to control leucocyte maturation, traffic and homing of lymphocytes and the development of lymphoid tissues.

- **Cancer Invasion and Metastasis: Cellular, Molecular and Clinical Aspects**
- **Chemokine Receptors**
- **Methylation of Proteins**

Chemotaxis

Definition

Chemotaxis is the movement of a microorganism or a eukaryotic cell in response to a chemical gradient, either towards or away from the substance. For example, the migration of an aggregation-competent *Dictyostelium amoeba* depends on chemotaxis towards cAMP as chemoattractant. In multicellular organisms, chemotaxis is critical in development as well as normal function. In host defense and inflammation, leucocytes show directional migration in response to certain chemotactic signals, like bacterial tripeptides (F-Met-Leu-Phe), complement 5a (C5a) or Chemokines. The phosphoinositides $\text{PtdIns}(3,4,5)\text{P}_3$ (PIP_3) and $\text{PtdIns}(3,4)\text{P}_2$ [$\text{PI}(3,4)\text{P}_2$] are key signaling molecules

for the control of directed movement of a cell in a chemical gradient.

- **Actin Cytoskeleton**
- **Cell Migration**
- **Cytokine Receptors**
- **Photoreceptors**

Chemotherapy

Definition

Chemotherapy is the employment of different chemical compounds to kill pathogenic cells. For advanced and metastatic tumors, chemotherapy is the most effective treatment, whereas treatment of the most locally confined malignant tumors is based on surgical and radiotherapeutic approaches. Chemotherapeutics are aimed at interfering of division and growth of cancer cells, particularly based on their fast growth as compared to the controlled growth behavior of normal cells; however, fast growing normal healthy cells might also be affected.

- **Multi-Drug Resistance**

Chenodeoxycholic Acid

Definition

Chenodeoxycholic acid, besides cholic acid, is one of the two primary bile acids produced from cholesterol in the liver. It inhibits the synthesis of cholesterol in the liver and reduces the biliary secretion of cholesterol.

- **Cholic Acid**
- **Peroxisomal Disorders**

Chiasma(ta)

Definition

Chiasma(ta) are a cytologically visible equivalent of chromosomal crossovers that are held in place by sister chromatid cohesion.

- **Meiosis and Meiotic Recombination**

Childhood Cerebral ALD

Definition

Childhood cerebral adrenoleukodystrophy (CCALD) is a severe neurologic disorder presenting in young boys in the middle of the first decade starting with behavioural abnormalities, decline in school performance and deterioration of vision and hearing ALD (Adrenoleukodystrophy).

► [Peroxisomal Disorders](#)

Chimeric Clones

Definition

The chimera is a mythic beast composed of parts of different animals and in the context of recombinant DNA a chimera is a clone consisting of insert DNA fragments not found together in nature. All library construction methods are liable to produce chimeras at some level but it has been a particular problem in YAC libraries.

► [YAC and PAC Maps](#)

Chimeras (Chimeric Clones)

Definition

The chimera is a mythic beast composed of parts of different animals. In the context of recombinant DNA technology, the term chimera is used to describe a clone consisting of insert DNA fragments not found together in nature. All library construction methods are liable to produce chimeras at some level. For example, chimeric mice contain cells of two different sources (usually, the original pluripotent stem cells from the embryo and the targeted embryonic stem cells of the host blastocyst aggregation chimerae).

► [Mouse Genomics](#)

► [Large-Scale Homologous Recombination Approaches in Mice](#)

► [Shotgun Libraries](#)

► [YAC and PAC Maps](#)

Chimeric Proteins

Definition

Chimeric proteins are fusion proteins that consist of two (normally separate) proteins or fragments thereof.

► [Two-Hybrid System](#)

ChIP on Chip

Definition

ChIP on chip refers to a method that is based on the hybridization of Chromatin ImmunoPrecipitation (ChIP) derived material, on a genomic microarray (chip) that allows the genome-wide localization of DNA-bound proteins.

► [ChIP Technologies, Basic Principles](#)

► [Drosophila as a Model Organism for Functional Genomics](#)

Chimeric Antibodies

Definition

Chimeric antibodies are hybrid ► [immunoglobulins](#) in which the original murine variable regions are preserved, and the constant regions are replaced with those of human immunoglobulin. This way, the human anti-mouse antibodies (HAMA) effects are reduced, and human effector functions for immune activation are gained. Such antibodies are produced by recombinant methods.

► [Monoclonal Antibodies](#)

► [Protein Interaction-Phage Display](#)

ChIP Technologies, Basic Principles

STEFAN WÖLFEL

Institut für Pharmazie und Molekulare Biotechnologie,
Abteilung Biologie, Arbeitsgruppe Bioanalytik und
Molekularbiologie, Ruprecht-Karls-Universität
Heidelberg, Heidelberg
wolfel@uni-hd.de

Definition

It is not absolutely sure when the expression chip technologies in the context of biomedical science was first used, but the fields of Genomics and Proteomics can hardly be imagined without the use of highly specialized (bio-)chips. (Bio-)chips are small tools typically used for a bioanalytic application. In most cases bio-chips are devices on which a biological molecule, complexes of molecules or even a complete living organism is immobilized on a solid surface and used to analyze the interaction with other biomolecules, complexes of molecules or living organisms. In contrast to other techniques based on immobilization on solid surfaces, the meaning also includes, that the carrier is rather small, in particular when compared to more traditional bioanalytical tools.

In most cases biochips contain more than one species of the immobilized molecules. These are than typically arranged in a geometric array, so that each immobilized molecule can be precisely defined by its position. This typical array design is also the reason that the term (bio-)chip and chip-technologies is used almost synonym to the term microarray. A microarray is a miniaturized array of molecules on a solid surface. If these molecules are biomolecules, the array is also a bio-chip. Chip technologies in bio- and bio-medical sciences including Genomics and Proteomics can be categorized in various ways: the arrangement in which molecules are placed on the chip, microarray (see below); the way they are manufactured, e.g. printed glass slides; the type of molecules they contain as probes, e.g. DNA-, cDNA-, protein-, antibody-chip, cell-chip; the application for which they are used, e.g. gene expression chip, SNP-chip; the way they are used and other technological specifications, e.g. microfluidics chip, microchromatography, BioMEMS (MEMS: micro electro mechanical systems). In recent years also the term lab on a chip became used, describing integrated solutions, that combine analytical applications with integrated fluidic systems that enable a simple work flow and no additional laboratory equipment (except for a dedicated workstation) to perform the designated bioanalytical task.

In the following, the basic design and the manufacturing of chips used in Genomics and Proteomics will be described.

Description

Many bio-chips have in common that probe molecules are immobilized on a solid surface (chip) and used for the detection of interacting molecules in a test solution. To enable the systematic analysis of large numbers of molecules in parallel, probe molecules, each specific for a single analyte, are arranged in a two dimensional array format, so that each position is defined by two coordinates x,y. Given the small micrometer scale

required to deposit hundreds and thousands of probes on a single chip, these types of chips are also called microarrays. The array arrangement ensures that each individual species of immobilized molecules is clearly identified by the coordination system, which is essential for the precise detection and measurement of the detected molecules.

The basis for the manufacturing of microarrays was the development of high precision tools that allow the precise deposition of probe molecules in the array format, as well as improvements for the synthesis of biological macromolecules, especially nucleic acids and peptides. In the latter case, one particular development was from the beginning connected with the fabrication of bio-chips (DNA-chips), light directed synthesis of oligonucleotides (1). This method allows to use photolithographic techniques developed in the computer industry to synthesise nucleic acid polymers (oligonucleotides) directly on a solid chip surface *in situ* (see below). For the versatility of microarrays it was, however, more important, that prefabricated biomolecules, i.e. the synthesis or preparation is performed independent of the microarray surface, could be produced and purified in large quantities. For the production of specific bio-chips, these molecules are then repeatedly deposited on the surfaces of chip substrates using high precision printing tools. This type of technology was in a first application used to spot and immobilise cDNA fragments on microscope glass slides (2).

Solid Support

The basis for all bio-chip and microarray applications is a reliable solid support on which probe molecules can be immobilized. The basic requirements for the solid support are i) compatibility with the detection system, ii) suitable reactive groups on the surface for the immobilization of the probe molecules, iii) compatibility with the immobilization protocol and finally iv) integration into an easy handling system.

- i For most optical detection systems utilising fluorescent dyes, glass surfaces are the preferred substrates. They are translucent, do not interfere significantly with a broad spectrum of light and do not show auto-fluorescence. For specialized applications plastic materials, e.g. polypropylene, may be equally suitable. Silicon wafers, the typical substrate for microelectronics- (computer-) chips are also used but require more specialized detection systems. Some detection systems use specific optical properties at the surface, that require additional surface optimization, e.g. wave guides.
- ii Appropriate reactive groups are not genuinely present on the surfaces described above. These reactive groups are mostly introduced in a two step

process. First reactive groups are generated on the solid support; second the surfaces are coated with thin layer of reactive molecules that contain highly specific reactive groups for immobilization of the probe molecules. This step is also often used to create more flexible, rough surfaces that facilitate accessibility, increases the surface area and provide more reactive groups for coupling of the probe molecules.

- iii Both, the solid support and the surface must be compatible with the immobilization protocol. Most spotting systems work well with any type of surface and solid support. However, immobilization with microfluidic systems is limited to optimal prepared structures, which include alternating hydrophilic and hydrophobic as well as reactive and non reactive areas.
- iv For routine use of chip technologies in research laboratories but especially in diagnostic applications the chip must be integrated in an easy handling system. In many integrated systems, the actual biochips are glass for compatibility with the optical detection system.

Further variations of the solid support contain microstructures, e.g. cavities to deposit cells, hydrophobic barriers, specific reactive positions, porous structures and microsieves.

Surface Chemistry

For efficient immobilization an appropriate surface chemistry is required. Glass surfaces and even more silicon surfaces are only poorly reactive with biomolecules. One technology is the application of a thin transparent gold layer. Gold adsorbs with high efficiency thiol groups (-SH), which can be easily introduced in biomolecules. The association of thiols with gold is very strong and withstands treatment with most aqueous buffer solutions commonly used in all subsequent analytical reactions with the biochip. To introduce alternative reactive groups to the surface linker molecules, that contain two reactive groups, are used. Common reactive groups suitable for immobilization reactions are amino (-NH₂), epoxy (-COC) and aldehyd groups (-COO). An alternative to the gold thiol chemistry is the use of silan-linkers. These linkers contain a silan group on one end and the desired reactive group for coupling of the probe molecule on the other end. The silan group attaches efficiently to the glass surface leading to an efficient coating with the specific reactive groups. The specific probes are then immobilized through groups that are complementary to the reactive groups on the surface. In the pioneering work of Schena and colleagues, cDNA microarrays on microscopic glass slides were prepared using a simple coating with poly-lysine. This single amino-acid

polymer provided a layer of reactive amino groups to which the spotted cDNA could be crosslinked and used as probes for hybridization to detect gene expression (2).

In addition to the creation of simple layers of reactive groups on the surface, various approaches were used to increase the capacity of planar surfaces. These approaches include coating of glass slides with biopolymer layers, e.g. agarose or gelatine, or with chemically defined matrices, like polyacrylamide or nitrocellulose. These surface layers provide a three dimensional matrix which is accessible for reaction by diffusion and – very important – a significantly increased number of reactive groups for the immobilization of probe molecules, which leads to an increased probe concentration in hybridization or other binding reactions.

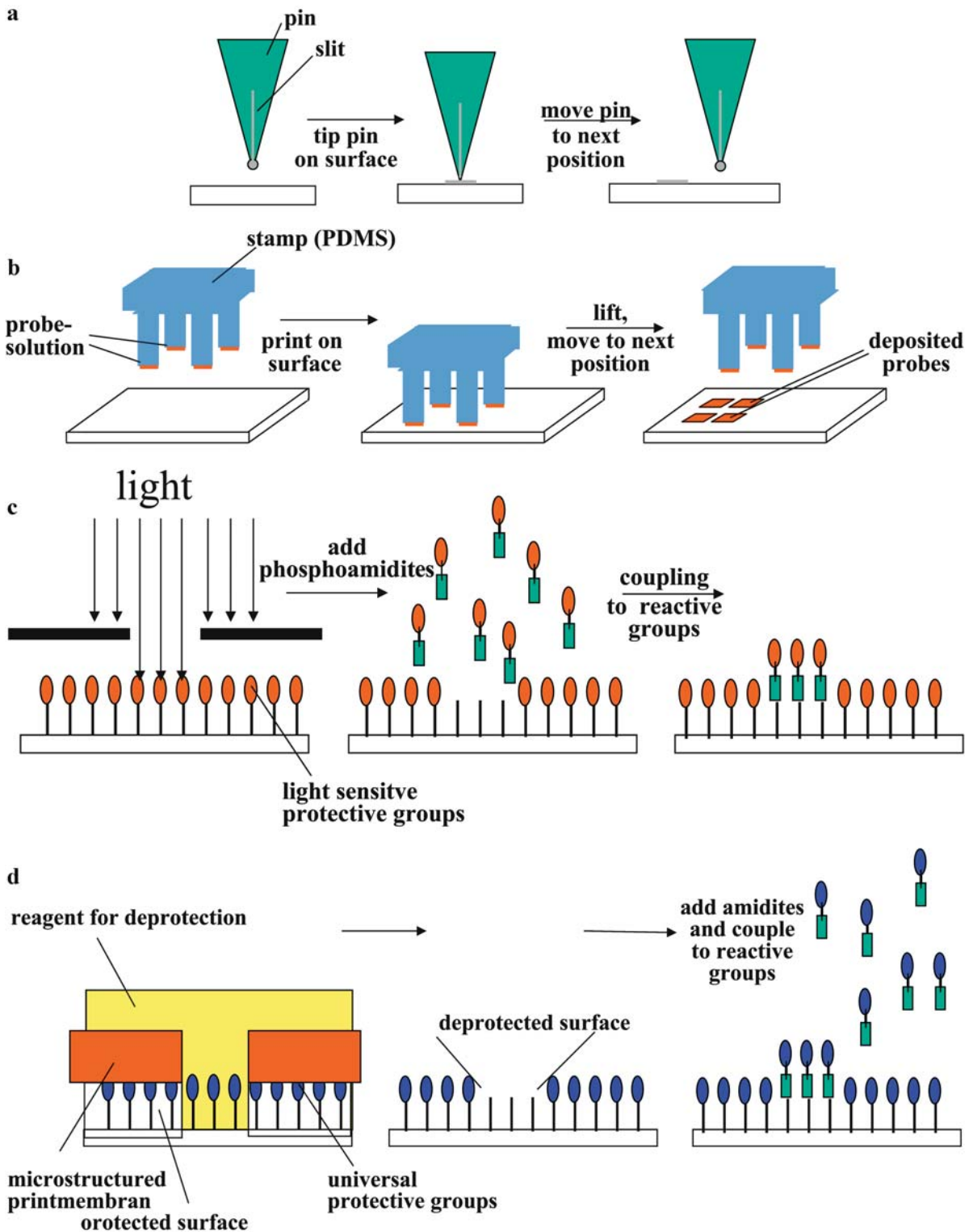
Despite these advantages most applications use the simpler linker based surface chemistry, which at least at present is better compatible with handling and detection methods.

Immobilization

Chips carrying microarrays are prepared following two main procedures. One approach is deposition of prefabricated molecules on the designated positions of the chip. The other is the *in situ* synthesis of the specific molecules on the chip surface. Both approaches are widely used and have their advantages and disadvantages. The first is more widely distributed at present. It uses simpler technology, dedicated robotic systems with sub-micrometer precision, that can be relatively easy set up in any research laboratory. This approach also has the advantage that any type of molecule and even living cells can be arranged on a chip in array format. The latter, *in situ* synthesis, is more suited for large scale manufacturing of identical chips, but requires at present high precision mask aligner technology and clean room facilities.

Deposition of Prefabricated Molecules

The most widely used method at present to deposit prefabricated molecules is contact tip printing using split pins. Split pins function in principal like the feather of a fountain pen. Tipped in a reservoir with liquid, the slit of the pin is filled by capillary forces. The pin is then moved to the desired position on the array substrate, where after contacting the surface it releases a tiny volume of liquid, typically a few nanoliters (nL) (Fig. 1a). The released volume depends on the slit in the pin, the surface of the array and the viscosity of the liquid. Surfaces with extended 3D structures, e.g. (bio-)polymer gels, will soak more liquid from the pin than single molecule reactive layers, leading to higher probe concentrations. Nevertheless, with identical surfaces, the passive dosing will result in



Chip Technologies, Basic Principles. Figure 1 Printing methods (a) Split pin contact tip printing (b) Micro contact printing (c) Photolithographic deprotection (d) Mikrowet Printing (μ WP).

roughly similar amounts of deposited probes in each spot (position). This is a critical point considering precise quantification, which is not obtained by this method.

An other approach to prepare microarrays is the use of precise nanoliter pipetting systems, e.g. piezo-electric microdispensers. These systems deliver the volumes to the chip surface with a much higher precision. At present these type of robotic dispensing is only used when smaller numbers of chips with not to many different features (specific properties of the chip system, which are used alternatively to probe in many microarray applications) but with high precision are needed, because the spotting process is more time consuming compared with the split pin technology.

For high throughput printing of micro arrays a printing technology called top spot printing was also developed. This method uses a printing cartridge with can be filled with hundred different liquids which is then delivered through fine capillaries to the chip surface. The release of drops from the microcapillaries is mediated by applying a short pressure pulse to the liquid reservoir (4). Information about the precision for quantification is not available for this method.

Printing in microscale is also possible using simple microstructured stamps typically made from PDMS. These stamps are made using microstructured templates prepared with photolithography and are then used to transfer thin films of liquid containing reactive molecules to chip surfaces. Like traditional stamps, this type of technology is best if only one type of liquid must be transferred (Fig. 1b). If no repeated deposition is required, high precision micropatterns of reactive groups, can be prepared in large number without any expensive equipment. This method is for example very suitable to deposit peptides, or protein layers to guide the growth of cells to specific positions and in specified form on glass or plastic substrates (5).

Microstructured PDMS but also other technologies allow to create micro fluidic networks that can be place on chip substrates in such a way that tightly sealed channels are generated. These fluidic channels are than filled with reactive molecules for deposition. Using micropatterned surfaces which contain defined reactive positions immobilization can be precisely directed to these positions and specific patterns of immobilized molecules are obtained (6).

In situ Synthesis of Probe Molecules

In situ synthesis of chip-size microarrays is mainly used for nucleic acid microarrays. In principal any type of polymeric probe molecule could be made with these protocols, if the appropriate reactive building blocks are available (1). Light directed chemical synthesis requires polymer synthesis building blocks with protective groups that are removed under the influence

of light (Fig. 1c). These are available for nucleic acids synthesis, but not for protein synthesis. (Nucleic acids require only four different building blocks (amidites), whereas peptide synthesis requires 20 building blocks.) The two other high resolution protocols for *in situ* synthesis, protection with photoresist (7) and micro wet printing (8) (Fig. 1d) can utilize any type of protective group chemistry and thus are suitable for the synthesis of any type of polymeric probes on chip microarrays.

Using light directed chemical synthesis in combination with large glass wafers and mask aligner technology, which – like in the computer industry – yields several hundred chips in parallel, the company Affymetrix Inc., USA (14) produces oligonucleotide arrays, that contain several hundred thousands (the number is still increasing) of oligonucleotide probes, 20–30 nucleotides in length. The photolithographic process allows to create tiny squares for each probe position in the low micrometer scale with nanometer precision. Oligonucleotide arrays fabricated in this way are used for gene expression profiling and genotyping. Readily assembled in a fluidic cartridge, these arrays are easy to handle and deliver highly reproducible results.

(There are also protocols for the *in situ* synthesis of peptide arrays, but although the synthesized peptide arrays are reduced in size, they still can not be considered as chips or microarrays. Nevertheless, peptide macro arrays are a high quality and reliable source for the pre-synthesis of peptides for spotting on other chip platforms (9).)

An alternative approach for the *in situ* synthesis of (micro-) arrays is the use of ink-jet type of printers (10, 11). The limitation here is that unlike the ink used for printing documents, the liquid reagents required for biopolymer synthesis can not be adjusted as easily to the viscosity requirements of these highly specialized dispensers.

Handling and Detection

Handling of chips in the laboratory as well as detection of signals in chips is at present far away from being standardized. The reason for this is, that for many laboratory applications using microscope glass slide based microarrays various low tech solutions can be used to obtain results. For reliable reproduction of the results it is required to constantly agitate (stir, move) the low sample volumes to enhance the poor and slow reactions between the analyte molecules and the immobilized probes. Some commercial microarray systems are preassembled in a microfluidic cartridge in which the on chip reactions are performed. Moreover, microarrays can be integrated into established laboratory reaction tubes or microtitre plates, e.g. ArrayTube® (Clontech Chiptechnologies, Jena, Germany) (12, 15).

The most widely used method for the detection of microarray results is the detection of fluorescence using dedicated high resolution microarray scanners. For this, fluorescent dyes are either directly coupled to the analyte molecules or are added in a final development step, e.g. through binding of fluorescently labelled streptavidin to biotin incorporated in the target molecules.

An alternative to fluorescence is the use of colorimetric detection. This includes various approaches, which lead to the specific deposition of dye (or silver) precipitates at the respective position on the microarray, which can be detected with CCD cameras or by optical scanners in the visible light range (13). Similarly chemiluminescence reactions can be used in the final detection step.

Clinical Applications and Therapeutical Consequences

In the context presented here, chip technologies are the highly parallel analytical tools for the detection of specific biomolecules. In this context they have already a significant impact in clinical research for the identification of molecular changes in disease processes and during therapy that may lead to the development of better diagnostic strategies. (This impact is discussed for specific chip technologies in the chapter ▶Peptide Chips) Considering the wider applications of chip technologies in cell biology or biomedical research, we may see some unexpected advances, that can not be predicted at present.

What can be clearly seen at present, is that chip applications will play a significant role in medical diagnostic. The revolution may be two fold. First we will see that genomics and proteomics knowledge, gained in research (also with chip technologies and bioinformatics) will change the approach to therapy decisions. In particular it will enable the medical profession to take more personal variances into consideration that may influence the decision which therapy will be used. At present we see that genomic variations that influence drug metabolism between individuals are better understood, which influence therapeutic effectiveness and undesired toxic reactions. Dedicated diagnostic chips can be used to identify the metabolic genotype, which than can be included in the selection of drugs and drug concentration.

First chip based diagnostic tools are already available on the market, a chip for the detection of mutations defining the risk for thrombosis (Ogham Diagnostics GmbH, Münster, Germany) (16) or a chip used for the analysis of mutation in cytochrome P450 2D6 and 2C19 (Roche Diagnostics) (17). The use of these diagnostic tools allows the physicians to suggest specific treatments and adjust the dosage of medication that is metabolized by the above mentioned

cytochromes, the activity of which is depending on the existing genotype.

Other potential applications are fast detection of pathogens and the resistance phenotype of the pathogen, and monitoring of the disease progress under a given therapy.

Diagnostic chips in particular as integrated micro (fluidic) devices, hold the promise that they may be used with easy to use small work stations outside a typical diagnostic laboratory. This would enable any person, who can collect the required diagnostic sample material (blood, tissue), to perform the diagnostic test.

References

1. Fodor SP, Read JL, Pirrung MC et al (1991) Light-directed, spatially addressable parallel chemical synthesis. *Science* 251:767–773
2. Schena M, Shalon D, Davis RW et al (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270:467–470
3. Rubina AY, Dementieva EI, Stomakhin AA et al (2003) Hydrogel-based protein microchips: manufacturing, properties, and applications. *Biotechniques* 34:1008–1014, 1016–1020, 1022
4. Gutmann O, Niekrawietz R, Kuehlewein R et al (2004) Non-contact production of oligonucleotide microarrays using the highly integrated TopSpot nanoliter dispenser. *Analyst* 129:835–840
5. Singhvi R, Kumar A, Lopez GP et al (1994) Engineering cell shape and function. *Science* 264:696–698
6. Delamarche E, Bernard A, Schmid H et al (1997) Patterned delivery of immunoglobulins to surfaces using microfluidic networks. *Science* 276:779–781
7. McGall G, Labadie J, Brock P et al (1996) Light-directed synthesis of high-density oligonucleotide arrays using semiconductor photoresists. *Proc Natl Acad Sci USA* 93:13555–13560
8. Ermantraut E, Schulz T, Tuchscheerer J et al (1998) Building highly diverse arrayed substance libraries by micro offset printing. In: Harrison J, Berg A van den (eds) *Micro Total Analysis Systems '98*. Kluwer Academic Publishers, Dordrecht, pp 213–216
9. Frank R (2002) The SPOT-synthesis technique. Synthetic peptide arrays on membrane supports – principles and applications. *J Immunol Methods* 267:13–26
10. Hughes TR, Mao M, Jones AR et al (2001) Expression profiling using microarrays fabricated by an ink-jet oligonucleotide synthesizer. *Nat Biotechnol* 19:342–347
11. Okamoto T, Suzuki T, Yamamoto N (2000) Microarray fabrication with covalent attachment of DNA using bubble jet technology. *Nat Biotechnol* 18:438–441
12. Wölfl S, Dummer A, Pusch L (2004) Analyzing Proteins and Protein Modifications with ArrayTube* Antibody Microarrays. In: Schena M (ed) *Protein Microarrays*. Jones and Bartlett Publishers, Sudbury
13. Taton TA, Mirkin CA, Letsinger RL (2000) Scanometric DNA array detection with nanoparticle probes. *Science* 289:1757–1760
14. ▶www.affymetrix.com
15. ▶www.clondiag.com

16. ► www.ogham.de
17. ► www.roche-diagnostics.com/products_services/amplchip_cyp450.html

Chirality

Definition

A molecule is chiral (and said to have chirality) if it is not identical to its mirror image, or more particularly, can't be mapped to its mirror images by rotations and translations alone.

► PNA Chips

CHK2

Definition

The CHK2 (checkpoint kinase 2) gene encodes the human homologue of the yeast Cds1 (for checking DNA synthesis 1) of Rad53 G2 checkpoint kinases, whose activation in responses to DNA damage prevents cellular entry into mitosis. The gene maps to 22q12 and heterozygous germline mutations occur in

► Li Fraumeni Syndrome.
 ► Cell Cycle - Overview

CHN

► Congenital Hypomyelinating Neuropathy

5-Chlorodeoxyuridine (CldU) & 5-Bromodeoxyuridine (BrdU)

Definition

CldU and BrdU are DNA intercalating agents. They immediately induce a senescence-like phenomenon in any type of mammalian cells probably due to a change in nuclea matrix structure. BrdU and CldU are widely used to determine DNA synthesis or cell proliferation.

For instance, BrdU can be incorporated into DNA in place of thymidine. Cells, which have incorporated BrdU into DNA, can be quickly detected using a monoclonal antibody against BrdU.

► Proteomics in Ageing

CHN

Congenital Hypomyelinating Neuropathy.

Cholera Toxin

Definition

Cholera toxin (CTX) is produced by *Vibrio cholerae* and ADP-ribosylates the G_sα subunit at a conserved arginine residue which is involved in GTP hydrolysis. Thus, CTX renders G_s constitutively active.

► G-Proteins

Cholesterol Efflux

Definition

Cellular cholesterol is moved to HDL via simple diffusion LCAT-mediated FC gradient force and ABCA1-mediated active transport of free cholesterol.

► High-HDL Syndrome

Cholesterol Modification

Definition

Cholesterol modification refers to the attachment of cholesterol to proteins, especially to secreted signalling proteins that are encoded by the hedgehog gene family. Cholesterol modification involves two autocatalytic steps: a precursor protein is endoproteolytically cleaved and the C-terminal product is modified with cholesterol. Attachment occurs through ester-bond formation between the carboxyl-group at the C-terminus of the

protein and the -OH group of cholesterol. Several hedgehog proteins are also palmitoylated at an N-terminal cysteine residue via an unusual amide-linkage (N-acylation).

► [Fatty Acid Acylation of Proteins](#)

Cholic Acid

Definition

Cholic acid, besides ► [chenodeoxycholic acid](#), is one of the two primary bile acids produced from cholesterol in the liver.

► [Chenodeoxycholic Acid](#)

► [Peroxisomal Disorders](#)

Chondroblast

Definition

Chondroblast is the committed precursor of the chondrocyte, but not yet surrounded by matrix.

► [Bone and Cartilage](#)

Chondrocyte

Definition

Chondrocyte(s) are the mature cartilage cells embedded in the cartilage matrix. They are derived from the mesenchyme.

► [Bone and Cartilage](#)

► [Bone Disease and Skeletal Disorders, Genetics](#)

Chondrodysplasia

Definition

A disturbance in the development of cartilage, primarily the long bones. This can result in arrested growth and dwarfism.

► [Bone Disease and Skeletal Disorders, Genetics](#)

Chorea Huntington

► [Huntington's Disease](#)

► [Predictive Genetic Testing](#)

Chorea Major

► [Huntington's Disease](#)

Chorionic Villus Sampling

Definition

CVS is a technique used in prenatal diagnosis, where a small amount of placental tissue representative of fetal DNA is removed.

Choroideremia

Definition

Choroideremia is an X-linked disease that manifests characteristic fundus findings of diffuse choroidal and retinal atrophy. Its pattern of visual function loss is quite similar to that of typical X-linked RP (► [Retinitis Pigmentosa](#)).

Chromatid

Definition

Chromatid refers to one of two daughter chromosomes after duplication but before separation from the other daughter chromosome (also called "sister chromatids"). The pair of chromatids are joined together with a centromere, and can be seen during the earlier stages of mitosis and meiosis, before the pair are separated and pulled to opposite ends of the nucleus.

► [Cell Cycle – Overview](#)

► [Centromeres](#)

Chromatin

Definition

Chromatin is the general term for DNA packed by histones. Chromatin determines the state of condensation and strongly influence the transcription of genes. When cells divide, the chromatin is seen as distinct chromosomes.

- ▶ Centromeres
- ▶ Chromatin Acetylation
- ▶ Chromatin Remodeling
- ▶ Fragile X Syndrome
- ▶ Muscle Development
- ▶ Nuclear Compartments
- ▶ Nucleosomes
- ▶ Methylation of Proteins
- ▶ RNA Polymerase I
- ▶ Transgene Silencing
- ▶ Transcription Factors and Regulation of Gene Expression

Chromatin Acetylation

ANN EHRENHOFER-MURRAY

Otto-Warburg-Laboratories, Max-Planck-Institute of Molecular Genetics, Berlin, Germany
ehrenhof@molgen.mpg.de

Synonyms

Histone acetylation

Definition

Chromatin is the basic organizational form of DNA in the eukaryotic nucleus. The repeat unit of chromatin is the nucleosome in which 146 bp of DNA are wrapped around the histone octamer consisting of two molecules each of the ▶ [histones](#) H2A, H2B, H3 and H4. Lysine residues within the amino-terminal regions of the histones (histone tails) can be posttranslationally acetylated at the ε-N position of the side chain in a process termed chromatin or histone acetylation. The position of acetyltable lysines within the histone tails is highly conserved among different organisms, H3 lysines 9, 14, 18 and 23, and H4 lysines 5, 8, 12 and 16 are targets for acetylation. Acetylation is carried out by histone acetyltransferases (HATs), and it can be reversed by histone deacetylases (HDACs). Two general classes of histone acetylation are recognized: a) acetylation of nucleosomal histones by A-type

nuclear HATs (HAT A) is linked to nuclear processes like transcription, replication and repair (see below) and b) cytoplasmic acetylation of free histones by B-type HATs is related to their deposition during ▶ [chromatin assembly](#). The chromatin acetylation state of a particular chromosomal region correlates with the transcriptional activity of the genes embedded in this region. Active genes generally lie in acetylated chromatin, whereas gene repression is accompanied by chromatin hypoacetylation. Histone acetylation may affect chromatin structure directly by altering DNA-histone interactions within and between nucleosomes, thus resulting in a more open higher-order chromatin structure. An alternative, non-exclusive model is that chromatin acetylation states, in combination with other modifications (▶ [histone methylation](#), phosphorylation, ubiquitination), present an interaction surface for other proteins (co-activators, co-repressors) that bring about gene activation or repression (1).

Characteristics

Transcription-related Chromatin Acetylation

Chromatin acetylation is generally required to overcome the inhibitory effect of chromatin to DNA accessory factors. The best-studied type of chromatin acetylation to date is acetylation coupled to transcription initiation (2). In this process, multisubunit protein complexes containing HATs are recruited by sequence-specific proteins to promoters. Two well-studied examples are yeast SAGA/human PCAF/ Gcn5 and the yeast NuA4/human Tip60 complexes, which both have roles in transcription initiation. The HAT complexes promote the association of TATA-binding protein (TBP) or other general transcription factors to the basal promoter. Promoter chromatin acetylation is also functionally connected to ATP-dependent ▶ [chromatin remodeling](#) (nucleosome sliding). Many promoters require both activities for gene activation, and the order of recruitment is a promoter-specific characteristic. Gene activation by acetylation is self-reinforcing in that both HAT complexes and chromatin remodeling complexes contain ▶ [bromodomain](#) motifs that bind to acetylated nucleosomes. Recognition and binding of effector proteins to acetylated chromatin is proposed to translate the “histone code” into a gene expression state (1).

Chromatin acetylation is also associated with ▶ [transcription elongation](#). The hyperphosphorylated, elongating form of RNA polymerase II is tightly associated with the elongator complex, which contains the HAT Elp3. Chromatin acetylation by Elp3 is proposed to facilitate the passage of the polymerase through chromatin.

In contrast to acetylation, chromatin deacetylation is generally correlated with gene repression. In analogy to HATs, HDACs are recruited to promoters and interfere with transcription initiation. In ▶ [gene silencing](#) and

► **heterochromatin** formation, HDACs deacetylate chromatin, which allows the binding of heterochromatic components that prevent transcription.

Chromatin Acetylation in DNA Repair and Replication

Functional connections exist between chromatin acetylation and DNA repair. The human HAT p300 interacts with proliferating cell nuclear antigen (PCNA), a processivity factor for the DNA polymerases δ and ϵ that functions in DNA replication and repair. Thus, p300 is recruited by PCNA to sites of DNA damage in order to acetylate chromatin in the vicinity of the damaged site. Also, the yeast NuA4 HAT complex plays roles in double-strand break repair and non-homologous end joining. Potential roles for HATs in DNA repair include improved access of repair proteins and DNA polymerases to chromatin, protection of broken DNA ends from loss or degradation and direct recruitment of repair factors, all of which are expected to facilitate damage removal and restoration of chromatin. HATs may likewise be required to facilitate DNA replication. The timing of replication initiation in yeast is advanced by chromatin acetylation, and the human HAT HBO1 interacts with ORC1, a component of the eukaryotic replication initiator, as well as with the replication-licensing factor MCM2.

Global Chromatin Acetylation

In addition to being targeted to promoters, genes or sites of DNA damage (see above), acetylation also occurs throughout the genome in a global and untargeted manner. A balance is maintained between acetylation and deacetylation, thus enabling the chromatin to be a dynamic structure. Global (de)acetylation serves to modulate basal transcription and allows the rapid reversal of changes and the return to a “ground state” after activation. Also, acetylation of specific lysine residues inhibits the binding of heterochromatic proteins to chromatin. Thus, particular HATs fulfill a boundary function in that they prevent the spreading of silencing proteins into euchromatic regions.

Histone Acetyltransferase Families

- The GNAT family includes the transcriptional activator Gcn5, human PCAF and Elp3 (Elongator complex) as well as the HAT B Hat1. In addition to the highly conserved acetyl-CoA binding site, Gcn5 and PCAF contain a bromodomain.
- The members of the MYST family of acetyltransferases also contain the acetyl-CoA binding motif. This family was named for its founding members MOZ, Ybf2/ Sas3, Sas2 and Tip60. Other members include *Drosophila* MOF (required for ► **dosage compensation**), yeast Esa1, which is the catalytic

subunit of NuA4 and human HBO1. MOF and Esa1 carry a ► **chromodomain** that in MOF has RNA-binding activity. MOZ additionally carries a PHD motif. Several members contain an atypical ► **zinc finger** that is required for HAT activity.

- p300 and CREB binding protein (CBP) are two highly related HATs that contain several discrete domains, including three putative zinc fingers, a bromodomain, an acetyl-CoA binding site and at least two independent regions that interact with multiple transcription factors (3).

Histone Deacetylase Families

- The HDAC family consists of the class I HDACs that are similar to yeast Rpd3, and the class II HDACs that resemble yeast Hda1. Four class I and five class II HDACs exist in humans. HDAC1 and HDAC2 are found in several multiprotein complexes: a) the Sin3 complex, a repressor complex that interacts with DNA binding factors, for instance MeCP2 (methyl-cytosine binding protein 2); b) the NuRD chromatin remodeling complex and c) the CoREST complex, which contains the SANT domain protein CoREST.
- The family of Sir2-like deacetylases (sirtuins) is structurally unrelated to the HDAC family. Sirtuins require NAD^+ as a cofactor for the deacetylation reaction. The founding member of the family, yeast Sir2, is required for transcriptional silencing at the silent mating-type loci, the telomeres and in the ribosomal DNA (rDNA) locus in yeast. Four homologs exist in yeast, Hst1-4 (Homolog of Sir Two) and several homologs have been identified in humans, *C. elegans* and *Drosophila*. Yeast and *C. elegans* Sir2 have been linked to aging, which may be mediated by the NAD^+ requirement of Sir2 as a link to metabolic processes and the suppression of rDNA recombination in yeast (4).

Cross-talk Between Chromatin Acetylation and Other Chromatin Modifications

Histones are subject to a diverse set of modifications other than lysine acetylation, including serine phosphorylation, lysine and arginine methylation and lysine ubiquitination. Though classically found on the histone tails, modifications have recently also been identified on the core region of histones. Individual modifications do not occur independently of each other, but rather, exert influences on each other. Thus, combinations of chromatin modifications regulate the binding of downstream proteins and ultimately modulate gene expression (5).

- A particular residue can carry only one modification at a time. For instance, lysine 9 of H3 can be either acetylated or methylated.

- Neighboring modifications influence each other. For example, phosphorylation of Ser10 of H3 by the kinase Snf1 is required for H3 lysine 14 acetylation by Gcn5.
- The modification of one histone tail regulates the modification of another histone tail (trans-tail interactions). Ubiquitination of lysine 123 in the core region of H2B regulates the methylation of lysine 4 on H3.
- Histone modifications can influence DNA methylation. Methyl-cytosine binding proteins interact with HDACs, and H3 histone methyltransferases are required for DNA methylation in *Neurospora* and *Arabidopsis*.

Clinical Relevance

Inappropriate chromatin acetylation can result in the misregulation of genes, which ultimately leads to proliferative and developmental diseases. The central role of acetylation in gene regulation also renders it a target for therapeutic intervention.

Chromatin Acetylation and Cancer

- Alterations in HATs occur in several cancer types. p300 classifies as a tumor suppressor gene, since mutations in p300 are found in epithelial cancers. Mutations in CBP are the cause for [▶Rubinstein-Taybi syndrome](#), a congenital malformation syndrome that is associated with an increased risk of cancer development.
- Several oncogenic translocations involving HAT genes are associated with leukemia. Fusions have been identified between MOZ and CBP, p300 or the transcription factor TIF2, which are postulated to misdirect HAT activity and thus result in inappropriate chromatin acetylation.
- HDACs interact with oncogenes and tumor suppressor genes. HDAC activity becomes redirected by translocations between an HDAC-controlled repressor and the DNA binding domain of a transcription factor. For instance, in translocations of PML or PLZF to retinoic acid receptor- α (RAR α), HDAC-mediated repression blocks myeloid differentiation.
- Several HDAC inhibitors suppress transformed cell growth, differentiation and apoptosis and are being clinically tested as anti-cancer drugs. Their precise mechanism of action remains to be determined, but may include increased expression of the cell-cycle inhibitor p21, which causes cell-cycle arrest in G1 and/ or G2.

Chromatin Acetylation in Other Diseases

Several disorders have been linked to chromatin acetylation and [▶epigenetic](#) control of gene expression. Fragile X syndrome, a hereditary disorder

associated with severe mental retardation, is caused by the expansion of an unstable CGG repeat in the 5' region of the *FMR-1* gene. Subsequent hypermethylation of the DNA in this region cooperates with chromatin hypoacetylation to inactivate *FMR-1*. Also, [▶Rett syndrome](#), a neurological disorder of female humans, is caused by mutations in the X-linked gene encoding MeCP2. MeCP2 mediates gene repression through interactions with the Sin3 HDAC complex. Thus, disruptions in epigenetic gene regulation are the likely molecular mechanism of this disease.

Furthermore, connections have been identified between neurological disorders and chromatin acetylation. Expanded polyglutamine-containing proteins interact with several chromatin regulators. For example, the polyglutamine portion of the Huntington's disease protein, huntingtin, directly binds and inhibits the HATs CBP and p300. Treatment with HDAC inhibitors arrests progressive neuronal degradation and lethality in *Drosophila* models of polyglutamine diseases. Another link is given by the observation that a cleaved portion of the Alzheimer protein APP forms an active HAT complex with Tip60 and thus probably has a function in gene regulation (6).

▶Polyglutamine Disease, the Emerging Role of Transcription Interference

References

1. Jenuwein T, Allis CD (2001) Translating the histone code. *Science* 293:1074–80
2. Naar AM, Lemon BD, Tjian R (2001) Transcriptional coactivator complexes. *Annu Rev Biochem* 70:475–501
3. Roth SY, Denu JM, Allis CD (2001) Histone acetyltransferases. *Annu Rev Biochem* 70:81–120
4. Grozinger CM, Schreiber SL (2002) Deacetylase enzymes: Biological functions and the use of small-molecule inhibitors. *Chem Biol* 9:3–16
5. Berger SL (2003) Histone modifications in transcriptional regulation. *Curr Opin Genet Dev* 12:142–148
6. Timmermann S, Lehrmann H, Polesskaya A et al (2001) Histone acetylation and disease. *Cell Mol Life Sci* 58:728–736

Chromatin Assembly

Definition

Chromatin assembly designates the deposition of histones on DNA after DNA replication or repair. Replication- and repair coupled chromatin assembly is catalyzed by protein complexes, for instance chromatin assembly factor-I (CAF-I).

▶Chromatin Acetylation

Chromatin Modifications

Definition

Proteins packaging the DNA can be enzymatically modified by acetylation, methylation, phosphorylation and ADPribosylation, and hence the DNA structure is changed secondarily.

►Chromatin Acetylation

►Transcription Factors and Regulation of Gene Expression

Chromatin Remodellers

►Chromatin Remodelling

►Non-Histone Chromatin Proteins

Chromatin Remodelling

LEONIE RINGROSE

Centre for Molecular Biology (ZMBH), University of Heidelberg, Germany

l.ringrose@zmbh.uni-heidelberg.de

Synonyms

The process of chromatin remodelling is also referred to as nucleosome remodelling, or more precisely, ATP-dependent nucleosome remodelling. The multiprotein complexes that carry out ATP-dependent nucleosome remodelling go by several names: nucleosome (or chromatin) remodelling machines, nucleosome (or chromatin) remodelling factors, nucleosome (or chromatin) remodelling enzymes, or simply remodellers.

Definitions

Chromatin

Chromatin is the eukaryotic cell's solution to the immense problem of genome packaging: how to fit up to 2 meters of DNA into the cell nucleus without losing access to the information it encodes. In chromatin, DNA is wrapped around histone proteins, forming nucleosomes. Each nucleosome contains 8 histone proteins (the histone octamer), that are wrapped by 1.67 turns of DNA, comprising 147 base pairs. The nucleosome is held together by weak interactions between histones and the DNA helix, at 14 sites along the length of the wrapped DNA. The sum of these interactions gives a remarkably stable structure. The

flexible N- terminal tails of the histones protrude from the globular, DNA-wrapped nucleosome core. The nucleosome is the basic repeating unit of chromatin. Nucleosomes are connected by linker DNA, whose length is typically in the range of 50 to 80 bp. In the nucleus, this fibre is folded into higher order structures whose architecture varies with gene activity (1).

Chromatin Remodelling

The term chromatin remodelling is sometimes loosely used to refer to large-scale global changes in the higher order structure of chromatin, or to local changes in the patterns of histone tail modifications, such as phosphorylation, methylation or acetylation. Although these processes are crucial for regulating gene activity and chromatin dynamics, they will not be covered here. This article will be limited to ATP-dependent nucleosome remodelling. This is itself an umbrella term, used to summarise a large number of energy-dependent alterations in the basic nucleosomal structure described above, catalysed by a large family of chromatin remodelling machines. ATP-dependent nucleosome remodelling is vital for ensuring that DNA wrapped in nucleosomes nevertheless remains accessible and open for business: be it gene regulation, replication, or repair (1, 2).

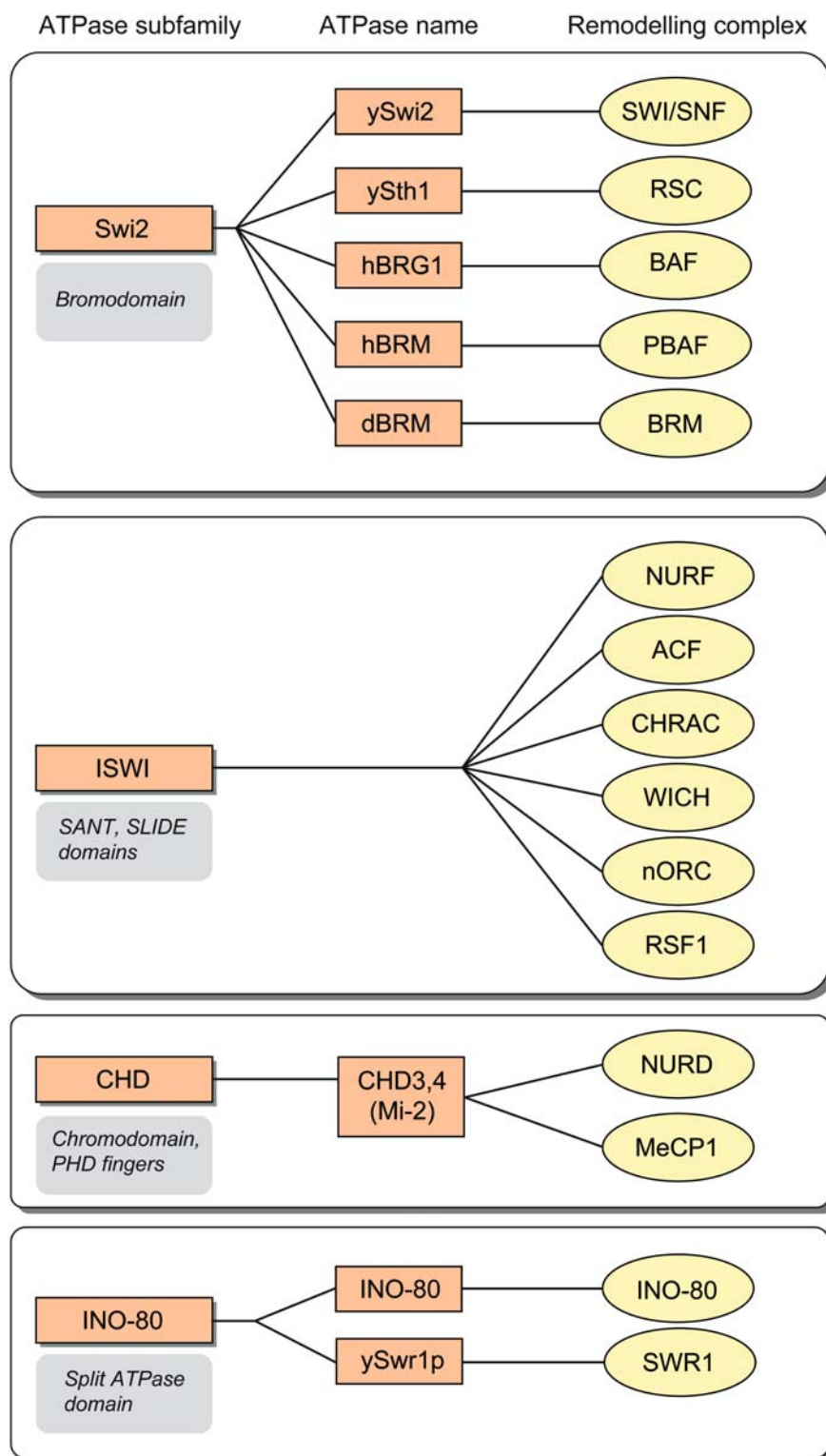
Chromatin Remodelling Machines

Chromatin remodelling machines are multiprotein complexes. All of them contain as their engine an ATPase subunit, which is accompanied by one or more other proteins that modulate its function. Different chromatin remodelling machines contain different ATPases (summarized in figure 1), which are related by the presence of a DEAD box helicase motif, and belong to the Swi2/Snf2 subfamily (2).

Characteristics

A Family Tree of Chromatin Remodelling Machines

Over the past decade, biochemical characterisation of chromatin remodelling machines has generated a bewildering array of acronyms for a menagerie of different complexes and their resident ATPases. There are four well-characterised classes of ATPase, which are united by their DEAD box motif and distinguished by their additional domains (Figure 1). Further differences are seen when isolated ATPases are compared in *in vitro* assays for nucleosome remodelling, in terms of what best stimulates their activity (free DNA or nucleosomes), and their requirement for histone tails (1). In addition, although each isolated ATPase can promote histone octamer sliding *in vitro*, they perform with different efficiencies in other assays for nucleosome mobility. Whether these differences truly reflect different reaction mechanisms for the



Chromatin Remodelling. Figure 1 A family tree of chromatin remodelling machines. Four subfamilies of chromatin remodelling ATPases are shown (orange boxes, left). All ATPases contain a DEAD-box helicase domain. Additional domains that characterize each subfamily are shown (grey boxes, left). Individual family members are listed (orange boxes, centre). Where ATPases have different names in different species, the species is indicated by y = yeast, h = human, d = *Drosophila*. Complexes containing each ATPase are shown (yellow ovals, right). For a comprehensive listing of complex subunits, see references 2, 3 and 5.

different ATPases, or simply kinetic variations on a common reaction theme, is currently a matter of debate (for example, see the opposing views in (1) and (3)). The ATPases do not work alone *in vivo*, but are accompanied by accessory proteins (Figure 1). Indeed, the presence of these proteins generally increases the *in vitro* activity of the ATPases by an order of magnitude or more, and can influence the outcome of the remodelling process. Several recent reviews give comprehensive listings of accessory proteins (1-3); however, new complexes are being discovered and named at an alarming rate. Thus, any list will probably be incomplete within a few years of its making.

Biological Processes: Remodelling Machines for Every Occasion

The main classes of remodelling complex shown in Figure 1 are not only different in terms of their resident ATPases and subunits, but also in terms of their main function in the cell.

The members of the 11 subunit yeast SWI/SNF complex were originally identified genetically as transcriptional activators. Later biochemical characterisation of the complex and its nucleosome remodelling activities lead to the idea that SWI/SNF is important for increasing promoter accessibility. The highly related RSC complex is similarly involved in transcriptional activation, although the two complexes target different genes. An important difference between SWI/SNF and RSC is that only the RSC complex members are essential for viability (2). In mammals, the situation is similar: two highly related complexes containing the BRM and BRG1 ATPases facilitate transcription of distinct sets of target genes. Knockout mice lacking BRM are viable, whilst BRG1 knockout is lethal (2, 4; see also clinical relevance, below). In *Drosophila*, the BRM complex is involved in maintaining active chromatin architecture to counteract the repressive effects of the Polycomb group proteins (2). In summary, the complexes containing Swi2 family ATPases appear to be mainly involved in loosening the repressive effects of nucleosomes during transcriptional activation.

In contrast to Swi2, the ISWI complexes are frequently found to be associated with chromatin assembly during DNA replication: another situation where nucleosomes have to be moved. Many of these complexes were originally identified in chromatin assembly extracts from *Xenopus* or *Drosophila*, as factors that enable newly assembled chromatin to space its nucleosomes in a regular array. The ISWI complexes achieve this by ATP-dependent nucleosome sliding, to “shuffle” nucleosomes into place. Although this sliding activity means that many of these complexes can facilitate transcription from chromatin templates *in vitro*, there

are several lines of evidence that their main *in vivo* role is not in transcriptional regulation, but in chromatin assembly. For example, the ISWI protein in *Drosophila* does not co-localise with RNA Polymerase II on polytene chromosomes (2). In mammals, the ACF complex, comprising ISWI and ACF1, localises to replication foci in pericentric heterochromatin (5). Highly related to ACF is the mammalian WICH complex, comprising of ISWI and WSTF. This complex localises to many sites in replicating chromatin (5; see also clinical relevance below). In summary, the main job of ISWI complexes appears to be in overcoming the repressive effects of nucleosomes at the replication fork.

The Mi2 containing complexes NURD and MeCP1 have yet another role that at first glance appears surprising for remodelling machines: they are involved in transcriptional repression. The NURD complex contains, among other subunits, a histone deacetylase. The *in vitro* deacetylation activity of NURD complexes is strongly stimulated by the ATPase Mi-2. *In vivo*, histone deacetylation is a hallmark of silent repressive chromatin, suggesting a scenario in which the NURD complex may first remodel nucleosomes to render their tails accessible, and then deacetylate, contributing to silencing. In addition, Mi-2 may shuffle nucleosomes into positions that are unfavourable for transcription. The dual action of the NURD complex is refined even further in the related MeCP1 complex, which contains the NURD subunits in addition to the MBD2 protein, which binds preferentially to DNA methylated at CpG positions. CpG methylation is associated with silenced genes. The MeCP1 complex is targeted by MBD2 to nucleosomes containing methylated DNA, where it remodels and deacetylates, thus reinforcing the local silencing marks in chromatin (2).

In yeast, the INO-80 complex is recruited to DNA double strand breaks, where it may be required to clear nucleosomes from the broken end (3). Finally, the recently characterized yeast SWR1 complex, containing the ATPase Swr1p, which is related by its split ATPase domain to INO-80, adds a new twist to the nucleosome remodelling family. This complex catalyses the exchange of histone H2A for its variant H2AZ (in yeast, Htz1), which acts as a buffer to limit the spread of silent heterochromatin, both at telomeres and other locations (3).

Mechanisms: Sliding, Jumping, Ejection and Exchange

Although the nucleosomal particle is a stable structure, it also has an inherent elasticity. Single molecule experiments have shown that by pulling in opposite directions on the two DNA ends, a nucleosome can be almost entirely unwrapped, but will nevertheless snap back to its intact structure once the tension is released. Indeed, even without such extreme experimental

disruption, nucleosomes “breathe” to a certain extent, transiently exposing stretches of their DNA, and the histone octamer will even slip back and forth along its associated DNA at elevated temperatures and salt concentrations (1, 2). The role of remodelling machines can thus be seen as facilitating and regulating this inherent dynamic behaviour.

How exactly do remodellers work? Mechanistic insights have come from *in vitro* studies, usually on single nucleosomes assembled with defined DNA fragments. All remodellers can catalyse sliding, and some can additionally catalyse jumping to another DNA segment (trans-displacement) ejection of histone octamers to entirely denude a DNA segment of nucleosomes, or exchange of histone variants.

The mechanism of nucleosome sliding is the best characterized to date. Early models for sliding envisaged a gradual movement of histone octamer along the DNA one base pair at a time, driven by the ATPase twisting the DNA helix, rather like a screw thread. These models have been laid to rest by studies of both ISWI and Swi2 class ATPases, showing that sliding proceeds in much larger steps: The remodeler attaches both to the histone octamer and to the DNA. It then lifts a section of DNA off the octamer surface, and in its ATP-dependent power stroke, pulls the DNA into a bulge (1, 6). Like a sailor hauling in a rope, the ATPase then attaches to a new site on the DNA, and the bulge is thus pumped around the nucleosome, to emerge on the other side. The net result is sliding of the histone octamer relative to its starting position on the DNA. The histone octamer itself does not appear to be disrupted. Although they both use this basic mechanism, the Swi2 and ISWI classes of remodellers show clear differences in sliding assays *in vitro*. Detailed studies of the molecular interactions of the two classes of remodellers with nucleosomes suggest that these different sliding outcomes may arise from differences in where the remodeler grabs the DNA, where it anchors on the histone octamer, and how fast it pumps DNA into the nucleosome (1, 6).

In vivo, nucleosome sliding by remodellers may provide enough access to DNA for gene activation and replication, as it would transiently expose DNA to competing factors. However, other nucleosome remodelling events such as trans-displacement, and ejection of nucleosomes, are less easy to explain by a sliding action. *In vitro*, the yeast SWI/SNF and RSC complexes can, under certain conditions, move a histone octamer to a different piece of free DNA. It is unclear whether this reaction occurs *in vivo*, since it would depend on the availability of free DNA and the concentration of remodeler (1, 2). More difficult to dismiss is the observation that *in vivo*, many active genes contain stretches of nucleosome-free DNA, showing that nucleosomes have been removed entirely

rather than merely pushed aside. It is not clear exactly what role is played by remodellers in this process. The best studied examples of nucleosome ejection are the yeast PHO5 and PHO8 promoters. Here, the histone chaperone Asf1 plays a role in dismantling nucleosomes upon activation, possibly as part of a remodelling complex (3).

The remodelling events described above all have in common that they disrupt histone – DNA contacts, leaving the histone octamer itself intact. A remarkable exception to this rule has recently been characterized. The yeast SWR1 complex, comprising the Swr1p ATPase and 12 other subunits (Figure 1), disrupts the histone octamer itself, exchanging an H2A- H2B dimer for an H2AZ- H2B dimer, resulting in exchange of the regular histone H2A for its variant H2AZ (called Htz1 in yeast). The SWR1 complex carries out an efficient exchange *in vitro*, whereas other remodellers perform poorly in this assay. Though the exact mechanism of histone exchange by SWR1 remains to be elucidated, the disruption of histone – histone interactions represents a new class of interaction that may be entirely distinct from nucleosome sliding (3).

Clinical Relevance

Several remodelling ATPases or subunits of remodelling complexes are mutated or deleted in human disease. This section covers those proteins whose reduction or disruption causes disease, and for which there is *in vitro* evidence of involvement in ATP-dependent nucleosome remodelling. However, in several cases, a direct causal link between loss of nucleosome remodelling and disease remains to be established.

The BRG1 ATPase (Figure 1) is a central regulator of the cell cycle and a tumour suppressor. The *BRG1* gene is mutated in several tumour cell lines including pancreatic, breast, lung and prostate cancer cell lines. *BRG1* $-/-$ mice die early in development, and *BRG1* $+/-$ mice are predisposed to tumours. There is clear *in vitro* evidence that BRG1 functions in chromatin remodelling, and facilitates transcription from chromatin templates. About 80 targets of BRG1 regulation have been identified by transcriptional microarray analysis (these may include both direct and indirect targets). A plausible scenario is that loss of BRG1 activity leads to loss of remodelling, and thus of transcription at a key set of genes that normally repress tumorigenesis (2, 4). The ATRX protein (α -thalassaemia, mental retardation, X-linked protein) has a long history in disease, and has long been suspected to play a role in ATP-dependent nucleosome remodelling. Recent biochemical evidence shows that the ATRX protein, in complex with the transcription factor Daxx, does indeed catalyze ATP-dependent nucleosome remodelling *in vitro*. The protein itself is huge, (2492 amino acids) and contains

a DEAD box helicase domain, in addition to a zinc finger, a PHD domain, and a plant homeodomain (7). This domain structure renders ATRX difficult to classify into any of the known families of ATPases shown in Figure 1. In humans, 25% of mutations that lead to ATRX syndrome are found in the helicase domain, and 65% are found in the zinc finger and PHD region. It is not yet clear whether loss of remodelling activity is the direct cause of all the dramatic phenotypes associated with ATRX syndrome, which include α -thalassemia, mental retardation, facial anomalies and gonadal and urogenital abnormalities. Complete absence of ATRX in null knockout mice is lethal at day 9.5. Further elucidation of the molecular mechanisms by which reduction or disruption of ATRX activity causes disease will require the identification of interaction partners and target genes (7).

Apart from the ATPases themselves, other subunits of remodelling complexes play a central role in disease, notably INI1 and WSTF. INI1 is an accessory protein shared by both the mammalian BRG1 and BRM complexes (Figure 1), and bears homology to the yeast SNF5 protein, a SWI/SNF subunit. INI1 is a potent tumour suppressor. Malignant rhabdoid tumours of the kidney and brain carry biallelic inactivating mutations of the INI1 gene. Mice in which the INI1 gene has been knocked out in selected cells develop rhabdoid tumours or lymphomas within 11 weeks, and complete knock-out of INI1 is lethal during early embryogenesis. However, it is not clear whether the tumour suppressor function of INI1 depends on its participation in chromatin remodelling complexes: the phenotype of the INI1 knockout is more severe than that of either BRG1 or BRM, suggesting that INI1 may have additional independent functions. Alternatively, it is possible that by removing INI1, both BRG1 and BRM complexes are disabled, giving more severe effects than the loss of a single complex (2, 4).

Williams-Beuren syndrome is a developmental disorder caused by deletion of a 1.6Mb region of chromosome 7, containing 15 genes, one of which encodes the WSTF protein (Williams syndrome transcription factor). Williams syndrome patients show a complex range of phenotypes, including vascular and heart defects, mental retardation and growth deficiency. The WSTF protein, together with the ATPase ISWI, forms the WICH complex (Fig. 1). This complex catalyses ATP-dependent nucleosome remodelling *in vitro*, to give regular spaced nucleosome arrays. *In vivo*, the WICH complex is recruited by PCNA to replication foci, suggesting a role in chromatin assembly. At present, it is not clear to what extent the symptoms of Williams syndrome patients are caused by loss of WICH chromatin remodelling activity. A mouse knockout of the gene encoding WSTF may throw light on this question (5).

- Chromatin Acetylation
- DNA Helicases

References

1. Längst G, Becker PB (2004) Nucleosome Remodelling: One Mechanism, Many Phenomena? *Biochim Biophys Acta* 1677:58–63
2. Becker PB Hörz W (2002) ATP-Dependent Nucleosome Remodelling. *Annu Rev Biochem* 71:247–273
3. Cairns B (2005) Chromatin Remodelling Complexes: Strength in Diversity, Precision through Specialization. *Curr Opin Genet Dev* 15:185–190
4. Gibbons RJ (2005) Histone Modifying and Chromatin Remodelling Enzymes in Cancer and Dysplastic Syndromes. *Hum Mol Genet* 14:R85–R92
5. Poot R, Bozhenok L, van den Berg DLC et al (2004) The Williams Syndrome Transcription Factor Interacts with PCNA to Target Chromatin Remodelling by ISWI to replication foci. *Nat Cell Biol* 6:1236–1244
6. Becker PB (2005) Nucleosome Remodellers on Track. *Nat Struct Mol Biol* 12:732–733
7. Tang P, Park DJ, Marshall Graves JA et al (2004) ATRX and Sex Differentiation. *Trends Endocrinol Metab* 15:339–344

Chromatin Remodelling Complexes

Definition

The unit structures of chromatin, the nucleosomes, may have to be changed in the course of gene activation. This remodeling of nucleosomes is mediated by remodeling complexes targeted to enhancer and promoter regions.

- Chromatin Remodelling
- Transcription Factors and Regulation of Gene Expression

Chromatin Remodelling Enzymes

- Chromatin Remodelling
- Non-Histone Chromatin Proteins

Chromatin Remodelling Factors

- Chromatin Remodelling

Chromatographic Procedures

Definition

Chromatographic procedures refer to separation methods, in which substrate mixtures are dissolved in a liquid or gaseous fluid known as mobile phase. The solution passes through a column filled with a solid matrix, the stationary phase. Individual solutes interact with matrix, which retards their flow-through depending on the properties of each solute.

► [SNP Detection and Mass Spectrometry](#)

Chromodomain

Definition

Chromodomains are highly conserved sequence motifs that have been identified in a variety of animal and plant species. A chromodomain harbours a methylated lysine residue. Chromodomain proteins appear to be either structural components of large macromolecular chromatin complexes, or proteins involved in remodelling the chromatin structure. Some chromodomains have a binding affinity for histone H3 methylated on lysine 9, whereas others have RNA binding capacity.

- [Chromatin Acetylation](#)
- [Chromatin Remodelling](#)
- [Methylation of Proteins](#)
- [Nucleosomes](#)

Chromophore

Definition

Chromophore designates a prosthetic group, in most cases covalently attached to a protein, which absorbs light in the ultraviolet to visual light range due to the interaction of its conjugated p-electron system with the host protein.

► [Photoreceptors](#)

Chromoprotein

Definition

Chromoprotein refers to the protein-chromophore complex that absorbs in the near UV and Vis range.

► [Photoreceptors](#)

Chromosomal Aberration

Definition

Chromosomal aberration refers to numerical or structural abnormalities of the ► [karyotype](#), such as the gain of a genome (triploidy, tetraploidy, etc.), chromosomal loss or gain (nullisomy, monosomy, trisomy, etc.), or a chromosomal rearrangement (deletion, insertion, translocation, inversion, ring chromosome formation, etc.).

- [Chromosomal Instability Syndromes](#)
- [Double-Strand Break Repair](#)
- [Microdeletion Syndromes](#)
- [Chromosome 21, Disorders](#)

Chromosomal Instability Syndromes

MARTIN DIGWEED

Institute of Human Genetics, Charité University
Medicine Berlin, Berlin, Germany
martin.digweed@charite.de

Synonyms

Chromosome breakage syndromes

Definition

Chromosomal instability ► [syndromes](#) (CIS) are autosomal recessive genetic disorders characterised by a particularly high rate of spontaneous and induced chromosomal breakage in mitotic cells observed in microscope preparations. Chromosomal breaks stem from unrepaired ► [DNA lesions](#) sustained before ► [mitosis](#) and many of the syndromes are caused by mutations in genes required directly or indirectly for the efficient repair of DNA damage. DNA damage

can be due to environmental agents or occur as part of normal physiological processes such as ►DNA replication and ►DNA recombination. However, not all DNA repair deficiencies lead to chromosomal instability.

Chromosomal instability reflects an increased mutation rate in somatic cells and this is responsible for the greatly increased cancer incidence amongst CIS patients. In addition, in some of the syndromes, the orderly ►rearrangement of immunoglobulin genes, which requires DNA cutting and rejoining, is also disturbed, resulting in congenital immunodeficiency.

Characteristics

Table 1 lists the major CIS disorders together with some of their specific features. Although chromosome breakage is common to all, there are differences in the type of aberration observed. Figure 1a shows some of the chromosome aberrations found in cells from CIS patients. In ►Fanconi anaemia (FA) there are frequent chromatid breaks (Fig. 1b) and translocations between chromosomes (Fig. 1c), nearly always non-homologous chromosomes. In contrast, the many translocations observed in Bloom syndrome (BS) are exclusively between homologous chromosomes. In ►ataxia telangiectasia (AT) and Nijmegen breakage syndrome (NBS) lymphocytes there are often inversions and translocations, here involving particular breakpoints on chromosome 7 and chromosome 14 and affecting the genes for the heavy chain of the immunoglobulins and the T-cell receptor. This is an indication that the rearrangement of these genes is disturbed and offers a possible explanation for the immunodeficiency characteristic of AT and NBS but not, for example, of FA. In addition to spontaneous chromosomal breakage, the cells of CIS patients show a specific increase in chromosomal damage after treatment with various ►genotoxins. FA cells are particularly sensitive towards bi-functional alkylating agents that can intercalate into DNA and form covalent crosslinks between the two strands. AT, NBS and ATLD cells are particularly sensitive to ionising irradiation and chemicals such as bleomycin, which produce similar DNA lesions, in particular, DNA double-strand breaks. This differential sensitivity indicates that different pathways of ►DNA repair are involved in the various CIS disorders.

Cellular and Molecular Regulation

Chromosome Breaks and DNA Double-Strand Break Repair

It is generally accepted that visible chromosome damage is the end result of a complex cellular process for dealing with a DNA double-strand break (DSB). During repair of DSBs, DNA ends are rejoined, not, however, necessarily in the correct arrangement. In fact

the repair of DSBs probably always involves exchange of chromosome material and even simple ►chromatid breaks probably represent incomplete translocations (1). Ionising radiation and some chemicals can cause chromosome breakage independently of the phase of the cell-cycle in which the primary DNA damage occurs. Other genotoxins, such as crosslinkers, require that the cell traverses S-phase before a DNA lesion is turned into a chromosome break. This suggests that DNA crosslinks are normally detected and repaired primarily during replication. Indeed many DSBs and anomalous DNA-replication intermediates are expected to occur during replication and will be processed by the enzymes generally considered to be DNA-repair enzymes.

DNA is the only macromolecule of the cell which, when damaged, is repaired rather than replaced. The mechanisms leading to repair of DNA lesions are amongst the oldest cellular processes and have been strongly conserved during evolution. There are several different repair pathways for various types of DNA lesion; in connection with chromosomal instability it is the repair of DSBs that is particularly relevant.

The eukaryotic cell has two principle pathways for repairing DNA DSBs, repair that requires little or no homology between the rejoined sequences and homology-dependent repair (Fig. 2). Homology dependent repair can be further subdivided into 1) ►gene conversion which requires a second DNA molecule containing a copy of the sequence to be repaired and 2) single strand annealing (and related mechanisms) which use homologous sequences within a single DNA molecule to effect repair. The extremely high content of repetitive DNA in the human genome gives this latter process a particular importance.

In non-homologous end joining (NHEJ), the double-strand break is repaired by more or less simple ligation. This requires preparation of the DNA ends and necessarily results in a deletion of genomic DNA. It is, however, independent of the cell-cycle and therefore more universally applicable. Since only 2% of the genome actually codes for proteins, the deletions incurred will be tolerated by most cells.

Single strand annealing (SSA) is similar to NHEJ except that the DNA ends are digested back to regions of homologous sequence, single stranded DNA derived from both sides of the original break can then form a double-strand and DNA polymerase can replace the missing sequence using the complementary strand as a template. The process also leads to deletions.

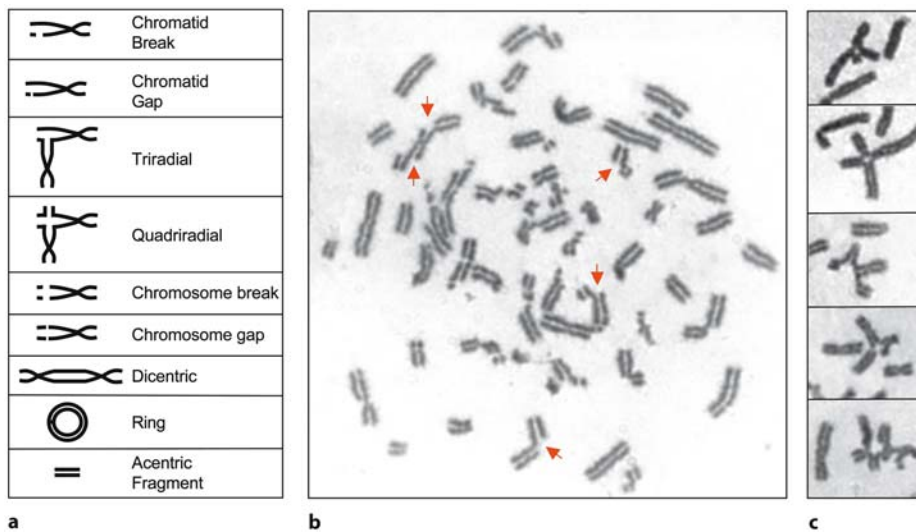
Gene conversion, a form of ►homologous recombination, is the only potentially error free pathway for repair of double-strand breaks. Here, a DNA single strand from the broken molecule invades a homologous DNA helix and searches for a sequence with which it can form complementary base pairs. In bacteria the RecA

Chromosomal Instability Syndromes. Table 1 Chromosome Instability Disorders

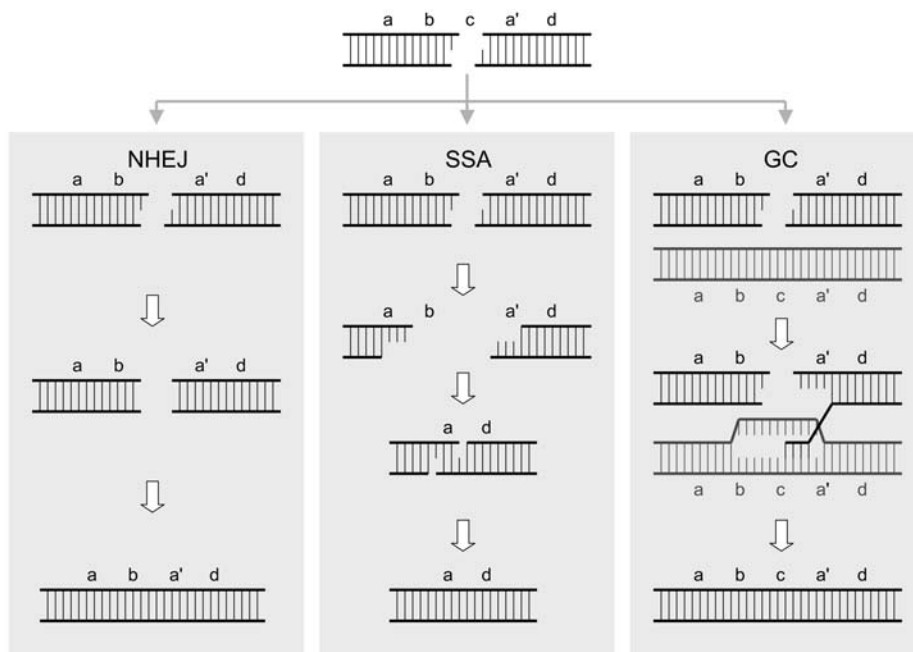
Syndrome	Ataxia telangiectasia (AT)	AT-Like Disorder (ATLD)	Nijmegen-Breakage-Syndrome (NBS)	Fanconi anaemia (FA)	Bloom-Syndrome (BS)	Werner-Syndrome (WS)	Rothmund-Thomson Syndrome (RTS)	Ligase I-Deficiency (LIGI)	Ligase IV-Deficiency (LIGIV)
OMIM ¹	208900	604391	251260	227650	210900	277700	268400	126391	601837
Locus	11q2.23	11q21	8q21	11 loci	15q25.1	8p11-12	8q24.3	19q13	13q33-34
Protein	Protein-Kinase	MRE11	Nibrin	Protein-Complex	RECQ2-Helicase	RECQ3-Helicase	RECQ4-Helicase	Ligase I	Ligase 4
Spontaneous chromosome damage	7/14-Translocation in lymphocytes	7/14-Translocation in lymphocytes	7/14-Translocation in lymphocytes	Recombination between non-homologous chromosomes	Recombination between homologous chromosomes	Mosaic, Translocation	Mosaic, Aneuploid and Isochromos.	?	Breaks
Sensitivity	γ-radiation	γ-radiation	γ-radiation	Crosslinkers	UV	4-NQO ²	4-NQO	γ-radiation, UV, alkylating agents	γ-radiation
Cancer predisposition	T-cell lymphomas	unknown	B-cell lymphomas	leukaemia	various	soft sarcomas	Osteo-sarcomas	Lymphoma?	?

¹Online Mendelian Inheritance in Man (► <http://www3.ncbi.nlm.nih.gov/Omim/>)

²4-nitroquinoline-1-oxide



Chromosomal Instability Syndromes. Figure 1 Chromosomal instability. (a) Schematic representation of some of the chromosome aberrations seen in lymphocyte chromosomes from CIS patients. (b) The photograph shows the lymphocyte chromosomes of a patient with Fanconi anaemia. This patient has bi-allelic mutations in the *BRCA2* gene. Only some of the chromatid breaks in this massively damaged metaphase are indicated. (c) Examples of the characteristic chromosome reunion figures seen in lymphocyte metaphases from FA patients.



Chromosomal Instability Syndromes. Figure 2 DNA double-strand break repair. The graphic shows a schematic representation of the three main double-strand break repair pathways, non-homologous end joining (NHEJ), single-strand annealing (SSA) and gene conversion (GC). In NHEJ, the DNA ends at the break at position “c” are processed and religated. In SSA, the DNA ends are resected to repetitive sequence “a” which allows complementary base pairing and resynthesis of the missing DNA strands. In GC, a homologous double helix (here in blue) is used as a template for DNA synthesis after invasion by a single strand originating at the double-strand break. Only in the GC pathway is the original “c” sequence restored.

protein is essential for this process. In mammalian cells, RecA is represented by at least 5 proteins of which ►*RAD51* is considered to be the primary functional homologue. The invaded helix generally remains intact and serves merely as a template for DNA synthesis by DNA polymerase. Thus a non-reciprocal exchange of information from the intact DNA to the damaged DNA is the result, a so-called gene conversion. Resolution of the exchange event occasionally results in a reciprocal exchange with a ►crossover event; this is, however, a relatively rare event in human cells.

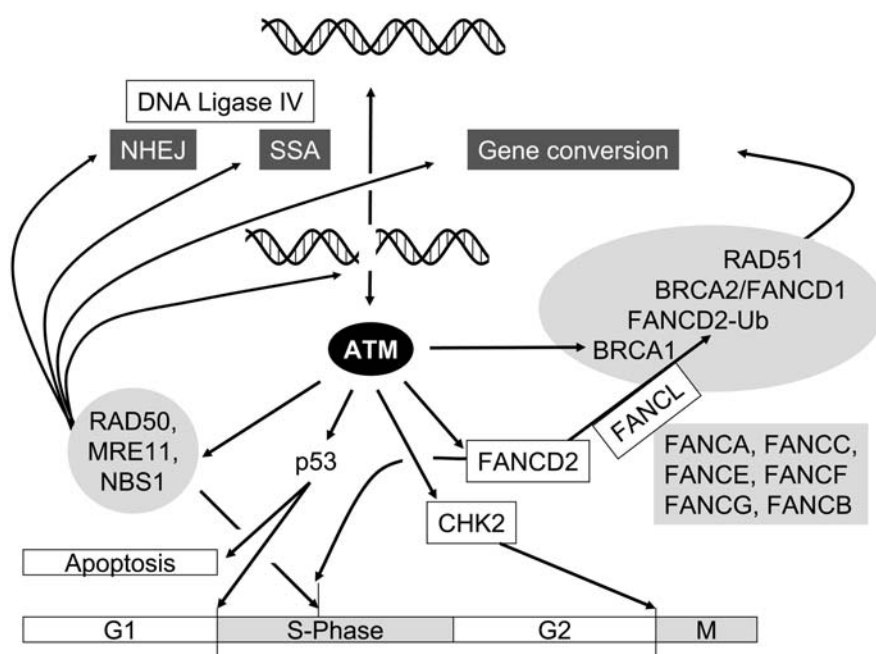
The abundance of repair pathways in the human cell reflects its homeostatic nature and ensures that DNA damage is repaired. On the other hand, it is clear that the pathways have specific inherent error rates, which necessitate the exact matching of the repair pathway with the type of lesion and its time of occurrence within the cell-cycle. It is becoming clear that some of the CIS disorders can best be explained, not by a failure to repair, but rather by the use of inappropriate repair pathways. Indeed, since the different repair pathways are competitors for the same unprocessed double-strand breaks it is understandable that even a modest inefficiency in one pathway will lead to increased use of another.

This seems to be the case at least for those FA cells in which ►*BRCA2* has bi-allelic mutations. For such

cells homology-dependent gene conversion has been shown to be reduced and homology-dependent SSA concomitantly greatly stimulated (2). Cells deficient in another breast cancer predisposition gene, ►*BRCA1*, show a similar pattern of chromosomal breakage and a shift in use of DNA repair mechanisms.

DNA-Repair and the Cell Cycle

The CIS gene, *ATM*, plays a role not only in the repair of DSBs but also in the control of the cell-cycle. Figure 3 shows the interactions of this gene product within the network of the CIS genes. ATM is a protein kinase with an enormous number and variety of target proteins, amongst others, p53, CHK2, BRCA1, NBS1 and the FA protein FANCD2 are all phosphorylated by ATM as a direct result of its induction by DSBs. A cellular characteristic of cells from AT patients and patients with the closely related NBS and ATLD disorders is their so called radioresistant DNA synthesis (RDS); the continued replication of DNA even though DNA lesions are present. This is an indication that the control of the cell-cycle is defective in these cells and indeed, ATM impinges directly on cell-cycle control *via* p53 at the border of G1-phase and S-phase, *via* the RAD50/NBS1/MRE11 protein complex and possibly FANCD2 during S-phase and *via* CHK2 at the border of G2-phase and mitosis. Not surprisingly, RDS



Chromosomal Instability Syndromes. Figure 3 The network of CIS proteins. A highly simplified schematic representation of the interactions of the gene products involved in maintenance of chromosomal stability. FANCA-FANCG, FANCL and BRCA2 (FANCD1) are the products of the 9 known FA genes, a complex of 7 of the FA proteins is required for the monoubiquitination of a further FA protein, FANCD2. The other proteins are described in the text.

is more pronounced in AT cells than in NBS or ATLD cells which are defective in only part of the network. The interaction of ATM with p53 links the detection of DSBs with programmed cell death or ►apoptosis. This is a particularly safe way of dealing with DNA damage and avoiding the risk of its fixation as a mutation. For multicelled organisms, apoptosis can prevent the development of a neoplasia. The phosphorylation of ATM target proteins does occur even in AT cells, albeit with delayed kinetics. In this case a further protein kinase, ATR, takes over the function of ATM. Mutations in ATR have been described in tumours and in a patient with Seckel syndrome, a CIS disorder with some similarities to FA.

Helicases and Ligases

Helicases, particularly the RecQ family of helicases, named after the *E.coli* helicase RecQ, are enzymes required for unwinding the DNA double helix. They are involved in many aspects of DNA metabolism and are essential for the progression of the replication fork during DNA replication. This may be especially important when the replication fork becomes stalled at a DNA lesion such as a DNA strand-break. Of the five RecQ helicases known in man, three are involved in CIS. In addition to chromosome instability and increased cancer risk, patients with Bloom syndrome, Werner syndrome and Rothmund-Thompson syndrome show growth retardation and premature ageing. Differences in the tissues prone to neoplasia are presumably reflections of tissue specificity in the expression of the helicase genes.

DNA ligases, required for the ATP-dependent rejoining of DNA ends are also involved in chromosome stability. DNA ligase I was mutated in a patient whose cells showed hypersensitivity towards a variety of genotoxins including UV, alkylating agents and ionising radiation. This presumably reflects the central role of DNA ligase I in all the DNA repair pathways. However, although this patient had an immunodeficiency, immunoglobulin gene rearrangement was undisturbed. This, and the finding that mutation in such a central gene is at all compatible with life, is yet another indication that loss of one enzyme can be compensated for by the use of other enzymes and pathways.

In contrast to DNA ligase I, mutations in DNA ligase IV do disturb immunoglobulin gene rearrangement and the immunodeficiency of CIS patients with mutations in this gene can be attributed to that disturbance. DNA ligase IV is one of the 5 proteins known to be involved in NHEJ, which is the mechanism involved in immunoglobulin gene rearrangement and is also the major repair pathway for ionising radiation-induced DSBs; ligase IV deficient cells are hypersensitive to ionising radiation.

Genetics

Most CIS disorders are inherited as autosomal recessive traits. The majority of the diseases are genetically homogeneous, however, FA shows extreme ►locus heterogeneity with at least 11 causative genes. The proteins coded by the FA genes interact with BRCA1 and BRCA2 in a multimeric nuclear complex that is involved in the repair of DNA damage, probably via the gene conversion pathway and RAD51 (3). Similarly, the NBS and ATLD gene products associate together with the RAD50 protein and this trimeric complex is activated by the ATM protein. This complex is apparently involved in DNA repair by both NHEJ and homologous recombination and also plays a role in the regulation of the cell-cycle.

Apart from the children of consanguineous parents, the majority of CIS patients are ►compound heterozygotes. Most mutations are found in single individuals or in a very small number of patients. Nijmegen breakage syndrome is an exception since over 90% of the patients are homozygous for a five base pair deletion in the *NBS1* gene (4). Founder mutations in the FA genes *FANCA* in South Africa and *FANCG* in Germany and in Japan have been described. Similarly, approximately 1% of the ►Ashkenazi Jewish population is the carrier of a founder deletion-mutation in *BLM*.

Most mutations in CIS genes lead to truncated proteins and the effect of null mutations in man is generally unknown. In mice, null-mutants for some CIS genes, such as *RAD51*, *MRE11* and *NBS1*, are embryonically lethal whilst others, such as the FA genes have a very mild phenotype. Mutations in the tumour suppressor gene, *BRCA2*, result in a high risk for breast cancer, inherited as an autosomal dominant trait. Breast cancer occurs if the second, wild type allele suffers a somatic mutation. The bi-allelic mutations in *BRCA2* found in FA-D1 patients are hypomorphic (5). Presumably the more damaging mutations are embryonically lethal in man in the homozygous state, as in *BRCA2* null-mutant mice.

Clinical Relevance

Diagnosis of the CIS disorders is routinely conducted by the examination of chromosomal breakage, particularly after a specific genotoxic treatment. However, since many CIS genes have been identified it is also possible to analyse these for mutations in potential CIS patients. This undertaking is complicated, however, by the lack of common mutations, the likelihood of compound heterozygosity and the extreme size of some of the genes, for example *ATM* with 65 exons and 9,171 base pairs of coding sequence and *BRCA2* with 27 exons and 10,257 base pairs.

If DNA from an index patient is available, molecular diagnostics for further family members are often more

efficiently achieved indirectly, by using polymorphic DNA markers to follow the segregation of chromosomes in the family.

► DNA Repair Mechanisms

References

1. Revell SH (1959) The accurate estimate of chromatid breakage and its relevance to a new interpretation of chromatid aberrations induced by ionizing radiations. *Proc R Soc London B* 150:563–589
2. Tutt A, Bertwistle D, Valentine J et al (2001) Mutation in Brca2 stimulates error-prone homology-directed repair of DNA double-strand breaks occurring between repeated sequences. *EMBO J* 20:4704–4716
3. Digweed M, Rothe S, Demuth I et al (2002) Attenuation of the formation of DNA-repair foci containing RAD51 in Fanconi anaemia. *Carcinogenesis* 23:1121–1126
4. Varon R, Vissinga C, Platzer M et al (1998) Nibrin, a novel DNA double-strand break repair protein, is mutated in Nijmegen breakage syndrome. *Cell* 93:467–476
5. Howlett NG, Taniguchi T, Olson S et al (2002) Biallelic inactivation of BRCA2 in Fanconi anemia. *Science* 297:606–609

Chromosomal Microdeletion Syndromes

► Microdeletion Syndromes

Chromosomal Non-Histone Proteins

► Non-Histone Chromatin Proteins

Chromosomal Translocation

Definition

Chromosomal translocations are structural irregularities that result from a change in position of a chromosomal segment, but do not alter the total number of genes present in the cell. Translocations are caused by a chromosomal break whose ends rejoin with a non-homologous chromosome, thus generating a novel chromosome carrying segments from each.

► Cre/Lox P Strategies

► Methylation of Proteins

Chromosome

Definition

A chromosome is a rod-shaped structure containing the genetic material that is found in the cell nucleus. Each chromosome consists of a single, long DNA molecule with a telomere at each end, and a centromere mediating attachment to the mitotic spindle. In general, all human chromosomes have similar structures. The DNA and protein are found in the arms of the chromosomes and the genes are located at specific sites. The two arms of the chromosome can usually be identified as the short and the long arm, separated by a narrow area on each chromosome called the centromere. Every species has a characteristic number of chromosomes that are visualized as separate entities only during mitosis (e.g. humans have 46 chromosomes). Autosomes are chromosomes that occur in equal number in both sexes. Women carry two X while men carry a Y and an X chromosome. X and Y collectively are called sex chromosomes.

► Chromosome Condensation

► Fragile X Syndrome

► Repeat Expansion Diseases

Chromosome 21, Disorders

SAMUEL DEUTSCH¹, STYLIANOS E. ANTONARAKIS²,
ALEXANDRE REYMOND³

¹Department of Genetic Medicine and Development,
University of Geneva Medical School,
Geneva, Switzerland

²University Hospital of Geneva, Geneva, Switzerland

³Center for Integrative Genomics, University of
Lausanne, Lausanne, Switzerland

samuel.deutsch@medecine.unige.ch,
stylianos.antonarakis@medecine.unige.ch,
alexandre.reymond@medecine.unige.ch

Definition

The use of an entire chromosome allows for the analysis of a contiguous piece of DNA large enough to contain many of the variable features of the ► [genome](#). The human chromosome 21 (Hsa21) has attracted considerable attention partly due to its small size (long arm 33.5 Mb, ~1% of the human genome) and partly

due to its association with Down syndrome. Hsa21 is continually used as a model for structural and functional genome studies. Several innovative studies have been done on features of this chromosome, such as establishment of the first ►genetic maps and ►physical maps, systematic exon trapping, sequencing, ►single nucleotide polymorphisms (►SNPs) and ►haplotype mapping and transcriptional activity (4, 6). Similarly, comparative genomics of Hsa21 combining computational, experimental and evolutionary analyses allowed us to provide evidence for a large number of potentially functional but non-genic conserved sequences (1). The development of the genomic infrastructure of Hsa21 and the quality of its ►annotation has greatly facilitated medical related research. To date, 22 Mendelian ►monogenic disorders have been mapped to the chromosome, and for 20 of these, the affected genes and their respective mutations have been identified. These disorders cover a wide range of medical conditions, from autoimmune disorders to cancer, and in some cases they have illustrated novel mutational mechanisms (Table 1).

In this chapter we describe the main features of ►trisomy 21, the monogenic disorders that map to this chromosome, the gene catalogue and other features of the sequence, and some of the tools that can be used to understand how chromosome 21 genes contribute to human pathology.

Characteristics

Trisomy of Chromosome 21—Down Syndrome (OMIM#190685)

Down syndrome (DS) was first described as a clinical entity in 1866 by John Langdon H. Down in a group of patients with a series of common phenotypes including mental retardation, characteristic facies, short stature and muscle hypotonia (2). ►Cytogenetic studies showed that DS was caused by trisomy 21 almost a century later (5). Extensive characterization has revealed a variety of phenotypes in DS individuals such as early-onset Alzheimer disease, congenital malformations in the heart (present in 30–40% of patients) and the gastrointestinal tract, immune system alterations and multiple dysmorphic features. However there is considerable variation in the expression of these phenotypes from patient to patient (3). DS is the most common cause of genetic mental retardation affecting 1 out of 700 live births in all human populations. Moreover there is a pronounced maternal age effect with risk increasing from about 1/1000 at the age of 30 to 9/1000 at the age of 40. In terms of the cytogenetic alterations present in DS individuals, one can distinguish four categories (see schematic representation in Fig. 1):

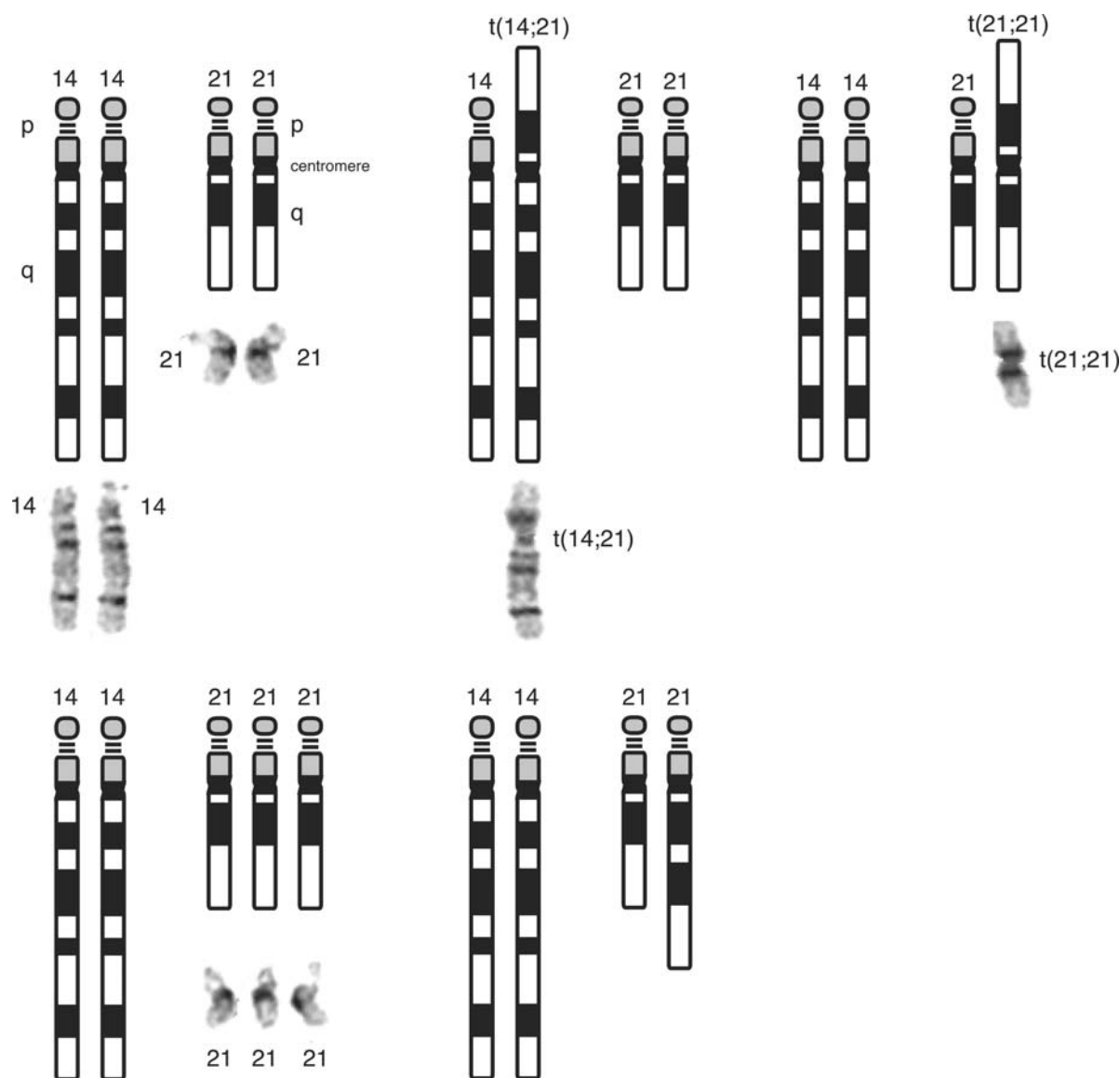
1. Free trisomy of Hsa21 is the most common cause of DS observed in 92.5% of cases. The extra chromosome originates from a non-disjunction event in maternal ►meiosis I (73% of the free trisomy cases), in maternal meiosis II (18%), in paternal meiosis (4%) or in mitosis (5%).
2. ►Translocation trisomy of Hsa21 with another ►acrocentric chromosome (usually Hsa14 or Hsa21) is observed in 4.8% of DS patients. The t(14;21) translocations originate primarily from *de novo* maternal germ cells rearrangement. The t(21;21) rearrangement is usually an ►isochromosome, generated by a duplication of the long arm of Hsa21 rather than a real robertsonian translocation. The t(21;21) ►duplication can be of maternal or paternal origin.
3. Partial trisomy is a very rare event in which patients present several of the DS phenotypes, but only a partial duplication of the long arm of Hsa21. The study of these rare patients allows definition of gene candidates for each DS phenotype.
4. Trisomy 21 ►mosaicism is observed in 2.8% of patients. Affected individuals have cells with three copies of Hsa21 and cells with two copies of Hsa21. The phenotype of these individuals is highly variable and thus impossible to predict. This can be caused by a mitotic non-disjunction event early in the postzygotic life, or by the early loss of an Hsa21 in trisomic cells.

Monogenic Disorders

There are a number of monogenic disorders that have been linked to Hsa21 (a complete list is available in Table 1). The first of these disorders to be characterized was leukocyte adhesion deficiency (LAD), an autosomal recessive disorder affecting neutrophil function that results in recurrent bacterial infections. The ITGB2 gene was shown to harbor mutations in patients with LAD. Since then additional monogenic disorders have been mapped to chromosome 21, and for the majority of these, the mutated gene has been identified (Table 1). Examples of these include (i) the autoimmune polyendocrinopathy syndrome (APECED), which was the first autoimmune syndrome to be characterized at the molecular level, (ii) progressive myoclonus epilepsy 1 (EPM1), which is caused in most cases by the expansion of a dodecamer repeat upstream of the CSTB gene, a novel mutation mechanism which results in the gene down-regulation and (iii) autosomal recessive congenital neurosensory deafness types 8 and 10 (DFNB 8/10) caused by another novel mutational mechanism, the insertion of a β -satellite repeat in the coding region of the Tmprss3 gene.

Chromosome 21, Disorders. Table 1 Hsa21 disorders

#	Gene Name	Mapping Position	SwProt	MIM Code	Description	Disorder
1	PRSS7	21q21.1	P05067	226200	serine protease-7 gene, encodes proenterokinase.	enterokinase deficiency
2	APP	21q21.3	P05067	104760	Amyloid a4 protein (beta-APP).	AD1
3	SOD1	21q22.1	P00441	147450	Superoxide dismutase (Cu-Zn).	ALS1
4	OLIG2	21q22.1	Q13516	606386	Prot. Kinase C-binding protein, T(14;21) T-cell ALL PRKCBP2, RACK17	
5	IFNGR2	21q22.1	P38484	147569	Interferon-gamma receptor beta chain, IFNGR2	Mycobacterial infection, familial disseminated
6	KCNE2	21q22.1	Q9Y6J6	603796	K channel, voltage-gated, Isk-related; MiRP1	Long QT syndrome
7	KCNE1	21q22.1	P15382	176261	Isk slow voltage-gated potassium channel protein.	Jarvell & Lange-Nielsen syndrome
8	RUNX1	21q22.1	Q01196	151385	Acute myeloid leukemia 1 protein (AML1), CBFA2	AML, t8;21; FPDMM (601399)
9	CLDN14	21q22.2	O95500	605608	Claudin-14	DFNB29
10	HLCS	21q22.2	P50747	253270	Biotin-protein ligase; Holocarboxylase synthetase; HCS	HLCS deficiency
11	TMPRSS3	21q22.3	P57727	601072	Transmembrane serine protease 3	DFNB10, DFNB8
12	CBS	21q22.3	P35520	236200	Cystathionine beta-synthase	Homocystinuria
13	CRYAA	21q22.3	P02489	123580	Alpha crystallin A chain (CRYA1)	AD cataract zonular cent; AR cataract
14	CSTB	21q22.3	P04080	601145	Cystatin B (Stefin B) (CST6)	EPM1
15	AIRE	21q22.3	O43918	240300	Autoimmune regulator	APECED
16	ITGB2	21q22.3	P05107	600065	Integrin beta-2 (CD18 antigen)	Leucocyte Adhesion Deficiency
17	COL18A1	21q22.3	P39060	267750	Collagen XVIII, alpha-1 chain	Knobloch syndrome, KNO
18	COL6A1	21q22.3	P12109	120220	Collagen VI, alpha-1 chain	Bethlem myopathy
19	COL6A2	21q22.3	P12110	120240	Collagen VI, alpha-2 chain	Bethlem myopathy
20	COL6A2	21q22.3	P12110	254090	Collagen VI, alpha-2 chain	Ullrich scleroatonic muscular dystrophy
21	?	21q21		602097		USH1E; Usher syndrome type 1E
22	?	21q22.3		236100		HPE1; Holoprocencephaly
23	?	21q21				AML type M7 and Transient Neonatal Leukemia
24	?	21q22.3				BPAD-21; Bipolar affective disorder
25	?	21q22.3				FCHL-21; Familial Combined hyperlipidemia, one locus
26	?	21q11-21				NSCLC 1; Deleted in Non-small cell lung cancer



Chromosome 21, Disorders. Figure 1 Disorders of Chromosome 21.

Susceptibility Alleles to Complex Common Phenotypes

In addition to the monogenic disorders, there are complex disorders for which there is considerable evidence of [linkage](#) to Hsa21 (Table 1). Complex disorders, usually refer to common diseases with a strong genetic predisposition that segregate in a non-Mendelian fashion. They are thought to be mediated by multiple allelic variations in more than one gene as well as non-genetic factors. For two such disorders, bipolar affective disorder (BPAD), and familial combined hyperlipidemia (FCHL), loci on chromosome 21 are thought to play important roles, however the gene or genes responsible remain elusive.

Cellular and Molecular Regulation Gene Catalogue

Initial annotation of the entire sequence of Hsa21 by a combination of experimental and bioinformatics approaches revealed 127 genes and 98 predicted transcripts (4). However experimental confirmation of the predicted transcripts and more sensitive *in silico* methods for gene detection, together with the completion of the sequencing of several genomes in particular that of the mouse permitted an update of this catalogue, which currently contains around 290 genes. Following characterization of the Hsa21 gene repertoire, the next goal is to uncover the genes' functions and to determine which ones are involved in DS and other disease

Chromosome 21, Disorders. Table 2 Examples of Hsa21 mouse orthologous genes phenotypes (a complete table of Hsa21 mouse orthologous genes phenotypes is available at the <http://www.medgen.unige.ch/ url>)

#	Gene Name	Mapping Position	SwProt	Description	Genotype	Phenotype
1	Aire	Mmu10	Q9Z0E3	Autoimmune regulator (Autoimmune polyendocrinopathy candidiasis ectodermal dystrophy protein)	-/-	Multiorgan lymphocytic infiltration, Circulating autoantibodies Infertility
2	App	Mmu16	(P05067)	Beta-amyloid precursor protein	-/-	Severely impaired in spatial learning and exploratory behavior Increased incidence of agenesis of the corpus callosum, Compromised neuronal or muscular function
3	Col6a1	Mmu10	Q04857	procollagen, type VI, alpha 1	-/-	Fiber necrosis and phagocytosis Pronounced variation in the fiber diameter Stimulated regeneration of fibers
					+/-	Similar, although milder, alterations
4	Cryaa	Mmu17	P02490	Alpha crystallin A chain	-/-	Smaller lenses, Develop lenses opacification
				lens opacity 18 (lop18) mouse <i>Cryaa</i> R54H	lop18/lop18	White cataract at weaning age, Develop large white nuclear cataract with mild cortical changes
5	Cstb	Mmu10	Q62426	Cystatin B (Stefin B)	-/-	Develop myoclonic seizures and ataxia Loss of cerebellar granule cells
6	Dnmt3l	Mmu10	Q9CWR8	DNA (cytosine-5-)methyltransferase 3-like, regulator of imprint establishment	-/-	Azoospermia Regions normally maternally-imprinted region are not methylated
					+/- progeny of -/- female	Die during midgestation
7	Dyrk1a	Mmu16	Q61214	Dual-specificity protein kinase Ortholog of the <i>Drosophila</i> minibrain (mnb) gene.	-/-	General growth delay, die during midgestation
					+/-	Increased neuronal densities in somatosensory cortex and thalamus Decreased neuronal densities in the superior colliculus Prewaning developmental delay

Chromosome 21, Disorders. Table 2 Examples of Hsa21 mouse orthologous genes phenotypes (a complete table of Hsa21 mouse orthologous genes phenotypes is available at the <http://www.medgen.unige.ch/> url) (Continued)

#	Gene Name	Mapping Position	SwProt	Description	Genotype	Phenotype
8	Kcne1	Mmu16	P23299	Potassium voltage-gated channel sub-family E member 1	-/-	Classic shaker/waltzer behavior Hair cells degenerate Lack of transepithelial potassium secretion in strial marginal cells and vestibular dark cells of the inner ear
9	Kcnj6	Mmu16	P48542	punk rocker (pkr) mouse <i>Kcne1</i> R67X mutation G protein-gated, inwardly rectifying K+ channel 2, GIRK2, KCNJ7 weaver (wv) mouse <i>Kcnj6</i> G156S mutation in conserved H5 domain of the channel	pkr/pkr -/- wv/wv	Distinctive head-tossing, circling, and ataxic phenotype Profoundly and bilaterally deaf Develop spontaneous seizures Reduced GIRK1 expression in the brain Neuronal differentiation blocked just after cessation of cell division, prior to cell migration and synaptogenesis Cerebellar granule cells fail to differentiate Reduced GIRK1 expression in the brain
10	Sim2	Mmu16	Q61079	Transcription factor Homolog of <i>Drosophila</i> single-minded gene (sim)	-/- wv/-	Milder cerebellar abnormalities than wv/wv mice, suggesting that wv mutation is gain-of-function Die before day three Lung atelectasis and breathing failure Rib protrusions, abnormal intercostal muscle attachments, diaphragm hypoplasia and pleural mesothelium tearing, Congenital scoliosis

More details are accessible through ► http://www.informatics.jax.org/searches/allele_form.shtml

Chromosome 21, Disorders. Table 3 Mouse models of DS

Model	Genotype	Neurological phenotype
Segmental trisomy		
Ts16	Trisomy 16	Reduced brain size and some structural changes
Ts65Dn	Trisomic for App-Znf295 (~86 known genes)	Learning and behavioral deficits, degeneration of basal forebrain cholinergic neurons Reduction of the cerebellar volume and granule cell density Reduced cell number and volume in the hippocampal dentate gyrus Reduction in excitatory (asymmetric) synapses in the temporal cortex at advanced ages Age-related degeneration of basal forebrain cholinergic neurons Astrocytic hypertrophy and increased astrocyte numbers
Ts1Cje	Trisomic for Znf295-Sod1 (~53 known genes)	Learning and behavioral deficits (less severe than in Ts65Dn)
Ms1Ts65	Trisomic for App-Sod1 (~33 known genes)	Learning deficits (less severe than in Ts1Cje)
Ts1Rhr	Trisomic for Mx1 to Cbr1	No cranofacial abnormalities
Single Genes		
TgSod1	Transgenic for human SOD1	Learning defects
TgPfkf	cDNA, highly over-expressed	-
TgS100β	2-12 copies	Astrocytosis, neurite degeneration
TgApp	YAC, low copy	Cognitive/behavioral defects
TgEts2	cDNA, highly over-expressed	-
TgHmgn1	2-6 copies	-
TgDyrk1a	YAC, 1-3 copies, and cDNA, highly over-expressed	Learning/memory defects
TgSim2	BAC, 1-2 copies, and cDNA, highly over-expressed	Behavioral defects

phenotypes. Among the known Hsa21 genes there are at least 10 kinases, 5 genes involved in ubiquitination pathways, 20 transcription factors, 5 cell adhesion molecules and 7 ion channels. Second generation annotation by functional genomics approaches will undoubtedly contribute to the understanding of the functions of Hsa21. To this end, the first chromosome-wide “Gene Expression Atlas” was produced using Hsa21 as a model. This resource contains the spatio-temporal expression pattern of all chromosome 21 mouse orthologous genes (7). These data can be used to refine candidate gene lists for both monogenic disorders and Down syndrome sub-phenotypes. The mouse is a well-characterized organism for modeling human disease states. As homologous

recombination in mouse allows the engineering of loss-of-function mutations in specific genes, we can study Hsa21 monogenic disorders and gene function using orthologous genes knockouts as exemplified in Table 2. Furthermore the study of these mouse phenotypes could help to unravel potential novel monogenic disorders that map on Hsa21 (Table 2 also describes phenotypes of mouse mutants for Hsa21 orthologous genes not yet linked to monogenic disorder).

Molecular Mechanisms of Pathogenesis in DS

One of the major current challenges for the chromosome 21 and DS related research is to understand which and how chromosome 21 genes contribute to the specific phenotypes of Down syndrome. The current

working hypothesis is that some but not all genes and non-genic conserved sequences in Hsa21, when present in three copies contribute to DS phenotypes. Some of the contributing genes and non-genic conserved sequences might exert a direct effect, and thus cause specific DS phenotypes, whereas others could have an indirect effect by altering downstream transcription pathways.

In addition, the incomplete penetrance of many DS phenotypes argues for allelic effects, i.e. that a specific DS trait is only present when the individual has inherited a certain combination of alleles in specific loci. In order to dissect each of these mechanisms three main experimental approaches are being taken:

1. Differential gene expression analysis between normal and trisomic cells and tissues. This approach should allow the identification of the genes and pathways that are deregulated in trisomic cells, and therefore might be implicated in pathology.
2. Development of mouse models of human trisomy 21, which can be either single gene trisomic mice, or segmental trisomic mice, described in detail in Table 3. For single gene trisomic mice, a candidate gene is over-expressed and its specific effects can be studied. For segmental trisomies large regions of mouse chromosome [▶syntenic](#) to Hsa21 are duplicated in order to model the human disease.
3. [▶Association studies](#) using single nucleotide polymorphisms (SNPs) in order to identify genetic variation that might contribute to specific DS phenotypes that are present only in a subset of patients.

Clinical Relevance

Understanding the molecular consequences of [▶aneuploidies](#) should lead to the discovery of genes responsible for several phenotypes, including the cognitive impairment of trisomy 21 patients, and provide targets for potential therapeutic interventions (8). Moreover identification of the molecular defect in monogenic disorders allows the discovery of the molecular pathophysiology, which in turn leads to the introduction of novel therapies and new accurate molecular diagnostics. Finally the discovery of susceptibility alleles for common disorders will also contribute to early diagnosis and therapeutic interventions.

References

1. Dermitzakis ET et al (2002) Numerous potentially functional but non-genic conserved sequences on human chromosome 21. *Nature* 420:578–582
2. Down JLH (1866) Observations on an ethnic classification of idiots. *London Hosp. Clin. Lect. Rep.* 3:259
3. Epstein CJ (1995) Down syndrome (trisomy 21). In: Scriver CR, Beaudet AL, Sly WS, Valle D (eds) *The Metabolic and Molecular Bases of Inherited Disease*, New York, pp 749–794

4. Hattori M et al (2000) The DNA sequence of human chromosome 21. *Nature* 405:311–319
5. Le Jeune J, Gautier M, Turpin R (1959) Etudes des chromosomes somatiques de neuf enfants mongoliens. *C.R.Acad.Sci.Paris* 248:721–722
6. Patil N et al (2001) Blocks of limited haplotype diversity revealed by high-resolution scanning of human chromosome 21. *Science* 294:1719–1723
7. Reymond A et al (2002) Human chromosome 21 gene expression atlas in the mouse. *Nature* 420:582–586
8. Antonarakis SE et al (2004) Chromosome 21 and down syndrome: from genomics to pathophysiology. *Nat Rev Genet* 10:725–738
9. OMIM (Online Mendelian Inheritance in Man), [▶http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM](http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM)

Chromosome Bands

Definition

Various chemical and physical treatments, followed by incorporation of dye, may be used to produce a pattern of specific transverse light and dark bands on condensed metaphase chromosomes. These bands correspond to clusters of GC- and AT-rich, and early- and late-replicating DNA, respectively, and thus reflect the underlying longitudinal structural and functional compartmentalisation of chromosomes.

[▶Nuclear Compartments](#)

Chromosome Breakage Syndromes

[▶Chromosomal Instability Syndromes](#)

Chromosome Condensation

ELLADA SAVVIDOU, MARGARETE M. S. HECK
Wellcome Trust Centre for Cell Biology, University of Edinburgh, Edinburgh, UK
elladasav@hotmail.com,
margarete.heck@ed.ac.uk

Definition

Chromosomal DNA, if stretched end-to-end, measures about 2 m in any one cell of the human body and therefore it must be highly folded to fit into a nucleus of

only 5 μm diameter. In addition to this, DNA has to condense even more prior to cell division in order to form the highly compacted metaphase **chromosome**. Several levels of packing enable DNA to achieve this packaging and a number of structural proteins are involved. The simplest level of packing involves the winding of DNA around octamers of core histones in a structure called the nucleosome. This shortens DNA about 7-fold relative to naked DNA. The folding of regularly spaced nucleosomes forms a shorter, thicker filament, called the 30 nm chromatin fibre, in which the DNA is about 40-fold more compact than when naked. The most highly condensed and least understood stage takes place prior to chromosome segregation when the DNA in metaphase chromosomes of higher eukaryotic cells is compacted nearly 10,000-fold in length. Each chromosome must not only be compacted, but also resolved into two individual sister chromatids separable in **mitosis**. The way that chromatin organizes into this higher-order structure of mitotic chromosomes is not well understood and the different models that have been proposed remain controversial.

Characteristics

First Level of Compaction: The Nucleosome

Interphase nucleosomal chromatin under low salt conditions appears (by electron microscopy) as “beads on a string” with a diameter of about 10 nm. This packaging results when 166 bp of DNA wind around an octamer of core histones in a left-handed superhelix following a helical path along the surface of the octamer. The histone octamer consists of an $[\text{H3}]_2:[\text{H4}]_2$ tetramer, flanked on either side by an H2A:H2B heterodimer. Each core histone has a compact domain of 70–80 amino acid residues and a flexible N-terminal tail of approximately 30 amino acids outside the nucleosome. These N-terminal tails allow the nucleosomes to pack into a 30 nm fibre, while specific modifications of the tails are crucial for modulation of chromatin structure. Histone H1, or other linker histones, bind to a variable amount of “linker” DNA that extends between adjacent nucleosomes. The N-terminal tails of the histones are also important for extra-nucleosomal interactions. The variants of linker histone differ from the core histones in that the globular central domain is flanked by unstructured basic domains at both the N- and C-termini. Histone H1 resides at the side of nucleosome where DNA enters and exits the nucleosome. The master cell-cycle regulatory kinase, CDK1 (cdc2: cyclin B), phosphorylates serine and threonine residues on the histone H1 tails during mitosis. Hyperphosphorylation of histone H1 was originally thought to be important for mitotic chromosome condensation but recent studies show that it is not required for this process. Mitosis-specific phosphorylation of histone H3 at serine-10 (by Aurora

B kinase) is also not necessary for mitotic chromosome condensation. The N-terminal tails of histones may also be acetylated or methylated, resulting in an epigenetic “histone code” important for gene expression or assembly of heterochromatin.

Above and Beyond the Nucleosome

Chromatin isolated from interphase nuclei has a diameter of 30 nm. This fibre has been variously modelled as either a solenoidal coiling, random or organized folding of the nucleosomal fibre – its actual structure remains undetermined. Fragility of the 30 nm fibre during preparation is probably responsible for the inability to ascertain its architecture unambiguously. However, the 40-fold length compaction within the 30 nm fibre is not nearly enough to account for the compaction of the genome in interphase nuclei or mitotic chromosomes.

Mitotic Chromosome Condensation

The mechanism for packaging the 30 nm fibre into the highly compact, characteristic rod-shaped mitotic chromosomes remains a fundamental question for modern cell biologists. The shortening of DNA by roughly 10,000-fold is essential for the proper physical separation of sister chromatids to opposite poles, e.g. in order to avoid cleavage of the genome by an ingressing cleavage furrow. In addition, chromosomes must be properly aligned *via* their **kinetochores** to ensure that one of each chromatid pair ends up in each new daughter cell. However, early approaches utilizing electron microscopy to examine isolated native mitotic chromosomes only served to confirm that the DNA is indeed very densely packed within chromosomes.

Several approaches have been utilized in an attempt to expose the factors underlying higher order chromosome structure. By depleting core histones from highly purified mitotic chromosomes (either with a high salt concentration or polyanions which compete with histones by charge) under conditions that maintain residual non-histone proteins, some of the first hints about large scale chromatin folding were obtained. Long extended loops of DNA (about 15–100 kb) surrounded a protein core that retained the general size and shape of a native chromosome. These striking images led to a model known as the “radial loop model” in which the folding of the chromatin fibre into loops is mediated by proteins of an insoluble chromosome “scaffold” that is essential for the final size and shape of a chromosome. Others though considered the scaffold to be an artefact, a result of inappropriate protein interactions created during preparation of this fraction. Nonetheless, this model was first proposed more than 25 years ago, and there is a significant amount of data supporting it. A:T-rich DNA sequences (SARs – scaffold attachment regions, or

MARS – matrix attachment regions) have been identified as the *cis*-sequences interacting with the proteinaceous component. The DNA intervening between two SAR sequences bound to the “scaffold,” separated by up to 100 kb of DNA, would form a chromatin loop.

Protein Components – ▶ Topoisomerase II and ▶ Condensin

A biochemical analysis of scaffolds showed there were two major components, Sc1 and Sc2, which later were identified as DNA topoisomerase II and SMC2 (a subunit of the condensin complex), respectively. In addition to this, characterization of chromosomes remodelled from sperm chromatin in *Xenopus* egg extracts, resulted in the identification of several protein components of these *in vitro* assembled structures. The mitotic chromosome fraction contained histones and linker histones, topo II α , chromokinesin, the ▶ **chromatin remodelling** ATPase ISWI, and a complex subsequently named condensin. These studies demonstrated that mitotic chromosomes are rich in ATPases, which are probably important to induce conformational changes in chromosomes and to support microtubule-dependent movement of chromosomes.

Topoisomerase II is an ATP-dependent enzyme that can alter DNA topology by passing one double-strand helix through another *via* a transient double-strand break, an activity necessary for the resolution of catenated sister DNA strands. Topo II is located both at the centromeres and axially along mitotic chromosome arms at the base of loops. All eukaryotic organisms have at least one topo II. Higher eukaryotes though, have two isoforms (topo II α and topo II β) that have similar catalytic activities *in vitro* but different localization *in vivo*. Topo II activity is required for the resolution of the catenated strands of DNA so that each chromatid of a pair becomes an individual unit, facilitating successful separation during anaphase. Indeed, mutations in topo II in both *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* were lethal in mitosis, with sister chromatids failing to segregate properly. Further analysis in *S. pombe* revealed that topo II was also required for the final stages of chromosome condensation, possibly in collaboration with the condensin complex, but its exact role in this process is not yet clear. In *Drosophila*, depletion of topo II by ▶ **RNAi** resulted in a 2.5-fold decrease in the level of chromatin compaction but overall chromosome condensation occurred without topo II. In the same study, chromosome morphology was abnormal and one or more chromosome arms frequently stretched out from the metaphase plate, suggesting an unexpected role for topo II in chromosome arm congression to the metaphase plate. Even though topo II seems to be important for the establishment of chromosome

condensation in many different organisms, in *Xenopus* egg extracts it seems not to be essential for the maintenance of chromosome condensation, suggesting that topo II might contribute to the early stages of this process. Interestingly, as chromosomes decondense at the end of mitosis, topo II α undergoes specific degradation, only to be resynthesized at the start of S phase if cells are actively cycling.

ScII was identified as SMC2, a structural maintenance of chromosomes protein, and a member of the condensin complex. This complex consists of two ▶ **SMC proteins** (SMC2 and SMC4) and three non-SMC subunits (CAP-D2, CAP-G and CAP-H). SMC2 and SMC4 form a heterodimer and the non-SMC subunits presumably play regulatory roles. The SMCs are present in all species examined and they constitute a family of highly conserved and ubiquitous proteins. All members are large polypeptides with a molecular mass of 115–165 kD and a similar structural organization. The proteins have two extended coiled-coil domains, which are separated by a central globular hinge region. The N- and C-termini flanking the coiled-coil domains are highly conserved among SMC proteins. The N-terminal domain contains a Walker A box (an ATP-binding motif). The C-terminal domain contains a “DA” box resembling a Walker B site (ATP hydrolysis motif) and it has been shown to bind DNA. Eukaryotic SMCs comprise at least six distinct subfamilies (SMC1-SMC6), which form 3 different heterodimers: SMC1/SMC3 (chromatid cohesion), SMC2/SMC4 (chromosome condensation), and SMC5/SMC6 (DNA repair). Recent data suggests that each of the SMC subunits of the ▶ **cohesin** and condensin complex folds intramolecularly and their heterodimerisation is mediated by a hinge-hinge interaction. The intramolecular folding of an SMC molecule allows the association of the N- and C-terminal domains and the formation of an ATP catalytic domain at each end of the heterodimer. The non-SMC subunits bind to one or both of the catalytic domains.

Much of the current understanding of condensin activity comes from the biochemical characterization of the complex from *Xenopus* egg extract. In this system it has been shown that all five subunits are necessary for DNA-stimulated ATPase activity, which reconfigures DNA in an ATP-dependent manner. Condensin introduces positive supercoils into relaxed circular DNA in the presence of topoisomerase I and converts nicked circular DNA into positively knotted forms in the presence of topoisomerase II *in vitro*. Recent observations of the condensin complex by spectroscopic imaging suggest that a single condensin complex organizes two oriented supercoils, thus inducing superhelical tension. A model that emerges from these observations is that each of the two end domains wraps DNA into nearly one superhelical turn

allowing a single condensin complex to create two supercoils. Although the way that DNA interacts with the two-armed structure of condensin is unknown, a recent study showed that the region that binds DNA might be in the coiled-coil arms and not in the catalytic domains. A similar model has been recently proposed for the action of cohesin. Another study suggests that condensin may form a DNA loop by holding two distantly located DNA segments.

Two models have been proposed to explain how condensin may act on chromatin but both models are highly speculative. The first model proposes that local positive supercoiling driven by condensin introduces a compensatory superhelical tension into the neighbouring region of DNA, which in turn acts as a driving force for the compaction of a chromatin fibre. The problem with this model is that there is no evidence for superhelical tension in metaphase chromosomes. The second model is based on the prediction that a single condensin complex introduces two positive supercoils by a trapping mechanism and forms a topological link between the condensin “clamp” and the trapped DNA strands. If a free translocation of the DNA strands were allowed without disrupting the topological link, a positive loop of DNA would be created. A continuing creation of positive loops by the action of condensin might contribute to mitotic chromosome organization.

Even though condensin was originally thought to be responsible for chromosome condensation (hence the name), recent genetic studies in different organisms suggest that this complex is required for the proper organization and segregation of mitotic chromosomes and not condensation *per se*. Yeast mutations in the condensin complex increased the average distance between two loci on a mitotic chromosome arm, in the same way depletion of condensin from *Xenopus* mitotic egg extracts resulted in less compact chromosomes. Condensin however seems to be more than a simple compactor and is essential for resolution between sister chromatids *in vivo*. In *Drosophila* SMC4 mutants, shortening of the longitudinal axis of the chromosomes was unaffected, but chromosomes failed to resolve into two distinct chromatids at prophase, and then showed chromosome bridging in anaphase. Similarly depletion of SMC4 by RNAi in *Caenorhabditis elegans* caused dramatic abnormalities in chromosome morphology at prometaphase, yet a high degree of compaction still occurred by metaphase. Mutation or RNAi of non-SMC condensin subunits also resulted in aberrant chromosome segregation (but apparently normal condensation). In *Drosophila*, depletion of the condensin subunit SMC4 disrupted the localization of topo II in a central axial structure and furthermore to led a reduction of topoisomerase II-dependent DNA decatenation activity *in vitro*.

Condensin appears to be required for proper chromosome organisation, thus facilitating the action of topo II in sister chromatid resolution. Recently in an ScII/SMC2 knockout in chicken cells, it was found that chromosome condensation was delayed but it could finally reach its normal levels. In this study, mitotic chromosomes were missing a number of scaffold proteins and they lacked structural integrity, indicating that condensin was necessary for the establishment of a robust chromosome structure. One of the most important results to date is the observation that when both topo II and condensin are disrupted by RNAi, chromosomes are still able to condense. There are clearly more components essential for mitotic chromosome condensation to be identified.

The supercoiling activity of the condensin complex has been shown in vertebrates to require mitosis specific phosphorylation of the three non-SMC subunits by CDK1, the major mitotic kinase. CAP-D2 is phosphorylated at its C-terminus probably directly by CDK1, while CAP-H is phosphorylated probably by CDK1 or other kinases. The hyperphosphorylation of the non-SMC subunits is proposed to have two dual roles. The first one is to activate the ATPase activity of SMCs and the second to allow the condensin complex to be associated with chromatin in a mitosis-specific manner.

Recently a second condensin (II) complex was identified in vertebrate cells. Both complexes share SMC2/SMC4, but contain different sets of non-SMC subunits. The two complexes may have distinct distributions along chromosome axes, but show a similar distribution at the centromere region. It has been proposed that the two complexes make distinct mechanistic contributions to mitotic chromosome architecture in vertebrate cells. Depletion of condensin I or condensin II produces weaker defects in chromosome morphology, while simultaneous depletion of both complexes causes the severest defect. Furthermore, the centromeric regions were structurally disorganised after the depletion of either condensin I or II subunits, suggesting a role for both complexes in centromere organization. Sequence comparisons suggest that condensin II should exist in other organisms, but experimental data will confirm if that is the case.

Along with the phosphorylation of condensin subunits by the master mitotic kinase CDK1, the phosphorylation of histone H3 at the onset of mitosis has been proposed to play a role in condensin recruitment and chromosome condensation. Histone H3 mutations in *Tetrahymena thermophila* show abnormal chromosome condensation and extensive chromosome loss. In *C. elegans*, *Drosophila* and human cells, Aurora B, a chromosome-bound kinase, is targeted to chromosomes along their length during prophase, becomes concentrated at the centromeric regions during

metaphase and is transferred to the spindle midzone when sister chromatids are separated in anaphase. This kinase is thought to interact with two other molecules, the inner centromere protein (INCENP) and survivin, which together are referred to as chromosomal passengers. Aurora B has been proposed to be the physiological kinase that phosphorylates histone H3 during mitosis. A model that connects the phosphorylation of histone H3 and chromosome condensation is that the modification of the histone may send a signal to initiate chromosome condensation. The phosphorylation could function as a receptor for the recruitment of proteins such as the condensin complex. Experiments in several organisms though, suggest that the phosphorylation of histone H3 is not responsible for the direct recruitment of condensin onto chromosomes and chromosome condensation is not affected when histone H3 phosphorylation is diminished by depletion of Aurora B. Future work will clarify the role of H3 phosphorylation in mitotic chromosome dynamics.

Dependence on Appropriate DNA Replication and Repair

The final condensation of mitotic chromosomes is dependent not only on mitotic events, but also on appropriate timing of DNA replication and the satisfaction of DNA-structure checkpoints. In *Drosophila*, mutations of ORC2/3/5, subunits of the origin recognition complex, result in the formation of irregularly condensed chromosomes in mitosis. Regions of euchromatin that were replicated abnormally late in these mutations showed a severe condensation defect in contrast to other regions of the chromosome. Mutations in RFC4 (subunit 4 of replication factor C) resulted in DNA-structure checkpoint defects, chromosome pulverization (mitotic catastrophe) and also prematurely separated chromosomes in mitosis – as results of interfering with different functions of RFC. Whether the condensation defect observed in these mutants is a direct effect on topo II or condensin, is currently unknown.

A Yin-yang Balance with Cohesion?

Sister chromatid cohesion is established between the two replicated sister chromatids during S phase by an evolutionarily conserved complex called cohesin. The cohesin complex (a heterodimer of SMC1 and SMC3 and two additional non-SMC subunits – Scc1 and Scc3) is essential for holding sister chromatids together until the onset of anaphase, ensuring the accurate 1:1 partitioning of chromatids. In yeast all cohesin dissociates from chromosomes at the beginning of anaphase when the Scc1 subunits are cleaved. In higher eukaryotes, most of the cohesin complex already dissociates from chromosome arms in prophase after

phosphorylation by the polo kinase with only a small amount of cohesin left at centromeric regions. Residual cohesion is lost at the onset of anaphase by the cleavage of the Scc1 non-SMC subunit. The dissociation of cohesin from chromosome arms in prophase could facilitate either the condensation or resolution of chromatids. Alternatively, condensation may help to release cohesion between chromatids. However, loss of Scc1 (and cohesion) by RNAi has no apparent effect on condensation. It is still plausible that a mechanism ensuring the correct balance between cohesin and condensin may exist. A structural and functional link between cohesion and condensation, if it exists, remains to be determined.

We have attempted to highlight some of the most important issues surrounding chromosome organization and condensation throughout the cell-cycle. The organization of a “scaffold” and its association with specific DNA sequences, along with the mechanisms by which chromosomes condense prior to mitosis or decondense after mitosis are poorly understood. Nevertheless, exciting progress is being made through diverse approaches. We also believe there are as yet, uncharacterized factors that play crucial roles in these events, which may even result in the proposal of new models for the dynamics of higher order chromosome architecture.

Clinical Relevance

Chromosome condensation (and segregation) is a necessary part of cell division and the formation of a multi-cellular organism. Not surprisingly therefore, aberrations in this process have extreme implications for human health. The currently accepted rate of aneuploidies in pre-implantation embryos is 20%, while for miscarriages it is 35%! Less than 1% of live births are aneuploid, most commonly with a missing or extra sex chromosome or an additional chromosome 21 (►Down syndrome). Quite obviously, anything but the diploid state is incompatible with normal life.

Aneuploidy is also a factor, albeit less well understood, in the development of cancer (many solid tumours have between 60–90 chromosomes instead of the normal 46 in human). As transformed cells exhibit defective division control, targeting actively dividing cells has long been a chemotherapeutic strategy (with the undesired side effect of damaging normal dividing cells). Intriguingly, however, a number of topoisomerase II inhibitors are used in the clinic to treat particular cancers. This suggests that other proteins important for chromosome organization and dynamics may also be useful therapeutic targets.

►Chromatin Remodelling

►Nucleosome

References

1. Belmont AS (2002) Mitotic chromosome scaffold structure: new approaches to an old controversy. *Proc Natl Acad Sci USA* 99:15855–15857
2. Bharadwaj R, Yu H (2004) The spindle checkpoint, aneuploidy, and cancer. *Oncogene* 23:2016–2027
3. Cobbe N, Heck MMS (2004) The Evolution of SMC proteins: Phylogenetic Analysis and Structural Implications. *Mol Biol Evol* 21:332–347
4. Hagstrom KA, Meyer BJ (2003) Condensin and cohesin: more than chromosome compactor and glue. *Nat Rev Genet* 4:520–534
5. Hassold T, Hunt P (2001) To err (meiotically) is human: the genesis of human aneuploidy. *Nat Rev Genet* 2:280–291
6. Nasmyth K (2001) Disseminating the genome: joining, resolving, and separating sister chromatids during mitosis and meiosis. *Ann Rev Genetics* 35:673–745
7. Pollard TD, Earnshaw WC (2002) *Cell Biology*. Saunders, Philadelphia
8. Swedlow JR, Hirano T (2003) The making of the mitotic chromosome: modern insights into classical questions. *Mol Cell* 11:557–569

Chromosome Duplication

► Gene Duplications

Chromosome Engineering

Definition

Chromosome engineering comprises of techniques that introduce defined chromosomal rearrangements (deletions, duplications, and inversions) into the mouse genome, in order to model human chromosomal disorders, such as those associated with deletions or duplications of chromosomal segments (for example, ► [Down syndrome](#) and Charcot-Marie-Tooth type 1A). The Cre-loxP chromosome engineering strategy provides a unique and unprecedented opportunity to manipulate the mouse genome.

► [Cre/loxP Strategies](#)
 ► [Mouse Genomics](#)

Chromosome Instability

Definition

Chromosome instability refers to signs of impaired structural integrity of a chromosome seen at metaphase,

such as breaks, loss of chromosome material, dicentric chromosomes, joining of chromosomes, acentric fragments, ring shaped chromosomes and other deviations from the normal structure.

► [Bloom Syndrome](#)
 ► [Chromosomal Instability Syndromes](#)

Chromosome Passengers

Definition

Chromosome passenger proteins are proteins that are present in the centromeric region of chromosomes in pro-metaphase cells, but translocate to the spindle midzone at the onset of anaphase, and finally collect at the spindle equator just prior to the telophase. It has been suggested that chromosome passenger proteins play a role in the control of cytokines. So far there are 5 known members: INCENP, survivin, Aurora B, TD-60, and borealin.

► [Chromosome Condensation](#)
 ► [Mitotic Recombination](#)
 ► [Mitotic Spindle](#)

Chronic Airflow Limitation

► [COPD and Asthma, Genetics](#)

Chronic B-Cell Lymphocytic Leukaemia

Definition

Chronic B-cell lymphocytic leukaemia is the most common leukaemia in adults in western countries and belongs, according to the WHO classification of tumours of the haematopoietic and lymphoid tissues, to the mature B-cell neoplasms.

► [Leukemia](#)

Chronic Myelogenous Leukaemia

Definition

Chronic myelogenous leukaemia belongs, according to the WHO classification of tumours of the

haematopoietic and lymphoid tissues, to the group of chronic myeloproliferative disorders. Genetically, the disease is characterized by the Philadelphia chromosome and its molecular equivalent, the BCR-ABL fusion gene.

- Catalytic RNAs
- Leukemia

Chronic Obstructive Airway Disease

- COPD and Asthma, Genetics

Chronic Obstructive Lung Disease

- COPD and Asthma, Genetics

Chronic Obstructive Pulmonary Disease

- COPD and Asthma, Genetics

Chronic Pancreatitis

Definition

Chronic pancreatitis is a chronic inflammatory disease of the pancreas that is mostly induced by alcohol abuse, but can also occur sporadically or be inherited.

- Microarrays in Pancreatic Cancer

Chylomicron

Definition

Chylomicrons are lipoproteins that transport absorbed lipids from the intestine to the liver for further processing.

- Tangier Disease

Cilia

Definition

Cilia are hair like structures that project from the apical surface of certain epithelial cells (e.g. renal tubular epithelial cells; hair cells of the inner ear). A cilium is composed of a specialized assembly of tubulin (see ►tubulin) and associated proteins. Nine outer doublet microtubules constitute tubulin, which are surrounded by a membrane domain contiguous to the plasma membrane, but populated by a specific pattern of membrane associated proteins (which include polycystins 1 and 2). The function of this structure is poorly understood, but ciliary deflection results in cell calcium responses, indicating a possible role in mechanosensation.

- Autosomal Dominant (Inherited Disorder)
- Polycystic Kidney Disease, Autosomal Dominant

CIN

Definition

CIN stands for chromosomal instability, which reflects the alteration in chromosome number and integrity that characteristically occurs in many tumor types.

- Chromosomal Instability Syndromes
- Colorectal Cancer

Circadian

Definition

From Latin circa dies, the term refers to the length of the endogenous period of a biological rhythm of approximately one day.

- Circadian Clocks

Circadian Clocks

ACHIM KRAMER

Institute for Medical Immunology, Laboratory of Chronobiology, Charité University Medicine, Berlin, Germany

achim.kramer@charite.de

Definition

► **Circadian** clocks are endogenous ► **oscillators** that drive daily rhythms in physiology, metabolism and behavior. They provide organisms, ranging from unicellular algae to humans, with an internal representation of local time, thus allowing them to anticipate daily recurring chances and challenges in their environment. The components of circadian clocks are highly conserved among all animal species and the fundamental mechanism of circadian timekeeping is cell-autonomous. Within almost every cell, a gene-regulatory network involving clock genes and clock proteins forms a molecular clockwork that regulates many cellular processes including gene expression, neuronal activity, secretion of bioactive molecules, metabolism, mitosis, apoptosis etc. These molecular rhythms ensure temporal coordination of tissue-specific physiology and global behavioral states such as the sleep-wake cycle.

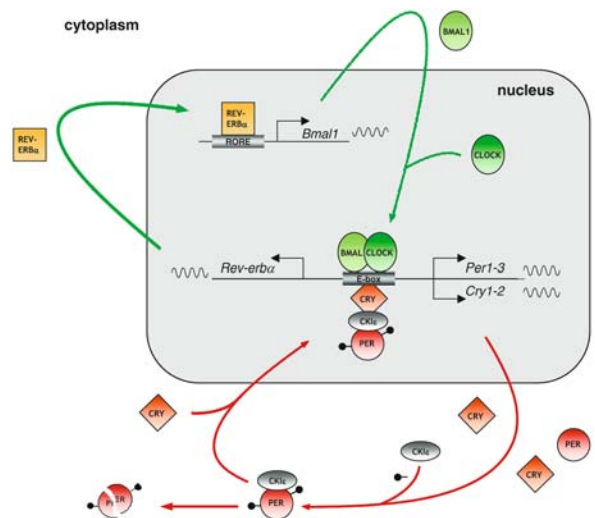
Characteristics

Circadian rhythms are self-sustained, endogenously generated oscillations, i.e. they persist in constant conditions indefinitely without any exogenous time information. The endogenous period of these rhythms is close to but not exactly 24 h. In contrast to other biological oscillations, the circadian period is temperature-compensated, i.e. it is almost invariable over a wide range of temperatures. Circadian rhythms are synchronized (entrained) to the light-dark cycle of the environment by a specialized input pathway (see below) to ensure a stable ► **phase** relation of the endogenous oscillator to the outside world. In mammals, circadian timing is organized in a hierarchy of multiple circadian oscillators. The master clock resides in the suprachiasmatic nucleus (SCN) of the hypothalamus controlling the phase of circadian clocks in other brain areas as well as peripheral organs. In this chapter, the circadian system of mammals is primarily described, for other organisms see (1).

Molecular Interactions

Molecular Basis of Circadian Rhythm Generation in Mammals

A cell-autonomous gene-regulatory network generates circadian rhythms in the mRNA and protein concentrations of clock genes with a delayed negative feedback as the fundamental mechanism (Fig. 1). The basic helix-loop-helix (bHLH) transcription factors CLOCK and BMAL1 activate as a heterodimer the transcription of the *Period* (*Per1*, *Per2*, *Per3*), *Cryptochrome* (*Cry1*, *Cry2*) and *Rev-erb* genes, by binding to ► **E-box** enhancer elements in their promoters. After a delay of several hours, PER and CRY proteins translocate into the nucleus and inhibit CLOCK/BMAL1 transactivation. As a consequence



Circadian Clocks. Figure 1 Schematic view of the gene-regulatory network involving interlocked negative and positive feedback loops that generate molecular circadian rhythms. In a negative feedback loop (red arrows), the transcription factor heterodimer CLOCK/BMAL1 activates the transcription of *Per*, *Cry* and *Rev-erb* genes via binding to E-box elements in their promoters. PER proteins are bound, phosphorylated and thereby destabilized by CK1ε. Complex formation with CRY proteins leads to nuclear accumulation of the PER/CRY/CK1ε, which inhibits the transactivational activity of CLOCK/BMAL1. In a positive feedback loop (green arrows), *Bmal1* transcription is activated (de-repressed) by PER/CRY/CK1ε, because this complex also inhibits REV-ERBa synthesis, which on its part represses *Bmal1* transcription via binding ROR response elements (RORE).

of inhibiting the activators of their own synthesis, PER and CRY levels fall, CLOCK/BMAL1 are no longer inhibited and the cycle restarts.

Not only are the *Per*, *Cry* and *Rev-erb* genes expressed rhythmically, but also *Bmal1* mRNA levels oscillate, with a phase opposite to those of the *Per* and *Cry* mRNAs. This is caused by the orphan nuclear receptor REV-ERBa that rhythmically represses *Bmal1* transcription via binding to ► **ROR response elements** in the *Bmal1* promoter.

The time delay between transcription of *Per* and *Cry* genes and their action as nuclear inhibitors of CLOCK/BMAL1 transactivation is crucial for generating circadian oscillations. Posttranslational processes like phosphorylation, degradation, complex formation and regulated nuclear import and export ensure that the peak of nuclear PER and CRY protein abundance occurs 4–6 h after the peak of their mRNA rhythm. As an example, casein kinase 1ε (CK1ε) binds, phosphorylates and thereby destabilizes PER proteins. Complex

formation with CRY proteins stabilizes PER2 and leads to nuclear accumulation of the PER-CRY-CKI ϵ complex. For a detailed review of current knowledge see (2).

Phenotype of Clock Gene Mutations in Mammals

Most clock gene products are not necessary for viability, but seem to be responsible solely for circadian timekeeping – evidence for clocks being advantageous but not essential to life. Table 1 shows the phenotypes of known mammalian clock gene mutants. PER and CRY proteins are essential for a functional clock, although there is some degree of redundancy in the function of the different PER and CRY family members within the core clock mechanism, which is not yet fully understood. Disruption of either *Per1* and *Per2*

together or both *Cry* genes causes immediate behavioral arrhythmicity when the double-knockout animals are placed in constant conditions. In addition, all molecular rhythms disappear. The same effect is observed, when the *Bmal1* gene is disrupted. Single gene knockouts of *Per* or *Cry* genes do not abolish oscillations immediately, but have an effect on the period of the rhythm.

Circadian Regulation of Transcription

The circadian clock regulates a significant proportion of an organism's entire transcriptome. In cyanobacteria every transcript oscillates with a period of approximately 24 h. In mammals, 6–10% of all transcripts are rhythmically expressed as revealed by global gene chip analysis (3). Most oscillating transcripts are tissue specific, only a few were found to be circadianly expressed in more than one tissue (including the components of the clock core mechanism). Some of the oscillating target genes (e.g. the neuropeptide arginine vasopressin) are directly regulated by CLOCK/BMAL1 *via* E-boxes in their promoters; others are indirectly regulated by rhythmically expressed transcription factors.

Circadian Control of Mitosis and Tumor Development

Cell division in many mammalian tissues including oral mucosa, intestinal epithelium and bone marrow is associated with specific times of day, suggesting an intimate coupling of the circadian clock with the cell cycle. In rapidly proliferating liver cells, the circadian clock controls the expression of cell cycle regulatory genes and thereby modulates the timing of mitosis, e.g. the expression of *Wee-1* is presumably directly regulated by CLOCK/BMAL1 (4). WEE-1 phosphorylates the cell cycle inhibitor CDC2 kinase and thus prevents the liver cells making the transition from the G₂ to the M stage of the cell cycle. Consequently, progression into mitosis is restricted to times of the day when WEE-1 levels are low.

Several epidemiological studies suggest that perturbations of the circadian timing system such as shift work may promote tumorigenesis in humans. In mice, destruction of the SCN accelerates tumor growth and genetic lesions of the *Per2* gene lead to tumor development (5). The mechanisms by which the circadian clock acts as a tumor suppressor are currently unknown.

Regulatory Mechanisms

Clock Cells in the Suprachiasmatic Nucleus

The SCN is a bilateral structure located just above the optic chiasm on the base of the third ventricle in the hypothalamus. It contains about 20,000 neurons, most of which are able to generate circadian rhythms in neuronal activity even when isolated in cell culture.

Circadian Clocks. Table 1 Circadian clocks

Clock gene mutation(s)*	Behavioral phenotype [§]
<i>Bmal1</i>	arrhythmic
<i>Clock</i> [#]	long period, then arrhythmic
<i>Per1</i> (or <i>Per1</i> and <i>Per3</i>)	short period, then arrhythmic
<i>Per2</i> (or <i>Per2</i> and <i>Per3</i>)	short period, then arrhythmic
<i>Per3</i>	short period
<i>Per1</i> and <i>Per2</i>	arrhythmic
<i>Cry1</i>	short period
<i>Cry2</i>	long period
<i>Cry1</i> and <i>Cry2</i>	arrhythmic
<i>Per2</i> and <i>Cry2</i>	rhythmic with normal period
<i>Per2</i> and <i>Cry1</i>	short period, then arrhythmic
<i>Rev-erba</i>	short period
<i>CKIϵ</i> [°]	short period
<i>Per2</i> (<i>Ser 662-Gly</i>) ⁺	family advanced sleep phase syndrome in humans ^{&}

* Loss-of-function mutations in mice unless indicated otherwise.

[§] Activity of homozygous mutants under constant dark conditions unless noted otherwise.

[#] Semi-dominant-negative autosomal mutation resulting in loss of transactivational activity.

[°] Semi-dominant autosomal mutation in hamster resulting in deficiency of CKI ϵ in phosphorylating PER proteins.

⁺ Autosomal dominant point mutation in human *Per2* resulting in less efficient phosphorylation of PER2 by CKI ϵ .

[&] Affected individuals have a 4-hr advanced sleep phase in entrained conditions.

Ablation of the SCN results in complete loss of all overt physiological and behavioral circadian rhythms. Transplantation of fetal SCN tissue into SCN-ablated hosts rescues some behavioral rhythms including locomotor activity rhythms. The genotype of the graft (Table 1) determines the restored endogenous circadian period. The SCN has direct neural and endocrine links to many neuroendocrine and behavioral control systems, especially in the hypothalamus, brain stem and pituitary. These connections regulate many overt circadian rhythms, such as the sleep-wake cycle, core body temperature rhythms, locomotor activity rhythms and rhythmic ►melatonin and corticosteroid secretion.

►Entrainment of Circadian Clocks by Light

To keep the endogenous oscillators synchronized with the outside world, a critical feature is their ability to respond to environmental stimuli. The endogenous period has evolved to be close to but not exactly 24 h to allow the entrainment process to modify the phase of the internal clocks until internal and external time are in congruence. This not only ensures a unique phase relation between internal clock and external light-dark cycle, it also enables an organism to measure day lengths and to adapt its physiology to seasonal changes. Light is the most potent stimulus to reset the mammalian clock in the SCN. While it has no effect on the clock during ►subjective day when light is present in a natural light-dark cycle, a light pulse administered at early ►subjective night causes a phase delay and a light pulse near the end of subjective night causes a phase advance. This differential sensitivity to light is commonly described in a phase-response curve.

Molecular Mechanism of Entrainment by Light

A specialized small fraction of dispersed retinal ganglion cells in the eye that contain the putative circadian photoreceptor melanopsin is intrinsically photosensitive (6). These ganglion cells transmit light signals along the retinohypothalamic nerve tract to SCN neurons. The principal neurotransmitters used are glutamate and pituitary adenylate cyclase-activating peptide (PACAP). Their release leads to a postsynaptic calcium influx that activates several kinase cascades ending up with the phosphorylation of CREB (cAMP response element binding protein). Phosphorylated CREB binds to specific sites in the promoters of the *Per* genes which leads to the rapid induction of their transcription. Light exposure in the early night induces *Per1* and *Per2* transcription, whereas only that of *Per1* is upregulated in the late night. Changes in nocturnal *Per* mRNA levels phase-shift the core clock oscillator.

Peripheral Oscillators

Cells in other brain areas as well as in peripheral organs, such as heart, liver, kidney, lung and skeletal

muscle contain circadian oscillators that regulate local physiology. Even cell-lines in culture such as rat-1 and NIH3T3 fibroblasts are capable of circadian gene expression after stimulation. Current evidence suggests that the SCN serves as a master synchronizer rather than a ►pacemaker and coordinates the phases of the multiple clocks thereby preventing an internal desynchronization (7), as is the case after traveling across time zones (►jet lag). Since peripheral cells of mammals are not light sensitive, they have to be entrained by non-photic cues. Feeding time is the dominant entrainment agent for peripheral clocks, while leaving the clock in the SCN unaffected. It is not yet known, whether the SCN synchronizes peripheral organs directly by humoral or neural means or indirectly by setting time windows for feeding behavior. Experimentally, many substances have been shown to reset clocks in cell culture, such as glucocorticoids, retinoic acid, endothelin-1, forskolin, fibroblast growth factor, calcium ionophores and butyryl cAMP. Some effects triggered by these components activate the same mechanisms as described for photic entrainment in the SCN including phosphorylation of CREB.

Circadian Control of the Sleep-Wake Cycle in Humans

A current model for the generation of the human sleep-wake cycle is the interaction between the circadian clock and a separate oscillatory process, called the sleep homeostat. The sleep homeostat monitors and reacts to the need for sleep, giving rise to the urge to sleep depending on prior amounts of sleep or wakefulness. The circadian clock in the SCN generates a rhythmic wake or arousal signal that increases during the day, peaks in the evening hours and decreases during the night with a minimum at early morning. In the absence of the SCN, the sleep-wake cycle loses its monophasic organization and is presumably solely driven by the sleep homeostat. By opposing the homeostatic increase or decrease in sleep need during the course of the day or night, respectively, the circadian system consolidates a monophasic sleep-wake cycle and ensures a virtually uninterrupted period of sleep at a defined circadian phase.

References

1. Dunlap JC (1999) Molecular bases for circadian clocks. *Cell* 96:271–290
2. Reppert SM, Weaver DR (2002) Coordination of circadian timing in mammals. *Nature* 418:935–941
3. Storch KF, Lipan O, Leykin I et al (2002) Extensive and divergent circadian gene expression in liver and heart. *Nature* 417:78–83
4. Matsuo T, Yamaguchi S, Mitsui S et al (2003) Control mechanism of the circadian clock for timing of cell division *in vivo*. *Science* 302:255–259

5. Fu L, Pelicano H, Liu J et al (2002) The circadian gene *Period2* plays an important role in tumor suppression and DNA damage response *in vivo*. *Cell* 111:41–50
6. Berson DM, Dunn FA, Takao M (2002) Phototransduction by retinal ganglion cells that set the circadian clock. *Science* 295:1070–1073
7. Yamazaki S, Numano R, Abe M et al (2000) Resetting central and peripheral circadian oscillators in transgenic rats. *Science* 288:682–685

two different mutations are in trans during a cis-trans complementation test; but shows wild-type phenotype when the same mutations are in cis. Synonymous with gene, originally defined as a functional genetic unit within which two mutations cannot complement. Now equated with the term gene, as the region of DNA that encodes a single polypeptide (or functional RNA molecule such as tRNA or rRNA).

Circumoral Numbness

Definition

Circumoral numbness refers to absent or reduced sensory perception around the mouth.

► [Hyper-and Hypoparathyroidism](#)

Cis-Acting Elements

Definition

Cis-acting elements are transcription factor binding elements typically present in the promoter region of a gene.

► [Tangier Disease](#)

Cis-Element

Definition

Cis-element is a regulatory portion of a gene sequence that blocks expression of gene elements downstream, or further upstream, along the DNA in the direction of processing by DNA polymerase. Cis elements do not code any proteins, and therefore their actions are limited to the genome on which they are physically located. These elements were discovered by Francois Jacob and Jacques Monod.

► [Alternative Splicing](#)

Cistron

Definition

The term was coined by Benzer for the smallest genetic unit that does not show genetic complementation when

CJD

Definition

Creutzfeldt-Jakob disease is the most common human prion disease, which can be idiopathic as in sporadic CJD (sCJD), hereditary as in familial CJD (fCJD) or acquired as in iatrogenic CJD (iCJD).

► [Prion Diseases](#)

CK2

Definition

CK2 is a highly conserved ► [serine/threonine kinase](#) that is ubiquitously expressed in both the cytoplasm and nucleus of eukaryotic cells.

► [RNA Polymerase III](#)

CKI

Definition

CKI stands for ► [CDK](#) (cyclin dependent kinase) inhibitory protein. Four members of the Ink4 family (p15, p16, p18 and p19) and three members of the Cip/Kip family (p21, p27 and p57) regulate CDKs in mammalian cells.

► [Cell Cycle – Overview](#)

CKI

► [Casein Kinase I](#)

CKII

► Casein Kinase II

c-KIT

Definition

The c-KIT proto-oncogene, also called stem cell factor receptor (SCF-R), encodes a membrane-bound receptor protein tyrosine kinase (RPTK) that belongs to the RTK subclass III family. c-KIT Ligand (KL) or SCF is expressed in haematopoietic, melanogenic and gametogenic precursors and derivatives, and in interstitial cells of Cajal. In haematopoiesis, KL supports the survival and self-renewal of haematopoietic stem cells, synergizes with erythropoietin in stimulating erythroid progenitor cell proliferation, as well as with thrombopoietin in promoting megakaryocyte progenitor cell growth and maturation. Furthermore, KL is a potent enhancer of proliferation, survival, chemotaxis and adhesion of mast cells. The human c-KIT gene is located on chromosome 4q11–12.

► Leukemia

Classification

Definition

Classification, also known as class prediction, displays a type of supervised analysis that tries to find features in the data, which allows the assigning of new objects to particular a priori given classes, e.g. assigning samples to ‘diseased’ or ‘normal’ classes by their expression profiles.

► Computational Diagnostics

► Microarray Data Analysis

Classification of Active Centers

CORNELIUS FRÖMMEL

Faculty of Medicine, Georg August University,
Göttingen, Germany
c.froemmel@med.uni-goettingen.de

Definition

An active center is the small region of a protein which is responsible for the distinct function. There are a large number of synonyms (active site, active center, binding cavity, binding cleft, binding groove, binding pocket; binding site) owing only slightly different meaning.

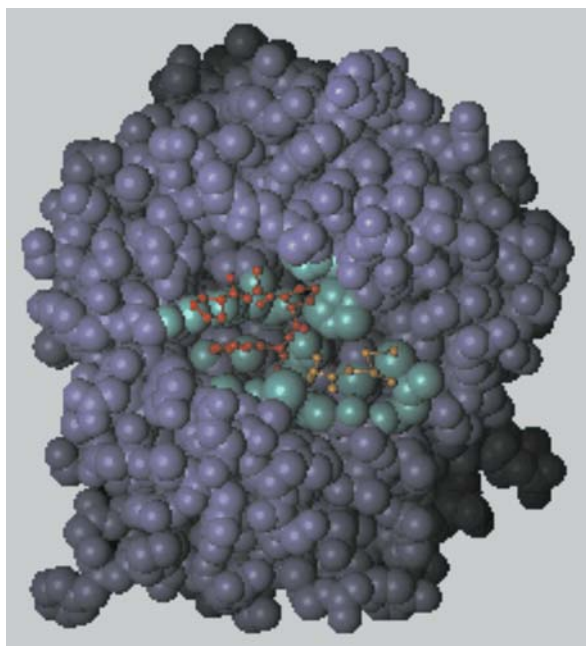
Description

The biological function of proteins typically depends on the structure of specific binding sites. At the binding site the protein molecule interacts specifically with another molecule, the second molecule is referred to as the ligand of this molecule. In other cases the interaction takes place between a protein molecule and electrons and photons (light), respectively. In all examples the site of physical interaction is called the active site. Due to the ligand binding the structure of the protein can be changed (► [induced fit](#)). In some cases only few atoms move, in others one can observe a strong conformational change.

The binding property of a protein typically depends on the presence of a small number of precisely positioned amino acid residues which are brought together by protein folding. They represent a small fraction of the number of amino acids present in the protein. The region of a protein that interacts with another small molecule usually consists of a pocket (cavity, depression, hole) formed by a specific arrangement of amino acids on the protein surface. These sites are located at the surface or show close contact to the water accessible surface of a protein molecule. If the ligand (substrate, effector) is not present, the binding site is (dynamically) filled by the solvent, normally by water molecules and ions. An exception are membrane proteins which show light harvesting/absorbing properties and light emitting protein (e.g. the green fluorescent protein). In this cases the light can cross protein structures and the active site is completely in the inner of the protein molecule. Active sites consist of tens of non-hydrogen atoms. They represent about 10% of the accessible surface. One ► [protein domain](#) contains mostly one binding site for small ligands.

In literal sense ‘active site’ and its synonyms are used to describe the interaction site between any low molecular weight ligand and a given protein. As ligands can serve metal ions (Na^+ , K^+ , Mg^{++} , Ca^{++} , Zn^{++} , $\text{Fe}^{++(+)}$, etc.), inorganic molecules like phosphate, carbonate and all thinkable organic substances. In case of metal ions the binding sites consists of only few residues (2–6 amino acids).

In a narrower sense the term ‘active sites’ is used for the description of the region of enzymes in which one or more substrate molecules are bound and can be transformed to the product (and vice versa). In nearly



Classification of Active Centers. Figure 1 FK506 binding protein (space filling model) with bound ligand (ball and stick model) (3). The FK506 binding protein is an enzyme (cis-trans-prolyl isomerase) catalysing interconversion of to distinct conformations of the peptide bond preceding a prolyl residue. The atoms located deeply in the binding cavity are coloured in green. They are in close contact with the ligand.

almost of the cases it is formed as a small pocket or hole (a depression on the water accessible surface). Non-covalent binding of ligands is the first step as well as of signal transduction for enzyme catalysis. According to the main steps of enzyme catalysis (binding followed by the chemical reaction) one can differentiate in many but not in all cases between the binding region of the active site and the reactive residues in the active site. Latter catalyse the chemical reaction transforming the substrate in the product in a reversible manner. The close neighbourhood of polar amino acid side chains in active centres can influence strongly their reactivity. The resulting reactivity is one molecular reason for the huge catalytic efficiency of enzymes. E.g. the well known triad in serine proteases (Ser His Asp) is responsible for very fast splitting of peptides. The reactivity is caused by spatial order of the three residues in the active side.

Reversibly bound substrates of enzymes are often co-enzymes which are common used to transfer small molecular pieces from one enzyme to another. In different proteins which bind identical co-enzymes one observe frequently that homologues domains are used.

The binding specificity and strength, respectively depends on the number of non-covalent and attracting interactions. Furthermore, only ligands that can bind tightly to a protein are those that fit precisely onto its surface (molecular complementarity). The specificity can be perfect (only one natural ligand is bound), broader (e.g. a distinct type of chemical compounds like sugars are bound) or very broad (all ligands with a distinct property like hydrophobicity are bound). The dissociation constant K_d shows values between 10 mM and 10^{-12} M and less. Structurally related molecules bind often to the same binding site (►molecular mimicry).

The extended interaction site between different protein molecules (mostly long living protein complexes) are not denoted as active site. In contrast in examples in which proteins are partner of signal transducing molecules (mostly short living complexes) or substrates of enzymes the term 'active site' is used to describe the interaction site.

Ligands can bind to a different place as the active site and influence e.g. the activity of enzymes. Such sites are called allosteric site or regulatory site. Due to their influence on enzyme activity they are important for the regulation of the metabolism.

For several of interactions and chemical reactions the set of ►proteinogenic amino acid residues are not suitable. Therefore active sites include further strongly bound compounds called ►prosthetic groups: e.g. to accept hydrogen flavin nucleotide or to bind electrons iron sulphur cluster or haeme. In another example the retinal molecule in rhodopsin is responsible for absorption photons.

Footprint of active sites on ►primary structure

Normally strictly conserved residues in homologues are predicted to be functionally relevant but this is strongly valid only for amino acid residues relevant in catalytic events. They can be found by sequence similarity searching methods. In the binding site there are more degrees of freedom (1).

Footprint of active sites on ►tertiary structure

In most cases the binding sites can be found by purely geometrical based methods (2) because they are small cavities on the protein surface.

Synonyms to 'active site'

- Active site
- Active center
- Binding cavity
- Binding cleft
- Binding groove
- Binding pocket
- Binding site

Therapeutical Consequences

The knowledge about function, specificity and structure of active site is relevant for in silico search of drugs (drug design). The main goal of drug design is to find low molecular weight substances which compete with natural ligands for specific binding (active) sites or allosteric sites to influence enzyme activity or signal transduction and can be used as therapeutics or diagnostics, respectively.

- Molecular Dynamics Simulation in Drug Design
- Structure Based Drug Design

References

1. Casari G, Sander C, Valencia A (1995) A method to predict functional residues in proteins. *Struct Biol* 2:171–178
2. Peters KP, Fauck J, Frömmel C (1996) The automatic search for ligand binding sites in proteins of known three-dimensional structure using only geometric criteria. *J Mol Biol* 256:201–213

Classification of Expression Profiles

- Computational Diagnostics

Clathrin

Definition

Clathrin is the major constituent of the proteinaceous, electron dense ‘coat’ of coated pits and coated vesicles formed from the plasma membrane, and from membranes on the trans- or exit side of the Golgi complex. Clathrin molecules are composed of three heavy and three light chains which stably associate into soluble triskelia. A number of adaptor proteins have been identified that recognize sorting signals, but also bind to clathrin and regulate the multimerization of individual triskelia into polyhedral cages. The polymerization of clathrin and adaptor proteins is thought to drive budding of a coated lipid bilayer. Clathrin-coated buds form at the plasma membrane and at the *trans*-Golgi network. Distinct flat clathrin lattices have been observed on endosomal membranes.

- Exocytotic Pathway
- Protein and Membrane Transport in Eukaryotic Cells
- Vesicular Traffic

Cl_{Ca}

Definition

Cl_{Ca} refers to a calcium activated chloride channel that is strongly expressed in many ion-transporting epithelia, including the kidney.

- Autosomal Dominant, Polycystic Kidney Disease

CldU

- 5–Chlorodeoxyuridine (CldU) & 5–Bromodeoxyuridine (BrdU)

Cleavable Cross-Linkers

Definition

Chemical cross-linking reagents that contain two or more reactive groups. These are connected by a spacer group containing molecules that can be cleaved by specific chemical reagents.

- Protein Interaction Analysis: Chemical Cross-Linking

Cleavage and Polyadenylation Specificity Factor

Definition

Cleavage and polyadenylation specificity factor (CPSF) refers to a protein complex that is involved in 3′ processing of pre-mRNA. It consists of at least four subunits, binds the AAUAAA signal, and is required for both cleavage of the pre-mRNA and polyadenylation.

- Polyadenylation

Cleavage Factor I (CF I)/Cleavage Factor II (CF II)

Definition

Cleavage factor I (CF I)/Cleavage factor II (CF II) designate protein complexes that are involved in 3′

processing of pre-mRNA. There are likely to be different versions of CF I, consisting of a common small subunit and any one of a number of related large subunits. CF I binds RNA, and CF II contains at least two subunits. Both are only required for cleavage of the pre-mRNA.

►Polyadenylation

Cleavage Stimulation Factor

Definition

Cleavage stimulation factor (CstF) refers to a protein complex that is involved in the 3' processing of pre-mRNA. It consists of three subunits, binds the downstream element of the 3' processing signal, and is only required for cleavage of the pre-mRNA.

►Polyadenylation

Cleft Lip Palate

MARY L. MARAZITA

Craniofacial and Dental Genetics, School of Dental Medicine, University of Pittsburgh, Pittsburgh, Pennsylvania, USA
marazita@sdmgenetics.pitt.edu

Synonyms

Orofacial clefts (OFC); Oral-facial clefts

Related Terms

- Cleft lip with or without cleft palate
- Cleft lip
- Cleft palate
- Hare-lip (archaic)
- Primary palate
- Secondary palate

Definition

Orofacial clefts refer to any fusion failure of the structures of the face and oral cavity. Clefts of the lip and palate are by the far the most common such anomalies, although clefts of other parts of the face may occur. Embryologically, orofacial structures develop in roughly two stages with development of the anterior structures followed by development of the posterior

structures. The anterior structures (the lip and the maxillary alveolus) are often referred to as the primary palate, with the posterior structures (the hard palate and the soft palate) referred to as the secondary palate. Orofacial clefts may be unilateral or bilateral, complete or incomplete, and may involve either or both of the primary and secondary palates. Thus, there is a great variety in the types of possible orofacial clefts, ranging from very minor defects such as a notch in the vermillion border or a cleft uvula, to a complete bilateral cleft of the primary and secondary palates. In etiological studies, orofacial clefts are usually grouped into either cleft lip with or without cleft palate (CL/P) or cleft palate alone (CP). In an individual, orofacial clefts may be an isolated anomaly or may occur as part of a ►syndrome.

Characteristics

Orofacial clefts can occur as part of Mendelian syndromes, as part of the phenotype resulting from chromosomal anomalies, or as the result of prenatal exposure to certain ►teratogens. Orofacial clefts demonstrate remarkable differences in frequency by gender and laterality. There is an approximate 2:1 ratio of males:females for CL/P, although slightly more females than males have CP. Within unilateral clefts, the ratio of left-sided to right-sided clefts is also about 2:1. Orofacial cleft birth prevalence shows a wide range, from about 1/500 births to about 1/2000 depending on population; in general, Asian and Amerindian populations have the highest frequencies and African-derived populations the lowest.

Over 300 syndromes exist in which orofacial clefts are part of the phenotype; about half of those are due to ►Mendelian inheritance of alleles at a single genetic locus. Much progress has been made in recent years in delineating Mendelian disorders and in gene discovery of such disorders (refer to ►OMIM on the ►NCBI web site for a catalog of such disorders). However, only a small proportion of individuals with orofacial clefts has



Cleft Lip Palate. Figure 1 Asian twin pair with unilateral cleft lip plus cleft palate.

Cleft Lip Palate. Table 1 Summary of genes and regions potentially involved in nonsyndromic orofacial clefts

Chromosomal Region	Candidate genes/regions ¹		Genome Scan ²	Notable Syndrome (MIM#, gene) or MIM designation (MIM#) ³	Chromosomal Rearrangements ⁴	Animal models ⁵
	Gene	L/A	L/A/M			
1p36-31	MTHFR	L,A	L		Deletion 1p36	SKI1, LHX8: K/O, E
1q32	IRF6	A	M	Van der Woude (MIM 119300, IRF6)	Deletion 1q Micro-deletion 1q32-41	IRF6: E
2p13	TGFA	A	L, M	OFC2 (MIM 602966)		TGFA: E
2q35			L, M			
3p25			A, M		Duplication 3p26-21	
3q26			L	Ectrodactyly-ectodermal dysplasia-clefting 3 (MIM 604292, TP63)		TP63: K/O, E
4p16	MSX1	L,A	A		Deletion 4p (Wolf-Hirschorn syn)	MSX1: K/O, E
4q31		L	L		Deletion 4q31-35	
5p15			L, A		Deletion 5p (cri du chat syndrome)	
6p23	F13A1	L,A	L, M	OFC1 (MIM 119530)		TFAP2A: K/O, E
6q25			L, M			
7p13			A, M			
7q21			L, A	Ectrodactyly-ectodermal dysplasia-clefting 1 (MIM 12990)		
8p21			L, M		Deletion 8p23	
8q23			L, M			
9q21			L, M	Basal cell nevus (MIM 109400, PTCH)		
10q25			L		Duplication 10p15-11	
11p12	PVRL1		L,A	Margarita Island (MIM 225060, PVRL1)		
12p11			L,A			
14q12	TGFB3	A	L, M			TGFB3: K/O, E
15q22			L, M			
16q			L		Trisomy 13	
17p11			L			
17q12	RARA	A	M			KCNJ2: K/O
18q23			M		Trisomy 18 Deletion 18q23	

Cleft Lip Palate. Table 1 Summary of genes and regions potentially involved in nonsyndromic orofacial clefts (Continued)

Chromosomal Region	Candidate genes/regions ¹		Genome Scan ²	Notable Syndrome (MIM#, gene) or MIM designation (MIM#) ³	Chromosomal Rearrangements ⁴	Animal models ⁵
	Gene	L/A	L/A/M			
19q13	APOC2/ BCL3	L,A	L	OFC3 (MIM 600757)	Translocation 19q13.3	
20p12			L,A			
Xq21			L	X-linked CP with ankylo- glossia (MIM 303400, TBX22)		TBX22: E

Candidate genes/regions: genes and regions with at least two positive reports of linkage (L) or association (A) in the literature. Genome scan: regions with positive linkage (L) or association (A) results with anonymous markers spaced ≤ 10 cM apart throughout the genome (one or more genome scans); M=positive meta-analysis results over all genome scans (4).

Notable syndrome or MIM designation: listed here are Mendelian syndromes that are mapped to regions with positive candidate gene or genome scan results. Also listed in this column are the MIM designations that have been assigned for three regions (OFC1, 2, 3). MIM# = identification number in OMIM.

Chromosomal rearrangements: regions of chromosomal rearrangements that include clefts in their phenotypic expression

Animal models: genes investigated in animal models with phenotypes that include clefting. K/O = knockout, E=expression studies.

a known etiology (3, 6). Most orofacial clefts are nonsyndromic and are considered **complex** traits. Given the public health importance of orofacial clefts, many etiologic studies have been conducted of nonsyndromic orofacial clefts.

The importance of inheritance in the etiology of cleft lip and cleft palate has been noted by scientists for more than 200 years. The first published description of a family with several affected members was in 1757 (7). Charles Darwin (1) pointed out a publication of “the transmission during a century of hare-lip with a cleft-palate”. Since those early publications, many studies have been undertaken using the tools of statistical and molecular genetics.

The **multifactorial threshold** model was proposed to explain many of the features of nonsyndromic orofacial clefts (such as the altered gender ratio); however, the predictions of that specific model could be rejected when tested in several populations. In the early years of the 20th century, several seminal works were published regarding the inheritance patterns of orofacial clefts, but until recently progress has been slow in determining the exact genes involved. **Segregation analyses** and statistical analyses of familial **recurrence risk** patterns were consistent with hypotheses of either **major locus** involvement, and/or relatively few loci [3–14; see (3, 4)] interacting to cause orofacial clefts. With statistical evidence that orofacial cleft family patterns were consistent with genetic inheritance, several groups began **linkage** and association studies in order to identify the genes contributing to the familiarity of orofacial clefts.

Cellular and Molecular Regulation

Our understanding of the developmental biology of orofacial clefting (both spontaneous and teratogen-induced) in the mouse model is under rapid development, and several pathways have been delineated [see (5) and (8) for reviews]. Our understanding of human orofacial development is less advanced, in part because many details from the mouse model do not seem to translate directly.

Instead, most researchers in human orofacial cleft etiology are attempting to identify susceptibility genes by linkage and association studies. Any identified loci would then provide insights into the human orofacial developmental process. Table 1 summarizes the results found to date [see (3, 4, 5, 8) for additional details] from linkage and association studies of candidate genes, from genome scans, and from chromosomal rearrangements.

Also summarized in the table are several Mendelian syndromes in which some of the affected individuals present as phenocopies of non-syndromic clefting: e.g. van der Woude syndrome, ectrodactyly-ectodermal dysplasia/clefting, basal cell nevus, and others. Each of these syndromes might have mutations in particular domains that could result in or modify non-syndromic CLP. For example, IRF6, the gene identified for van der Woude syndrome also shows strong **allelic association** with CL/P (although not CP, (9)).

A recent CL/P genome scan meta-analysis found genome-wide significant results for 15 regions (4). The chromosomal positions with the most consistent evidence for involvement in orofacial clefting

include 1q, 2p, 2q, 3q, 4p, 4q, 6p, 14q, 17q and 19q (see Table 1). It is important to note that, like many other complex disorders, the results have not been consistent across all studies. Further progress is most likely to be made by using meta-analysis methods to combine statistical results across studies in order to set priorities as to genes and/or regions to investigate by fine mapping, sequencing and other molecular genetic techniques.

Clinical Relevance

Clefts of the lip and palate are the most common craniofacial birth defects and are also among the most common of all birth defects, with birth prevalence ranging from 1/500 to 1/2000 depending on the population. While the severity of orofacial cleft anomalies varies, multidisciplinary treatment is often necessary and may include craniofacial surgery, specialized dental and orthognathic treatment, speech and hearing intervention, and educational, psychological and social assessment and intervention. The multidisciplinary nature of cleft care was realized even in the first recorded surgical repair of a cleft lip (in the annals of the Chin dynasty in China, about A.D. 390; translated by Khoo (2)—detailed post-operative instructions were listed for optimal results).

Orofacial clefts represent a significant public health problem due to the significant lifelong morbidity and complex etiology of these disorders. The extensive psychological, surgical, speech, and dental involvement emphasize the importance of understanding the underlying causes of CL/P, in order to optimize treatment planning, to predict the long-term course of any affected individual's development, to improve recurrence risk estimation, and to provide pre-reproductive counseling.

References

1. Darwin C (1845) The variation of animals and plants under domestication. 2nd edn, vol 1. Appleton, New York, p 466
2. Khoo B-c (1966) An ancient Chinese text on a cleft lip. *Plast Reconstr Surg* 38:89–91
3. Marazita ML (2002) Genetic etiologies of facial clefting. In Mooney MP, Siegle MI (eds) *Understanding Craniofacial Anomalies*, Wiley-Liss Inc., New York, pp 147–161
4. Marazita ML, Murray JC, Lidral AC et al (2004) Meta-analysis of 13 Genome Scans Reveals Multiple Cleft Lip/Palate Genes with Novel Loci on 9q and 2q. *Am J Hum Genet* 75:161–173
5. Mooney MP, Siegle MI (2002) *Understanding Craniofacial Anomalies*, Wiley-Liss Inc., New York
6. Murray JC (1995) Invited editorial: Face facts: Genes, environment, and clefts. *Am J Med Genet* 57:227–232
7. Trew CJ (1757) Sistens plura exempla palati deficientis. *Nova Acta Physico-Medica Academiae caesariae Leopoldion-Carolinae* 1:445–447

8. Wyszynski DF (2002) *Cleft Lip and Palate, from Origin to Treatment*. Oxford University Press, New York
9. Zuccherro TM, Cooper ME, Maher BS et al (2004) Interferon regulatory factor 6 (IRF6) gene variants confer risk for isolated cleft lip and palate. *N Engl J Med* 351:769–780

Cleidocranial Dysplasia

Definition

Cleidocranial dysplasia is characterized by mutation in RUNX2. It is a transcription factor essential for osteoblast differentiation, hypoplastic clavicles, open fontanelles, supernumerary teeth and short stature.

► **Bone Disease and Skeletal Disorders, Genetics**

Clinical Bioinformatics

► **Computational Diagnostics**

Clinical Gene Transfer

KLAUS CICHUTEK, MATTHIAS SCHWEIZER, CHRISTIAN BUCHHOLZ, EGBERT FLORY
Paul-Ehrlich-Institute, Division of Medical Biotechnology, Langen, Germany
cickl@pei.de

Definition

► **Medicinal products**, also termed drugs or medicines, are used for the treatment, *in vivo* diagnosis or prevention of diseases in or on human subjects. Gene therapy and somatic cell therapy using genetically modified cells are best summarized under the term clinical gene transfer and involve the treatment of human subjects with gene transfer medicinal products, the latter defined in § 4 of the German Drug Law as follows.

Gene transfer medicinal products (GT-MPs) are medicinal products

- which consist of or contain viral or non-viral vectors or nucleic acid used for genetically modifying human somatic cells *in vivo* (*in vivo* strategy)
- or which consist of *ex vivo* genetically modified autologous, allogeneic or xenogeneic cells (*ex vivo* strategy)
- or which consist of replication competent, recombinant microbes used for other purposes than prevention or therapy of the infectious disease they cause (*in vivo* strategy)
- and which are used for treatment, *in vivo* diagnosis or prevention of diseases in human subjects.

This definition is in agreement with the variety of GT-MPs covered by the current European Note for guidance on the quality, preclinical and clinical aspects of gene transfer medicinal products (CPMP/BWP/3088/99). Viral vectors (replication-incompetent viruses able to transfer a therapeutic gene), non-viral vectors (DNA encompassing the therapeutic gene mixed with ►transfection reagents), “naked” DNA or conditionally replicating microbes (recombinant adenovirus, recombinant Salmonella) are currently used for genetically modifying a cell *in vivo* or *ex vivo*.

The application of antisense nucleic acid itself which does not encompass a gene and which is used without the intention of genetically modifying cells, e.g. ►antisense RNA for the treatment of cytomegalovirus-induced retinitis, is not considered to belong to the field of clinical gene transfer, but rather to molecular medicine in general.

Characteristics

During clinical gene transfer, nucleic acids encompassing one or several (trans)genes are transferred into cells in order to restore normal cellular functions or to add new ones (1). The nucleic acid construct (encompassing promoter, enhancer and other regulatory sequences, one or more cDNAs including a polyadenylation signal and a coding region or a region specifying an RNA product) is termed an ►expression construct, ►expression vector or transfer vector. The ►transgene product is normally a protein with a specific function or an antigen. It may also be a specific RNA transcribed from a transgene, e.g. in ribozymes which mediate inactivation of a target RNA within the modified cell.

The vehicle (also termed viral or non-viral vector or ►vector particle) which allows transfer of the expression vector or nucleic acid (and thereby the transgene(s)) into cells determines the structure of the expression vector and can be of viral or non-viral origin. Viral vector particles, non-viral vectors and free nucleic acids (e.g. naked DNA) which are currently in clinical use are summarized in Table 1 (7). ►Plasmid DNA used as naked DNA or as part of a non-viral vector is

produced from *E. coli* bacteria. Viral vectors are generally produced by mammalian packaging cells in culture. Retroviral and lentiviral transfer vectors can be delivered by using the respective vector particles enveloped with homologous retro- or lentiviral envelope proteins. If heterologous proteins such as the vesicular stomatitis virus G-protein (VSV-G) or heterologous retro- or lentiviral envelope glycoproteins are used as envelope proteins, such vector particles are termed pseudotype vectors. If naked nucleic acids or non-viral vectors (5) are applied, a specific transfer or transfection method rather than a vehicle is used for delivery of the expression construct, which is normally encompassed by a bacterial plasmid. Naked DNA inoculated intramuscularly or intra-dermally is spontaneously taken up by mammalian cells *in vivo*. Delivery can be improved by using instead of needle inoculation, a gene gun (gold particles coated with the plasmid DNA) or a jet gun (plasmid/helium gas under pressure). Future vector applications will involve targeting genes or their expression into specific cell types or tissues (2).

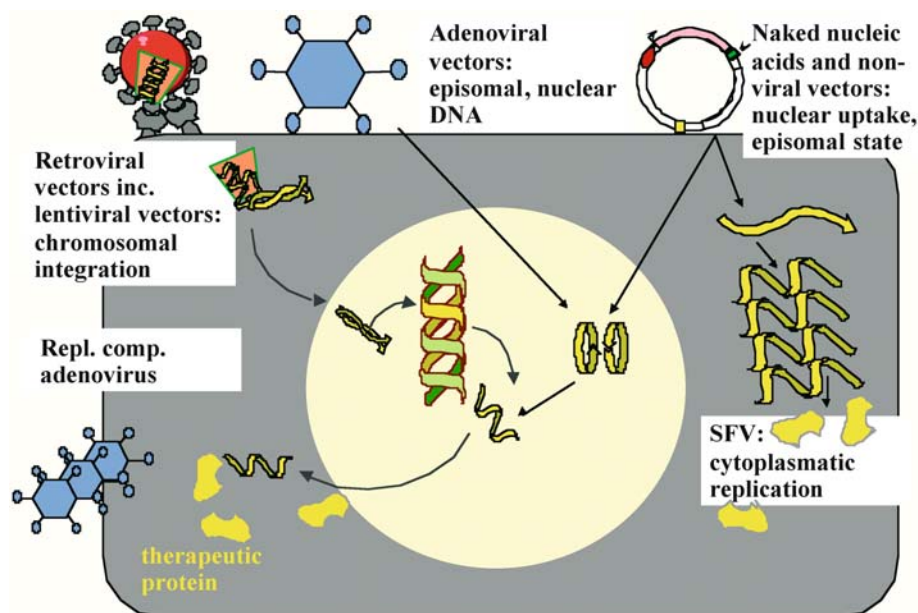
Clinical gene transfer may follow the *in vivo* or the *ex vivo* strategy (6). Examples of *in vivo* strategies include intra-muscular injections of non-complexed, so-called naked DNA, intra-tumoral inoculation of non-viral vectors, infusion of adenoviral vectors *via* the intra-hepatic artery or intra-muscular injection of adeno-associated virus (AAV) vectors. The transgene is thus delivered to human somatic cells *in vivo* and subsequently expressed. Any ►transduction of germ line cells is considered to be a risk which has to be avoided. This risk is assessed by preclinical experiments in animals and suitable biomonitoring during clinical trials.

During *ex vivo* gene transfer, mammalian cells which have been genetically modified in cell culture are delivered to humans. The ►genetic modification may involve viral or non-viral vectors, whatever is considered most suitable. Human autologous cells such as modified CD34-positive cells, encompassing hematopoietic stem cells or progenitor cells, may give rise to a large number of modified daughter cells *in vivo* which may have a long life span. Allogeneic cells may be used as tumor vaccines and are then irradiated before use and readily deleted *in vivo*. They are used in order to induce an immune response against residual tumor cells and their presumed or defined tumor-specific antigens. The use of xenogeneic primary cells in humans is not considered safe because of the theoretical risk of transmitting xenogeneic infectious agents to humans and will therefore not be used in the foreseeable future. A border line application has been the *in vivo* inoculation of genetically modified retroviral packaging cells. These xenogeneic cells produced retroviral

Clinical Gene Transfer. Table 1 Viral vectors currently in clinical use

Origin of vector or carrier nucleic acid	Type	Chromosomal integration	Replication competence	Heterologous envelope proteins
Plasmid DNA	naked (non-complexed)	no (i.m.)	no	not applicable
	liposome or other transfection agent	mostly no	no	not applicable
Artificial chromosome		no	no	not applicable
Retrovirus	MLV	yes	no	yes
	Lentivirus	yes	no	yes
Adenovirus	deletions in E1, E3 or E4, E2ts, or combinations	no	no	no
	guttled (fully deleted)	no	no	no
	no transgene	no	yes	no
Adeno-Associated Virus	wt	yes	no	no
Poxvirus	MVA	no	no	no
	ALVAC	no	no	no
	Vaccinia	no	yes	no
Alphavirus	Semliki Forest Virus (SFV)	no	no	no
Herpes virus	Herpes simplex virus	no	no	no

MLV, Murine Leukemia Virus; MVA, Modified Vaccinia Ankara; ALVAC, Avian pox virus vector



Clinical Gene Transfer. Figure 1 Genetic modification of a cell by free, so-called naked nucleic acid, non-viral vectors or commonly used viral vectors. Informative examples are illustrated. SFV, Semliki Forest Virus.

vectors *in vivo* which were supposed to deliver suicide genes, e.g. the Herpes simplex virus-derived thymidine kinase gene, into human tumor cells in the brain. After subsequent prodrug (ganciclovir) treatment, the delivered genes in the tumor cells and in the packaging cells were able to convert the prodrug into a cell-toxic drug at local sites, thus killing the transgene containing cells *in vivo* and the cells directly in their neighborhood (suicide gene transfer). Examples for recombinant micro-organisms used in clinical gene transfer are replication competent adenoviruses not carrying any transgenes. They are inoculated, e.g., into local head-and-neck tumors and, in conjunction with conventional chemotherapy, local tumor regression has been observed.

Clinical Relevance

Clinical Gene Transfer (Overview)

A variety of clinical trials currently show promising results (see Table 2). It is becoming clear that each disease needs the development of a particular gene transfer method and regimen. Standard use of the first licensed gene transfer medicinal products can be foreseen within the next few years.

More than 7,000 human subjects have probably been treated with gene transfer medicinal products worldwide, most of them in the USA. Within Europe, probably the highest number of clinical gene transfer studies were planned or have been carried out in Germany with a total of about 300 patients that have been treated ([►http://www.pei.de](http://www.pei.de); [►http://www.zks.uni-freiburg.de/dereg.html](http://www.zks.uni-freiburg.de/dereg.html)). A description of regulatory aspects of clinical gene transfer in Germany is available ([►http://www.pei.de/themen/themlink.htm](http://www.pei.de/themen/themlink.htm) → Genterapie). A general perspective of the field's attempts can be obtained from websites ([►http://www.wiley.co.uk/genetherapy](http://www.wiley.co.uk/genetherapy); [►http://www4.od.nih.gov/oba/rac/clinicaltrial.htm](http://www4.od.nih.gov/oba/rac/clinicaltrial.htm)).

Most clinical gene transfer protocols are aimed at the treatment or prevention of cancer, cardio-vascular disease, infectious disease such as AIDS or monogenic disorders. The vectors most often used *ex vivo* are MLV-derived oncoretroviral vectors, whereas adenoviral and poxviral vectors have mostly been used *in vivo*. An increasing number of studies involves the use of non-viral vectors or naked DNA.

Gene Therapy of Monogenic Diseases

The idea of gene therapy was born with the thought that defective genes could be replaced with functional ones in order to treat genetic diseases. A mutation of the gamma-c-chain gene underlies the X-linked form of the severe combined immunodeficiency syndrome (SCID-X1). As a result, immunologically relevant receptors like the interleukin (IL)-2 receptor are not functional and T and B lymphocytes are not developed, due to an early differentiation block. Newborn babies suffering from SCID-X1 therefore have a severely impaired immune system and a limited life expectancy, unless HLA-identical bone marrow transplantation is available. In an attempt at clinical gene therapy, autologous CD34-positive cells retrovirally transduced with the functional gamma-c chain gene were reinfused. As a result, functional T and B cells developed, and for a time period of more than a year, on nine of eleven treated patients, mostly less than one year old, beneficial treatment was achieved. This was the first established success of gene therapy (3). For how long the correction will prevail remains to be seen. If the earliest blood stem cells able to bring about life-long repopulation were genetically modified, the immunological correction would be expected to last life-long. However, a repeated treatment by gene therapy would also be possible.

Clinical Gene Transfer. Table 2 Examples of clinical gene transfer approaches currently showing promising results

Disease	Therapeutic gene	Administration route/vector	Benefit
SCID-X1	<i>Gama-c-chain (IL-2R)</i>	Blood stem cells/ MLV vector	15 babies benefited, 18 treated, 3 leukemias
PAOD	VEGF	i.m./plasmid DNA	Improved vascularization
Head and neck tumors	Cell-dep.repl.adeno-virus, no transgene	Tumor cells	Local tumor regression
Leukemia, graft versus leukemia effect, GvH disease treatment	<i>HSV-tk</i>	T cells/MLV vector	Successful GvH disease treatment
Hemophilia B	Factor IX	i.m./AAV vector	Improved plasma levels

PAOD, Peripheral Artery Occlusive Disease; SCID, Severe Combined Immunodeficiency Disease; i.m., intra-muscular administration

The treatment of hemophilia B is attempted by using mini-cDNAs of the human factor IX gene that allow functional reconstitution of the blood clotting system. On-going trials using intra-muscular injection of adeno-associated virus have shown that detectable increases in the serum level of factor IX can be achieved. However, ongoing clinical trials will have to achieve a further increase in factor IX levels, the time span of clinical benefit and the absence of serious adverse reactions when repeated AAV injections are performed.

Other monogenic disease targets include enzyme or protein deficiencies such as Gaucher disease or immune deficiencies such as adenosine deaminase deficiency (ADA) or chronic granulomatous disease (CGD).

Cancer Gene Therapy

Cancer is a target for various kinds of innovative treatments aiming either at molecular mechanisms underlying the malignant cell transformation, at tumor cell ablation or at enhancing the immune functions able to deplete tumor cells including metastases. Gene therapy using tumor suppressor genes has been attempted, and has progressed to later phases of clinical development. Direct tumor cell ablation by local injection of replicating adenoviral vectors in head and neck cancer has provided local tumor regression, particularly when used in conjunction with chemotherapy. Mutant herpes viruses have also been used in brain tumor injections to lyse the tumor cells *in vivo*. Here, the replicating or conditionally replicating viruses seem to provide the advantage of spreading to some extent *in vivo*, which allows improved tumor cell killing. By such methods, the so-called needle-track transduction observed, for example, during intra-tumoral injection of replication incompetent adenoviral vectors carrying the tumor suppressor gene p53 has been avoided. The latter strategy was aimed at inducing tumor cell killing by apoptosis.

Attempts to enhance the immune recognition of tumors have been made by injecting vectors intra-tumorally which carry foreign MHC genes such as B7.1 or B7.2 or cytokine or interleukin genes such as IL-2 or GM-CSF with little success. In addition, autologous or allogeneic tumor cells modified with immune enhancing or tumor-specific antigen genes were also used as therapeutic tumor vaccines. Here, the only hopeful trial has been a phase I trial where autologous tumor cells were adenovirally transduced with GM-CSF and re-injected. It will take further improvements to lead such attempts to a success.

Clinical Gene Transfer in Cardio-Vascular Disease

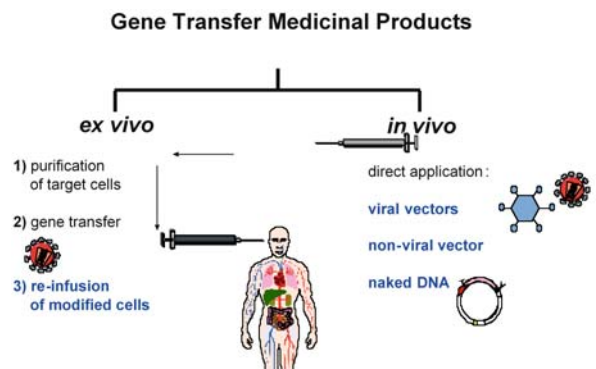
Local intra-muscular injection of plasmid DNA encompassing a gene able to potentiate adaptation to hypoxia was shown *in vivo* to lead to improved

vascularization of ischemic tissue. Vascular endothelium growth factor (► **VEGF**) or fibroblast growth factor (FGF) genes have been directly injected at local sites by catheter application or needle-mediated intra-muscular injection of plasmid DNA or adenoviral vectors. This approach is aimed at the treatment of critical limb ischemia or myocardial vascularization problems. Formation of new micro-circulatory vessels has been observed and clinical trials of phase III will have to show objectively curing or improvement of disease.

Restenosis following balloon catheter-mediated cardiac vessel dilatation is probably caused by the proliferation of smooth muscle cells following injury of the vessel endothelium. Adenoviral or plasmid DNA-mediated transfer of the inducible nitric oxide synthase (iNOS) gene are thought to reduce muscle cell proliferation and to induce apoptosis. Adenoviral vectors or non-viral vectors have been used for gene transfer and efficacy will have to be shown.

Prevention and Therapy of Infectious Disease

Chemotherapy of infectious diseases such as AIDS has become a true therapeutic option during the last five to ten years. Combinations of powerful small molecules aiming at inhibiting crucial steps in the replication cycle of HIV-1 have proven to reduce the virus load to sometimes undetectable levels. Due to the necessity for long-term treatment and massive side effects connected with this, gene therapy of HIV infection is an option that still has to be developed. Retroviral transfer of HIV inhibitory genes into peripheral lymphocytes or CD34-positive human cells *ex vivo* has been attempted, so far with very little success. The genes used expressed (i) decoy RNA with multiple Rev or Tat Responsive Elements, so-called poly-TAR or poly-RRE sequences, (ii) scFv able to mediate intra-cellular inactivation of viral gene products, (iii) trans-dominant negative mutants of viral proteins such as RevM10 or (iv) ribozyme RNA able to intra-cellularly cleave viral



Clinical Gene Transfer. Figure 2 *Ex vivo* and *in vivo* application of gene transfer medicinal products (GT-MPs; printed bold).

RNA. Further genes under development will aim at preventing entry or integration of HIV. It will have to be shown whether such therapeutic interventions will provide a therapy option compared to combinations of new chemical or biotechnological drugs.

Prevention of infectious disease is best provided by prophylactic vaccines. Attempts with naked DNA or vectored vaccines to prevent infection of animals with a variety of micro-organisms have been reported. Clinical gene transfer mostly aims at developing vaccines against HIV-1 infection, malaria, hepatitis B and influenza A virus infections. It has been reported that pox virus vectors such as ALVAC or MVA in combination with naked DNA as a priming method and possibly followed by booster injection of subunit vaccines may be most effective in humans. Clinical trials to establish the induction of a suitable immune response and possibly leading to some extent to prevention of infection are on-going.

Clinical Gene Transfer for the Treatment of Other Diseases

Clinical gene transfer may also be applied to other diseases where a gene transfer approach seems promising, although the contribution of genetic factors to the disease has not been clearly defined. ► **Rheumatoid arthritis**, for example, may benefit from a reduction of the inflammatory reactions in the joints which is presumably caused by a cascade of events leading to over-expression of a variety of proinflammatory cytokines. Monoclonal antibodies able to reduce the amount of free tumor necrosis factor alpha (TNF) have been successfully applied to severe cases. Although TNF may not be the initial cytokine priming the inflammatory reactions observed, reduction of its local concentration is obviously able to reduce inflammation and local pain. Clinical gene transfer involves the expression, e.g., of interleukin-1 receptor antagonist by synovial cells in the joints. This is accomplished either by local application of adenoviral vectors or by *ex-vivo* transduction of explanted autologous synovial cells and reimplantation. Clinical gene transfer is applicable to all human diseases where a genetic approach seems potentially helpful. In most cases, it provides for the most natural expression of therapeutic molecules which is the production by human somatic cells *in vivo*.

Current Expectations Concerning Clinical Gene Transfer

Unfortunately, three of the successfully treated SCID-X1 patients (see above) so far developed a leukemia-like lymphoproliferative disorder about three years after treatment. Current state of knowledge led to the interpretation that the leukemia-like diseases were a consequence of chromosomal integration of the retroviral vector. Within the leukemic cells of all three

patients, the vector encompassing the therapeutic gamma c-chain gene was found integrated within the LMO-2 locus and led to over-expression in the leukemic cells, probably *via* a so-called enhancer mechanism. Additional vector-mediated gene activations were found in the third leukemia case. Translocations involving the LMO-2 gene have been found to be associated with some naturally occurring human leukemias. Additional genetic events promoting tumor development may include the influence of the therapeutic gamma c-chain gene on intra-cellular signal transduction and/or differentiation and chromosomal abnormalities that have been observed in the leukemic cells, but have not yet been functionally defined. The latter may have arisen during the massive *in vivo* expansion of the genetically corrected hematopoietic progenitor cells in an almost fully depleted lymphoid cell compartment characteristic for SCID-X1. Depending on whether tumors will be observed as a consequence of retroviral vector use in gene therapy of other diseases, the leukemia development may be classified as being associated with the specific SCID-X1 gene therapy or with other clinical retroviral treatment strategies.

Following infusion of more than 10^{11} adenoviral vector particles per kg into the intra-hepatic artery one of 17 patients treated with the aim of correcting the monogenic disease OTC (ornithine transcarbamylase deficiency) died. Although analysis of the trial led to the detection of shortcomings in good clinical practice, it is believed that the systemic use of a very high number of adenoviral vector particles led to a toxic reaction in the patient Gelsinger which ultimately was the cause of the lethal outcome of the treatment. Clinical trials involving the systemic use of adenoviral particles are therefore carried out only by using a lower number of adenoviral particles measured with the help of adenovirus reference material available. Non-systemic uses of adenoviral particles are not considered unsafe.

Fortunately, a current assessment of the clinical perspectives of gene therapy can be based on an increasing range of clinical experiences. It is no longer believed that gene therapy will replace conventional therapies in a large number of diseases. However, gene therapy will provide valuable therapies in diseases, where current treatment is insufficient or only palliative. It may also provide new therapy options for widespread diseases, but each disease will necessitate the development of a particular molecular and gene transfer strategy. The improvement of gene transfer efficiency alone will not suffice. With the improvement of the clinical gene transfer approaches, the frequency of known serious adverse reactions will increase and theoretical risks will be monitored as real ones, as observed during the development of other novel

therapies. Clinical gene transfer still holds the promise of giving us a new tool to perform evidence-based medicine using current knowledge of the molecular biology underlying a disease.

References

1. Balicki D, Beutler E (2002) Gene therapy of human disease. *Medicine (Baltimore)* 81:69–86
2. Buchholz CJ, Stitz J, Cichutek K (1999) Retroviral cell targeting vectors. *Curr Opin Mol Ther* 1:613–621
3. Cavazzana-Calvo M, Hacein-Bey S, Yates F et al (2001) Gene therapy of severe combined immunodeficiencies. *J Gene Med* 3:201–206
4. Friedmann T, Roblin R (1972) Gene therapy for human genetic disease? *Science* 175:949–955
5. Li S, Ma Z (2001) Nonviral gene therapy. *Curr Gene Ther* 1:201–226
6. Pfeifer A, Verma IM (2001) Gene therapy: promises and problems. *Annu Rev Genomics Hum Genet* 2:177–211
7. Robbins PD, Ghivizzani SC (1998) Viral vectors for gene therapy. *Pharmacol Ther* 80:35–47

Clinical Marker Gene

Definition

A gene is a good clinical marker if the distributions of its expression values in two diagnostic entities can be well separated.

► [Computational Diagnostics](#)

Clonal Analysis

Definition

Clonal analysis designates the follow up of the fate or the developmental potential of single cells using either *In Vivo*- or cell culture techniques.

► [Neural Crest Cells and their Derivatives](#)

Clones

Definition

Clones refer to one or more genetically identical copies of a cell or organism, and as such would apply to any

bacterial colony. In the context of molecular biology, it refers to a clone of cells genetically engineered to contain some fragment of DNA, and by extension to the fragment of DNA itself.

► [YAC and PAC Maps](#)

Cloning

Definition

In the context of recombinant DNA technology, cloning (or: molecular cloning, DNA cloning, gene cloning) defines a process by which a DNA fragment of interest is transferred from one organism to a self-replicating genetic element, such as a bacterial plasmid. The DNA of interest can then be propagated in a foreign host cell.

In contrast, reproductive cloning is a technology used to generate an animal that has the same nuclear DNA as another currently or previously existing animal. The term might also be used to describe “therapeutic cloning”, also called “embryo cloning,” which is the production of human embryos for use in research. The goal of this process is not to create cloned human beings, but rather to harvest stem cells that can be used to study human development and treat disease.

► [Protein Microarrays as Tool for Protein Profiling and Identification of Novel Protein-Protein Interactions](#)

Cloning Vector

Definition

Cloning vector refers to a DNA molecule that is designed and constructed to be able to accept fragments of foreign DNA, and propagate them (as a new construct) in a suitable host strain. Sequencing vectors are a specific form of general cloning vectors, normally with a high copy number, and with a specific sequence (the priming site) adjacent to the cloning site.

► [Shotgun Libraries](#)

CLP

► [Common Lymphoid Progenitor](#)

CLSM

- Confocal Laser Scanning Microscope

some criteria or metric. The objective is to discover interesting new segments, whatever their number.

- Computational Diagnostics
- EST Mining for Expression Analysis
- Microarray Data Analysis

Cluster Analysis

Definition

Cluster analysis is the clustering, or grouping, of large data sets on the basis of similarity criteria for appropriately scaled variables that represent the data of interest. Similarity criteria among the several clusters facilitate the recognition of patterns and reveal otherwise hidden structures. Cluster analysis uses a set of statistical methods to group variables or observations into strongly inter-related subgroups.

- DNA Microarrays/DNA Arrays

Cluster of Differentiation

Definition

Cluster of differentiation (CD) is a nomenclature for a group of cell-surface molecules, that are recognized by a given set of monoclonal antibodies. This cluster of differentiation then received a number, for example CD1, CD2 etc., which stands only for the order of discovery. In general, each CD is associated with one or more functions, which were discovered through the effects on cell or tissue function of the antibodies that define it. CD cell surface molecules are used to differentiate between leucocyte subsets.

- Interleukins and Their Receptors, a Role in Regulating Hematopoietic Stem Cells
- Proteomics in Cardiovascular Disease

Clustering

Definition

Clustering is the process of creating clusters (using k -means, hierarchical clustering, decision trees, etc.), such that all objects of a cluster are similar according to

Clustering of Expression Profiles

- Computational Diagnostics

CMAP

Definition

CMAP is the compound muscle action potential that is evoked by repetitive stimulation and may reveal a myasthenic phenotype.

- Hereditary Neuropathies, Motor and/or Sensory
- muscle
- Neuropathies

CML

- Chronic Myelogenous Leukaemia

CMS

- Congenital Myasthenic Syndromes

CMT

- Charcot-Marie-Tooth Disease

c-Myc

Definition

The term c-Myc designates a proto-oncogene product which is involved in controlling metazoan cell growth and division, and is frequently activated in human cancers.

►RNA Polymerase III

interactions with sorting motifs present in the cytoplasmic domains of cargo proteins.

►Exocytotic Pathway

►Vesicular Traffic

Coactivator

Definition

Coactivator is a factor required for transcription that does not bind DNA, but is required for DNA-binding activators to interact with the basal transcription factors.

►Nuclear Hormone Receptors: Regulators of Gene Transcription and Cellular Signaling

Coactivator Complexes

Definition

Coactivator complexes achieve the connection between DNA-bound transcription factors and the preinitiation complex.

►Transcription Factors and Regulation of Gene Expression

Coat Proteins

Definition

Coat proteins are complex sets of interacting cytoplasmic proteins that can be recruited to intracellular membranes in a selective, compartment-specific fashion. There, they assemble to deform this “donor” membrane into buds and vesicles. Coat proteins select cargo during vesicle budding through selective

Cocaine

Definition

Cocaine is classified as a psychostimulant, as are other drugs of abuse such as amphetamine. Although these stimulants have similar behavioural and physiological effects, there are some major differences in the basic mechanisms of how they work at the level of the nerve cell. Cocaine is quickly removed and almost completely metabolized in the body.

►Addiction, Molecular Biology

Cockayne Syndrome (CS)

Definition

Cockayne syndrome (CS) is an autosomal recessively inherited disorder, which arises from mutations in the CSA and CSB genes. The CS complementation group protein (CSB) is at the interface of transcription and DNA repair and is involved in transcription-coupled and global genome-DNA repair, as well as in general transcription. CS is associated with abnormal UV hypersensitivity, growth retardation, and psycho-neural abnormalities. The disease develops by an age of 6 months. It is not a cancer prone disease.

►DNA-Repair Mechanisms

Coding Regions (of the Human Genome)

Definition

Coding regions (of the human genome) are regions in a genome that contain information for RNA transcription coded in triplets of nucleobases.

►SNP Detection and Mass Spectrometry

Codon

Definition

Codon describes one of the 64 possible combinations of nucleotide triplets that code for an amino acid or stop sequence.

- Base Excision Repair
- Duchenne Muscular Dystrophy
- tRNA

Codon Redefinition

Definition

Codon redefinition is a programmed event during mRNA translation that leads to the incorporation of non-standard amino acids at termination codons, usually UGA. Examples are the incorporation of selenocysteine or pyrrolysine.

- Ribosomes

Coefficient of Variance in Percent (%CV)

Definition

CV is a measure for assessing the reproducibility of analytical methods. $\%CV = (\text{standard deviation} / \text{average}) * 100$

- Mass Spectrometry: Quantitation

Coenzyme Q

Definition

Coenzyme Q (ubiquinone) is a coenzyme in the mitochondrial respiratory chain. It has a side chain made up of 10 isoprene units. Its synthesis can be inhibited by HMG-CoA reductase inhibitors (statins).

- Nucleotide Biosynthesis

Cohesin

Definition

Cohesin is a multiprotein protein complex that is required to hold together the sister chromatids of replicated chromosomes. It consists of two SMC proteins (SMC1 and SMC3), and two non-SMC subunits (Scc1 and Scc3). In higher eukaryotes, cohesin dissociates from chromosome arms in prophase and from the centromeres at the onset of anaphase, while in yeast, it dissociates from chromosomes at the onset of anaphase.

- Chromosome Condensation
- Meiosis and Meiotic Recombination

Coiled-Coil Domain

Definition

A coiled-coil domain is a heptad repeat of amino acids, with hydrophobic (H) residues at first and fourth positions, and polar (P) residues elsewhere. Coiling results from two right-handed α -helices around one another with a slight left-handed superhelical twist. Coiled-coil sequences adopt dimeric, trimeric and antiparallel tetrameric conformations. A coiled-coil is a quaternary protein structure of two or three α -helices, which permits the highly coordinated dimerization and heteromer formation of proteins, thereby mediating protein/protein interactions

- Autosomal Dominant (Inherited Disorder)
- Leucine Zipper Transcription Factors: bZIP Proteins
- Molecular Motors
- Polycystic Kidney Disease, Autosomal Dominant

Co-Immunoprecipitation

Definition

Co-immunoprecipitation is a process where macromolecules that associate with each other in the cell, are isolated together using antibody-based affinity purification methods.

- Proteomics in Human-Pathogen Interactions

Colchicine

Definition

Colchicine is an alkaloid prepared from the dried corms and seeds of *Colchicum autumnale*, the autumn crocus. Colchicine blocks tubulin polymerisation by binding to heterodimeric β -tubulin. Thus, it inhibits the migration of granulocytes into areas of inflammation and the release of proinflammatory agents. It is used to treat attacks of gout.

► [Cytoskeleton](#)

Colinearity Rule

Definition

The colinearity rule describes the exact point-by-point relationship between the order of amino acids along the polypeptide chain, and the order of the corresponding codons along the polynucleotide chain of the nucleic acids.

Collagen

Definition

Collagens represent a major protein family of the extracellular matrix. Collagens consist of three polypeptide chains, each called α -chain, having repeating domains of glycine-X-Y sequences. This triplet sequence is also known as the collagen sequence. Vertebrates have at least 27 collagen types, assembled from 42 distinct α -chains. Collagens are numbered with roman numerals in the order of their discovery (collagen types I–XXVII). In addition, there are some 20 additional proteins with collagen-like domains.

► [Extracellular Matrix](#)

Collecting Duct

Definition

Collecting duct designates structures of the collecting duct system that drain the nephrons of urine and that derive from the branched ureteric bud.

► [Kidney](#)

Colonies

Definition

Colonies are masses of progeny of a (generally) single microbial cell on the surface of a solidified growth medium (agar plate), which have grown large enough to be seen.

► [YAC and PAC Maps](#)

Colony Stimulating Factors

Definition

Colony stimulating factors (CSF) are proteins that induce the proliferation and/or differentiation of hematopoietic (progenitor) cells towards one or more mature hematopoietic cell types. These factors were first identified in assays where the production of hematopoietic colonies (group of hematopoietic cells) was used as read-out.

► [Growth Factors](#)

► [Interleukins and Their Receptors, a Role in Regulating Hematopoietic Stem Cells](#)

Colorectal Cancer

JÜRGEN BEHRENS¹, BARBARA LUSTIG²

¹Nikolaus-Fiebiger-Center for Molecular Medicine, University Erlangen-Nürnberg, Erlangen, Germany

²Clinic for Abdominal, Endocrine and Thoracic Surgery, Clinical Center Nürnberg, Nürnberg, Germany

jbehrens@molmed.uni-erlangen.de

Definition

Colorectal cancer (CRC) is a malignant tumor disease arising from the epithelium of the colon and rectum.

Characteristics

CRC is the second leading cause of cancer-related deaths in industrialized Western countries. CRCs are treated by surgical resection and by neoadjuvant and adjuvant chemotherapy and radiotherapy. The long-term survival of CRC patients depends on the tumor stage and the presence of distant metastases. Thus, the 5-year-survival rate exceeds 90% in the UICC stage I (limited invasion without regional lymph node

metastasis), but is below 20% in the UICC stage IV (presence of distant metastasis).

CRC develops from normal colorectal epithelium as a result of the accumulation of genetic alterations in defined oncogenes and tumor suppressor genes. In addition, loss of genomic stability leading to chromosomal and/or microsatellite instability (the ►CIN and ►MIN phenotypes, respectively) is an inherent feature of CRC. Much of what is known about the genetics of CRC has been derived from studies on hereditary colorectal cancer syndromes, of which the two best characterized are ►familial adenomatous polyposis (FAP) and ►hereditary nonpolyposis colorectal cancer (HNPCC). FAP is an autosomal dominant inherited disease characterized by the development of hundreds to thousands of polyps in the colon, which eventually transform into cancer. In 1991, the *APC* (►adenomatous polyposis coli) tumor suppressor gene was identified by positional cloning of the FAP locus at chromosome 5q21 (see below). While FAP comprises only around 0.1–1% of all colorectal cancers, HNPCC accounts for about 5% of the cases. Underlying this hereditary form of colorectal cancer is a defect in the DNA mismatch repair system (also known as mutation ►mismatch repair/MMR system). This system consists of a set of proteins that recognizes and repairs base-pair mismatches that occur during DNA replication. Germ-line mutations in *hMSH2* and *hMLH1* account for about 45–70% of the cases in HNPCC. Somatic mutations of the MMR system are also found in approximately 15% of sporadic colon cancers. The most prominent genetic consequence of mismatch repair deficiency is the occurrence of microsatellite instability (MSI), which reflects the deletion or addition of nucleotides in repetitive DNA sequences, leading to frame shifts and protein truncations when present in coding sequences. As an example, mutations of the TGFβ receptor occur frequently in MSI high tumors resulting in escape of tumor cells from growth inhibition by TGFβ. Other hereditary forms of CRC include the ►Peutz-Jeghers syndrome characterized by mutations of the tumor suppressor gene *LKB1* and ►juvenile polyposis in which the *PTEN* and *SMAD 4* tumor suppressors are mutated in an autosomal dominant fashion.

Molecular Interactions

A genetic model of colorectal tumorigenesis by Fearon and Vogelstein proposes that the acquisition of sequential mutations in ►oncogenes and ►tumor suppressor genes leads to progression from single crypt lesions to small benign tumors (adenomatous polyps) to malignant cancer (1). Mutations of the *APC* tumor suppressor gene are thought to initiate the clonal outgrowth of a single cell. *APC* has been considered to be a gatekeeper of colorectal tumorigenesis as it was

found to be mutated in the earliest premalignant lesions of sporadic CRC such as aberrant crypt foci. Additional mutations such as in the oncogene *K-ras* allow these cells to overgrow their sister cells and further mutations in the tumor suppressor gene *p53*, as well as other genes yet to be identified, lead to progression to carcinomas. In the following, we will focus on the signaling mechanisms and genetics of *APC* because mutations in *APC* are highly specific for CRC and initiate the process of CRC tumorigenesis.

Functional Role of APC in ►Wnt Signaling

The first clues to the function of *APC* were provided by the identification of the cytoplasmic protein β-catenin as a binding partner of the APC protein. Subsequently it was shown that APC acts as a major negative regulator of the Wnt signal transduction pathway by inducing degradation of β-catenin. In the absence of Wnt factors, a multiprotein complex (termed the “β-catenin destruction complex”) consisting of APC, the scaffold proteins axin or conductin/axin2 and the glycogen synthase kinase 3β (GSK3β) causes the ubiquitin-mediated proteasomal degradation of β-catenin. ►Axin/conductin interacts with APC and β-catenin and, importantly, recruits GSK3β to the destruction complex. GSK3β phosphorylates β-catenin at specific residues in its amino terminus and thereby earmarks the protein for ►ubiquitination. Wnts are secreted factors that bind to and activate Frizzled receptors leading to the inhibition of GSK3β. As a result β-catenin is stabilized, it enters the nucleus and teams up with ►HMG box transcription factors of the ►LEF/TCF family (below referred to as TCF). The TCF/β-catenin complexes activate target genes by binding to specific promoter elements and so transmit the Wnt signal into the nucleus (2). Mutations in components of the β-catenin destruction complex, most frequently of APC, lead to aberrant stabilization of β-catenin and thus to constitutive, Wnt-independent activation of TCF/β-catenin signaling.

Molecular Consequences of APC Mutations

The APC protein binds to β-catenin via specific repeated sequences called 15 amino acid and 20-amino acid repeats, which occur three and seven times respectively in the central domain of the APC protein. In addition, APC binds to the RGS domain of axin or conductin/axin2 via three so-called ►SAMP repeats. In colorectal tumors mutations of *APC* occur mostly in a mutation cluster region that is located closely 5′ to the sequence encoding the first SAMP repeat. These mutations generate stop codons or frameshifts leading to the deletion of the C-terminal half of the APC protein and thereby removing the interaction sites for axin/conductin (3). In addition, truncated APC also lacks several nuclear export sequences (NES) thought to be

important for the APC-mediated export of β -catenin from the nucleus. In summary, APC mutations disrupt the ability of APC to bind axin/conductin and β -catenin, which leads to increased TFC/ β -catenin interaction and constitutive activation of the Wnt pathway in the absence of exogenous Wnt factors.

APC interacts with a variety of other partners thereby regulating different cellular functions. APC is involved in ►cell migration, chromosomal stability, ►cell-cycle regulation and ►cell adhesion. For instance, APC binds to the APC-stimulated guanine nucleotide exchange factor (ASEF), which can activate the small G-protein rac and plays a role in cell migration. Truncated, but not wild type, APC appears to activate ASEF and hence stimulate cell migration, which could be important for the invasiveness of APC mutated cells. APC is also connected to microtubules and plays a role in the correct establishment of the mitotic spindle. Disturbance of APC function by mutations could lead to ►chromosomal instability (CIN), which is a hallmark of the majority of CRC.

APC Genotypes and Phenotypes in CRC

In FAP, patients harbor germline mutations in one allele of *APC* and lose the second *APC* allele through loss of heterozygosity. *APC* mutations lead to stop codons or truncating frameshifts thereby removing the β -catenin regulatory domains. Depending on the specific mutation sites of the first and second allele, patients develop extra-abdominal manifestations. The most common and well known are congenital hypertrophy of retinal pigment epithelium (CHRPE), which in some cases can be used to identify FAP patients, desmoid tumors, polyps in the upper gastrointestinal tract and thyroid carcinomas. Specific clinical associations are summarized by syndromes, Gardner syndrome (FAP, desmoid tumors, osteomas) and Turcot's syndrome (FAP, brain tumors).

In contrast to the "classical" FAP the "attenuated FAP" (AFAP) is characterized by fewer adenomas that develop at a later age and tend to be more right-sided. Interestingly, the *APC* mutations in AFAP result either in an ultra-short or a very long APC protein.

APC is mutated in about 80% of sporadic carcinomas, leading to a truncated APC protein. Around 60% of the *APC* mutations are found in the mutation cluster region. *APC* can also be inactivated by hypermethylation at CpG sites of the *APC* promoter; these have mostly been detected in sporadic CRC (4).

Catenin Mutations

β -Catenin is mutated in about 10% of all sporadic colon carcinomas by point mutations or in-frame deletions of the serine and threonine residues that are phosphorylated by GSK3 β . Like APC mutations, these mutations result in stabilization of β -catenin and activation of Wnt

signaling. β -Catenin and APC mutations are mutually exclusive, possibly reflecting the fact that both components act in the same pathway. β -Catenin mutations are more frequent in other tumor types such as liver tumors than in CRC.

Axin and Axin2/Conductin Mutations

Germline mutations of *conductin/axin2* associated with oligodontia and colorectal neoplasias in the affected families have been described. Mutations suggesting a tumor suppressor role of axin and the related conductin/axin2 gene have also been identified in various entities such as medulloblastomas, endometrioid ovarian carcinomas, hepatomas and hepatocellular carcinomas.

Mutations in Microsatellite Instable Tumors

Mutations of components of the Wnt signaling pathway also occur in microsatellite instable tumors but at a lower frequency than in the microsatellite stable counterparts. It appears that *β -catenin* is more often mutated than *APC*. In addition, 25% of microsatellite instable tumors show mutations in the *conductin/axin2* gene and *TCF4* mutations occur in about 40%, the consequences of which are not clear yet.

Regulatory Mechanisms

Wnt Target Genes and Cancer Development

Several of the known Wnt target genes have a potential role in CRC development. Among them are regulators of cell-cycle progression, such as *c-myc* and *cyclin D1*, which are directly controlled by TCF/ β -catenin, as well as growth factors and their receptors, such as FGF18 and the c-met tyrosine kinase. c-Met is the receptor for the mesenchyme-derived scatter factor/HGF, which acts as an epithelial growth factor but also promotes cell motility and invasion. Wnt signaling can prevent apoptosis by up-regulating anti-apoptotic proteins such as the caspase inhibitor, survivin. TCF/ β -catenin complexes can also up-regulate VEGF (vascular endothelial growth factor) and may thus induce ►angiogenesis, indicating that aberrant Wnt signaling is not only important for the initial expansion of the transformed cell compartment as implied by the loss of the "gate-keeper" function of APC, but might also lead to the acquisition of properties required for tumor progression. In line with this, several proteases capable of degrading extracellular matrix such as matrilysin/MMP7 and MMP-26 are Wnt targets that could aid the tumor cells in invasion and metastasis. Metastasis formation could also be promoted by up-regulation of Wnt targets such as CD44 and NrCAM. That TCFs control cell differentiation was revealed by the finding that EphB2 and EphB3 receptors were down-regulated by dominant-negative TCF, while their ligand ephrinB1 was up-regulated. Further analysis in

EphB knockout mice showed that this differential regulation served to prevent the intermingling of cells within the intestinal epithelium. Whether this is also relevant for human colon tumor formation remains to be determined.

Feedback Control

The Wnt pathway, like other signal transduction cascades, includes mechanisms for feedback control. The scaffold component of the β -catenin destruction complex, *conductin/axin2* is a direct target gene of Wnt signaling and might play a role in normal embryonic development where it could serve to attenuate the Wnt signal. Conductin is massively over-expressed in early colorectal adenomas and carcinomas as well as in other tumors. Whether this up-regulation has a functional role in tumors remains to be determined.

Negative control of the pathway is also achieved through the induction of dominant-negative forms of TCF-1. *Tcf-1* knockout mice develop spontaneous intestinal adenomas and mammary adenocarcinomas. When *Tcf-1*^{-/-} mice were mated to *min* mice, which carry a mutated *apc* allele and are prone to the development of intestinal adenomas, the number of adenomas was significantly increased in the progeny. Conversely, LEF-1 appears to be involved in a positive feedback loop that may be important in colon tumors. The *LEF-1* gene is activated by TCF/ β -catenin complexes but in this case, activation leads to transcription of full-size LEF-1 that is capable of binding to β -catenin. These LEF-1/ β -catenin complexes could further boost transcription of Wnt target genes.

Mouse Models

Mouse models for intestinal tumorigenesis have been established by introducing germ-line mutations in the *apc* gene or by mutational activation of β -catenin. Unlike the human disease, in all of these models tumors develop mainly in the small intestine. The *min* mouse was generated by chemical mutagenesis that introduced a chain-terminating mutation in murine *apc*. Other mouse models of FAP have been created by gene targeting of *apc*, such as APC Δ 716 and APC 1638N. In another approach a deletion mutant of exon3 encoding the phosphorylation sites of β -catenin was generated using the loxP/Cre system; this also led to intestinal tumorigenesis. Tumorigenesis in *apc* deficient mice could be accelerated in compound mice having additional deletion of the *smad2* or *smad4* tumor suppressor genes. Conversely, tumor formation was reduced when the genes for cyclooxygenase 2 or the multidrug resistance gene *mdr1* were additionally mutated (5).

β -Catenin Localization

In colorectal carcinomas the nuclear staining for β -catenin often shows a heterogeneous pattern with

strong nuclear enrichment at the invasion front and mainly cytoplasmic and membrane staining in the central tumor area. This indicates that high levels of nuclear β -catenin play a role in the transition of the tumor cells to the invasive state. The molecular basis for this differential distribution of β -catenin is not known, but it can be speculated that signals coming from the mesenchyme that surround the invasive tumor cells play a role (6).

Therapeutic Interference with Wnt Signaling

As ►Wnt Signaling is crucial for CRC development, interference with TCF/ β -catenin mediated transcription seems a valid option for rational cancer therapy. Small molecule inhibitors that prevent the TCF/ β -catenin interaction have recently been identified (7).

References

1. Kinzler KW, Vogelstein B (1996) Lessons from hereditary colorectal cancer. *Cell* 87:159–170
2. Seidensticker MJ, Behrens J (2000) Biochemical interactions in the wnt pathway. *Biochim Biophys Acta* 1495:168–182
3. Polakis P (2000) Wnt signaling and cancer. *Genes Dev* 14:1837–1851
4. Fodde R, Smits R, Clevers H (2001) APC, signal transduction and genetic instability in colorectal cancer. *Nat Rev Cancer* 1:55–67
5. Lustig B, Behrens J (2003) The Wnt signaling pathway and its role in tumor development. *J Cancer Res Clin Oncol* 129:199–221
6. Brabletz T, Jung A, Kirchner T (2002) Beta-catenin and the morphogenesis of colorectal cancer. *Virchows Arch* 441:1–11
7. Lepourcelet M, Chen YN, France DS et al (2004) Small-molecule antagonists of the oncogenic Tcf/ β -catenin protein complex. *Cancer Cell* 5:91–102

Combined Pituitary Hormone Deficiencies

►Hypothalamic and Pituitary Diseases, Genetics

Commitment

Definition

Commitment designates the fixation of the prospective fate of cells.

►Muscle Development

Common (Multifactorial) Diseases

Definition

Common (multifactorial) diseases mean frequently occurring disorders (>1% in the population) arising as the result of the interaction of several to many genes, and usually in addition to non-genetic factors.

- ▶ Common Diseases, Genetics
- ▶ Repeat Expansion Diseases

Common Diseases, Genetics

M. JULIA BROSNAN, ANNA F. DOMINICZAK
BHF Glasgow Cardiovascular Research Centre,
University of Glasgow, Scotland, UK
mjb8n@clinmed.gla.ac.uk
anna.dominiczak@clinmed.gla.ac.uk

Definition

Evidence from numerous family studies has shown that the common diseases such as essential hypertension, atherosclerosis and cancer are caused by varying combinations of genetic and environmental factors. Since common diseases are, by definition, those present at 1% or higher in the general population, the resultant morbidity and mortality puts a significant strain on health budgets. Consequently, much effort has been put into elucidating which genes are involved and how they may interact with each other and the environment. In recent years a number of the technologies developed in the interrogation of the more simple monogenic disorders have assisted in the identification of the chromosomal locations and in a few cases genes containing mutations, which contribute to the susceptibility to a particular common disease. These technologies are now being applied in the identification of the genes contributing to the development of the common more complex diseases.

Characteristics

The common diseases are those present at 1% or higher in the general population and are caused by the complex interactions of multiple genes located on several chromosomes. Identification of susceptibility genes along with the unravelling of their interaction with environmental factors presents one of the most exciting challenges in current medical research.

Cellular and Molecular Regulation

Connecting phenotype with genotype is the fundamental aim of all human genetics. Almost 3 billion base pairs of DNA, containing 30,000–40,000 protein-coding regions are housed in the 46 chromosomes of the human genome. The coding regions constitute less than 5% of the genome – the function of the remaining 95% is unclear. The capacity to determine which DNA sequences cause which specific traits was elusive until genome-wide linkage analysis using anonymous DNA polymorphisms was first suggested. The starting point of gene identification is ▶ [linkage analysis](#). Families in which the disease phenotype segregates are analysed using a series of genetic polymorphisms. The earliest and most successful disease in which linkage analysis led to the gene was cystic fibrosis. At that time the DNA markers were ▶ [restriction fragment length polymorphisms](#) (RFLP) and simple sequence repeats such as ▶ [microsatellites](#). Nowadays such endeavours benefit increasingly from ▶ [single nucleotide polymorphisms](#), which have been made available through the sequencing of the human genome. Linkage analysis results in the identification of a chromosomal region which contains the disease causing gene(s), albeit many others as well. The genes are then searched for using ▶ [positional cloning](#) in which the target region of the chromosome is more thoroughly interrogated using a variety of molecular techniques.

To date many of the genes underlying monogenic disorders have been identified using linkage analysis and positional cloning (e.g. cystic fibrosis and Duchenne muscular dystrophy). However, the ▶ [candidate gene approach](#) has had its own successes and identified the ApoE gene as being involved in late onset Alzheimer's disease (1). The genetic paradigm underlying common diseases is often more complicated but the analysis has undoubtedly benefited from the advances in technology, in particular the concept of genomics, implemented in the investigation of the monogenic diseases as described in the rest of this essay.

Common Cancers

The term “cancer” represents a heterogeneous group of growth disorders, caused by a number of interacting genetic and environmental factors, which may occur in any organ of the body. The common cancers include those of the lung, breast and prostate. Cancer occurs when cell division, normally under tight regulation, gets out of control. An extensive network of signalling molecules dictates which cells divide, how often and when. Mutations in one or more of these signals, caused by either a genetic predisposition or an environmental effect such as smoking or a combination of both, can trigger cancer. Recently three different mechanisms have been defined (i) transformation of a

Common Diseases, Genetics. Table 1 Monogenic Forms of Hypertension and Hypotension

Hypertensive	Syndrome	Mode of inheritance ¹	Enzyme/Location
	Apparent mineralocorticoid excess	Recessive	11 β -hydroxysteroid dehydrogenase
	17 α -hydroxylase deficiency	Recessive	17 α -hydroxylase
	11 β -hydroxylase deficiency	Recessive	11 β -hydroxylase
	Liddle's syndrome	Dominant	β or the γ subunit of ENaC
	Hypertension and brachydactyly	Dominant	Chromosome 12p12.2-11.2
	Glucocorticoid remediable aldosteronism.	Dominant	Aldosterone synthase and 11 β -hydroxylase
	Gordon's Syndrome	Dominant	Chromosome 12 and 17. Wnk1 and Wnk4
	Pregnancy related hypertension	Dominant	Mineralocorticoid receptor
Hypotensive			
	Bartter's syndrome type 1	Recessive	Apical Na-K-2Cl cotransporter
	Bartter's syndrome type 2	Recessive	ATP-sensitive K ⁺ channel ROMK
	Bartter's syndrome type 3	Recessive	Cl ⁻ channel <i>CLCNKB</i>
	Recessive PHA1	Recessive	ENaC
	Aldosterone synthase deficiency	Recessive	Aldosterone synthase
	21-hydroxylase deficiency	Recessive	21-hydroxylase
	Gitelman's syndrome	Recessive	Thiazide-sensitive Na-Cl cotransporter
	Dominant PHA1	Dominant	Mineralocorticoid receptor

¹ The inheritance in all of these diseases is autosomal

normal gene into an **▶oncogene** (ii) dysfunction of a **▶tumour suppressor gene** and (iii) impairment of a **▶DNA repair** pathway. Lung cancer shows a particularly high association, over 80%, with environmental effects such as cigarette smoking. Small cell lung carcinoma, on the other hand, has been associated with a deletion of part of chromosome 3. Prostate cancer is one of the leading causes of cancer death in men and yet little is known about the genetic predisposition. Numerous studies indicate that a family history may be responsible in up to 10% of all prostate cancers. Identification of a susceptibility locus on chromosome 1 called HPC1 (hereditary prostate cancer 1) presents a promising breakthrough. The next steps will include cloning the gene and characterising it in model systems.

Common Cardiovascular Disorders

Diseases of the cardiovascular system include essential hypertension, coronary artery disease and stroke. Like cancer they are multi-factorial and dissection of the contribution of genetics and environmental factors is

ongoing. Earlier studies of the simpler monogenic forms of hypertension may provide some clues. Molecular genetic studies have identified mutations in several genes that cause **▶Mendelian forms of hypertension** and/or hypotension (2). All of these genes, shown in Table 1, act to alter net renal salt reabsorption in the kidney. Accordingly, efforts to identify genes predisposing to essential hypertension have largely focused on candidates in pathways implicated in blood pressure control. Candidate gene studies have typically compared the prevalence of hypertension among individuals of contrasting genotypes at the candidate locus. Among such studies, common variants in genes of the renin-angiotensin system, the same pathway implicated in Mendelian forms of hypertension, have provided some evidence of effect. For example, a variant in the *angiotensinogen* gene is associated with higher circulating angiotensinogen levels and hypertension in several but not all populations. This variant is in strong **▶linkage disequilibrium** with another variant near the **▶transcription** start site that has been proposed to result in

Common Diseases, Genetics. Table 2 The Contrasting Role Of Genetics And Environment In The Development Of The Three Forms Of Diabetes

	Clinical features	Environmental Component	Genetic component	Causative/Candidate Genes
MODY	Develops in children		+++	Glucokinase
Type 1	Insulin dependent	+	++	Several implicated
Type 2	Non-insulin dependent	++	+	Several implicated

increased transcription of the gene. Interestingly, a mouse model featuring increased expression of the *angiotensinogen* gene demonstrates elevated blood pressure, indicating this as a plausible mechanism. Variants in other genes in the same physiological pathway, such as the ACE (►[angiotensin converting enzyme](#)) gene, have also been investigated with contrasting results in different populations. Poor reproducibility in these studies suggests that these variants impart singly, at most, very modest effects and the true picture will emerge when it is possible to look at the combined effects of several mutations simultaneously.

More recently, several studies have performed a ►[genome-wide analyses](#) of linkage, comparing the inheritance of each chromosome segment to phenotypes such as hypertension or blood pressure. In addition to human studies, investigation of naturally occurring and genetically engineered animal models of disease hold promise for understanding the long-term regulation of blood pressure.

To date, although genetic studies have identified several genes in which rare mutations impart large effects on blood pressure, providing new insight into the causes of hypertension, genetic studies of the general population have until recently been relatively disappointing and suggested that there are no single loci that account for a large fraction of the variation in blood pressure in the general population. However, the recently published British Genetics of Hypertension (BRIGHT) study, the largest blood pressure genome scan performed, in which 2010 severely hypertensive sibling pairs were genotyped, identified one chromosomal region with a significant LOD score at 3.21 on chromosome 6q and three suggestive loci scores on 2q, 5q and 9q.

Further integration of the data obtained from well-designed genetic, molecular, clinical and epidemiological studies could disclose subsets of patients in whom specific combinations of genetic and environmental factors raise blood pressure and lead to more individualised treatment.

Common Endocrine Disorders

Family and twin studies have shown that diabetes has high ►[heritability](#) but for many years an understanding

has been elusive. The first step towards progress was the careful classification of the different types of diabetes (Table 2). Type I and type II diabetes are totally different diseases with different causes and totally different genetics. Maturity onset diabetes of the young (MODY) is a rare variant having the clinical features of type II but the age of onset of Type I. Type II results from a decreased number of pancreatic β cells and/or the development of insulin resistance. Obesity, age and a lack of physical exercise along with unknown genetic factors contribute in the pathogenesis. MODY is highly heritable and in some families has been shown to be caused by a mutation in the gene encoding glucokinase, an enzyme key to glucose metabolism and control of insulin secretion on chromosome 7.

►[Type I diabetes](#) is a complex trait in which mutations in several genes contribute to the disease. About 10 ►[loci](#) in the human genome have been identified which confer susceptibility to type I diabetes. Chromosome 6 harbours at least one of these at a region referred to as IDDM1 (insulin dependent diabetes mellitus). Although it is not yet clear how a mutation at this locus acts to increase susceptibility to diabetes, it is thought that the expression of antigens encoded by genes in this region may be involved. Other possible loci include a gene at IDDM2 on chromosome 11 and the gene for glucokinase, which has been confirmed as causative in MODY.

Common Immunologic Disorders

Complex diseases of the immune system include ►[asthma](#), which now affects over 1 in 10 children. It is a chronic inflammatory disorder of the airways caused by oedema (fluid in the intracellular space) and the influx of inflammatory cells into the walls of the airways. Family and twin studies suggest that asthma has a strong genetic component (60–70%). Genomic screens and positional cloning studies to find the genes responsible are now beginning to come to some agreement with genes located on chromosomes 5, 6, 7, 11, 14 and 12 implicated to contribute. Typical candidate regions include those encoding key molecules such as those seen to be involved in asthma including ►[cytokines](#) and ►[growth factor receptors](#).

Several genetic studies have been performed to find asthma susceptibility genes in large families. The researchers were interested in whether a particular allele of a given polymorphism shows up more frequently in affected individuals *versus* controls and found multiple polymorphisms in the ADAM33 gene, which were associated with an increased risk of asthma. ADAM33 is a membrane bound metalloproteinase whose family members are known to induce cell shedding of proteins such as cytokines and cytokine receptors. However, the link between ADAM33, cell-surface shedding and asthma has still to be established (4). Other studies have revealed the strong association between asthma susceptibility and severity. Laitinen et al. performed a genome scan of a Finish population and found linkage to a region on chromosome 7 for three phenotypes, asthma, high levels of immunoglobulin E and a combination of the two with the strongest being the IgE level. IgE is the best marker for allergies (5).

Clinical Relevance

Molecular biological investigations using well-defined model systems of common diseases, in particular rodents, have already presented us with candidate pathways and target genes for the therapeutic manipulation of common diseases. The publication of the sequence of the human genome and the identification of functional **▶SNPs** will aid in the identification of those genes, which, in combination with environmental effects contribute to the pathogenesis of the disease. The identification of contributory genes will allow screening possibilities and those individuals harbouring combinations putting them at risk may be offered screening opportunities such that the appropriate life-style changes can be implemented to ease the impact of these diseases. In addition, the pattern of SNPs ascribed to an individual will act as a genetic signature of drug efficiency and lead to individualised medicines (pharmacogenetics).

Once a human disease gene has been characterised, molecular genetic tools can be used to design novel therapies. Gene therapy is the genetic modification of the cells of a patient in order to combat disease. This broad definition encompasses many different possible approaches including transfer of cloned human genes, double stranded human gene segments, oligonucleotides and artificial genes such as anti-sense genes. At present most gene therapy aims to alleviate the symptoms of the disease, in a risk-free manner, but in the future will not only be based on improving the clinical situation but also on replacing the combination of mutated genes underlying the disease.

Future Directions

Whilst the first draft of the human genome has brought with it a tremendous step forward in our capacity to

investigate the genetics of common diseases, it will not answer all the questions. In each disease the identification of genes may provide a target for genetic screening and in some cases gene therapy. However, the complexity of the system means that the identification of genes does little to define protein expression and activity underlying the pathogenesis of the disease. There is little doubt that the careful application of model systems and innovative technological developments will signpost the way to the Holy Grail of polygenic diseases.

Model Systems

Model systems, in particular genetically manipulated animals, will be invaluable as a resource in understanding the genetics of common diseases. Inbred strains of mice and rats, selected for a particular disease, have underpinned the identification of key contributory genes. In human genetics, the determination of linkage is essentially a statistical process and requires absolute knowledge of the status of affected individuals. Where misdiagnosis, caused by unrecognised **▶phenocopy**, results in the inclusion of individuals within a disease group, linkage analysis can fail even in a large cohort. Consequently the ability for a more absolute diagnosis in a model system such as the rat or the mouse allows increased power in the statistical approach. Furthermore the relatively short life span of a rodent model means that increased number of individuals can be analysed. In addition the use of transgenic rats and “**▶knock-in** and **▶knock-out**” mice allows the definition of primary events and those that occur as adaptations to the disease.

Technological Advances

The **▶polymerase chain** reaction (PCR) has revolutionised our capacity to examine DNA. PCR provides the capacity to generate large quantities of DNA, from minimal starting material, with a sequence identical to that of the contributing patient. In the past restriction fragment length polymorphisms (RFLP) which identify differences in the sites for restriction enzymes between the DNA of different individuals and more recently automated SNP analysis and sequencing using high speed robotics and automated throughput have been important technologies which will accelerate the discovery of contributing genes. The bioinformatics revolution brings with it the “omics” era consisting of genomics, transcriptomics (analysis of the complement of mRNAs transcribed from a cell’s genome), proteomics and metabonomics (the totality of small metabolite molecules in an organism). Together these offer to fill the gap in our understanding between genes and pathogenesis and in the development of effective strategies for early prognosis and treatment (6). This may be useful in defining the genes and

subsequently the mechanisms and perhaps more importantly point us in the right direction to develop the best therapeutic approaches to treat these common diseases.

References

1. Roses AD (1994) Apolipoprotein E affects the rate of Alzheimer's disease expression: beta amyloid burden is a secondary consequence dependent on ApoE genotype and duration of disease. *J Neuropathol Exp Neurol* 53:429–437
2. Lifton RP, Gharavi AG, Geller DS (2001) Molecular mechanisms of human hypertension. *Cell* 104:545–556
3. Caulfield M, Munroe P, Pembroke J et al (2003) Genome-wide mapping of human loci for essential hypertension. *Lancet* 361:2123
4. Ahmadi KR, Goldstein DB (2002) Multifactorial diseases: asthma genetics point the way. *Curr Biol* 12: R702–R704
5. Laitinen T, Daly MJ, Rioux JD et al (2001) A susceptibility locus for asthma-related traits on chromosome 7 revealed by genome-wide scan in a founder population. *Nat Genet* 28:87–91
6. Collins FS, Green ED, Guttmacher AE et al (2003) A vision for the future of genomics research. *Nature* 422:835–847

Common Lymphoid Progenitor

Definition

The common lymphoid progenitor (CLP) is an immature hematopoietic progenitor cell and expresses the IL-7 receptor. The CLP gives rise to B- and T-cells.
► [Interleukins and Their Receptors, a Role in Regulating Hematopoietic Stem Cells](#)

Comparative Genomic Hybridization

Definition

Comparative genomic hybridization (CGH) defines an array based method for determining genome-wide changes in chromosome copy number or gene amplification events.

► [Genomic Information and Cancer](#)

► [Matrix-CGH \(Comparative Genomic Hybridization\)](#)

Comparative Modeling

► [Homology Modeling](#)

Compartment Modeling

Definition

Compartment modelling designates a mathematical technique for the quantitative description of tracer kinetic processes.

► [PET](#)

Complement System

Definition

Complement system refers to a set of plasma or tissue proteins that act together to attack extracellular forms of pathogen. Complement also binds to [immunoglobulins](#) and C-reactive protein (CRP).

► [Inflammatory Response](#)

Complementary Base

Definition

The Watson-Crick model of the double helix of DNA is based on the discovery of Erwin Chargaff, that only purin bases (adenine, guanine) of one polynucleotide strand interacts with pyrimidine bases (cytosine, thymine) of the other strand through hydrogen bonding, thus forming the double helix. Adenine interacts with thymine, guanine interacts with cytosine, and the interacting bases are called complementary.

Complementation

Definition

When it is shown genetically that two (or more) genes control a phenotype, the genes are said to form a

► **complementation** group. Host strain/selection marker combinations rely on the concept of complementation: in the host strain an essential gene, without which cells are not viable, is complemented, i.e. its function is replaced by a gene on the expression plasmids. Therefore, only cells carrying this vector survive on selective media.

► **Recombinant Protein Expression in Yeast**

Complementation Studies

Definition

Complementation studies are aimed at resolving the genetic homo- or heterogeneity of a particular set of disorders, based on fusion of patient fibroblast cells followed by biochemical or cell biological studies.

► **Peroxisomal Disorders**

Complex Diseases/Complex (Multigenic) Traits

Definition

Complex (multigenic) traits/diseases are diseases that are influenced by multiple loci (genes), each of which is known as a quantitative trait locus (QTL). Thus, complex traits are characterized by a genetic component that is not strictly Mendelian (dominant recessive or sex-linked). Additional to genetic variation, environmental factors play an important role. Complex diseases are often characterized by clinically heterogeneous presentation.

► **Atopy Genetics**

► **COPD and Asthma Genetics**

► **Crohn's Disease**

► **Mendelian Forms of Human Hypertension and Mechanisms of Disease**

► **SNP Detection and Mass Spectrometry**

Complex Numbers

Definition

The mathematical entities (of dual real and imaginary components) that can be used to express wave amplitude and phase quantities and simplify notation.

► **X-Ray Crystallography—Basic Principles**

Complex Product

Definition

The product of two complex numbers $(a + ib)(a' + ib')$. The product of a complex number with its complex conjugate $(a + ib)(a - ib)$ is the squared amplitude $(a^2 + b^2)$ of the number and is thus a real number. The diffraction intensities $I(h, k, l)$ are the real squared amplitudes of the complex structure factors $F(h, k, l)$.

► **X-Ray Crystallography—Basic Principles**

Complex-Type

Definition

Complex-type refers to an N-glycan with lacking terminal non-reducing mannose residues and other terminal non-reducing residues, such as sialic acid, galactose, fucose, and/or N-acetylglucosamine.

► **Glycosylation of Proteins**

Compound Heterozygote

Definition

Compound heterozygote is an individual who is compound heterozygous if (s)he has two different mutated alleles at a particular locus.

► **Acute Intermittent Porphyria**

► **Chromosomal Instability Syndromes**

► **Genetic Screening in Populations**

► **Hereditary Diseases Genetic Basis**

► **High-HDL Syndrome**

Compulsive Drug Seeking

Definition

Compulsive drug seeking occurs when the addict loses control of drug-taking behaviour despite being aware of the adverse health, social or legal consequences. Drug use pervades all life activities.

► **Addiction, Molecular Biology**

Computational Diagnostics

RAINER SPANG, FLORIAN MARKOWETZ
Max Planck Institute for Molecular Genetics,
Berlin Center for Genome Based Bioinformatics,
Berlin, Germany
spang@molgen.mpg.de

Synonyms

Clinical bioinformatics; Classification and clustering of expression profiles

Definition

Computational diagnostics deals with statistical models and algorithms to analyze large-scale genomic datasets for diagnostic and prognostic purposes. The fundamental question of computational diagnostics is “What can be learned about the disease of an individual patient from genomic data?” This differs from classical bioinformatics, where the focus is on functional genomics: “What can be learned about the function of a certain gene?”

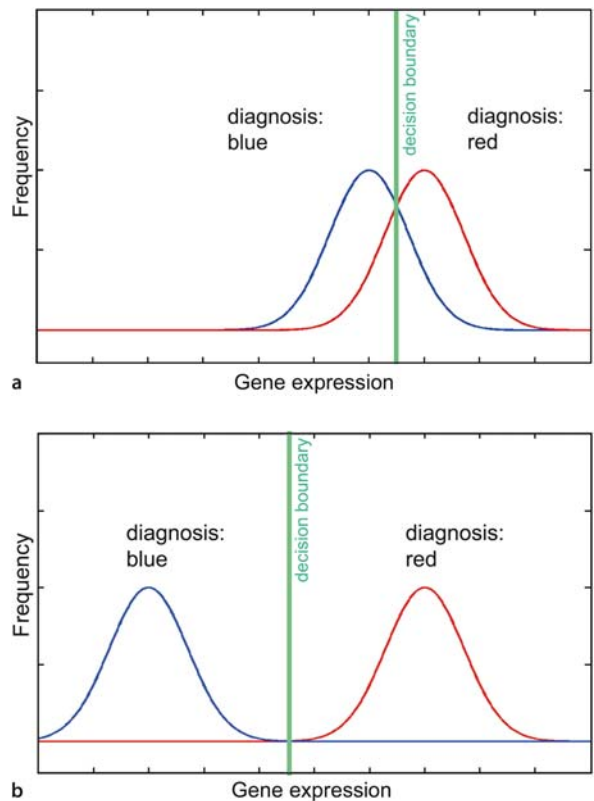
The data used in computational diagnostics is genome scale phenotype data like mRNA expression profiles, protein expression profiles and methylation screens.

► **Association studies** of genotype data like single nucleotide polymorphisms are part of statistical genetics and will not be considered here.

The goal of computational diagnostics is to detect patterns that are predictive for certain disease entities or outcomes. Statistical analysis of microarray data is currently the most applicable approach. The data used in microarray studies is very high dimensional. This requires advanced statistical modeling. The two major problems are (1) screening a huge number of genes for potential clinical markers and (2) learning general features of the given data without ► **overfitting**. The solution lies in statistical methods for feature selection and regularization techniques from ► **machine learning**.

Characteristics

In clinical studies, the analysis of genomic data divides into clinical functional genomics and computational diagnostics. Clinical functional genomics investigates the role of certain genes in molecular disease mechanisms (3). Computational diagnostics explores large-scale genomic data from an individual patient for improved diagnosis and therapy decision-making. We focus here on three central issues in the statistical analysis of microarray data, screening for marker genes, deriving predictive ► **signatures** and detecting new disease entities.

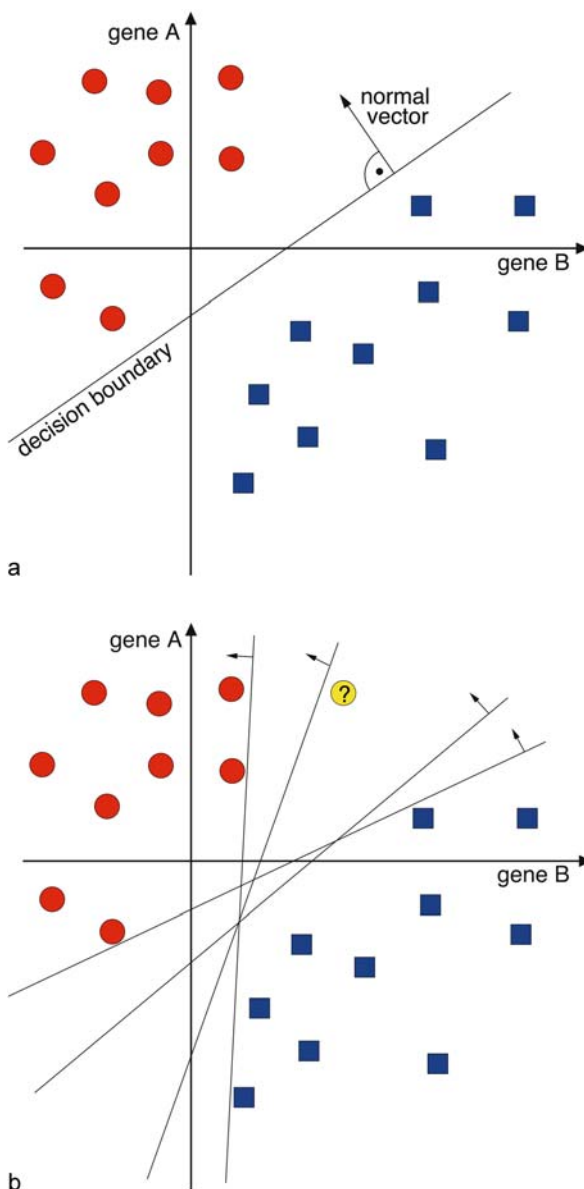


Computational Diagnostics. Figure 1 The distribution of expression levels in two disease groups for a single gene. In (a) one observes a large overlap. The gene is more or less useless as a marker gene, even if the difference between the red and blue distributions can be detected by statistical tests. (b) shows two clearly separated distributions; this gene is an ideal marker gene.

Marker

The commonly used approach to find “interesting” genes is to apply a test for differential gene expression like the t-test and extract genes with low p-value (5). However, genes found significant in a multiple testing scenario need not be valuable clinical markers. The left plot in Fig. 1 gives a schematic view of the distribution of single gene expression levels in two disease groups (blue/red). The difference between the two distributions can be insignificant in statistical tests for small sample sizes and significant for large sample sizes, but due to the large overlap the gene can be useless as a clinical marker.

There is a difference between screening for evidence to disprove that a gene’s expression levels are essentially the same in two disease groups and screening for genes that can serve as clinical markers. The goal of the first one is to improve our understanding of the role of certain genes in a certain disease (functional genomics) and the goal of the second is to find ► **clinical marker**



Computational Diagnostics. Figure 2 (a) two disease entities (red and blue) can be separated by combining gene A with gene B. Points facing in the direction of the normal vector of the hyperplane are classified as red, points on the other side of the hyperplane are classified as blue. (b) Overfitting: many different hyperplanes separate the red points from the blue points. They disagree in their class prediction for the new patient (yellow). This shows that separation of the training set does not guarantee good generalization performance.

genes for improved diagnosis and prognosis (computational diagnostics).

For evaluating the quality of a gene as a diagnostic marker the performance of a decision rule (green line) needs to be evaluated. A gene is a good clinical marker, if the two distributions can be separated by a decision

boundary with only a small number of misclassifications. The right plot of Fig. 1 shows the ideal scenario for a diagnostic marker. Unfortunately, genes with such a classification power are hardly ever found.

Signatures

The solution is often to combine single genes into multi-gene expression patterns, so-called signatures, as is shown in the left part of Fig. 2. Note that both gene A and gene B are poor markers as single genes, but in combination they separate the diagnostic entities perfectly. The signatures are to be defined such that they distinguish the patients with the disease from the patients without the disease. Mathematically, signatures are hyperplanes in high-dimensional spaces separating one class (with disease) from the other class (without disease). A signature is used to predict the class of a new patient by comparing the patient profile to the normal vector of the hyperplane; patients in the direction of the normal vector are predicted as one class, patients on the opposite side of the hyperplane as the other class.

Overfitting

One would expect that finding a separating hyperplane is a difficult task. However, if the number of genes is bigger than the number of patients, there always exists a hyperplane separating the training data without error. But flawless separation of the study data does not necessarily lead to good diagnostic performance in the future. This phenomenon is called overfitting; a hyperplane, which correctly identifies the training set, may be useless to diagnose new patients. In the right part of Fig. 2 we show an example where two groups can be separated by a large number of hyperplanes with conflicting diagnoses for a new patient.

Regularization

The problem of overfitting is addressed by strategies of regularization (2). Regularization means “making the classification problem harder”. This can be done in several ways. The most widely used approach is gene selection; limit the number of genes in the predictive signatures. A second widely used strategy is to choose a classifier that maximizes the margin of separation between the samples of the two classes. Both methods improve diagnostic performance ability.

Evaluation

Excellent performance on the training data does not guarantee good generalization to new patients. It is important to evaluate the performance of a diagnostic signature on independent test data. This can be done by either splitting the data into a test and a training set or by ►crossvalidation. It is important to note that gene selection is part of classifier training. The test data has to be classified using the genes selected only on the

training data, it must not be included into the gene selection procedure. In crossvalidation, gene selection has to be repeated anew in each step (5).

► Clustering

In prediction, we generalize a structure (a class distinction) from training examples to a new specimen. A different task is to find hidden structures in the data (2, 3, 5). The question here is “Does a known disease entity split into several sub-entities, that can be distinguished reliably based on genomic data?” This is called clustering or ► **unsupervised learning**, in contrast to prediction, which is called ► **supervised learning**. The difference between supervised and unsupervised learning is that in the first we need information about the structure of the data (given by the class labels) and in the second we infer this structure from the data. Particular diligence is needed to prove the validity of the results in unsupervised learning.

► Molecular Symptoms

Clustering patients by gene expression profiles can find subtypes of a known disease. These techniques find homogeneous subgroups in the data. Microarrays contain information about thousands of genes; not all of them matter to describe a split between two subclasses. Therefore, one clusters patients according to subsets of genes. This leads to the concept of a molecular symptom: a certain subset of genes that are characteristic for a certain subset of patients. With two-way clustering methods, patients and genes are clustered at the same time (5). In the methods described so far one gene or one patient was a member of only one cluster. More advanced methods use overlapping clusters. One patient can have several distinct molecular symptoms and the same gene can play a role in more than one symptom.

Clinical Relevance

The molecular view of diseases brings clarity to previously muddy diagnostic categories. Detecting new disease entities leads to more stratification options and refined clinical studies. The higher the diagnostic resolution, the more successful the treatment. Given a disease with several subtypes, the overall most successful treatment is not necessarily optimal for each subtype. First diagnosing the subtype and then using the optimal treatment will increase the overall success rate, although no novel therapies have been developed (4).

Identifying prognostic and predictive factors is a labor-intensive process, often requiring the combined expertise of hematologist, oncologist, pathologist and cytogeneticist. Gene expression assays have the potential to supplement cell surface, histological or gross pathological features with a molecular characterization of the abnormal cell. Microarray analysis is now

tested in many fields of medicine. We chose three studies (1, 6, 7), which demonstrate that expression profiling can accurately identify known prognostically important cancer subtypes.

► Leukaemia

► **Acute lymphoblastic leukaemia** is a heterogeneous disease, with individual leukemia subtypes differing in their response to chemotherapy. Yeoh et al. (7) used oligonucleotide microarrays to analyze the pattern of genes expressed in leukemic blasts from 360 pediatric ALL patients. Distinct expression profiles identify each of the prognostically important leukemia subtypes. In addition, another ALL subgroup is identified based on its unique expression profile.

► Breast Cancer

West et al. (6) analyzed expression profiles from 49 breast tumor samples and assessed the validity of their diagnostic model in predicting the status of tumors in crossvalidation determinations. The practical value of this approach relies on two features: (1) they assess relative probabilities of clinical outcomes for future samples and (2) they provide an honest assessment of the uncertainties associated with predictive classifications on the basis of the selection of gene subsets for each validation analysis.

Lymphoma

Alizadeh et al. (1) gave an example for the detection of new disease entities from gene expression profiles. In this study two types of diffuse large B-cell lymphoma (DLBCL) were identified. DLBCL was long known to be clinically heterogeneous; 40% of patients respond well to current therapy, whereas the remainder succumb to the disease. Alizadeh et al. argue that this difference reflects the molecular heterogeneity in tumors. Using hierarchical clustering they split the data into two homogeneous subtypes. The comparison of these gene expression subgroups with survival data of the patients shows significant differences in the overall survival.

References

1. Alizadeh AA et al (2000) Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. *Nature* 403(6769):503–11
2. Hastie T, Tibshirani R, Friedman J (2001) *The Elements of Statistical Learning*, Springer, New York
3. Kohane IS, Kho AT, Butte AJ (2003) *Microarrays for an integrative genomics*. The MIT Press, Cambridge MA
4. Spang R (2003) Diagnostic signatures for personalized medicine. *BIOSILICO* 1(2):64–68
5. Speed T (2003) *Statistical analysis of gene expression microarray data*. Chapman & Hall/CRC, New York
6. West M et al (2001) Predicting the clinical status of human breast cancer by using gene expression profiles. *PNAS* 98 (20):11462–7

7. Yeoh EJ et al (2002) Classification, subtype discovery, and prediction of outcome in pediatric acute lymphoblastic leukemia by gene expression profiling. *Cancer Cell* 1(2):133–43

target gene in a spatially and temporally restricted fashion.

► [Transgenic and Knockout Animals](#)

Concordance (Rate)

Definition

Concordance (rate) defines the percentage of sharing the same (disease) phenotype in two first degree relatives, especially in monozygotic or dizygotic twins. A significantly higher concordance rate in monozygotic twins (sharing 100% of their genome) compared with dizygotic twins (sharing approximately 50%) indicates a genetic component for disease expression.

- [Atopy Genetics](#)
- [Genetic Predisposition to Multiple Sclerosis](#)
- [Manic Depression](#)

Condensin

Definition

Condensin is a multiprotein complex that consists of two SMC subunits (SMC2 and SMC4), and three non-SMC subunits (CAP-D2, CAP-G and CAP-H). Condensin is required for chromosome organization and segregation and it only associates with chromosomes at prophase, when chromosomes condense and dissociate in anaphase when chromosomes start to decondense – though individual subunits show different behaviour.

- [Chromosome Condensation](#)

Conditional Gene Knockout

- [Cre/loxP Strategies](#)
- [Transgenic and Knockout Animals](#)

Conditional Gene Targeting

Definition

Conditional gene targeting is a recently developed transgenic technique that allows the mutation of a

Conditional Knockout (KO)

Definition

Conditional Knockout (KO) refers to a genetic modification that is manifested upon a predetermined trigger. The trigger can be external, e.g. tetracycline, or upon the expression of a transgene initiating a specific recombination event. This technique can be used to inactivate genes in a tissue-specific (spatial) and/or temporal-regulated approach.

- [Large-Scale Homologous Recombination Approaches in Mice](#)
- [Transgenic and Knockout Animals](#)

Conduction-System Disease

Definition

Electrical impulses that initiate contraction of the ► [heart](#) muscle arise from the sinus node, and are propagated through the atria to the atrioventricular node and subsequently to the ventricles. Gene mutations that cause defects of the conduction-system can result in sinus bradycardia (1st to 3rd degree), atrioventricular conduction delay, and supraventricular arrhythmias.

- [Familial Dilated Cardiomyopathy](#)

Cone-Rod Dystrophies

Definition

Cone-rod dystrophies are marked by unusually progressive loss of ► [cones](#) of the retina followed by photophobia, loss of colour vision and visual acuity; usually rods that are neighbouring cones in the retina are affected later in life by a so-called “by-stander-effect”.

- [Retinitis Pigmentosa](#)

Cones

Definition

Cones are one type of photoreceptor cells in the retina (see also ► [rods](#)). Compared to the rods, cones have a shorter and cone-like outer segment. There are three variations of cones in the human retina with photopigment of either 555 nm (long wavelength sensitive cone), 535 nm (middle wavelength sensitive cone), and 445 nm (short wavelength sensitive cone). Cones have their highest density (70.000 per square mm) in the macula, and they are responsible for colour vision and fine spatial resolution.

► [Retinitis Pigmentosa](#)

Configuration

Definition

The configuration of a molecule comprises of all the orientations of stereogenic centers in a molecule that can exist in different configurations. In molecular mechanics, a configuration is defined by the complete set of coordinates and momenta of all particles in the system.

► [Molecular Dynamics Simulation in Drug Design](#)

Configuration Space

Definition

Configuration space refers to the entirety of configurations accessible to a system.

► [Molecular Dynamics Simulation in Drug Design](#)

Confluent Monolayer/Confluence

Definition

Confluent monolayer refers to cells in tissue culture, e.g. epithelial cells, which form a cohesive sheet comprising of a single cell layer filling the entire surface area of the bottom of the culture dish. Cells have then reached confluence.

► [Desmosomes](#)

Confocal

Definition

Confocal refers to an optical imaging scheme that results in very small measurement volumes. The focal spot of a high resolution objective is projected onto a small field aperture (pinhole) in the image plane, thereby restricting z-resolution and suppressing out-of-focus light.

► [Confocal Laser Scanning Microscope](#)

► [FCS](#)

Confocal Laser Scanning Microscope

Definition

Confocal Laser Scanning Microscope (CLSM) is an imaging instrument that selectively excludes image noise above and below a plane of focus. This is accomplished by filtering out-of-focus information with a pinhole aperture. A field of view is selectively illuminated by a laser beam that is processively moved across the field by rotating mirrors.

► [FRAP](#)

Conformation

Definition

Conformation defines different spatial arrangements of the atoms in a molecule that may arise by rotation around the bonds. The 3D-coordinates of all atoms are completely characterized by the conformation. This also contains information about the stereochemistry (configuration) of all stereotopic atoms. The conformation of a biological molecule is closely connected with its function. The transition from one conformation to another occurs without any breaking of covalent bonds. To account for the dynamic nature of proteins, the term conformation is sometimes used in a wider sense to refer to an ensemble of arrangements that interconvert rapidly at ambient temperatures, and are therefore distinguishable.

► [Genetic Code](#)

► [Molecular Dynamics Simulation in Drug Design](#)

► [Protein Folding](#)

► [QSAR](#)

Conformation of Double-Stranded DNA

Definition

The B-form DNA is the most common conformation of right-handed, double-stranded DNA, and contains 10 base pairs per turn perpendicular to the DNA axis.

► [Protein/DNA Interaction](#)

locus. Subsequent analysis of the component phenotypes expressed in each of the resultant congenic strains, potentially allows a detailed characterization of the disease component contributed by each susceptibility gene. The focus is thus simplified from studying a polygenic model to examining a series of essentially monogenic models.

► [SLE Pathogenesis Genetic Dissection](#)

Conformational Analysis

Definition

Conformational analysis is the exploration of the accessible conformations of a molecule, frequently with the task of identifying the energetically most favorable conformation.

► [Molecular Dynamics Simulation in Drug Design](#)

Congenic Strain

Definition

A congenic strain bears one or more genetic intervals on selected chromosomes (usually including disease susceptibility loci) that have originated from another strain. Typically, a congenic strain may carry (bear) specific disease susceptibility loci on an otherwise normal genetic background.

► [SLE Pathogenesis Genetic Dissection](#)

Conformational Order

Definition

Conformational order refers to a non-random distribution of protein conformations in an ensemble of protein species, without requiring that the spatial arrangement of atoms be similar in all conformations of the ensemble.

► [Protein Disulfide Bonds](#)

Definition

Congenital describes properties or a defect that are present from birth onward, but is not necessarily genetic.

► [Bone and Cartilage](#)

► [Heritable Skin Disorders](#)

Congenic Dissection

Definition

Congenic dissection is a strategy for identification of disease causing genes. In animal models, congenic dissection separates the multiple genes mediating a polygenic disease into a collection of ► [congenic strains](#), each carrying a different disease susceptibility

Congenital Disorders in Glycosylation

Definition

CDGs (Carbohydrate-deficient Glycoprotein Syndromes) are a newly discovered group of inherited disorders, also called carbohydrate-deficient glycoprotein syndromes (CDGSs). The current identified types of CDG are Type 1a, 1b, 1c, 1d, 1e and IIa. Most of these types have a special physiognomy, neurological problems, liver and/or intestinal problems. CDGs are characterized by a deficiency of glycans or glycan structures in secretory glycoproteins, lysosomal enzymes, and probably also membrane glycoproteins.

► [Glycosylation of Proteins](#)

Congenital Dyserythropoietic Anemia

Definition

Congenital dyserythropoietic anemia defines a heterogeneous group of inborn anemias, in which formation of red blood cells in the bone marrow is abnormal.

► [Hemochromatosis](#)

swelling. Causes include chronic hypertension, cardiomyopathy and myocardial infarction.

► [Proteomics in Cardiovascular Disease](#)

Congenital Heart Disease

Definition

Congenital heart disease describes a hereditary disease affecting the development or function of the heart.

► [Drosophila Model of Cardiac Disease](#)

Congressional Movement

Definition

Congressional movement designates the movements of chromosomes during the prometaphase as they congregate toward the equatorial plate of the Mitotic Spindle. The chromosomes oscillate between the cell equator and the spindle poles as mitotic forces pull against the sister kinetochores, eventually aligning them midway between the spindle poles.

► [Centromeres](#)

► [Mitotic Recombination](#)

► [Mitotic Spindle](#)

Congenital Hypomyelinating Neuropathy

► [Hereditary Neuropathies, Motor and/or Sensory](#)

Conjugal Pair

Definition

Conjugal pair describes both parents that are affected by the disease of interest.

► [Genetic Predisposition to Multiple Sclerosis](#)

Congenital Myasthenic Syndromes

Definition

Congenital Myasthenic Syndromes (CMS) are a heterogeneous group of diseases caused by genetic defects affecting neuromuscular transmission.

Connective Tissue

Definition

Connective tissue is a general term for mesodermally derived tissue, which is rich in extracellular matrix (collagen, elastic fibers, proteoglycan etc.) and surrounds other more highly ordered tissues and organs.

► [Marfan Syndrome](#)

Congestive Heart Failure

Definition

Congestive heart failure is a condition of ineffective pumping of the heart leading to an accumulation of fluid in the lungs and venous system. Typical symptoms include shortness of breath with exertion, difficult breathing when lying flat, and leg or ankle

Connexin Channelopathies

Definition

Connexin channelopathies are mutations in ► [connexins](#) detected at the genomic and proteomic level that lead

to pathologies caused by gap junction channel dysfunction.

- ▶ CMT
- ▶ Intermediate Filaments

Connexins

Definition

Connexins comprise of a multigene family of membrane proteins with over 20 members identified in human and murine genomes. Connexins oligomerise into hexameric connexons with the subunits arranged around a central pore, forming ▶ [gap junctions](#) between adjacent cells.

- ▶ Gap Junctions
- ▶ Intermediate Filaments

Consanguinity

Definition

Consanguinity refers to a genetic relationship. Consanguineous individuals have at least one common ancestor in the preceding few generations.

- ▶ [High-HDL Syndrome](#)

Consensus Map

Definition

Consensus map denotes the location of all consensus sequences in a series of multiple aligned proteins or polynucleotides.

- ▶ [Protein Databases](#)

Consensus Sequence (Sequon)

Definition

Consensus sequence (sequon) defines the most commonly occurring amino acid or nucleotide at each

position of an aligned series of proteins or polynucleotides. Within a polypeptide, the consensus sequence is commonly found to direct a specific post-translational modification at one of the residues within the sequon.

- ▶ [Glycosylation of Proteins](#)
- ▶ [Protein Databases](#)

Conservation

Definition

Conservation is the substitution of one amino for another that preserves the physicochemical properties of the original residue. For example, when a hydrophobic amino acid residue is replaced by another hydrophobic residue.

- ▶ [Protein Databases](#)

Conserved Synteny

Definition

Conserved synteny refers to a linkage of two orthologous genes that is maintained in the different species.

- ▶ [Mutagenesis Approaches in Medeka](#)

Constitutive Expression

Definition

Constitutive expression means the continuous transcription of a gene in an organism.

- ▶ [Recombinant Protein Production in Mammalian Cell Culture](#)

Constitutive Mutation

Definition

Constitutive mutation describes a change in the genetic information resulting in the inability of a gene product to be positively or negatively regulated, thus it cannot be turned on or turned off, respectively.

- ▶ [Bone and Cartilage](#)

Constitutive Splicing

Definition

Constitutive splicing designates a mandatory inclusion of an exon in mRNA.

► [Alternative Splicing](#)

Contact Inhibition

Definition

Contact inhibition refers to growth arrest that is induced in cultured cells that have reached confluence on a cell culture dish. Cultured transformed cells or tumour cells have lost this property.

► [Ras Signalling](#)

Contact Sites

Definition

In the more general cell biological sense, contact sites are those parts of cell membranes where interacting/neighbouring cells contact each other (e.g. contact site between antigen presenting cells and T-cells).

In mitochondria, contact sites are ultrastructurally defined sites where the inner and the outer membrane of mitochondria are observed in direct apposition to each other, with no apparent space in-between. Such contacts may be necessary for a number of central mitochondrial functions such as the channeling of metabolites, coordinated fusion and fission of mitochondria, and protein transport into mitochondria.

► [Mitochondria – Biogenesis and Structural Organization](#)

Contig Maps

► [YAC and PAC Maps](#)

Contiguous Gene Syndromes

► [Microdeletion Syndromes](#)

Contractures

Definition

Contractures describe loss of joint motion due to changes in the muscles ligaments and tendons.

► [Duchenne Muscular Dystrophy](#)

Convergence

Definition

For a property monitored during simulations in the context of structure-based drug design, convergence is reached when its trajectory average neither displays a drift as a function of time, nor erratic behavior due to the occurrence of rare events affecting its value.

► [Molecular Dynamics Simulation in Drug Design](#)

Conversion Technology

Definition

Conversion technology means conversion of diploid human chromosome complement to a haploid state through fusion to rodent cells.

► [Hereditary Nonpolyposis Colorectal Cancer](#)

Convolution

Definition

The mathematical operation that creates a function $C(x)$ by summing the overlap of one function, e.g. $f(x')$ with a reversed and shifted second function $g(x-x')$. This operation for example generates a function that describes a crystal when one function describes the ideal crystal lattice and the second function describes the content of the unit cell.

► [X-Ray Crystallography—Basic Principles](#)

Cooperativity

Definition

Cooperativity of ligand binding is the phenomenon that binding constants of ligands can be influenced positively or negatively by the presence of a ligand. Increase in the binding constant is called positive cooperativity, decrease negative cooperativity.

► [Differential Scanning Calorimetry](#)

COPD

Definition

COPD comprises of a group of lung diseases characterized by limited airflow with variable degrees of air sack enlargement and lung tissue destruction. Emphysema and chronic bronchitis are the most common forms of chronic obstructive pulmonary disease.

► [COPD and Asthma, Genetics](#)

COPD and Asthma, Genetics

JIAN-QING HE, ANDREW SANDFORD
University of British Columbia McDonald Research
Laboratories / iCAPTURE Center, Vancouver, Canada
jhe@mrl.ubc.ca, asandford@mrl.ubc.ca

Synonyms

Chronic obstructive pulmonary disease (COPD): chronic obstructive lung disease, chronic obstructive airway disease, chronic airflow limitation, Asthma: bronchial asthma.

Definition

Chronic obstructive pulmonary disease (COPD) is characterized by decreased maximal expiratory flow. The Global Initiative on Obstructive Lung Disease (GOLD) has defined COPD as the progressive development of airflow limitation that is not fully reversible. COPD encompasses chronic obstructive bronchitis with obstruction of small airways and emphysema, with enlargement of air spaces due to destruction of lung

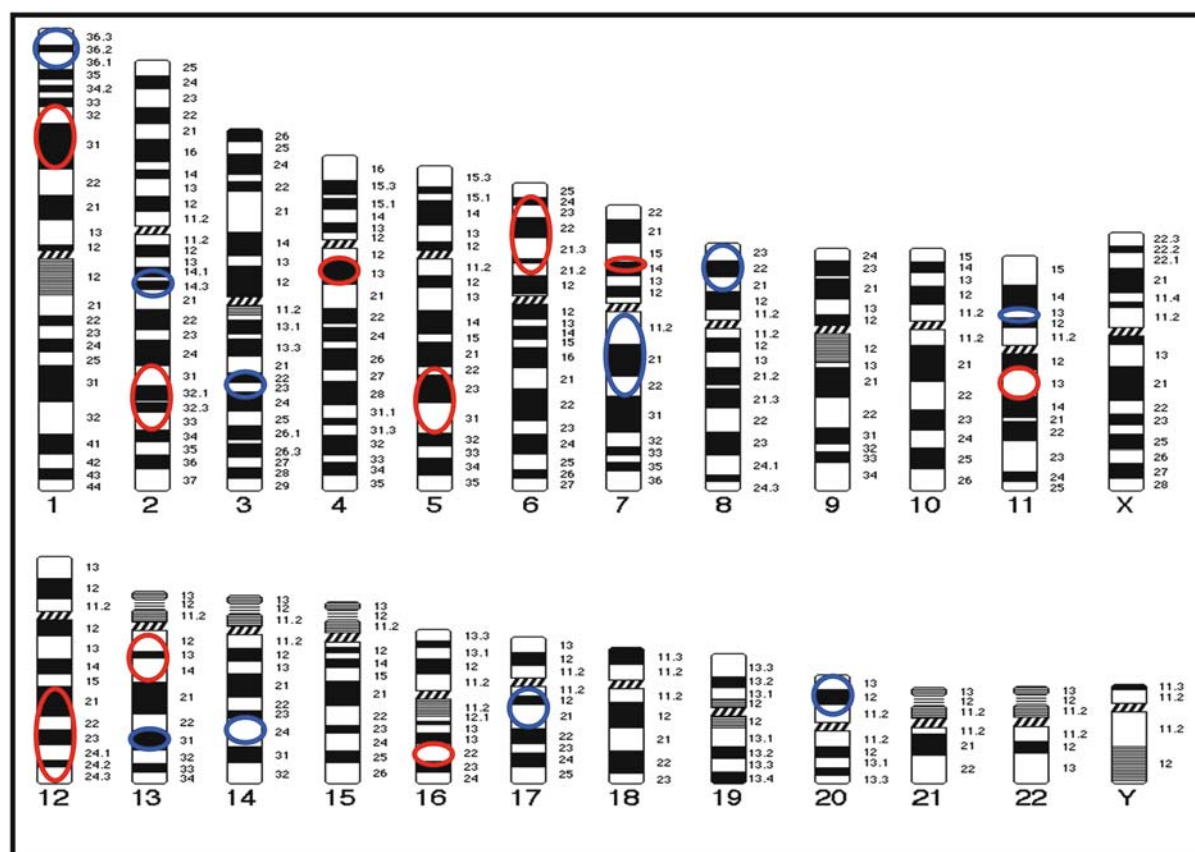
parenchyma and loss of lung elasticity. The airflow obstruction may be accompanied by airway hyperreactivity that is partially reversible. Emphysema is defined as a progressive destruction of alveolar wall leading to enlargement of airspaces without obvious fibrosis and with loss of normal architecture. Chronic bronchitis is defined as the presence of a productive cough that cannot be explained by other causes for at least three months in each year over two consecutive years. The major environmental risk factor for COPD is cigarette smoking. Asthma is a chronic inflammatory pulmonary disorder characterized by generalized reversibility of airflow obstruction caused by a multiplicity of stimuli. It is manifested physiologically as a widespread narrowing of the air passages, which may be relieved spontaneously or as a result of therapy and clinically by paroxysms of shortness of breath, cough and wheezing. Both COPD and asthma are ► [complex diseases](#), because they are likely to result from the interactions of multiple genetic and environmental factors. The genetics of COPD and asthma involve the study of the function, ► [polymorphisms](#) and expression of genes involved in ► [phenotypes](#) associated with these diseases.

Characteristics

Evidence of Genetic Background

Severe α_1 -antitrypsin (α_1 AT) deficiency, which follows a simple ► [Mendelian inheritance](#) pattern, has been known to be a genetic risk factor for COPD for four decades. However, convincing evidence of other genetic factors for COPD has been provided by epidemiological studies (1). There is an increased risk of COPD within the families of COPD ► [probands](#) but without clear Mendelian inheritance. Lower forced expiratory volume in one second (FEV_1), chronic bronchitis and COPD are more prevalent among the first-degree relatives of cases after correction for other risk factors such as smoking habits and α_1 AT deficiency. Twin studies have found estimates of ► [heritability](#) for FEV_1 that range from 0.5–0.8. The prevalence of COPD and similarity in lung function decrease with increased genetic distance.

The evidence for a genetic risk for asthma comes from similar epidemiological studies. First, asthma prevalence is especially high in individuals with a positive family history of asthma. Second, twin studies show that the concordance rates for asthma are higher in identical twins than in non-identical twins, irrespective of environmental risk factors. However, the fact that concordance in identical twins is not complete and the significant increase in incidence of asthma during the past two decades indicate that other factors than genetics play a major role in asthma pathogenesis. Currently, it is believed that both genetic and environmental factors contribute to asthma, with a heritability of about 60%.



COPD and Asthma, Genetics. Figure 1 Results of genome screens for asthma and its associated phenotypes. Regions that have been replicated are shown by red circles, regions that have not been replicated are shown by blue circles.

Characteristics of Genetic Studies of COPD and Asthma

Two major approaches have been used to identify susceptibility genes, full **genome screens** and **candidate gene association studies**. A genome screen involves searching the entire **genome** for regions that harbor disease-susceptibility genes by **linkage analysis** using affected families. Two advantages of genome screens are that novel genes can be identified in the pathogenesis of a disease and the approach is not confounded by **population stratification**. However, one disadvantage is the requirement for family data. This requirement makes genetic study of COPD difficult because of the late age of onset of the disease. Only two groups have reported full genome screens in COPD and related phenotypes. However, there have been more than 11 full genome screens for asthma and its related phenotypes (2). There is considerable consensus concerning regions of genetic **linkage** that are relevant to asthma. The linkages that have been replicated are 1p31-36, 2q4-32, 4q13, 5q23-31 (the cytokine gene cluster), 6p21-24 (the major

histocompatibility complex), 7p14, 11q13-31 (near the β chain of the high affinity receptor for IgE), 12q21-24, 13q12-14 and 16q21-23. Three regions of **linkage**, which although not replicated are statistically highly significant ($p < 0.001$), are 3q21-22, 14q24 and 17q12-21. The results of the eleven genomic screens are shown in Fig. 1 with the regions that have and have not been replicated highlighted. Recent studies have shown that evidence of **linkage** to asthma could depend on exposure to an environmental factor, such as environmental tobacco smoke.

Most of our knowledge of COPD genetics is derived from candidate gene association studies. Candidate genes for COPD include genes involved in protease-antiprotease and oxidant-antioxidant balance, inflammation and airway defense. Table 1 lists candidate genes that have been associated with COPD. In addition to severe α_1 AT deficiency, the results of different association studies have identified heterozygosity for α_1 AT, as well as polymorphisms in the tumor necrosis factor alpha (*TNF*), vitamin D binding protein (*GC*), microsomal epoxide hydrolase

COPD and Asthma, Genetics. Table 1 Candidate genes that have been associated with COPD. Bold text indicates those associations that have been replicated

Category	Candidate gene (symbol)	Phenotype
Antiprotease	α_1-antitrypsin (<i>SERPINA1</i>)	COPD, emphysema, rate of decline of lung function, early onset of COPD
	Tissue inhibitors of metalloproteinases-2 (<i>TIMP2</i>)	COPD
	α_1-Antichymotrypsin (<i>SERPINA3</i>)	COPD
	α_2 -Macroglobulin (<i>A2M</i>)	COPD
Protease	Matrix metalloproteinase-1 (<i>MMP1</i>)	Rate of decline of lung function
	Matrix metalloproteinase-9 (<i>MMP9</i>)	Emphysema
	Matrix metalloproteinase-12 (<i>MMP12</i>)	Rate of decline of lung function
Antioxidant and xenobiotic metabolizing enzyme	Heme oxygenase 1 (<i>HMOX1</i>)	Emphysema
	Microsomal epoxide hydrolase (<i>EPHX1</i>)	COPD, emphysema, rate of decline of lung function
	Glutathione S-transferase M1 (<i>GSTM1</i>)	Emphysema, chronic bronchitis, Lung function growth
	Glutathione S-transferase P1 (<i>GSTP1</i>)	COPD, Lung function growth
	Cytochrome P4501A1 (<i>CYP1A1</i>)	Emphysema
Inflammatory mediator	Tumor necrosis factor, alpha (<i>TNF</i>)	COPD, emphysema, Chronic bronchitis
	Vitamin D Binding Protein (<i>GC</i>)	COPD
	IL-1 β /IL1RN haplotypes	Rate of decline of lung function
	Interleukin-13 (<i>IL13</i>)	COPD
Airway defense	Cystic fibrosis transmembrane conductance regulator (<i>CFTR</i>)	COPD
	Beta defensin-1 (<i>DEFB1</i>)	COPD
	Surfactant proteins A, B and C (<i>SFTPA, B, C</i>)	COPD
Others	ABO blood group antigen secretor status (<i>ABO</i>)	Lung function, COPD

(*EPHX1*) genes and *ABO* blood group antigen secretor status as likely susceptibility factors for COPD. As for asthma, more than 200 association studies have been conducted, and associations that have been replicated with asthma and asthma-related phenotypes are listed in Table 2. Among them, the interleukin-4 (*IL4*)/*IL13* pathway is frequently reported as being associated with asthma or asthma-related phenotypes; these genes include *IL4*, *IL13*, *IL4RA* and signal transducer and activator of transcription 6 (*STAT6*). In addition, arachidonate 5-lipoxygenase (*ALOX5*) and granulocyte-macrophage colony-stimulating factor (*CSF2*) polymorphisms have

been associated with asthma, but these associations have yet to be replicated. Recently, two novel genes, a disintegrin and metalloproteinase domain 33 (*ADAM33*) and PHD finger protein 11 (*PHF11*) genes have been identified as asthma genes by whole genome screens, followed by fine mapping and association studies (3, 4). Van Eerdewegh and colleagues (3) performed genetic [linkage analysis](#) on 460 pairs of siblings from affected families and identified a locus on chromosome 20q, which was not linked to asthma in previous whole genome screens but was linked to asthma and bronchial hyperresponsiveness in this population. They assessed 135 SNPs of 23

COPD and Asthma, Genetics. Table 2 Candidate genes for asthma and related phenotypes that have been replicated

Chromosome location	Candidate gene (symbol)	Phenotype
5q23	Interleukin-4 (IL4)	Atopy, asthma, IgE
5q31	Interleukin-13 (IL13)	Asthma, IgE
5q31	Beta2-adrenoceptor (ADRB2)	Asthma, IgE, BHR, treatment response
5q31.1	Monocyte differentiation antigen CD14 (CD14)	IgE
5q35	Leukotriene C4 synthase (LTC4S)	Asthma
6p21.3	HLA DRB1	Atopy, allergy, asthma, IgE
6p21.3	HLA DQB1	Atopy, asthma, IgE
6p21.3	HLA DPB1	Asthma, IgE
6p21.3	Tumor necrosis factor (TNF)	Asthma
6p12.3	Platelet-activating factor acetylhydrolase (PLA2G7)	Asthma
11q13	β chain of the high affinity receptor for IgE (MS4A2)	Atopy, asthma, IgE
11q23.1	Clara cell-specific 16-kd protein (SCGB1A1)	Asthma
12q24.2	Nitric oxide synthase 1 (NOS1)	Asthma
12q13	Signal transducer and activator of transcription 6 (STAT6)	Asthma
16p12.1	Interleukin-4 receptor A (IL4RA)	Atopy, asthma, atopic dermatitis

genes in this region and identified the *ADAM33* gene as associated with asthma. Zhang and colleagues localized the underlying quantitative-trait locus in a comprehensive SNP map, based on a previously identified association between total IgE level and a 13q34 [▶microsatellite](#). They replicated the association to IgE and attributed it to several [▶alleles](#) in a single gene, *PHF11*. These alleles are associated with asthma, especially severe asthma (4).

The main problem of the association study approach has been lack of replication. When comparing association studies, three possibilities need to be borne in mind, a false negative study due to lack of power, a false positive original report and true differences between study populations. Three common reasons for inconsistency in association studies cannot be overemphasized. Firstly, there is a risk of false positive and false negative results due to population stratification. Recently, several approaches such as “[▶genomic controls](#)” have been advocated to attempt to correct for population stratification (5). Secondly, multiple comparisons inevitably result in false positive associations; this is particularly true for asthma association studies, since numerous asthma-related phenotypes have been reported. Thirdly, different phenotypes of COPD and

asthma in different studies also contribute to the difficulties of comparison among association studies.

Cellular and Molecular Regulation

DNA sequence variations in the human genome are numerous and there are different ways to categorize them. Variations can be divided to two categories according to frequency, variants and polymorphisms. Variants are variations with a minor allele frequency less than 1%, while polymorphisms are those with a minor allele frequency equal to or more than 1%. When classified according to mechanisms, the most common variations are [▶single nucleotide polymorphisms](#) (SNPs). On average, SNPs occur about every 500–1,000 base pairs. Other types of variations such as differences in repeat number, [▶insertions](#), [▶deletions](#) and duplications also occur but are much less frequent. DNA sequence variations can cause disease by various mechanisms. A common mechanism is loss of function of the protein. A classical example is severe α_1 AT deficiency and emphysema. α_1 AT is the most abundant circulating proteinase inhibitor. α_1 AT is a highly polymorphic gene with over 100 variations. Most of these are the results of SNPs that lead to single or sometimes double amino acid changes. Based on their

migratory distance on isoelectric focusing analysis, they are categorized as normal α_1 AT that migrates in the middle of the gel (M allele) and S and Z variations that migrate more quickly owing to changes in their overall charge. S and Z alleles are caused by amino acid substitutions from Glu to Val at amino acid position 264 and from Glu to Lys at amino acid position 342, respectively. In the Caucasian population, the frequencies of M, S and Z alleles are >95%, 2–3% and 1%, respectively. A small percentage of people inherit a null allele, which leads to complete absence of α_1 AT production. Individuals with MS and MZ **▶genotypes** have ~80% and 60% of normal α_1 AT levels, respectively. Heterozygous PI SZ is rare and individuals with this genotype have α_1 AT levels ~40% of normal. Individuals with two Z alleles or one Z and one null allele are referred to as PI Z. PI Z individuals have approximately 15% of normal plasma α_1 AT levels as 85% of the protein is retained within the rough endoplasmic reticulum of the hepatocyte. The reduced plasma levels are inadequate to protect the lung from the attack of proteinases. Thus the PI Z individuals are susceptible to early onset emphysema. This observation prompted the proteinase-antiproteinase hypothesis of lung injury in emphysema, which is still a dominant view of COPD pathogenesis.

Some sequence variations cause disease through a gain of function, where the protein takes on some new, deleterious or enhanced function. Examples include the coding region SNPs of the IL4 receptor alpha (*IL4RA*) subunit and asthma. The combination of the V75 allele with R576 resulted in expression of an IL4RA with enhanced sensitivity to IL4. There was also a significant association of V75/R576 with atopic asthma.

Many functional alterations are located in the coding sequence of genes, which influence the expression of a phenotype as discussed above. However, DNA sequence variations in regulatory region of genes may also change function. For example, a (GT)_n dinucleotide repeat in the heme oxygenase-1 (*HMOX1*) gene shows length polymorphism and could regulate the level of gene transcription. A large size (GT)_n repeat may reduce *HMOX1* inducibility by reactive oxygen species in cigarette smoke, resulting in the development of emphysema. Another example is the T allele of the C-589T polymorphism in the *IL4* **▶promoter** that was associated with increased luciferase reporter gene activity. The T allele was also associated with increased total IgE, asthma or asthma severity.

DNA sequence variations in other regions of genes also could change function. For example, SNPs within introns and the 3' untranslated region (3'UTR) may affect **▶alternative splicing**, splicing efficiency or messenger RNA turnover. For example, several of the SNPs in the *ADAM33* gene that were associated with

asthma were located in non-coding regions. However, there are no functional data available as to whether the identified sequence variations result in functional changes (3).

Clinical Relevance

Studies of genetic factors that could potentially play a role in determining treatment response led to the emergence of pharmacogenetics and later to pharmacogenomics. Furthermore, a deeper understanding of the genetics that underlies disease pathogenesis will lead to new therapeutic approaches.

Pharmacogenomics uses genome-wide approaches to elucidate the inherited basis of differences between persons in responses to drugs. Genetic polymorphisms could play a direct or indirect roles in drug responses. The best examples are the β_2 adrenoreceptor (*ADRB2*) and *ALOX5* genes.

The human *ADRB2* is an intronless gene on the long arm of chromosome 5 (5q31-33). There are at least seventeen SNPs within the promoter and coding regions. Four single amino acid substitutions have been associated with altered expression, down-regulation, or coupling of the receptor in response to β_2 -agonists. *In vitro* studies have indicated that the *ADRB2* SNPs are associated with physiologically relevant effects. *In vivo*, a study documented that Arg16 homozygotes had almost complete desensitization after continuous exposure to a β_2 -agonist, with vasodilation decreasing more than four fold after 1.5 hours of exposure, whereas Gly16 homozygotes had no significant desensitization. Similarly, another report demonstrated that in patients who receive long-term inhaled β_2 -agonist, those with the homozygous Arg16 genotype had a gradual decline in morning lung function, which was not seen in those with the homozygous Gly16 genotype. These data are consistent with a clinical study that asthmatics homozygous for the Arg16 allele showed a significantly increased number of exacerbations when treated with β_2 -agonist. Individuals with the Arg16 homozygous genotype have therefore been shown to have an altered pharmacologic response to β_2 -agonists. The genotype may be used to predict patients' response to long term therapy with inhaled β_2 -agonist and for those patients predicted to have no beneficial response, other therapeutic modalities may be chosen to avoid any delay in the optimal treatment.

An insertion/deletion polymorphism that affected Sp1 and Egr-1 transcription factor binding sites was found in the 5-lipoxygenase (*ALOX5*) promoter region. The wild type allele consisted of five tandem repeat Sp1 binding sites. Reporter gene experiments demonstrated that alleles consisting of a deletion of one or two binding sites or addition of a sixth binding site resulted in reduced gene expression. It was shown that

individuals with decreased 5-lipoxygenase expression represented a subset of asthma patients that responded poorly to the 5-lipoxygenase inhibitor. Those asthma patients with three, four or six tandem repeats responded poorly to a 5-lipoxygenase inhibitor. This was confirmed by clinical data. Individuals receiving high-dose 5-lipoxygenase inhibitor were genotyped for the promoter polymorphism. The patients who did not have a wild type 5-repeat allele showed no benefit from 5-lipoxygenase inhibitor in terms of percentage change in FEV₁ following treatment. The explanation was that leukotrienes play little role in modulating asthma in those patients who have the mutant promoter and thus they do not benefit from treatment with leukotriene synthesis inhibitors.

It is anticipated that the study of the genetics of COPD and asthma will lead to novel or better understanding of mechanisms of these diseases, thereby, leading to the identification of different types of COPD and asthma. Classification of COPD and asthma according to their underlying pathogenesis may lead to new and better treatments and may help to tailor the most appropriate individual treatments in the future.

References

1. Sandford AJ, Pare PD (2000) Genetic risk factors for chronic obstructive pulmonary disease. *Clin Chest Med* 21:633–643
2. Cookson W (2002) Genetics and genomics of asthma and allergic diseases. *Immunol Rev* 190:195–206
3. Van Eerdewegh P, Little RD, Dupuis J et al (2002) Association of the ADAM33 gene with asthma and bronchial hyperresponsiveness. *Nature* 418:426–430
4. Zhang Y, Leaves NI, Anderson GG et al (2003) Positional cloning of a quantitative trait locus on chromosome 13q14 that influences immunoglobulin E levels and asthma. *Nat Genet* 34:181–186
5. Bacanu SA, Devlin B, Roeder K (2000) The power of genomic control. *Am J Hum Genet* 66:1933–1944

COPI Vesicles

Definition

COPI (coat protein I) vesicles are formed by the action of COPI coat proteins, which are involved in a retrograde transport pathway that recycles proteins and lipids between the *cis*-Golgi and the endoplasmic reticulum. The COPI coatomer is a complex of seven subunits (α , β , β' , γ , δ , ϵ and ζ). The γ -subunit is thought to be responsible for cargo recognition, the β -subunit has been shown to interact with ARF1. COPI coat proteins also have complex roles in transport within the Golgi complex and in maintaining the normal structure of the interphase Golgi complex.

- Exocytotic Pathway
- Protein and Membrane Transport in Eukaryotic Cells
- Vesicular Traffic

COPII Complex/COPII Vesicles

Definition

COPII (coat protein II) complex is involved in vesicular transport from the endoplasmic reticulum (ER) to the Golgi apparatus. COPII complex is necessary to form COPII vesicles, which are essential for the exit of newly synthesized proteins from the ER and their delivery to the Golgi complex. COPII vesicles are formed from ribosome-free (“smooth”) regions of the ER, at so-called ER exit sites, defined by COPII coat proteins.

- Nuclear Pore Complex
- Vesicular Traffic

Core Binding Factor

Definition

Core binding factor (CBF) is a heterodimeric transcription factor consisting of a DNA-binding α subunit of the runt domain family, which heterodimerizes with the non-DNA-binding β subunit. CBF regulates transcription from a number of genes involved in haematopoietic differentiation.

- Leukemia

Core Promoters

ANANDA L. ROY

Departments of Pathology and Biochemistry, Programs in Immunology and Genetics, Tufts University School of Medicine, Boston, MA, USA
ananda.roy@tufts.edu

Synonyms

Basal promoter

Definitions

Cell extrinsic signals are transduced to the nucleus through a series of complex biochemical steps, resulting in either amplification or attenuation of eukaryotic gene expression (1). The expression of genes is often controlled at the level of copying of information from

DNA to RNA called ►**transcription**. The process of transcription is regulated by regions of DNA called the ►**promoter** that lies outside the actual “message coding region” of the gene. The nucleus of a eukaryotic cell contains three DNA-dependent ►**RNA polymerase** enzymes, which carry out the actual process of transcription. The enzyme RNA polymerase I (Pol I) transcribes genes coding for ribosomal RNA (rRNA), RNA polymerase II (Pol II) transcribes genes coding for messenger RNA (mRNA) and RNA polymerase III (Pol III) transcribes genes coding for transfer RNA tRNA. Although the polymerase enzymes have the intrinsic ability to carry out transcription, they cannot do so in a specific manner unless they interact with a set of proteins that are commonly present in all nuclei and are called the ►**general transcription factors** (GTFs) or basal factors (2). These factors are necessary and sufficient (in addition to RNA polymerase) to form the transcriptional machinery at the core promoter and direct accurate initiation of transcription *in vitro* (2, 3). Each polymerase interacts with a set of GTFs that is unique to each class of genes, with the notable exception of ►**TATA binding protein** (TBP), which is commonly required for most if not all genes. The resulting “transcriptional or preinitiation complex” for each polymerase specifically recognizes a DNA region near the transcription start site (TSS) called the core promoter in the corresponding genes. The steps in transcription that follow this initial nucleation step include ►**initiation**, ►**promoter clearance**, ►**elongation** and finally ►**termination** (2, 3).

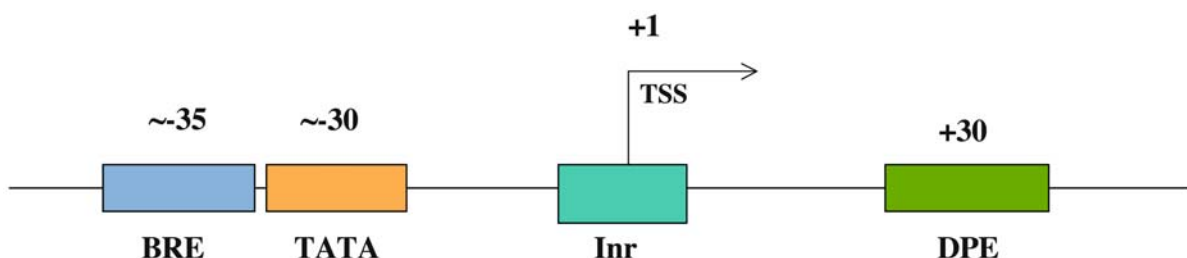
Characteristics

The promoter region of a eukaryotic protein-coding (mRNA) gene is arbitrarily divided into two segments, a proximal region (core promoter) and a distal region (enhancer) relative to the TSS (Fig. 1). The

sequence-specific, multiprotein transcriptional complex formation not only ensures high fidelity of gene expression but also allows for multiple points of regulation. Accordingly, it has been demonstrated that the gene-, lineage- or stage-specific ‘regulatory’ transcription factors interacting at enhancer elements can alter either the rate or the stability of transcriptional complex formation at the core promoter, thereby controlling the expression of the corresponding gene (2, 3). Although originally thought to be inert, it is now clear that the core promoter (and the corresponding transcriptional complex) directly contributes to spatial and/or temporal regulation of transcription. Hence, while a core promoter and enhancer elements can be swapped from different genes *in vitro* and in transient transfection experiments, a heterologous promoter may not faithfully recapitulate proper transcriptional response *in vivo* (4). A core promoter region in the protein-coding genes is comprised approximately of seventy nucleotides abutting the TSS and is defined as the minimal region of naked (non-chromatin) DNA required to direct accurate initiation of transcription *in vitro* when presented with the RNA polymerase and the corresponding GTFs (2, 3, 4). Although an independent core promoter may not exist *in vivo*, it nevertheless directs accurate transcription initiation from DNA templates *in vitro*. Hence, it has been particularly useful to define the minimum sets of GTFs or basal factors necessary to form the transcriptional complex (also called the ►**transcriptional machinery**) and ascertain the sequence of events that result in the preinitiation complex assembly (2).

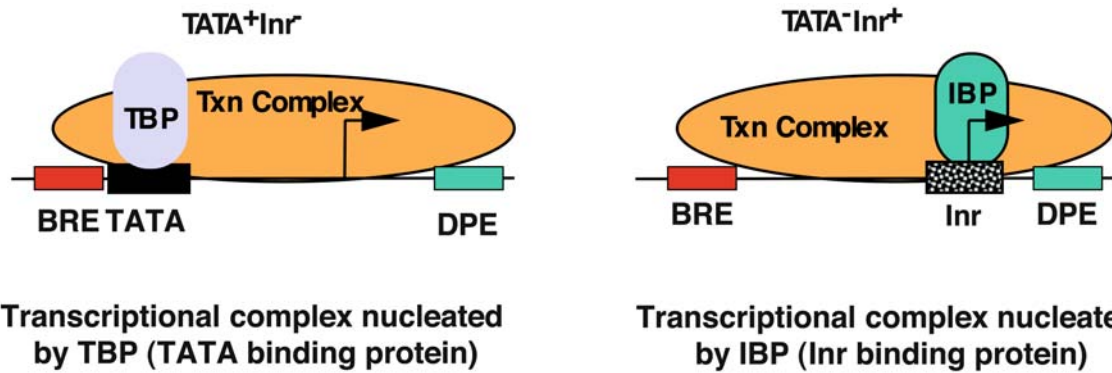
Core Promoter Elements

There are several key genetic elements within a core promoter. They are the TFIIB recognition element (►**BRE**), the ►**TATA box**, the pyrimidine rich initiator (►**Inr**) element and the downstream promoter



Eukaryotic Core Promoter Elements

Core Promoters. Figure 1 Eukaryotic core promoter elements and their position with respect to the transcription start site (TSS) indicated by the bent arrowhead. These are TFIIB recognition element (BRE), TATA box, initiator element (Inr) and downstream promoter element (DPE). Different classes of promoters may have only a subset of these elements allowing for greater combinatorial responses.



Core Promoters. Figure 2 Mechanism of formation of the transcriptional (Txn) complex in TATA-containing, Inr-less (TATA⁺Inr⁻) or TATA-less, Inr-containing (TATA⁻Inr⁺) core promoters. While binding of TBP initiates the nucleation event in TATA⁺Inr⁻ promoters, binding of IBP is believed to be the nucleation event in TATA⁻Inr⁺ promoters. The IBP could be TAFs or independent proteins like TFII-I or RNA pol II. For the sake of simplicity binding of factors to the BRE and DPE is not shown.

element (►DPE) (Fig. 1) (5). Among these, the TATA box and Inr element are more extensively characterized.

TATA Box

The TATA box (also known as the Goldberg-Hogness box) is usually present approximately 30 bases upstream of the transcription start site of a protein-coding gene and has the consensus sequence of either TATAAA or TATATA. The TATA box is conserved in all eukaryotes as well as in the archaea. In yeast, the position of the TATA box is variable and it can be situated 40–120 bases upstream of the transcription start site. Although initially thought to be highly conserved and required for all eukaryotic protein-coding genes, recent database analysis suggests that in fact more than 60% of eukaryotic genes lack a TATA box (5). However, when present, it directs accurate initiation of transcription (start site selector); mutations in a TATA box result in severely decreased transcription both *in vitro* and *in vivo* (5). The TATA box is recognized by the TATA binding protein, TBP, which is a subunit of a larger protein complex, ►TFIID (Fig. 2). The TATA-mediated basal (activator independent) transcription starts with TATA recognition by the TBP, a step that is sufficient to nucleate the assembly of additional GTFs and RNA polymerase II into a functional transcriptional complex (4). Crystal structure analysis of TBP bound to the TATA box revealed that this binding induces a bend or loop in DNA such that the upstream (enhancer) and downstream (core promoter) regions are brought into closer proximity (2). This step ensures the proper communication between the distinct transcriptional components that is essential for physiologically relevant transcriptional responses. Besides TBP, TFIID also contains TBP associated factors,

called ►TAFs. TAFs serve multiple roles and generally behave as core promoter discriminatory factors (5). There appears to be considerable flexibility in the TATA sequence such that a number of A/T combinations are recognized by TBP and function as TATA boxes. Moreover, a TATA box also functions generally in a heterologous promoter context so that it can be swapped between promoters.

Inr Element

Many of the eukaryotic core promoters instead of a TATA box contain a pyrimidine-rich Inr element encompassing the TSS, the consensus of which is YYA⁺NT/AYY (where A⁺ signifies the transcription start site and Y stands for a pyrimidine base). The Inr element is well conserved among eukaryotic promoters and in archaea (5). It is also found in parasites. There are some viral promoters that tend to be strong promoters and contain both a TATA box and an Inr element (4). The two core promoter elements can act in a synergistic fashion when separated by only 25 to 30 base pairs. However, the Inr element can function independently and in a heterologous promoter context as well, at least *in vitro* (5). The TATA-less but Inr containing core promoters show relatively weak transcriptional activity in *in vitro* reconstituted systems, making it more difficult to ascertain the precise biochemical mechanism for Inr-directed basal transcription (2). Inr-mediated basal transcription appears to require several factors, including TBP associated factors (TAFs) that may not be required for TATA-directed basal transcription (2, 3, 4, 5). Several pathways have been proposed to account for Inr-dependent binding and transcription. One pathway involves direct recognition of the Inr by one or more

TAF components and the adjacent DPE when present, followed by stable TFIID binding and subsequent preinitiation complex assembly (2, 3, 4, 5). A second pathway proposes that recognition of the Inr by independent Inr binding proteins (IBPs, e.g. TFII-I and YY1), followed by secondary interactions with TFIID or associated factors nucleates assembly of the GTFs at the core promoter (3, 4, 5). The third pathway involves recognition of Inr by RNA polymerase II in the absence of both TAFs and IBPs (Fig. 2). These observations could reflect diversity in core promoter Inr elements and corresponding interactions, especially in the light of the loose consensus sequence for such elements (2, 3, 4, 5).

Downstream Promoter Element

The DPE was first found in *Drosophila* promoters and since then has also been shown to be present in core promoters of human genes. It appears to be predominantly present in promoters that lack a TATA-box (5). The DPE, which is present in the intragenic +30 region, functions in conjunction with the Inr element and a mutation in any one element results in greatly diminished basal transcriptional activity due to loss of TFIID binding. Hence DPE does not function as an independent core promoter element (5). The loose consensus of this element is A/GGA/TC/TG/A/C and it appears to be recognized by two of the TAF components of the TFIID complex (5).

TFIIB Recognition Element

This core promoter element, the function of which is not at present fully determined, is present upstream of the TATA box in mammalian and archaeal promoters (5). It is not clear whether it can act independently as a core promoter element in the absence of either a TATA box or an Inr. The BRE, whose consensus is G/CG/CG/ACGCC, is recognized by the GTF, TFIIB through its C-terminal end (5). Because the yeast and plant TFIIB lacks this part, BRE may not be important for gene regulation in these organisms. Interestingly, while BRE augments transcription in archaeal promoters, it inhibits transcription in eukaryotic promoters (5).

PSE and Other Putative Core Promoter Elements

Apart from the well-characterized core promoter elements described above for the protein-coding genes, core promoter elements are also present in genes transcribed by RNA pol III. The best described is the proximal sequence element (PSE) which is located approximately 50 nucleotides upstream of the transcription start site. PSE is recognized by a multiprotein complex and can function synergistically with a TATA box (5). There are also some promoters for protein-coding genes that lack both TATA and Inr elements.

Transcription initiation from these promoters is inaccurate and starts at multiple sites. It is currently unclear how the transcriptional machinery assembles at these promoters. An intragenic sequence motif, called the MED-1 (multiple start site element downstream) that is distinct from the DPE appears to play a significant role in some of these promoters (4, 5).

Clinical Relevance

The presence of multiple core promoter elements, especially in the protein-coding genes, suggests different mechanisms for the formation of the transcriptional complex. This flexibility in transcription initiation process allows for multiple regulatory inputs that are necessary for tight developmental control of higher eukaryotic organisms. Moreover, it also allows for rapid and variable response to extracellular signals that could be detrimental to the organism. Given that the core promoter serves as a “landing pad” for the transcriptional machinery, mutations in the core promoter elements or cognate factors can have profound physiological effects ranging from total loss of gene expression to abnormally high gene expression.

References

1. Levine M, Tjian R (2002) Transcriptional regulation and animal diversity. *Nature* 424:147–151
2. Roeder RG (1996) The role of general initiation factors in transcription by RNA polymerase II. *Trends Biochem* 21:327–335
3. Lee TI, Young RA (2000) Transcription of eukaryotic protein-coding genes. *Annu. Rev Genet* 34:77–137
4. Novina CD, Roy AL (1996) Core promoters and transcriptional control. *Trends Genet* 12:351–355
5. Smale ST, Kadonaga JT (2003) The RNA polymerase II core promoter. *Annu Rev Biochem* 72:449–479

Coreceptor

Definition

Coreceptor is a cell surface molecule which enables and modulates the interaction between ligand and receptor. For example, the T-cell surface molecules CD4 and CD8 are coreceptors that foster the interaction between the T-cell receptor and MHC molecules on antigen-presenting cells. Other coreceptors trigger membrane fusion and entry mediated by the Env protein of some retroviruses.

► Retroviruses

Corepressor

Definition

Corepressor is a small molecule that triggers repression of transcription by binding to a regulator protein.

► [Nuclear Hormone Receptors: Regulators of Gene Transcription and Cellular Signaling](#)

Corneal Arcus

Definition

Corneal arcus refers to a cloudy opaque arc or circle around the edge of the eye.

► [Familial Hypercholesterolemia](#)

Corneal Opacity

Definition

Corneal opacity refers to a fine deposit of granular dots in the layers of cornea, especially in the periphery. Usually, the cornea has a dull, pale, cloud-like appearance.

► [HDL](#)
► [Dyslipidemia](#)

Coronary Artery

Definition

Coronary arteries are the blood vessels which transport blood to the heart muscle.

► [Heart](#)

Corpus Callosum

Definition

Corpus callosum is a thick C-shaped midline brain structure consisting of (myelinated) nerve fibers connecting the right and left cerebral hemispheres.

► [Brain](#)
► [Hypothalamic and Pituitary Diseases Genetics](#)

Correlation

Definition

Correlations are informations that are obtained from NMR experiments revealing that two nuclei are either close in space or within the chemical structure.

► [3D Structure by NMR](#)

Correlation Analysis

Definition

Correlation analysis designates a general mathematical formalism to temporally compare signals with themselves (autocorrelation), or with each other (cross-correlation), with respect to the information they carry.

► [FCS](#)

Correlative Microscopy

Definition

Correlative Microscopy is a way to correlate fluorescence information with electron microscopy information. It requires a reporter system that provides a signal that can be detected both by light microscopic and electron microscopic imaging methods.

► [Electron Tomography](#)

Cortical Column

► [Neocortex/Cortical Column](#)

Cortical Granules

Definition

Cortical granules are small membrane-bound vesicles found in large numbers in the cortex of unfertilized eggs that contain a variety of lytic enzymes.

► [Mammalian Fertilization](#)

Cortical Reaction

Definition

Cortical reaction describes a fusion of cortical granules with egg plasma membrane, with the release of granule contents in response to fertilization and activation of eggs.

► [Mammalian Fertilization](#)

Cortical Tuber

Definition

Cortical tuber is a distinctive form of brain ► [hamartia](#) in which there is architectural disarray with disruption of cortical lamination and giant cells commonly seen in tuberous sclerosis.

► [Tuberous Sclerosis](#)

Corticotropin-Releasing Hormone

► [CRH](#)

Cosmid

Definition

Cosmid refers to a cloning vector that can be introduced into *E. coli* cells by *in vitro* packaging (up to 35–50 kb) into defective lambda bacteriophage and transfection, but maintained as plasmids.

► [C. Elegans Genome, Comparative Sequencing](#)

► [YAC and PAC Maps](#)

COSY

Definition

COSY refers to two-dimensional correlation spectroscopy. This technique provides correlations between

nuclear spins that are neighbours in the chemical structure. For biological macromolecules, COSY yields correlations for nuclei separated by up to three, and occasionally four chemical bonds.

► [3D Structure by NMR](#)

Co-Transcriptional Capping

Definition

Co-transcriptional capping denotes a reaction in which capping occurs together with transcription, most likely at the initiation step, as distinguished from pre- or post-transcriptional capping.

► [RNA Capping](#)

Co-Translational Modification

Definition

Co-Translational Modification is a protein modification occurring during translation of a protein, typically considered in the context of protein biosynthesis in the endoplasmic reticulum.

► [Glycosylation of Proteins](#)

Coulomb Electrostatics

Definition

Coulomb electrostatics describes the interaction between two point charges described by Coulomb's law, which states that the energy of the interaction is proportional to the magnitude of the charges and inversely proportional to their distance.

► [Molecular Dynamics Simulation in Drug Design](#)

Covalent Coupling

► [Protein Interaction Analysis: Chemical Cross-linking](#)

COX

Definition

COX stands for cytochrome c oxidase. It is an enzyme complex in the respiratory chain that is composed of protein subunits encoded by the mtDNA, as well as subunits encoded by the nuclear DNA.

► **Mitochondrial Myopathies**

CpG

Definition

CpG is a dinucleotide. Cytidine is joined by a 5' to 3' phospho-diester linkage to guanine as part of a normal DNA strand.

► **CpG Islands**

► **Heritable Skin Disorders**

CpG Islands

CHRISTOPH PLASS, LAURA J. RUSH
The Ohio State University, Columbus, USA
plass-1@medctr.osu.edu, rush.61@osu.edu

Synonyms

HpaII tiny fragments (HTFs); HTF islands

Definition

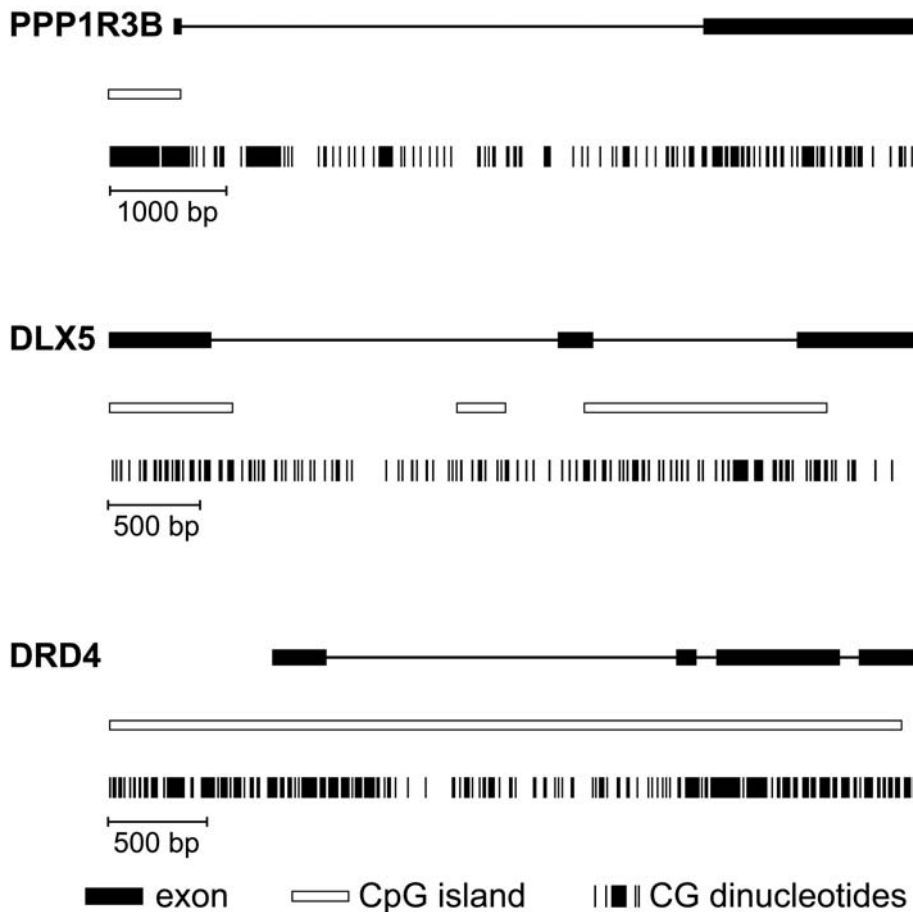
CpG islands are short stretches of DNA with an unusually high GC content and a higher frequency of ► **CpG** dinucleotides (cytosine 5' to a guanine, separated by a phosphodiester bond) compared to the rest of the genome. CpG islands account for about 1–2% of the genome and are usually located in the 5' regulatory regions of genes. They are associated with all ► **house-keeping genes** and up to 40% of ► **tissue-specific genes**.

Characteristics

With the accumulation of raw sequencing data, it became obvious that the four bases, adenine (A), cytosine (C), guanine (G), and thymine (T) are not evenly distributed in the genome. Normal DNA has an average GC content of 40% and an AT content of 60%, and contains only 25% of the number of CpG dinucleotides that would be expected. Early work by

Bird et al. in 1985 (1) identified stretches of genomic sequence characterized by an unusually high number of *HpaII* restriction sites (restriction site: C[^]CGG). These sequences were initially called “*HpaII* tiny fragments (HTFs)” or “► **HTF islands**”. Careful inspection of the sequences in the *HpaII* tiny fragments indicated that the ratio of CpG dinucleotides is higher than in the rest of the genome. These stretches of DNA with a high GC content and a frequency of CpG dinucleotides that is close to the expected value are now called CpG islands. The following three criteria, established by Gardiner-Garden and Frommer in 1987 (2), are commonly used to define CpG islands. First, the sequence must be at least 200 base pairs long. There is no upper limit on the length and some CpG islands are several kilobase pairs long. Second, the GC content is above 50%, while the rest of the genome is about 40%. Third, the CpG ratio (observed/expected) is above 0.6, while the rest of the genome has a CpG ratio of 0.2. Recently Takai and Jones (3) have developed more stringent criteria based on analysis of CpG islands in the completed drafts of chromosomes 21 and 22. These criteria (> 500 base pairs, GC content ≥55%, and CpG ratio of 0.65) allow for higher accuracy of predicting unknown genes while at the same time excluding ► **Alu repetitive elements**. The investigators have constructed a web-based program with which one can analyze any sequence of interest for CpG island characteristics (► <http://www.uscnorris.com/cpgislands/>). In addition, the program allows one to adjust the various criteria mentioned above, to apply more or less stringency as desired.

The human genome contains about 29,000 CpG islands (4) and it is estimated that there are somewhat fewer in the mouse (5). The majority of CpG islands are located in the 5' region (promoter and/or exon 1) of all housekeeping genes and a large number of tissue-specific genes. However, CpG islands are sometimes noted in the 3' ends of genes or intronic regions (Fig. 1). The preferential location of CpG islands in 5' regions of genes can be used for the identification of novel genes. Rare cutting ► **restriction enzymes** with GC rich recognition sequences such as *NotI* (GC[^]GGCCGC), *AscI* (GG[^]CGCGCC), *BssHII* (G[^]CGCGC) and *EagI* (C[^]GGCCG) can be used for the restriction mapping of large genomic clones. Clusters of those restriction enzyme cutting sites would indicate the presence of a CpG island. It is unknown why the mouse has fewer CpG islands than the human when the estimated number of genes in both genomes is expected to be very similar. One possible explanation is that the rate of CpG dinucleotide loss due to ► **deamination** (see below) is higher in the mouse than it is in the human. In general, CpG islands are present in the early replicating, less condensed and GC rich R-bands of chromosomes.



CpG Islands. Figure 1 Examples of various locations of CpG islands within genes. Shown are three genes with differing configurations of CpG islands. *PPP1R3B* has a CpG island located in the 5' region only. *DLX5* has three CpG islands interspersed throughout the length of the gene. A very large CpG island encompasses the entire *DRD4* gene. Exons are depicted by solid rectangles. CpG islands are depicted as open rectangles. CG dinucleotides are depicted as vertical hash marks.

Preservation of CpG Islands

The preservation of CpG islands in vertebrate genomes is inversely related to [DNA methylation](#) and a process called deamination. DNA methylation in vertebrate genomes is found mainly in CpG dinucleotides. CpG dinucleotides that are located within CpG islands are usually unmethylated, while CpG dinucleotides located outside of CpG islands are methylated at the 5' position of the cytosine ([5-methylcytosine](#)). 5-methylcytosine is sometimes referred to as the “5th base”. Methylation of CpG dinucleotides makes these sites vulnerable to spontaneous deamination leading to a transition of the 5-methylcytosine to thymine, a process that is believed to cause the depletion of CpG dinucleotides from the genome. Unmethylated CG dinucleotides in CpG islands are protected from deamination, resulting in CpG island preservation during evolution. However, a small fraction of CpG

islands are methylated in normal tissues as a result of important developmental processes.

Normal Developmentally Regulated CpG Island Methylation

- Most CpG islands on the inactive X-chromosome of females are densely methylated. This normal process of [X-chromosome inactivation](#) is linked to the transcriptional silencing of those genes and is thought to play a role in dosage compensation of the gene product. Examples of such genes include phosphoglycerate kinase 1 (*PGK1*), glucose-6-phosphate dehydrogenase (*G6PD*), and the androgen receptor (*AR*). Exceptions are found in a few CpG islands in genes that escape X-inactivation (e.g. *STS*, *ZFX* or *UBE1*).
- Some CpG islands become methylated in other normal processes including cell differentiation and

aging. The result of this methylation is the selective inactivation of genes in specific tissues or at certain developmental stages (e.g. estrogen receptor).

- Genes that are expressed exclusively from either the paternal or the maternal allele are called imprinted genes (►Imprinting). These genes have CpG island methylation on only one allele (e.g. *H19* and *IGF2*). While methylation usually occurs on the inactive allele, in some instances CpG island methylation has been found on the active allele. Allele-specific methylation in CpG islands was used as a tag for the identification of novel imprinted genes in the mouse using the restriction landmark genomic scanning (►RLGS) technique.

The transfer of a methyl group to a cytosine residue is carried out by DNA ►methyltransferase (Mtase) enzymes. Methylation of normally unmethylated CpG islands is important because of the effect this has on gene expression. Specifically, methylated CpG islands have been correlated with the transcriptional silencing of the associated genes. The detailed molecular process controlling this inactivation is not entirely clear. Protein complexes that contain *Sin3*, histone deacetylases and the methyl CpG-binding protein *MeCP2* bind to methylated promoters and induce the deacetylation of histones, which mediates the formation of transcription-repressing chromatin. In *in vitro* experiments, re-expression could be achieved by adding either 5-aza-2'-deoxycytidine, a demethylating agent, or trichostatin A (TSA), which inhibits histone deacetylases. In certain cases these two agents have a synergistic effect on restoration of gene transcription.

Clinical Relevance

In general, two types of methylation events in CpG islands can be distinguished, the developmentally regulated process of CpG island methylation discussed above and aberrant CpG island methylation in cancer.

Aberrant CpG Island Methylation in Cancer

- ►Hypermethylation of CpG islands in various cancers has been observed and correlates with the transcriptional inactivation of tumor suppressor genes and other cancer related genes (6). It was shown that methylation in a CpG island can serve as one of the two "hits" needed for the inactivation of a tumor suppressor gene. While CpG island methylation in some tumors is restricted to a small number of CpG islands, other tumors show a methylation phenotype with up to 10% methylated CpG islands (7). A subset of CpG islands is methylated in a tumor-type specific manner, while other CpG islands

can be methylated in many different tumor types. Many genes associated with methylated CpG islands can be reactivated in cell lines by *in vitro* demethylation using 5-aza-2'-deoxycytidine. Consequently, this drug is being tested in clinical trials with various types of malignancies and has demonstrated some effectiveness in restoring gene expression.

- ►Hypomethylation of normally methylated CpG dinucleotides has also been demonstrated in tumors. The mechanism(s) by which global hypomethylation paradoxically occurs concurrently with gene-specific hypermethylation is not known. Hypomethylation has been postulated to have an effect on oncogene activation (*HOX11* gene in T-cell acute lymphoblastic leukemia) and on transcriptional activation of mobile genetic elements such as long interspersed nuclear elements (LINEs) and to contribute to chromosomal instability. Any or all of these events would confer a selective growth advantage on a neoplastic cell.

References

1. Bird A, Taggart M, Frommer M et al (1985) A fraction of the mouse genome that is derived from islands of nonmethylated, CpG-rich DNA. *Cell* 40:91–99
2. Gardiner-Garden M, Frommer M (1987) CpG islands in vertebrate genomes. *J Mol Biol* 196:261–282
3. Takai D, Jones PA (2002) Comprehensive analysis of CpG islands in human chromosomes 21 and 22. *Proc Natl Acad Sci USA* 99:3740–3745
4. Venter JC, Adams MD, Myers EW et al (2001) The sequence of the human genome. *Science* 291:1304–1351
5. Antequera F, Bird A (1993) Number of CpG islands and genes in human and mouse. *Proc Natl Acad Sci USA* 90:11995–11999
6. Baylin SB, Herman JG, Graff JR et al (1998) Alterations in DNA methylation: a fundamental aspect of neoplasia. *Adv Cancer Res* 72:141–196
7. Costello JF, Fruhwald MC, Smiraglia DJ et al (2000) Aberrant CpG-island methylation has non-random and tumour-type-specific patterns. *Nat Genet* 24:132–138

CpG Methylation

Definition

CpG methylation is an addition of a methyl group to the cytosine-5 position of CpG dinucleotides, often correlated with a change in expression level of nearby genes.

►CpG Islands

►X-Chromosome Inactivation

CPSF

► Cleavage and Polyadenylation Specificity Factor

Craving

Definition

Craving is the uncontrollable desire for a drug of abuse. Craving is a multi-dimensional phenomenon. This hunger for drugs might become an important gateway to relapse. However, there is evidence that relapse occurs without any obvious craving, and that even strong craving signs are not necessarily followed by relapse.

► Addiction, Molecular Biology

CRC

► Calcium Release Channel

CRE

Definition

CRE stands for “cyclic AMP responsive element”. CREs are a common feature of all cAMP-responsive gene promoters and were first identified in the somatostatin promoter as an inducible enhancer of genes that can be transcribed in response to increased cAMP level. Its 8 bp plandromic sequence – TGACGT-CA – is conserved in other genes regulated by cAMP, such as *junB*, *zif268*, and *nur77*, and provides the binding site for the CRE binding protein ► [CREB](#).

Cre

Definition

Cre is a member of the integrase family of site-specific recombinases that catalyses recombination between *lox* sites. Cre cleaves DNA at one *loxP* site, and ligates it to

the cleaved DNA of a second *lox* site, to generate a contiguous strand.

► Cre/LoxP Strategies

► Mouse Genomics

► Transgenic and Knock-out Animals

Cre/loxP Strategies

BRIAN SAUER

Stowers Institute for Medical Research, Kansas City, Missouri, KS, USA

bls@stowers-institute.org

Synonyms

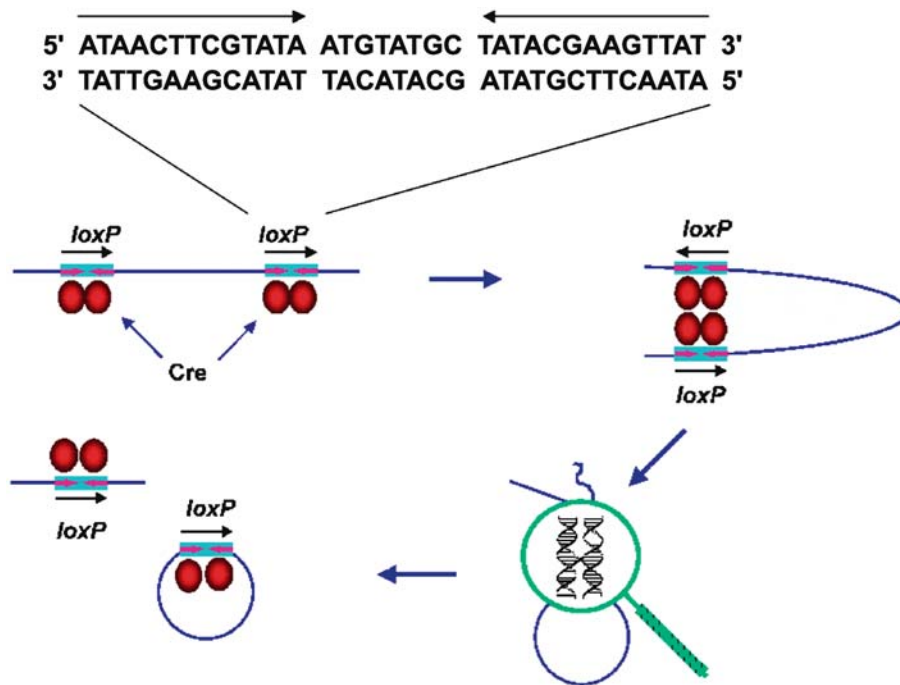
Conditional gene knockout; genome engineering; site-specific recombination

Definition

Accurate animal models of human disease are invaluable both in understanding the genetic contribution to disease and in the rigorous evaluation of therapeutic treatments. A remarkably useful molecular tool for making and inducing precisely designed mutational alterations in the genome in a tissue-specific manner is the site-specific DNA recombinase called Cre. The Cre recombinase from *Escherichia coli* ► [bacteriophage P1](#) efficiently catalyzes conservative site-specific DNA recombination at the 34-bp nucleotide *loxP* sequence in the P1 genome. The finding that Cre can also catalyze efficient DNA recombination in higher eukaryotic cells led to the development of site-specific DNA recombination as an important research tool for genetic and chromosome engineering in gene-modified animals (1). A particularly powerful strategy is the design of conditional mutations by targeted excision of a desired gene in the genome in a temporally and spatially controlled manner (2). Similar strategies can be used to design conditional gain-of-function mutations exhibiting temporal and spatial specificity. Cre-based recombination strategies also permit efficient site-directed integration of exogenous DNA into a pre-selected site on the genome, as well as permitting the design of precise chromosome translocations, deletions and duplications that facilitate the genetic manipulation of the eukaryotic genome.

Characteristics

The 38 kD Cre protein is encoded by the *cre* (cyclization recombination) gene of phage P1 and is a member of the Int family of tyrosine site-specific DNA



Cre/loxP Strategies. Figure 1 Site-specific DNA recombination by Cre recombinase. The DNA sequence of the 34 bp *loxP* site is shown, with the two 13 bp inverted repeats indicated by thin arrows above. Two Cre protein molecules bind to each *loxP* site, with the orientation of the site indicated by a horizontal arrow. Recombination proceeds through a Holliday junction intermediate, highlighted by the green magnifying glass.

recombinases (3). The *cre* gene benefits P1, which maintains itself in lysogens as a unit copy extrachromosomal plasmid, primarily by increasing plasmid stability. By catalyzing DNA recombination at a 34 bp site on the P1 genome called *loxP* (locus of X-over of P1) Cre resolves plasmid dimers that arise by DNA replication and/or homologous DNA recombination. Such dimers if unresolved would prevent partitioning of a monomer P1 genome to each daughter cell at cell division and thus lead to increased plasmid loss in a lysogenic bacterial population. Cre also cyclizes the terminally redundant, linear P1 virion genome upon infection of a bacterial host by recombination at *loxP* sites present in the terminal redundancy.

The *loxP* site consists of two 13 bp inverted repeats, each of which serves as a binding site for the Cre protein, flanking an 8 bp core region. The core region is nonpalindromic and so imparts an overall directionality to the site. It is also the region within which DNA recombination occurs (Fig. 1). Upon binding of two Cre molecules per *loxP* site synapsis of two Cre-bound *loxP* sites occurs followed by DNA recombination. Recombination is conservative and proceeds through a transient [Holliday junction](#) intermediate. Excision of a circular DNA molecule results if the two *loxP* sites are directly oriented on a DNA molecule, inversion

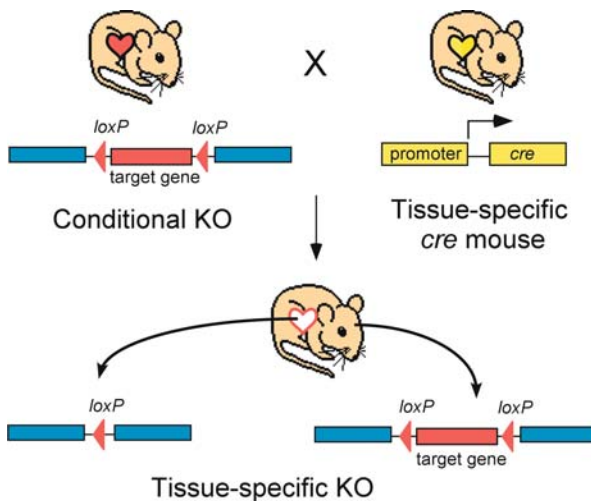
occurs if the two *loxP* sites are oppositely oriented. Intermolecular DNA recombination can also be catalyzed by Cre and results in integration of a *loxP*-containing circular molecule into another. If both molecules are linear, intermolecular recombination generates a translocation. Unlike many members of the Int family, Cre requires neither host factors nor a supercoiled DNA substrate for recombination. These minimal biochemical requirements probably contribute to Cre's ability to catalyze efficient DNA recombination in a wide variety of eukaryotic cells, notwithstanding Cre's prokaryotic origin. Because the 34 bp *loxP* site is sufficiently large to be unlikely to occur naturally in the genomes of higher eukaryotic cells, efficient Cre-mediated recombination can be targeted to just those sites in the genome into which *loxP* has been placed. Gene-modified and transgenic mice have become important in developing useful animal models of human disease and in gaining a detailed understanding of mammalian gene function. Cre/loxP and related strategies play a crucial role in the design, construction and genetic manipulation of these mice (3, 4). By targeting DNA recombination events at *loxP* sites placed into the genome, a designated gene can be either ablated or activated simply by controlling expression of Cre recombinase activity. With such conditional

mutations of either endogenous genes or transgenes, a mutant phenotype can be targeted to a particular tissue or to a particular time during development. In addition, recombinase strategies allow the generation of novel chromosomes useful in genetic analysis, removal of unwanted vector DNA sequences from the genome after gene targeting and efficient targeted integration of exogenous DNA's into the genome. Several functional mutant derivatives of Cre and of the *lox* recombination site have also proved useful in these strategies, as well as other site-specific recombinases, such as Flp from the yeast *Saccharomyces cerevisiae*.

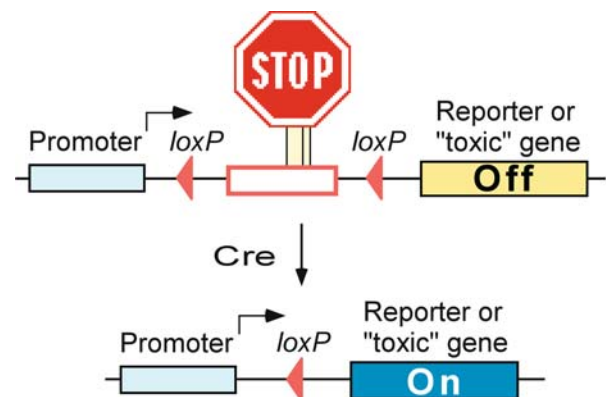
Conditional Gene Targeting

There are two components to recombinase-mediated conditional gene targeting in mice: 1) a *loxP*-targeted mouse and 2) a Cre delivery system, commonly a second mouse that expresses Cre in a pre-determined manner. A conditional loss-of-function allele of an endogenous gene in a mouse is generated by first designing a DNA construct carrying a genomic copy of that gene which has been modified by flanking that gene, or a critical region of it, by two directly repeated *loxP* sites. One copy of the gene in the genome is then replaced with the conditional allele by homologous gene targeting in ►embryonic stem (ES) cells. Injection of these modified ES cells into a mouse ►blastocyst leads to the production of a chimeric mouse carrying these modified cells. Breeding this

chimera generates progeny gene-modified mice having the desired conditional ►allele of the target gene. By having placed the *loxP* sites in non-critical regions at the genomic locus the homozygous modified gene is designed to give an unaltered and normal phenotype indistinguishable from wild type. However, introduction of Cre recombinase, for example by mating with a Cre mouse, results in excision of the *loxP*-flanked target gene in those cells, and only those cells, expressing Cre and thus produces a mouse in which the target gene, or any other unwanted bit of DNA, has been removed from the genome. Because various Cre transgenic mice can be designed with different promoters the *cre* transgene can be expressed in a tissue-specific or developmentally specific manner in each of these different Cre mouse lines. By choosing the appropriate Cre mouse with which to mate the conditionally mutant mouse, gene ablation can be targeted in whatever developmental and/or directed tissue-specific manner required because gene ablation mirrors the spatial and temporal expression specificity of the Cre mice used (Fig. 2). Conditional gene disruption thus permits the determination of the function of a gene in a particular tissue or type of cell without disturbing its function in other cells in the body. Moreover, because many genes are either essential for viability or fertility, this binary strategy allows the generation and propagation of a (conditional) ►knockout mouse for any gene in the genome.



Cre/loxP Strategies. Figure 2 Conditional gene knockout. The conditional knockout (KO) mouse carries a target gene flanked by directly repeated *loxP* sites (red arrowheads). When crossed with a *cre* mouse (here diagrammed to express only in the heart) progeny mice are produced in which Cre-mediated recombination has ablated the target gene specifically in the heart. In surrounding tissues the gene is not deleted.



Cre/loxP Strategies. Figure 3 Conditional gene activation. A synthetic DNA sequence called STOP is interposed between a promoter and the structural gene to be controlled. STOP is a DNA sequence designed to block gene expression and is flanked by two directly repeated *loxP* sites (red arrowheads). In mice having this construct STOP can be deleted in a tissue-specific fashion, activating downstream gene expression, by mating with the appropriate *cre* mouse.

Classical genetic approaches to elucidate gene function have relied not only on ►[loss-of-function](#) mutations but also on ►[gain-of-function](#) mutations. Using recombinase strategies, conditional gain-of-function gene mutations can be designed to pinpoint turn-on of a gene to a particular tissue at a pre-specified time. A common strategy is to interpose a *loxP*² STOP cassette between the promoter and the reporter or other gene to be regulated (Fig. 3). In such a transgene construct two *loxP* sites flank “STOP,” a DNA sequence designed to prevent downstream gene expression. This *loxP* transgenic mouse (►[Transgenic and Knock-out Animals](#)) is crossed with a Cre mouse having a desired pattern of expression to excise the STOP cassette specifically and only in those cells expressing Cre, thus turning on expression of the downstream gene. Conditional oncogene expression has been used to develop tissue-specific tumor models, and conditional expression of a toxic gene such as diphtheria toxin A-chain has been used to specifically ablate particular dopamine receptor neurons in the brain to better understand their role. Such recombinase-dependent gene activation strategies can be particularly valuable for cell lineage analysis and fate mapping because only brief transitory expression of Cre is sufficient to activate a reporter gene that will be expressed not only in the cell in which recombination has occurred but also in all of its descendants.

There are several ways to specify the spatial and temporal pattern of conditional gene ablation or activation. Tissue-specific promoters can specify the cells in which Cre is expressed and development-specific promoters can give some specification of timing. Greater temporal control of Cre expression has been achieved by placing the *cre* gene under the control of inducible promoters such as the MX promoter or synthetic “Tet” promoters. These two inducible promoters can be regulated in mice by dosing with interferon or tetracycline analogues, respectively. A second way to exert precise temporal control over Cre-mediated recombination has been to use a modified *cre* gene that encodes a variant protein having fused to Cre a ligand binding domain (LBD) derived from a steroid receptor. In eukaryotic cells this Cre-LBD fusion protein is inactive in the absence of ligand. Recombination is thus made dependent upon dosing the mouse with a synthetic steroid designed to activate Cre-LBD but not disturb normal signaling from endogenous steroid receptors. Thus, tissue-specific expression of the Cre-LBD protein followed by dosing of the animal at a desired time with the appropriate ligand (such as the estrogen analogue RU486) provides a remarkable strategy to specify in both time and space the ablation of a gene from the genome in a living organism. Lastly, the *cre* gene can be incorporated into a viral vector that is then used to infect a conditional knockout mouse,

with the tropism and/or infection route of the viral vector specifying the particular tissues in which Cre-mediated recombination is to occur.

DNA Integration and Genome Engineering

A major confounding problem in the generation of transgenic animals is the variability and unreliability of transgene expression between animals carrying the same transgenic construct but at different integration sites. This commonly necessitates the generation of several transgenic animals to obtain one that expresses the transgene in a suitable fashion. Variability occurs because injection of DNA into fertilized ►[zygotes](#) for transgenic animal production results in random integration of the transgene into the genome. Due to ►[position effects](#) on gene expression, transgenes at dissimilar chromosomal positions in different transgenic lines show variation in the level and tissue specificity of expression. In principle, this problem can be minimized by precise targeting, or knocking-in, a transgene to a particular chromosomal locus by ►[homologous recombination](#) in ES cells. Unfortunately the frequency of homologous targeting is too low for this to be a rapid strategy for this purpose. Cre-mediated integrative (intermolecular) targeting of a transgene to a chromosomal *lox* cassette can be a hundred times more efficient, however. When cells carrying a *lox* integration target in the genome are co-transfected with a *lox* transgene plasmid and a Cre transient expression construct, intermolecular recombination by Cre results in integration of a single copy of the transgene into the genome precisely at the genomic *lox* target. Although placement of the *lox* target, for example by homologous recombination, may still be a slow step, subsequent Cre-mediated targeting is precise and rapid, thus allowing one to quickly build a series of mutant alleles for a gene at a defined locus. Such transgene genomic placement by Cre is a powerful tool for genetic analysis. It has already become a useful method for constructing cell lines and is becoming increasingly useful for various types of genetic manipulation in mice.

The high efficiency in eukaryotic cells of Cre-mediated recombination between *lox* sites that are quite far from each other or even on different chromosomes has led to the ability to design novel chromosomes. Large chromosomal deletions, duplications and inversions can be generated in a precise fashion by first placing *lox* sites at defined chromosomal sites (5), for example by homologous targeting in ES cells. This is followed by Cre-mediated recombination to produce the desired novel chromosome rearrangement and mice are then generated from cells harboring the desired rearrangement. Such novel chromosomes are invaluable genetic tools. Large chromosomal deletions facilitate gene mapping, both deletions and duplications are useful for phenotypic determination of ploidy effects such as occur

in a variety of human genetic disorders, for example ►[Down Syndrome](#), and inversion chromosomes can serve as genetic ►[balancer chromosomes](#). ►[Chromosomal translocations](#) have often been found in various specific human cancers and these too can be modeled in the mouse by genome engineering using Cre (6).

Clinical Relevance

Recombinase strategies using Cre and other DNA recombinases, either singly or in combination, are of great utility in the genetic design of precise mouse models of human disease. In addition, Cre recombinase has been exploited in the design and genetic manipulation of various viral vectors, a number of which can be used for gene therapy in patients. For example, in adenovirus-based vectors Cre-mediated removal of a “stuffer fragment” vastly increases the vector’s capacity to carry therapeutic genes. The major current application for Cre/*loxP* strategies, though, is as a superb research tool for the genetic manipulation of DNA and chromosomes in living organisms, especially gene-modified mice, and to thus provide deeper insight into the roles of specific genes in specific cells and into the roles of specific cells in the body.

►[Limb Development](#)

►[Transgenic and Knock-out Animals](#)

References

1. Sauer B, Henderson N (1988) Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1. *Proc Natl Acad Sci USA* 85:5166–5170
2. Gu H, Marth JD, Orban PC et al (1994) Deletion of a polymerase beta gene segment in T cells using cell type-specific gene targeting. *Science* 265:103–106
3. Sauer B (2002) Chromosome manipulation by Cre-*lox* recombination. In Craig NL, Craigie R, Gellert M, Lambowitz AM (eds) *Mobile DNA II*. ASM Press, Washington, pp 38–58
4. Metzger D, Feil R (1999) Engineering the mouse genome by site-specific recombination. *Curr Opin Biotechnol* 10:470–476
5. Ramirez-Solis R, Liu P, Bradley A (1995) Chromosome engineering in mice. *Nature (London)* 378:720–724
6. Smith AJH, De Sousa MA, Kwabi-Addo B et al (1995) A site-directed chromosomal translocation induced in embryonic stem cells by Cre-*loxP* recombination. *Nat Genet* 9:376–385

C-Reactive Protein

Definition

CRP is an acute-phase protein featuring a homopentameric structure and calcium-binding specificity for

lyso-phosphatidylcholine. The wide distribution of phosphatidylcholine in polysaccharides of pathogens and in cellular membranes, allows CRP to recognize a range of pathogenic targets as well as membranes of apoptotic and necrotic host cells. CRP is an inflammatory marker.

►[Inflammatory Response](#)

Creatine Kinase

Definition

Creatine Kinase (CK) is a dimeric enzyme that is present in skeletal ►[muscle](#), cardiac muscle and the brain, and catalyses the ATP-dependent phosphorylation of creatine. CK levels in the blood stream can be elevated due to damage to the muscle caused by exercise injury or heart infarction and the muscular dystrophies.

►[Duchenne Muscular Dystrophy](#)

CREB

Definition

Cyclic AMP response element binding protein (CREB) is a 43 kD basic leucine zipper transcription factor that binds the ►[CRE](#) (cyclic AMP response element) consensus motif TGACGTCA. CREB was originally isolated as the protein that bound CRE in the somatostatin gene. CREB was first shown to be activated by elevated levels of cAMP, which results in its phosphorylation on serine-133 by PKA. CREB phosphorylation can be achieved by signalling cascades (e.g. the MAP kinase signalling cascade), eventually triggering transcription of the somatostatin gene as well as other genes that contain the CRE sequence in their promoter.

►[Wnt/Beta-Catenin Signaling Pathway](#)

Creb Binding Protein

Definition

Creb binding protein (CBP) is a transcriptional co-factor that modulates cyclic AMP signaling and

functions as a co-activator for the e.g. β -catenin/TCF complex.

► [Wnt/Beta-Catenin Signaling Pathway](#)

Cre-ERT

Definition

Cre-ERT is a modified form of Cre in which the ligand-binding domain of a mutated human estrogen receptor (ERT) that recognises tamoxifen has been added to Cre. *LoxP* recombination in cells in mice carrying the Cre-ERT transgene will not occur until exposed to tamoxifen (either by injection or in the food/drinking water).

► [Cre/loxP Strategies](#)

► [Mouse Genomics](#)

Creutzfeldt-Jakob Disease

► [CJD](#)

CRH

Definition

Corticotropin-releasing hormone (CRH) is expressed in the nucleus paraventricularis of the hypothalamus, and drives the stress hormone system by activating synthesis and release of corticotropin at the pituitary and in turn corticosteroid from the adrenal cortex. CRH is also expressed in many other ► [brain](#) locations not involved in neuroendocrine regulation, e.g. the pre-frontal cortex and amygdala. CRH has been implicated in the regulation of a wide range of behaviors including arousal, motor function, feeding, and reproduction. Behavioral pharmacology provides evidence that CRH overexpression accounts for many signs and symptoms characteristic of depression and anxiety disorders. CRH-type 1 receptors convey the CRH signal into cellular circuitries.

► [Addiction, Molecular Biology](#)

Cri-Du-Chat Syndrome

Definition

Cri-du-chat syndrome is a severe disorder (1:20,000 newborns) caused by terminal 5p15.2 deletion and characterized by growth retardation, round face in infancy contrasting with growth deficiency of the body, facial hemangiomas, strabism, heart defect in 30% of patients, hypotonia and severe mental retardation. The characteristic cat-like cry in infancy also occurs in the rare, mild variant with deletion of only 5p15.32.

► [Microdeletion Syndromes](#)

Crisis

Definition

Crisis (cellular crisis) is a proliferative barrier that is characterized by a state of chromosomal instability and widespread cell death. It is also called mortality stage 2 (M2). Crisis is associated with excessively shortened ► [telomeres](#). Some cells can survive crisis and acquire the constellation of genetic alterations needed for malignant transformation.

► [Telomerase](#)

Cristae

Definition

Cristae designates invaginations of the mitochondrial inner membrane harboring all protein complexes required for respiration and ATP synthesis.

► [Mitochondria – Biogenesis and Structural Organization](#)

Cristae Junctions

Definition

Cristae junctions are narrow tubular structures of the mitochondrial inner membrane connecting the inner boundary membrane with the cristae membrane.

► [Mitochondria – Biogenesis and Structural Organization](#)

Critical Micellar Concentration

Definition

Critical micellar concentration characterizes a concentration at which detergent molecules assemble spontaneously into micelles to bury the hydrophobic moiety in a hydrophobic core.

►Two-dimensional Crystallization of Membrane Proteins

Crk

Definition

Crk stands for Chicken retroviral kinase. It is the protein product of the *crk* gene from chicken retroviruses CT10 and ASV-1. The cellular homologues, the Crk adapters, mediate different cytoplasmic signals.

►Signal Transduction: Integrin-Mediated Pathways

►Tyrosine Kinases

cRNA

Definition

cRNA stands for complementary RNA. It is an RNA that is synthesized from a cDNA template by *in vitro* transcription.

►DNA Microarrays/DNA Arrays

Crohn Disease

STEFAN SCHREIBER

Institute for Clinical Molecular Biology and
Department of General Internal Medicine,
Christian-Albrechts-University, Kiel, Germany
s.schreiber@mucosa.de

Synonyms

Crohn disease is a heterogeneous syndrome that belongs to the group of inflammatory bowel disease 1 (OMIM 266600; IBD 1). Crohn disease can be differentiated from ulcerative colitis (OMIM 191390)

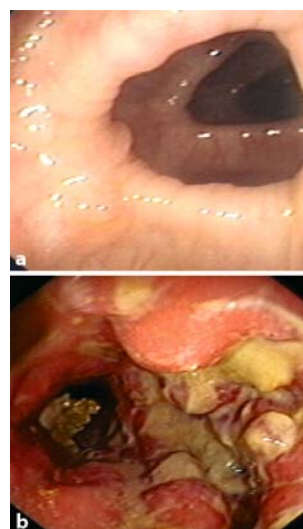
by clinical, endoscopic, radiological and histopathological parameters.

Definition

Typical presentation of Crohn disease includes inflammation in the terminal ileum. However, this typical presentation with ileitis is seen only in 70% of cases. 25% of patients present with disease confined only to the colon (Crohn colitis) and in 5% exclusive involvement of the upper gastrointestinal tract is seen. The diagnosis is mostly based on a suggestive clinical course, a typical distribution of affected areas throughout the gastrointestinal tract ("skip lesions") and a typical type of endoscopic lesions (Fig. 1). Histological characteristics (i.e. granuloma formation) are only seen in a minority of the patients.

Characteristics

Crohn disease was first described as a clinical entity in 1920. It is a chronic relapsing complex disease, that can present with diarrhoea, right lower quadrant pain, weight loss, malaise, fever and a host of extraintestinal manifestations in joints (arthralgia, arthritis), skin (pyoderma gangrenosum), vessels (erythema nodosum), liver (autoimmune hepatitis, rarely primary sclerosing cholangitis), eyes (uveitis, episcleritis), pancreas (hyperamylasemia and pancreatitis), lung (pneumonitis) and heart (rarely myocarditis). Complications include the development of intestinal strictures, fistula, osteoporosis and an increased risk for colonic cancer. Since the second World War a growing ►incidence of cases was observed in Western civilisations but is now also observed in Asian countries



Crohn Disease. Figure 1 Endoscopic presentation of Crohn disease (b) in comparison to normal colon (a).

including Japan, Korea and China. “Hot spots” of incidence and ►prevalence have been described in Western Canada and Iceland. Lifetime prevalence in Northern countries has been estimated at up to 0.5%. Typical age of onset peaks in the 2nd and 3rd decades of life, but the disease can also manifest in young children or in the elderly. Genetic anticipation has been proposed in Crohn disease. However, large epidemiologic studies suggest that the increase in incidence since World War II due to an accumulation of confounders has been confused with genetic anticipation.

Epidemiological studies have identified a significant genetic contribution to the aetiology of inflammatory bowel disease. Familial clustering is seen, with estimates of sibling relative risk (λ_s) ranging from 15 to 35 for Crohn disease and 5 to 17 for ulcerative colitis. Simple Mendelian models of inheritance are not supported by segregation analyses of inflammatory bowel disease. Concordance of Crohn disease in identical twins is seen in up to 56% in comparison with 4% in dizygotic twins. Therefore, disease concordance data in monozygotic and dizygotic twins suggest that both genetic and environmental factors are involved. Disease characteristics (subphenotypes) cluster in families and show concordance in identical twins, too. These observations support a complex immunogenetic model for inflammatory bowel disease, whereby genetically susceptible individuals harbour an aberrant response to yet unidentified environmental influences that most likely are included in the life-style of industrialized societies.

Cellular and Molecular Regulation

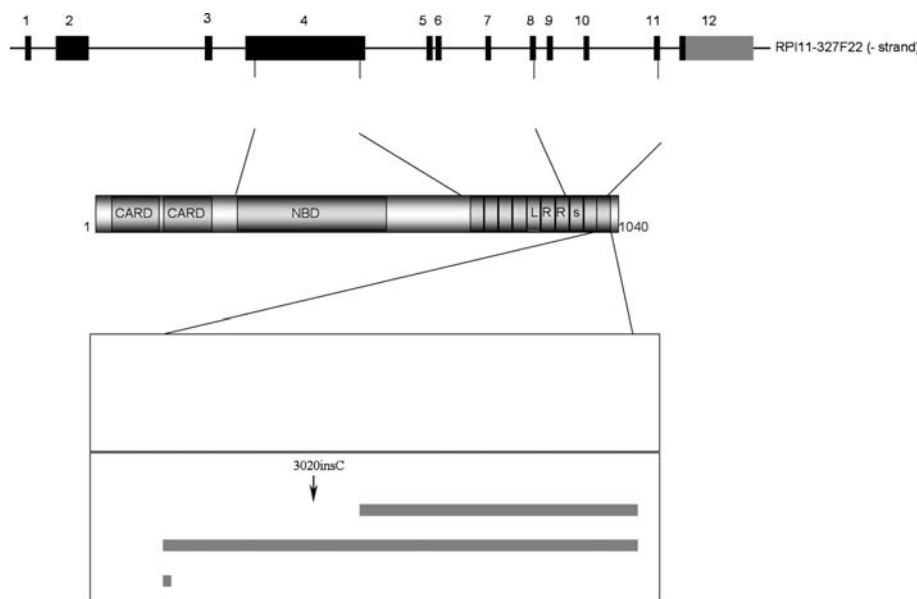
Attempts to localise susceptibility genes for inflammatory bowel disease through genome-wide linkage studies have identified putative loci on many human chromosomes with eight of them having been listed in

OMIM thus far (chromosomes 16p, 16q, 12, 6p, 14q11/12, 5q31, 19p13, 1q36 – Table 1); the original finding of the linkage on chromosome 16 has been replicated in several independent populations and by the IBD International Genetics Consortium. Recently three groups showed that sequence variations within the ►CARD15 (NOD2) gene (OMIM 605956) on chromosome 16q12, were strongly associated with susceptibility to Crohn disease but not to ulcerative colitis. P268S, R702W and 3020insC (Fig. 2) were found as the main disease associated variations with an ODDs ratio for the development of CD ranging between 2.4 (heterozygotes) and 40 (compound heterozygotes and homozygotes). The three main variations (P268S, R702W, 3020insC) result in amino acid exchanges or in a truncation, respectively, of leucine rich repeats at the C-terminal part of the CARD15 protein. A series of additional “private” mutations has been found in patients with Crohn disease in this part of the gene. Taken together it appears that variants in the CARD15 gene alone could explain the risk for development of Crohn disease in up to 15% of cases (compound heterozygotes and homozygotes) and may contribute to another 15% in concert with other disease genes (heterozygotes).

CARD15 is a member of a family of genes (Apaf-1, Ced-4, CARD-4/NOD1) implicated in activation of NF κ B, pro-inflammatory ►cytokine induction and ►apoptosis. They represent regulatory proteins with a central nucleotide-binding oligomerization domain (NOD) and N-terminal caspase recruitment domains (CARD) that are involved in programmed cell death and immune responses. A high degree of homology is apparent to a group of disease resistance (R) genes in plants that encode cytosolic and membrane-bound proteins mediating recognition of pathogens. Membrane-bound R proteins are homologous to

Crohn Disease. Table 1 Susceptibility loci for Crohn disease

Name	Chromosome	Sample	Publications
IBD1 (CARD15)	16q12	F/USA/D	Hugot Nature 2001, Cho Nature 2001, Hampe Lancet 2001
IBD8	16 q/p	D	Hampe PNAS 2002
IBD2	12p13.2-q24.1	UK	Duerr AJHG 1998, IBD Consort AJHG 2001, Satsangi Nat Gen 1996
IBD3	6p	D	Hampe AJHG 1999, Dechairo EJHG 2001
IBD4	14q11/12	USA	Duerr AJHG 2000, Mah IBD 1999
IBD5	5q31	Ca	Rioux Nat Gen 2001, Rioux AJHG 2000
IBD6	19p13	Ca	Rioux AJHG 2000
IBD7	1p36	USA	Chop PNAS 1998, Cho Hum Mol Gen 2000



Crohn Disease. Figure 2 Structure of the CARD15 gene and its protein product (reprinted from Cuthbert et al., 2002).

►**toll-like receptors** (TLRs) because they contain leucine-rich repeats (LRR) and recognize pathogenic components at the cell surface.

CARD15 is expressed in monocytes, B-cells and activated epithelial cells. It encodes the protein NOD2 that demonstrates a strictly intracellular distribution. Expression of *CARD15* mRNA and protein is regulated synergistically by ►**TNF** and interferon- γ . Preliminary functional evidence for *CARD15* as a susceptibility gene for Crohn disease was suggested by its ability to activate NF κ B following exposure to bacterial lipopolysaccharides (probably more specifically muramyl dipeptide). Most importantly, activation of NF κ B was impaired in cells transfected with *CARD15* constructs carrying the disease-associated variations. It is therefore currently assumed that *CARD15* variations may impair intestinal barrier function and lead to a defect in epithelial defence against the commensal flora. In support of the hypothesis, an increased presence of intracellular *E. coli* has been described earlier in colonic epithelial cells of patients with Crohn disease. A more detailed molecular definition of the role of *CARD15* in mucosal barrier function remains to be established.

The disease susceptibility locus on chromosome 5q31 could be resolved to a disease-associated haplotype called "IBD5". An interaction between *CARD15* and IBD5 in the risk for Crohn disease has been suggested. An independent association between ulcerative colitis and IBD5 was described recently in a large series of

patients. However, identification of putatively causative variants in a disease gene is still pending in this (IBD5) and other chromosomal regions.

Hierarchical mapping using the TDT in European CD patients identified a disease-associated haplotype block in the suggestive linkage region on chromosome 10q 51. This block contained exactly one gene, *DLG5*, which led to the conclusion that it represented a susceptibility gene for inflammatory bowel disease. This association has recently been replicated in Japanese CD patients albeit with different susceptibility alleles from those in European populations. Further replication in populations of European descent will be necessary. In contrast to the strong genetic effect of NOD2, where the associated variant confers a relative risk greater than 40 for homozygotes and compound heterozygotes, the risk of *DLG5* appears to be much smaller, with observed odds ratios of approximately 1.5. *DLG5* that codes for a protein involved in cell-cell contacts, is a disease gene that could be easily integrated into a hypothetical mechanism of a defect in barrier integrity as a main cause for Crohn disease. The haplotype structure of 5q31 has rendered the identification of the respective disease gene more problematic than on 10q and 16q. A strong linkage signal on 5q31 led to the identification of a large haplotype block that could not be resolved further by means of genetic epidemiological methods. Instead, Peltekova and colleagues used (7) a combination of functional and genetic evidence to implicate variants

around the *SLC22A4* and *SLC22A5* genes in the etiology of Crohn disease. Although the identification of the disease-associated variants in these genes has evoked clear functional consequences, the exact mechanism by which alteration of the encoded transporter proteins OCTN1/2 contribute to disease is still unknown.

Clinical Relevance

Disease-associated variants in the *CARD15* gene are associated with development of a particular subphenotype of Crohn disease. In particular, homozygosity or compound heterozygosity of 3020insC is associated with the “typical” presentation of the disease with ileal inflammation. It also appears that Crohn disease manifests at a younger age in patients presenting with a combination of *CARD15* variants and the IBD5 haplotype.

The example of *CARD15* also shows the heterogeneity of population from different ethnicities but also within the European white population. It underlines the importance of population representative assessment (i.e. through biobanks). The haplotypes and causative variants that are strongly associated with Crohn disease in white populations of European descent are completely absent in Asian populations and almost absent in African Americans. Even within Europe a great variation is seen as to the population frequencies and to the relative risk for carriers. Any future concept for a medical use must therefore take the specific population background into account.

Crohn disease is a chronic relapsing disorder. On average 40% of patients in remission relapse during 12 month of follow up. In some cases primary chronic active disease is developed despite adequate therapy. Short-term therapy with glucocorticoids is a hallmark of anti-inflammatory management of acute disease. However, it leads in 30% of cases to glucocorticoid-refractory or -dependent chronic active disease with additional complications through the long-term use of glucocorticoids. To date only azathioprine, methotrexate and infliximab (i.e. a monoclonal antibody directed against ▶*TNF-alpha*) are effective for remission maintenance.

Outcome of a therapy with infliximab has raised particularly interest in pharmacogenetic explorations. The targeted intervention (infusion of a monoclonal antibody which is directed against TNF) induces remission in approximately 40% of patients with active Crohn disease, while improvement is seen in another 20% and no response in the remainder. The availability of several large trials that were conducted for approval of the drug in the EU and in the USA facilitated a pharmacogenetic investigation with a design characterized by a high statistical power. Unfortunately neither variants in the TNF/TNF-receptor system nor variants

of the *CARD15* gene were associated with therapeutic outcome to anti-TNF therapy with infliximab.

References

1. Cuthbert AP, Fisher SA, Mirza MM et al (2002) The contribution of NOD2 gene mutations to the risk and site of disease in inflammatory bowel disease. *Gastroenterology* 122:867–874
2. Hampe J, Cuthbert A, Croucher PJP et al (2001) An insertion mutation in the NOD2 gene predisposes to Crohn Disease in the German and British populations. *Lancet* 357:1925–1928
3. Hampe J, Grebe J, Nikolaus S et al (2002) Association of NOD2 (*CARD 15*) genotype with clinical course of Crohn disease: a cohort study. *Lancet* 359:1661–1665
4. Hugot JP, Chamaillard M, Zouali H et al (2001) Association of NOD2 leucine-rich repeat variants with susceptibility to Crohn disease. *Nature* 411:599–603
5. Stoll M, Corneliussen B, Costello CM et al (2004) Genetic variation in *DLG5* is associated with inflammatory bowel disease. *Nat Genet* 36:476–80
6. Mascheretti S, Hampe J, Croucher PJ et al (2002) Response to infliximab treatment in Crohn disease is not associated with mutations in the *CARD15* (*NOD2*) gene: an analysis in 534 patients from two multicenter, prospective GCP-level trials. *Pharmacogenetics* 12: 509–515
7. Peltekova VD, Wintle RF, Rubin LA et al (2004) Functional variants of OCTN cation transporter genes are associated with Crohn disease. *Nat Genet* 36:471–5
8. Schreiber S et al (2005) Genetics of Crohn disease, an archetypal inflammatory barrier disease. *Nat Rev Genet*, in press

Cross-Bridge

Definition

The term cross-bridge is used to characterize the pear-shaped motor domain of myosin projecting in regular spacings from the thick filament. These domains interact with the actin of the thin filaments during muscle contraction.

▶*Muscle Contraction*

Cross-Fostering Studies

Definition

Cross-fostering studies compare the risk of the disorder in adoptees who have affected biological parents but

unaffected adopting parents, and in adoptees with unaffected biological parents but affected adopting parents.

►Schizophrenia Genetics

Crossing-Over

Definition

Crossing-over refers to a reciprocal exchange (recombination) between non-sister chromatids of a pair of homologous chromosomes and involves breakage and reunion of chromatids.

►Chromosomal Instability Syndromes

►Genetic Epidemiology

►Meiosis and Meiotic Recombination

Cross-Linking Patterns

Definition

Cross-linking patterns denote the pattern of proteins that are observed on a polyacrylamide gel after treatment with a ►cross-linking reagent.

►Protein Interaction Analysis: Chemical Cross-Linking

Cross-Linking Reagents

Definition

Cross-linking reagents are chemical compounds that are used to covalently couple two or more molecules.

►Protein Interaction Analysis: Chemical Cross-Linking

Cross-Talk

Definition

Cross-talk refers to the functional interaction between individual signal transduction pathways or cascades which may affect signalling positively or negatively.

►Ras Signalling

Crossvalidation

Definition

Crossvalidation is a method for the evaluation of classification techniques: the study data are split into several subsets; each is used in turn as a test set for a classifier trained on the other data sets.

►Computational Diagnostics

CRP

►C-Reactive Protein

Cryo Electron Crystallography

Definition

Cryo electron crystallography uses electrons for diffraction. Since electrons are diffracted much stronger than X-rays, very tiny crystals can be studied. Accordingly, cryo electron crystallography has been applied to extremely thin, so-called two-dimensional crystals, composed of membrane-bound proteins. However, the experimental requirements for this method are more demanding than for usual protein crystallography, and only limited resolution of the structures can be obtained.

►Structure-Based Drug Design

Cryo-Electron Microscopy: Single-Particle Reconstruction

ROLAND BECKMANN

Institute for Biochemistry, Charité University Medicine Berlin, Germany
roland.beckmann@charite.de

Definition

Cryo-electron microscopy (cryo-►EM) in combination with ►single-particle reconstruction refers to the combined application of two methods used for

structure determination of macromolecular assemblies: (i) ►**cryo-EM** describes the usage of transmission electron microscopy for the visualization of a specimen in a thin layer of ►**vitreous ice** at very low temperature and (ii) single-particle reconstruction describes two- (2D) and, preferably, three-dimensional (3D) reconstruction of the specimen structure by application of image processing techniques to the obtained two-dimensional EM images of the so-called ►**single particles**. Importantly, the term ‘single particles’ in this context refers to macromolecular assemblies as isolated and unordered particles with identical structure (1).

Description

Characteristics

In recent years it has become clear that many of the key functions of the cell depend on large macromolecular assemblies also termed molecular machines. They are often very large and flexible, which makes structure determination by the classical techniques of X-ray diffraction or NMR very difficult and sometimes impossible. In these cases cryo-EM in combination with single particle reconstruction turned out to be an excellent tool. An important advantage of cryo-EM is that the specimen can be visualized in a fully hydrated state without the need for crystal formation. Thus, under conditions that are very close to physiological conditions, a wide range of specimens can be studied in different functional and dynamic states or with a variety of ligands bound (1).

The images initially obtained by cryo-EM represent ►**projections** of the specimen. That means that the 3D structural information of the entire particle is combined along the optical axis of the microscope resulting in a 2D image. A large number of these 2D images of the specimen in different orientations can be combined in order to reconstruct the 3D structure by a process called ►**back-projection**. In contrast to X-ray and NMR techniques, however, the resolution achieved by the different cryo-EM techniques is limited. It ranges, dependent on the kind of specimen and reconstruction method, from more than 30 Å to atomic resolution. Well-ordered specimens such as two-dimensional (2D) or helical crystals allow for atomic or near atomic resolution. Unordered specimens such as single particles are more difficult to resolve at high resolution. Particles with very high symmetry, such as icosahedral viruses, are an exception and can be reconstructed more easily to a resolution under 10 Å. Particles with low or no symmetry, which represent the majority of cellular assemblies, yield reconstructions in a resolution range of 10–30 Å. A clear limit of the method is the size of the particles. The molecular weight should be at least a few hundred kD, preferably higher. The larger and more rigid the particle, the easier it is to determine its structure by single particle reconstruction.

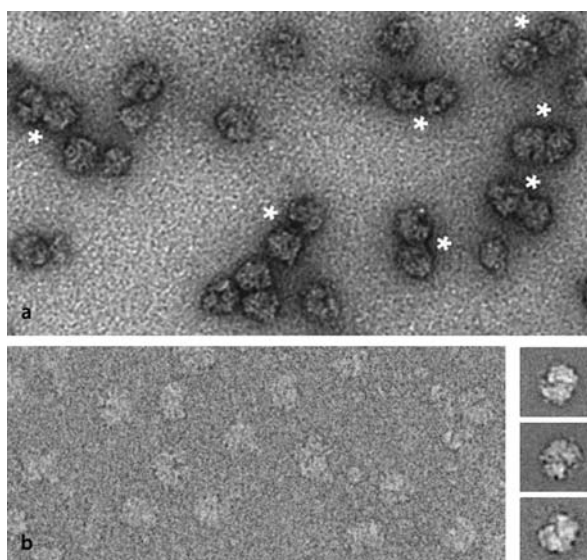
The reason for the differences in resolution described lies in the fact that cryo-EM images are extremely noisy, so that all reconstruction methods require averaging of many images. Superposition of the same structural elements for averaging, however, is much easier with an ordered specimen, as in crystals or symmetrical particles, than with randomly oriented unordered particles. In the latter case, one has to determine the relative orientation of every single particle before averaging and 3D-reconstruction is possible. The accuracy of this process leads to limits of the method regarding the achievable resolution and the size of the studied particle.

The low signal to noise ratio of cryo-EM images is a result of the image formation. This is dependent on the interaction of the electron beam with the specimen, which, unfortunately, causes damage to the specimen at the same time. This reduces the applicable electron doses to 5–40 electrons/Å². The classical method of negative staining uses the salts of heavy metals to air-dry and form a cast around the specimen. This cast is relatively resistant to high electron doses and provides high contrast information about the contour of the particle. However, no information about the density distribution inside the particle can be obtained and the particles are often distorted by the staining procedure. In cryo-EM, the specimen is embedded in vitreous ice and the interaction of the beam with the structure itself provides the weak contrast that can be enhanced by averaging and used for 3D reconstruction (Fig. 1).

Determination of a 3D Structure

The first and very decisive step on the way to structure determination by single particle reconstruction (Fig. 2) is the isolation of a homogenous population of particles. Notably, very small amounts (a few µg) and imperfect homogeneity of the isolated material still allow successful structure determination. However, especially at higher resolution, heterogeneity of conformational states or low ligand occupancy can be limiting. When recognized, heterogeneous particle populations can sometimes be sorted computationally. In the best case, however, the procedure starts with a sample that is as homogeneous and well characterized as possible.

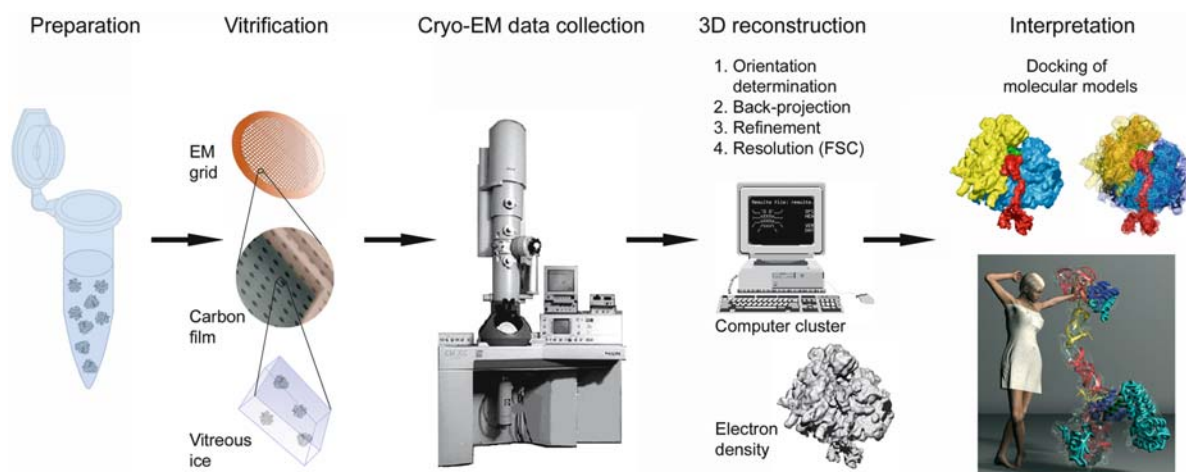
In the next step the sample has to be vitrified. This is done by application of the sample to a holey carbon film supported by an EM grid, blotting away excess sample and flash freezing the grid in liquid ethane that is cooled by liquid nitrogen. The very small mass of the grid allows for a sufficiently fast cooling rate to prevent ice crystal formation. As a result the particles are preserved in a very thin (30–100 nm) layer of vitreous ice.



Cryo-Electron Microscopy: Single-Particle Reconstruction. Figure 1 (a) Electron micrograph of negatively stained 80S ribosomes from yeast (1% uranyl acetate). Asterisks indicate dimeric translating ribosomes. (b) Electron micrograph of 80S ribosomes embedded in vitreous ice (left) and averaged images of 80S particles in three distinct orientations (right). Note the differences in contrast.

The third step is the data collection on a cryo-electron microscope usually yielding dozens to hundreds of micrographs with thousands to hundreds of thousands of particle images. The grid is inserted and kept in the vacuum of the microscope using a holder (or microscope stage) that maintains temperatures below -150°C . This can be achieved by cooling with liquid nitrogen (in standard cryo-microscopes) or even liquid helium. Micrographs are taken in the so-called ‘low-dose mode’ which exposes the area of interest only once to the electron beam. In this way the electron dose can be controlled in order to avoid radiation damage. Note that to increase the contrast of the images, they must be taken with the sample in a position just a little bit below the true focus of the objective lens. This leads to a distortion (or blurring) of the images described by the so-called point spread function (or in reciprocal space by the contrast transfer function, CTF), which needs to be corrected for later on, when reconstructing the 3D structure.

The fourth step is the first image-processing step. After the micrographs have been digitized and single particles selected, the relative orientation of every particle has to be determined in a process called alignment. The orientation can be described by five parameters, two shifts in the plane and three so-called



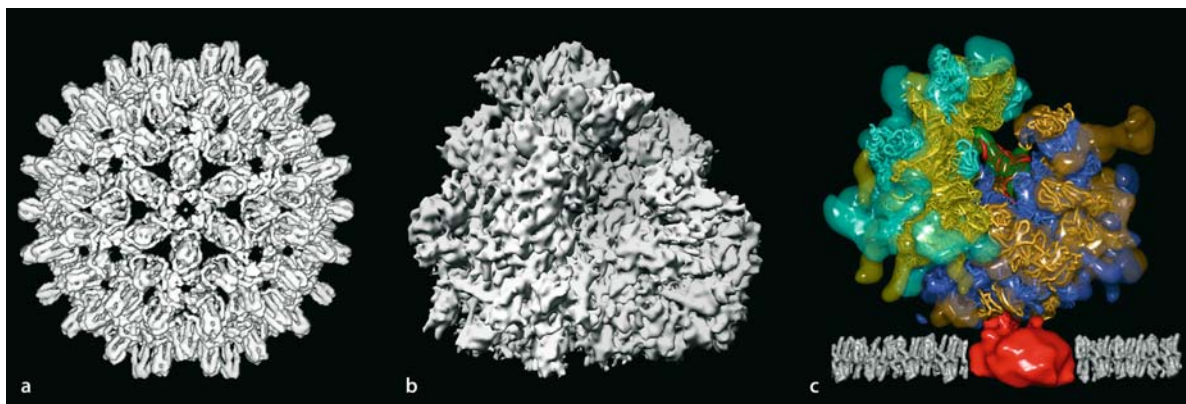
Cryo-Electron Microscopy: Single-Particle Reconstruction. Figure 2 Scheme of structure determination using cryo-electron microscopy (cryo-EM) in combination with single particle reconstruction. Preparation of the sample is followed by shock freezing (vitrification) on EM grids (usually copper) covered with a holey carbon film. The particles are captured in a thin film of vitreous ice and visualized in a transmission EM at very low temperatures (between -180°C and -260°C , maintained by liquid nitrogen or liquid helium). The images are digitized and then processed using large computer systems. After the determination of particle orientations (using several distinct methods, see text) they can be back-projected in order to generate a 3D density map. After refinement of the map the resolution is determined on the basis of the Fourier shell correlation (FSC). In many cases atomic models can be docked into the electron density map, which allows interpretation on a molecular level. Here, the mammalian signal recognition particle bound to the 80S ribosome is shown in red and transparent with docked atomic models (5).

Eulerian angles. It makes an important difference whether the structure is known (to a first approximation) or has to be determined *de novo*. If the particle belongs to a family with similar structures, one can use a known structure as a reference for the so-called ‘projection matching’ method. In this case, projections of the reference structure are generated computationally and compared with the experimentally obtained projections using cross correlation functions. The parameters of the best fitting reference projection are then assigned to every individual experimental projection. For *de novo* structure determination three different methods are commonly used to determine parameters. (i) Tomography creates a series of images of the same sample area tilted at different angles. The known tilt angles describe the geometrical relation between the different views for individual particles providing the needed parameters. (ii) The ‘random conical tilt’ method uses pairs of micrographs of a tilted and a non-tilted sample area in order to establish the geometrical relation between particles that are oriented in the same way in the plane of the grid. (iii) The ‘common line’ or ‘angular reconstitution’ approach needs only untilted images and takes advantage of the fact that in reciprocal space every two projections (2D) of the same 3D object share a common line. These shared lines can be identified and used to reconstitute an angular system in which all experimental projections are positioned. However, the handedness of an object is

not revealed by this method. Often, multivariate statistical methods such as principal component analysis or correspondence analysis are used in order to sort the particles and to group similar particle projections. The grouped images can then be averaged to enhance the signal to noise ratio as a prerequisite for finding common lines or to select a group of particles with a certain orientation for the random conical tilt approach.

In the fifth step, the obtained orientation parameters can be used for back-projection in order to reconstruct a first 3D structure. This first structure subsequently serves as a new reference in the so-called refinement of the reconstruction. In most cases, the refinement follows the projection matching method in an iterative way. Usually, during this step (or already on the original images) the detrimental effects of the CTF are corrected; this is obligatory when attempting to reach higher resolutions (below 15 Å).

In the sixth step, the resolution of the reconstruction is assessed by randomly dividing the dataset into two halves and determining the correlation of the resulting two reconstructions in reciprocal space as a function of spatial frequency (► [Fourier shell correlation](#), (FSC)). The resolution limits are obtained by finding the spatial frequency corresponding to a certain correlation cutoff; commonly a cutoff of 0.5 is used, which corresponds to a signal to noise ratio of 1. The interpretation of the structure is then guided by this limit.



Cryo-Electron Microscopy: Single-Particle Reconstruction. Figure 3 Examples of symmetrical (icosahedral) and asymmetrical cryo-EM single particle reconstructions. (a) Surface representation of an electron density map of the hepatitis B virus (HBV) capsid at 7.4 Å resolution generated by cryo-EM (2). Note the rod-like features representing the α -helical secondary fold of the capsid protein, which assembles in an icosahedral structure with 60-fold symmetry. (b) A surface representation of the asymmetrical 80S ribosome from wheat germ (Halic et al., unpublished) is shown at a resolution of 8 Å. (c) A molecular model of the translating 80S ribosome from yeast in complex with the protein-conducting channel (Sec61 complex, in red) is shown (3). Although the EM-map has only 15 Å resolution, a molecular interpretation is facilitated by docking of homology models into the separated densities for ribosomal protein and RNA (4). The gray features contacting the Sec61 complex indicate the likely position of the lipid bilayer.

In a seventh step, it is possible in many cases to dock atomic structures or molecular models of components (obtained by X-ray, NMR or homology modeling) into the EM-density of molecular assemblies, thereby positioning them in their functional context. The molecular models can either be docked as rigid bodies or, in the case of recognizable and significant conformational changes, the orientation of individual domains can be adjusted in a procedure called flexible ►**docking**. This often leads to a situation where a low-resolution structure can be reliably interpreted on a molecular level, exceeding the resolution limit of the EM density several-fold. In these instances, the identification of domains or specific regions of a macromolecule that are involved in interactions or dynamic behavior can be facilitated.

Clinical Applications

The technique described has been successfully applied to a wide variety of samples. Some of them have direct clinical relevance, such as the structures of viral pathogens. Most samples, however, represent macromolecular assemblies and machines of the cell in different functional states, for instance RNA polymerases, nuclear pore complexes, spliceosomes, ribosomes, chaperones, proteasomes, ion channels, ATPases etc.

A prominent example of the structure of a pathogen is the hepatitis B virus (HBV). The structure of the HBV capsid (2) was one of the first viral structures resolved by cryo-EM at a sufficiently high resolution to allow assignment of α -helical secondary structure (Fig. 3). In this case, the 60-fold icosahedral symmetry of the capsid structure could be exploited for the reconstruction and revealed the unusual fold of the HBV capsid protein.

The most prominent example for a non-symmetrical particle investigated by cryo-EM and single particle reconstruction is the ribosome (Fig. 3). For this technique it represents a perfect sample because it is large (several MD), dense (due its high RNA content), almost spherical (leading to random orientations) and displays a definite dynamic behavior. Many cryo-EM structures of 70S and also of 80S ribosomes in complex with tRNAs, translation factors and other ligands contributed significantly to the understanding of such fundamental processes as translation, protein targeting and protein translocation (3, 4, 5). Notably, a resolution of below 10 Å can be achieved for this asymmetrical particle. Moreover, the ribosome is an excellent example of combining cryo-EM with other structural methods in order to create molecular models by docking atomic models of subunits into EM density. The first molecular model (Fig. 3) for a eukaryotic ribosome (from *S. cerevisiae*) was generated in this way (3, 4).

References

1. Frank J (2002) Single-particle imaging of macromolecules by cryo-electron microscopy. *Annu Rev Biophys Biomol Struct* 31:303–319
2. Böttcher B, Wynne SA, Crowther RA (1997) Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. *Nature* 386:88–91
3. Beckmann R, Spahn CMT, Narayanan E et al (2001) Architecture of the protein-conducting channel associated with the translating 80S ribosome. *Cell* 107:361–372
4. Spahn CMT, Beckmann R, Narayanan E et al (2001) Structure of the translating 80S ribosome from *Saccharomyces cerevisiae* – tRNA-ribosome and subunit-subunit interactions. *Cell* 107:373–386
5. Halic M, Becker T, Pool M, Spahn CMT et al (2004) Structure of the signal recognition particle interacting with the elongation arrested ribosome. *Nature* 427:808–814

Cryo-EM

Definition

Cryo-electron microscopy (EM) describes a transmission electron microscopy method performed at low temperature of unstained, frozen-hydrated specimens in order to stabilize frozen samples in the vacuum of an electron microscope and to reduce radiation damage. Samples are usually embedded in ►**vitreous ice**.

►**Cryo-Electron Microscopy: Single-Particle Reconstruction**

►**EM Tomography**

Cryptic Splice Signals

Definition

Cryptic splice signals are sequences similar to functional splice signals that can be activated as active signals by mutations.

►**Repetitive DNA**

Cryptochrome

Definition

Cryptochromes comprise a diverse group of flavoproteins found in all kingdoms of life.

►**Photoreceptors**

Crystal Lattice

Definition

The periodic arrangement of unit cells in a crystal (the three dimensional grid of corner points of the unit cell).

► [X-Ray Crystallography—Basic Principles](#)

Crystal Packing

Definition

The arrangement of molecules of a crystal, usually referring especially to the packing interactions between neighboring molecules. These interactions can influence the geometry of the molecule.

► [X-Ray Crystallography—Basic Principles](#)

Crystalline Disorder

Definition

Deviations from perfect crystalline order, deriving for example from thermal motion, molecule heterogeneity, crystal growth defects, damage, etc. Crystalline disorder limits the diffraction of the crystal and lowers the resolution of the diffraction data.

► [X-Ray Crystallography—Basic Principles](#)

CsA

► [Cyclosporin A](#)

CSF

► [Colony Stimulating Factors](#)

CstF

► [Cleavage Stimulation Factor](#)

C-TAD

Carboxy-terminal transactivation domain.

► [Hypoxia Inducible Factors](#)

CTCF

Definition

The CTCF binding factor (CTCF) is a zinc finger protein guiding enhancer, enhancing activities to specific promoters by blocking enhancer activation of other promoters. This protein is involved in imprinting, X chromosome inactivation and tumor suppression.

► [Transcription Factors and Regulation of Gene Expression](#)

CTD

Definition

The term CTD describes the carboxyl-terminal domain of the largest subunit (Rpb1) of RNA Polymerase II. The CTD contains the phosphorylatable heptapeptide Tyr-Ser-Pro-Thr-Ser-Pro-Ser, repeated 26 or 27 times in yeast and 52 times in mammals.

► [Transcription Elongation](#)

CTL

Definition

Cytolytic T Lymphocytes (CTL) are a subpopulation of white blood cells. They recognize and kill cells that present epitopes of intracellular antigens through

► [MHC class I molecules.](#)

► [DNA-based Vaccination](#)

Current-Clamp Recording

Definition

Current-clamp recording is the commonly used protocol in patch clamp analysis. A certain current is injected to the cell while changes in membrane potential are measured.

► [Patch Clamping](#)

CVS

► Chorionic Villus Sampling

Cyclic AMP Receptor

Definition

Cyclic AMP receptor (CAR) is a Dictyostelium seven-transmembrane cAMP morphogen receptor upstream of Dictyostelium ► [Glycogen synthase kinase 3](#) (GSK3).

► Wnt/Beta-Catenin Signaling Pathway

Cyclin

Definition

Cyclins are a family of proteins that bind to and activate cyclin-dependent kinases, thus regulating the progression of cells in the cell cycle. Cyclins were originally discovered as proteins that are periodically expressed during the cell cycle.

► Cell Cycle – Overview

Cyclin D1

Definition

Cyclin D1 belongs to a family of proteins involved in progression to cell division.

► Cell Cycle – Overview

► Cell Division

► Hyper- and Hypoparathyroidism

Cyclin-Dependent Kinase

Definition

Cyclin-dependent kinases (Cdks) are closely related catalytic subunits of serine/threonine kinases, which require association with the positive regulatory cyclin for activation to control progression through the cell cycle. In yeast, a single Cdk (CDC28 in budding yeast and CDC2 in fission yeast) acts throughout the cell cycle with change in cyclins appropriate for each step. The situation is more complicated in mammalian cells.

Cyclin D/Cdk4 or 6 and cyclin E/Cdk2 are the G1 Cdks that phosphorylate Rb proteins to release E2F transcription factors. Cyclin E/Cdk2 and cyclin A/Cdk2 are the S phase promoting Cdks, and cyclin B/Cdk1 is the mitotic Cdk. In addition, cyclin A/Cdk1 appears active during S, G2 and M phases, although the biological consequences have yet to be clarified.

► Cell Cycle – Overview

► DNA Replication Initiation

► Growth Factors

► Rho, Rac, Cdc42

► Senescence

► Serine/Threonine Kinases

Cycling Genes

Definition

Cycling genes are genes which show oscillating rostrocaudal expression in the presomitic mesoderm during early embryonic development.

► Somitogenesis

Cyclooxygenases

Definition

Cyclooxygenases are enzymes that metabolize arachidonic acid to generate prostaglandin H₂ (PGH₂). PGH₂ is the precursor of many other prostaglandin metabolites such as PGE₂, thromboxan and PGI₂. Amongst many other activities, prostaglandins regulate inflammatory responses. Cyclooxygenases are the targets of aspirin-like drugs.

► Cytokines

Cyclopamine

Definition

Cyclopamine is a teratogenic steroidal alkaloid isolated from lily *Veratrum californicum*. The progeny of ewes feeding on this plant in the 1950's displayed a severe defect characterised by cyclopia, absent midline facial structures and a failure of forebrain division (holoprosencephaly). The ► [Hedgehog Signalling Pathway](#) is specifically inhibited by cyclopamine.

Cyclophilins

► Immunophilins and Cyclophilins

Cyclosporin A

Definition

Cyclosporin A is a cyclic undecapeptide that belongs to a family of secondary metabolites from the fungus imperfectus *Beauveria Nivea*. It has immunosuppressive, anti-inflammatory and anti-chemotactic properties. It shows strong affinity for ►Peptidyl Prolyl Isomerases of the cyclophilin family.

► Immunophilus and Cyclophilus
► Peptidyl Prolyl Cis/Trans Isomerases

Cysteine Proteases

Definition

Cysteine proteases constitute an important class of enzymes involved in the formation and hydrolysis of the peptide amide bond; these proteases require a cysteine in the active site for activity. Cysteine proteases have vital roles in mammalian cellular turnover and apoptosis, they are also very important in the life cycle of many parasites. A few examples of cysteine proteases are given: caspases, cathepsin F, and SUMO.

► Limb Girdle Muscular Dystrophies
► Sumoylation

Cystic Fibrosis (CF)

Definition

Cystic fibrosis, a common autosomal recessively inherited disease of the lung, intestines and pancreas, is caused by mutations in the gene encoding the cystic fibrosis transmembrane conductance regulator (►CFTR) protein, which functions as a chloride ion channel in the cells of secretory epithelia. The $\Delta F508$ CFTR mutation, which is carried by ~90% of Caucasian CF patients, results in the production of a misfolded but functionally active CFTR protein that is retained in the endoplasmic reticulum and targeted for degradation.

► Epistasis in Cystic Fibrosis

Cystic Fibrosis Transmembrane Conductance Regulator

Definition

Cystic fibrosis transmembrane conductance regulator (CFTR) is a protein that mediates and regulates a plasma membrane chloride conductance pathway required for transepithelial fluid secretion in many organs including the lung, intestines and pancreas. Loss of function of this protein results in ►cystic fibrosis.

► Autosomal Dominant (Inherited Disorder)
► Polycystic Kidney Disease, Autosomal Dominant

Cystine Knot

Definition

Cystine knot defines a structural conformation of a protein due to an unusual arrangement of six cysteines and the formation of three disulfide bonds.

► Neutrophilic Factors

Cytochrome P450

Definition

Cytochrome P450 (CYP) is a hemecontaining superfamily of enzymes that function as terminal oxygenases in cytochrome P450-dependent monooxygenase systems, and the reduced complex with carbon monoxide absorbs maximally at 450 nm.

CYP450 enzymes catalyze xenobiotics including drugs. They are also involved in the biosynthesis of steroid hormones, bile acids, fatty acids, and eicosanoids.

► Pharmacogenomics

Cytochrome C

Definition

Cytochrome c is an essential component of the mitochondrial respiratory chain. It is a soluble protein,

localized in the intermembrane space and loosely attached to the surface of the inner mitochondrial membrane. Cytochrome c carries electrons, which are transferred to molecular oxygen by the cytochrome oxidase. In the process of apoptosis, cytochrome c is released into the cytoplasm and induces, together with other molecules, caspase activation.

► Apoptosis

► Apoptosis, Regulation and Clinical Implications

Cytofluorometry

► Flow cytometry

Cytogenetic

Definition

Cytogenetic refers to the study of human chromosomes and karyotypes.

► Chromosome 21, Disorders

Cytokine Cluster Region

Definition

The cytokine cluster region is a region on chromosome 5q31 that contains several genes encoding cytokines (i.e. IL4, IL5, IL13, and IL3).

► Atopy Genetics

Cytokine Receptors

Definition

These heterodimeric receptors bind cytokines and signal by activating Janus kinases, which activate the STAT transcription factors.

► Cytokines

► Transcription Factors and Regulation of Gene Expression

Cytokines

MICHAEL KRACHT, HELMUT HOLTMANN
Medizinische Hochschule Hannover, Hannover,
Germany
kracht.michael@mh-hannover.de
holtmann.helmut@mh-hannover.de

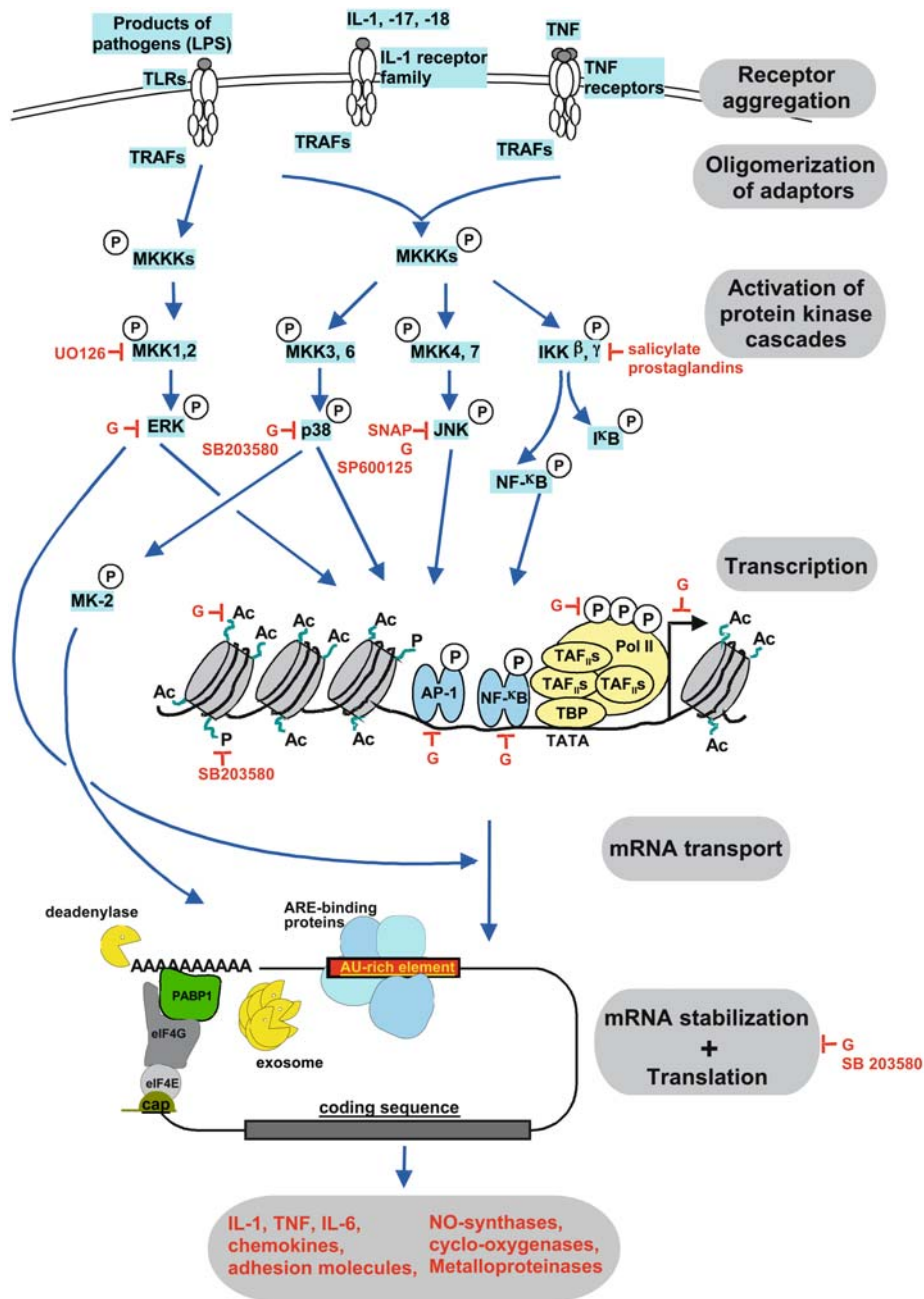
Definition

Cytokines are a large and heterogeneous group of small polypeptides, most of them with a molecular mass between 8–30 kD. They are mostly soluble extracellular mediators (some exist in membrane-associated form) and function through high affinity cell surface receptors. Released by cells of hematopoietic origin (reflected in the terms “lymphokines” and “monokines”) as well as by a wide variety of non-hematopoietic cells, cytokines are critical for the normal functioning of the immune defense, which they coordinate by communicating between participating cells.

As cytokines are not a structurally defined entity, the following terms are used as synonyms or for subgroups: lymphokines, monokines, interleukins, chemokines, interferons, growth factors.

Characteristics

Historically, many cytokines were discovered by means of their immunoregulatory properties, hence the term interleukins. However, cytokines regulate not only functions of the specific immune system, but also many aspects of infection and inflammation. At the cellular level they control proliferation, differentiation, migration, functional activity and apoptosis. As a whole, they are synthesized by a broad variety of different cell types. This also holds true for many cytokines individually (e.g. interleukin (IL)-6, IL-8, interferon (IFN)- β , for further information on individual cytokines see ► <http://www.copewithcytokines.de> (1)), whereas synthesis of others is restricted to one or a few cell types (e.g. IL-2, IL-4, IFN- γ). While some cytokines are continuously formed and play a role in normal homeostasis, for most of them biosynthesis is rapidly and highly inducible by changes in the cell's microenvironment. Cytokines often have a short half-life and production is restricted to potentially pathological conditions. Uncontrolled synthesis results in chronic inflammatory or other diseases. All cytokines act locally in an autocrine or paracrine manner, but some may also act systemically. Examples of the latter are the induction of fever by IL-1 and tumor necrosis factor (TNF) and the induction of acute phase proteins



Cytokines. Figure 1 Schematic representation of mechanisms that regulate cytokine-dependent gene expression. Proinflammatory cytokines (IL-1, IL-17, IL-18, TNF), but also ligands of toll-like receptors (TLRs) induce the aggregation of their plasma membrane receptor chains. The cytosolic portions of these receptors trigger oligomerization of adaptor proteins such as TRAF molecules, which in turn activate protein kinase cascades. Subsequently, a sequence of reversible phosphorylation steps results in activation of the MAP kinases JNK, p38 and ERK and in activation of the transcription factor NF-κB. MAP kinases induce and activate AP-1 proteins. Activity of AP-1 proteins and of NF-κB is modulated by phosphorylation. Modifications of histone H3 and H4 allow remodeling of the chromatin structure in promoter regions to facilitate binding of AP-1 and of NF-κB transcription factors to DNA. NF-κB and AP-1 in conjunction with components of the basal transcriptional machinery (TBP, TAF_{II}s) and many additional proteins form a large multimeric protein complex whose composition is specific for each individual gene (7). This complex ultimately allows rapid and efficient recruitment of RNA polymerase II. The newly formed RNA is capped, spliced, polyadenylated and exported to the cytoplasm. Half-life and in some cases translation is limited by AU-rich elements in the 3' untranslated regions of many cytokine and other rapidly inducible genes. This restriction is

in the liver by IL-6. Cytokines can therefore also be regarded as hormones. Cytokines are pleiotropic, i.e. they control multiple biological responses (2, 8). There is considerable functional overlap between certain cytokines (e.g. TNF and IL-1). In general, activation of a cell by a cytokine results in reprogramming of the cell's gene expression profile.

Cytokines in Medicine

Knowledge on cytokines has been exploited to improve the therapy of many diseases, basically in two ways:

Application of Cytokines as Drugs

Recombinant cytokines can be used to boost immune reactions during infection and cancer or to substitute cytokine deficiencies. In general, treatment with cytokines has turned out to be problematic due to their pleiotropic nature of function. Thus tumor necrosis factor, apart from its antitumoral effect, has severe systemic side effects as manifestations of its activity, which largely precludes its use as an anticancer drug. Similar toxicity has been observed for several other inflammatory cytokines.

Interferons induce the expression of antiviral, anti-proliferative and immunomodulatory genes. They are applied in the treatment of viral diseases, including chronic hepatitis B and C, certain malignancies and (IFN- β) multiple sclerosis.

Human recombinant colony-stimulating factors (GM-CSF, G-CSF) promote the differentiation of pluripotent bone marrow stem cells to leukocytes. They are effective in the treatment of congenital and acquired neutropenias during chemotherapy of cancer or bone marrow transplantation. Interleukin 2, a growth factor and activator of T-lymphocytes, is applied in the therapy of metastasizing renal carcinoma and melanoma. Erythropoietin is physiologically produced in the kidney and regulates proliferation of committed progenitors of red blood cells. It is used to supplement erythropoietin in

severe anemias due to end stage renal disease or treatment of cancer with cytostatic agents.

Interference with Cytokine Action

Inhibition of cytokine synthesis has turned out to be a major component of the activity of anti-inflammatory drugs, glucocorticoids being a prominent example. Because the molecular mechanisms of the actions of many cytokines have been worked out in detail, rational strategies for interfering with their biological activities have been or are being developed. The action of cytokines can be suppressed pharmacologically by different means.

1. Inhibition of their synthesis by drugs. The immunosuppressants cyclosporine and tacrolimus inhibit T-lymphocyte activation and IL-2 gene expression. Sirolimus inhibits IL-2-dependent progression of activated T-lymphocytes through the cell-cycle. All three drugs suppress clonal expansion of antigen-activated T-lymphocytes and are used to prevent rejection after organ transplantation.
2. Prevention of their interaction with cell surface receptors by soluble receptors, anti-cytokine or -cytokine receptor antibodies or natural antagonists. Etanercept, a chimeric molecule consisting of two extracellular domains of the TNF receptor I linked by Fc portions of human IgG1, binds with high affinity to extracellular TNF and reduces TNF activity. It is used to treat Crohn's disease and rheumatoid arthritis (4, 5). Side effects include immunosuppression and increased risk of infections.
3. Blockade of specific events in the intracellular signaling pathways (protein kinase inhibitors). Imatinib (Glivec[®]) is used to treat chronic myeloid leukemia in Philadelphia-chromosome positive patients. In these patients translocation of parts of chromosomes 9 and 22 results in the expression of a fusion protein with increased tyrosine kinase activity, called Bcr-Abl. Imatinib is a small

found to be alleviated—further boosting expression—by some of the signaling pathways activated by inflammatory inducers, including the p38, JNK and ERK as well as PI-3 kinase pathways. Likely targets are proteins that bind to the AU-rich elements and control mRNA metabolism (9). These proteins presumably act by interfering with the closed loop structure of cytoplasmic mRNA, formed by proteins that bind to its cap and poly(A) tail, limit access of degrading enzymes and facilitate translation.

All together these mechanisms allow rapid and tightly controlled synthesis of many pro-inflammatory proteins (e.g. many other cytokines and chemokines). Central steps in these pathways are inhibited by glucocorticoids (G), protein kinase inhibitors (SP600125, SB203580, SB20129), NO-donors (SNAP; S-nitro-N-acetyl-DL-penicillamine), salicylates and the anti-inflammatory cytokine IL-10, explaining the well-known anti-inflammatory properties of these agents.

Abbreviations: Ac, acetylation; AP-1, activating protein-1; ARE, AU-rich element; eIF, eukaryotic initiation factor; ERK, extracellular-regulated kinase; I κ B, inhibitor of NF κ B; JNK, JUN-N-terminal kinase, LPS, lipopolysaccharide; MAPK, mitogen-activated protein kinase; MKKK, MAPK kinase kinase; MKK, MAPK kinase; MK-2, MAPK-activated protein kinase 2; P, phosphorylation; Pol II, polymerase II; PABP1, poly(A)-binding protein 1; TAF_{II}, TBP-associated factor; TBP, TATA Box-binding protein, TLR, toll-like receptor; TRAF, TNF-receptor associated factor.

molecular weight inhibitor selective for the tyrosine kinase activity of Bcr-Abl. It thereby inhibits the Bcr-Abl induced cell-cycle progression and the uncontrolled proliferation of tumor cells.

Clinically used cytokines or cytokine antagonists are proteins. Repeated application, especially of those of non-human origin, bears the risk of recognition by the immune system and antibody production against them with subsequent loss in effectiveness of the drug. Therefore the proteins are often genetically engineered, i.e. humanized to be as homologous as possible to human proteins.

Structural/functional groups of cytokines (examples):

interferons	(IFN- α , β , γ)
hematopoietic cytokines/ colony-stimulating factors	(IL-3, IL-5, IL-7, GM-CSF, G-CSF, M-CSF)
IL-1 family	(IL-1 α , IL-1 β , IL-18, IL- 1RA)
TNF family	(TNF- α , TNF- β , CD40 li- gand, CD95 ligand, TRAIL)
chemokines	(IL-8, MIP-1- α , RANTES)
immunoregulatory cyto- kines	(IL-2, IL-4, IL-10, IL-12)

Molecular Interactions

The interaction partners of cytokines are receptors on the plasma membrane to which they bind with high affinity and specificity. These receptors fall into different classes, those with enzymatic tyrosine kinase activity (e.g. the receptors of certain hematopoietic growth factors (M-CSF, SCF)), G-protein-coupled receptors (chemokine receptors) and several groups that bind to other intracellular adaptor proteins (IL-1-, TNF-, interferon- and hematopoietic cytokine-receptor families). Ligand-dependent receptor clustering activates several intracellular signaling pathways. These pathways consist of consecutive interactions between protein components that in part involve enzymatic activities. The activity of many proteins in a given cytokine-activated pathway is regulated by reversible phosphorylation. The ►MAPK (mitogen-activated protein kinase) cascades (Fig. 1) and the ►JAK/STAT and ►PI-3 kinase pathways are intensively studied examples. Certain receptors (of the TNF-R family) can activate ►caspases; a family of proteases that execute ►programmed cell death. Other protein synthesis-independent effects of cytokines include cytoskeletal changes and modulation of cell surface receptors and adhesion molecules. Most cytokine-activated intracellular pathways ultimately turn on expression of genes whose products change the biological function of the

same or a neighboring cell. These include cytokines (hence the term “cytokine network”), cell surface receptors and adhesion molecules and enzymes involved in degradative processes (e.g. collagenases) and in formation of small molecular weight mediators (e.g. ►cyclooxygenase I and II, inducible NO-synthase).

Receptors for a number of cytokines have been demonstrated to exist in soluble forms, derived from their cell surface counterparts either by alternative splicing or by proteolytic cleavage of the extracellular domain. The soluble receptors are able to bind their ligands, which can have fundamentally different consequences; in the case of TNF, they can sequester the ligand and prevent it from acting on the cell surface receptors, a principle used in pharmacotherapy (see above). Soluble IL-6 receptors on the other hand can bind the ligand and then associate with a signal-transducing β -chain shared with other receptors (gp130). This can occur even on cells lacking the specific α -chain and result in responsiveness of these cells to IL-6 (4).

Regulatory Mechanisms

Most cytokines are rapidly induced in response to several groups of external cues that include:

bacterial, viral, fungal and parasite structures that are recognized by members of the Toll-like receptor family (3).

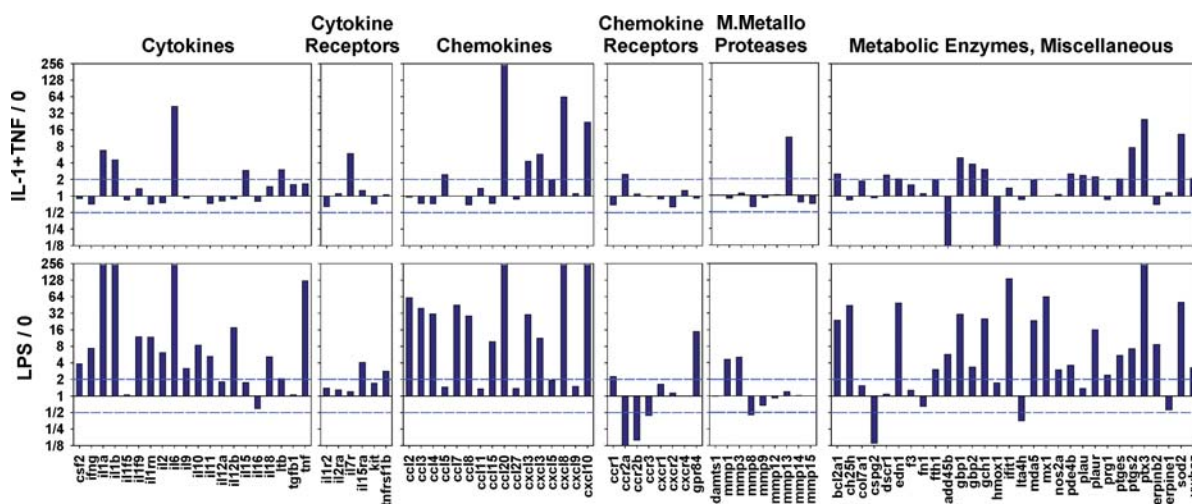
nonspecific physicochemical stress/cell damage

molecules formed/activated in response to the above stimuli:

- released cytokines (e.g. TNF, IL-1)
- activated complement components
- structures of the specific immune system (antibodies, T cell receptors).

The intracellular pathways activated by these stimuli are manifold, among them those depicted in Fig. 1. They have to enable stringent spatial, temporal and qualitative control of the response, necessitated by the potentially hazardous consequences of unscheduled cytokine activities. Absence of cytokine formation under normal conditions is the consequence of low promoter activity and rapid turnover of the mRNAs. Several thousand fold increases in cytokine formation are achieved within minutes by simultaneous activation of transcription, stabilization of the mRNA and increase in translational efficiency (7). Several mechanisms restrict the response and let the system return to its basal state, e.g. down regulation of receptors, induction of phosphatases that inactivate components of the signaling pathways, etc. Systemically the release of glucocorticoids is a major feedback mechanism that dampens the response at multiple levels, as depicted in Fig. 1.

The sets of cytokines induced by the different conditions and mechanisms imposing selectivity of induction are



Cytokines. Figure 2 Distinct and common patterns of genes induced by cytokines or pathogenic stimuli. Human epithelial cells were treated with IL-1 (10ng/ml) plus TNF (20ng/ml) for 4 hours or left untreated. Then, total RNA was isolated and transcribed into cDNA followed by fluorophore-labeled cRNA synthesis. The cRNA mixtures from both samples were hybridized to DNA microarrays containing oligonucleotide probes for 136 human genes with known relevance to inflammation. Fluorescence intensity bound indicates the amount of the corresponding mRNA expressed. Results are depicted as ratio of expression comparing treated with untreated cells. The lower panel shows an analogous experiment comparing gene expression of untreated peripheral blood mononuclear cells with cells that were treated for 3 hours with lipopolysaccharide (LPS, 100ng/ml), a cell wall component of Gram-negative bacteria. The blue lines indicate twofold regulation. Note that LPS induces many cytokine and chemokine genes, among others IL-1 and TNF. Cytokines (e.g. IL-1 and TNF) and LPS also induce metabolic enzymes such as MnSOD (sod2) or prostaglandin synthases (ptgs2, ptges) and NO-synthases (nos2a), which produce prostaglandins and NO, crucial mediators of acute inflammation. Persistent induction of destructive enzymes (e.g. matrix metalloproteases) contributes to the tissue damage observed in chronic inflammation (data provided by Michael Kracht and Oliver Dittrich-Breiholz).

under investigation. This involves DNA microarray-based approaches (6). An example is shown in Fig. 2.

- Crohn Disease
- Cytokines
- DNA-based Vaccination
- Growth Factors
- Interleukins and Their Receptors, a Role in Regulating Hematopoietic Stem Cells
- Mesenchymal Cells
- Morbus Wegener
- Psoriasis, Molecular Basis
- Rheumatoid Arthritis

References

1. Horst Ibelgaufts Cytokines Online Pathfinder Encyclopaedia. Free online information at ► <http://www.cope-withcytokines.de>
2. Baggiolini M (2001) Chemokines in pathology and medicine. *J Intern Med* 250:91–104
3. Barton GM, Medzhitov R (2003). Toll-like receptor signaling pathways. *Science* 300:1524–1525
4. Economides AN, Carpenter LR, Rudge JS et al (2003) Cytokine traps: multi-component, high-affinity blockers of cytokine action. *Nat Med* 9:47–52

5. Feldmann M (2002) Development of anti-TNF therapy for rheumatoid arthritis. *Nat Rev Immunol* 2:364–371
6. Huang Q, Liu D, Majewski P et al (2001) The plasticity of dendritic cell responses to pathogens and their components. *Science* 294:870–875
7. Kracht M, Saklatvala J (2002) Transcriptional and post-transcriptional control of gene expression in inflammation. *Cytokine* 20:91–106
8. Nathan C (2002) Points of control in inflammation. *Nature* 420:846–852
9. Bevilacqua A, Ceriani MC, Capaccioli S et al (2003) Post-transcriptional regulation of gene expression by degradation of messenger RNAs. *J Cell Physiol* 195:356–372

Cytokines

Definition

Cytokinesis is the process in which the cytoplasmic content, organelles and cytoplasm are separated. It is usually part of cell division.

- Cell Division

Cytolysis

Definition

Cytolysis is the pathological breakdown of cells by the destruction of their outer membrane.

► [Heritable Skin Disorders](#)

Cytolytic T Lymphocytes

► [CTL](#)

Cytoplasm

Definition

The compartment in the cell that is delineated by the plasma membrane in eukaryotic cells, exterior to the nucleus. The cytoplasm contains the mitochondrial network and several other important components of the cell, also called cytosol.

► [Cell Polarity](#)

► [Mitochondrial Myopathies](#)

Cytoskeletal Linkers

Definition

Cytoskeletal linkers are proteins that have the capacity to associate with more than one cytoskeletal system, enabling them to integrate two or all three cytoskeletal elements (actin filaments, microtubules and intermediate filaments).

► [Actin Cytoskeleton](#)

► [Hemidesmosomes](#)

► [Cytoskeleton: Microtubules and Intermediate Filaments](#)

Cytoskeleton

Definition

Cytoskeleton is a dense network of various filamentous proteins in eukaryotic cells, which are essential for cell shape, cell division or cell migration. Actin filaments, microtubules, and intermediate filaments form the major components of the cytoskeleton. They interact with many other proteins in the cell. Recently, a number of inter-► [cytoskeletal linkers](#) have been described.

► [Actin Cytoskeleton](#)

► [Cardiac Signaling: Cellular, Molecular and Clinical Aspects](#)

► [Cell Polarity](#)

► [Cytoskeleton: Microtubules and Intermediate Filaments](#)

► [Heritable Skin Disorders](#)

► [Integrin Signalling](#)

► [Intermediate Filaments](#)

► [Limb Girdle Muscular Dystrophies](#)

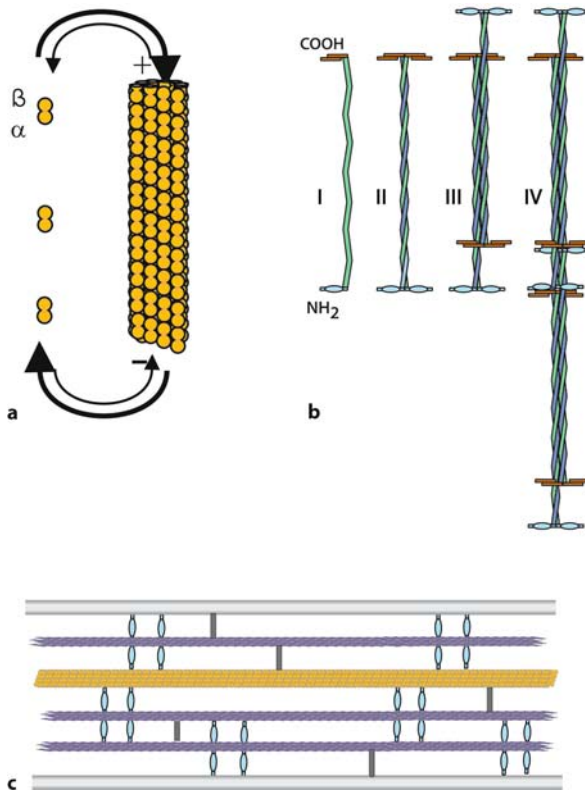
Cytoskeleton: Microtubules and Intermediate Filaments

DIETER K. MEYER

Institute for Experimental and Clinical Pharmacology and Toxicology, Albert Ludwig University Freiburg, Freiburg, Germany
dieter.meyer@pharmakol.uni-freiburg.de

Definition

► [Microtubules](#), microfilaments (F-actin) and ► [intermediate filaments](#) form the cytoskeleton. These polymers made from homo- or hetero-meric subunits are highly dynamic structures, which adapt the cellular architecture in response to intra- and extra-cellular signals. They are essential for cell morphology and attachment as well as motility. They play an important role in proliferation and organize intracellular compartments and transport. In addition, cytoskeletal elements anchor receptors and ion channels in the cell membrane and take part in the subsequent signal transduction pathways. In this chapter, only microtubules and intermediate filaments will be discussed.



Cytoskeleton: Microtubules and Intermediate Filaments. Figure 1 Microtubules and intermediate filaments are composed of monomers. (a) Microtubules. α - and β -subunits form dimers, which are added at the plus end of a microtubule and released from its minus end during the process of treadmilling. During the process of dynamic instability, dimers are added and released from the plus end. (b) Intermediate Filaments. Two monomers of an intermediate filament (I) form a coiled-coil dimer (II). Different colours are used for better resolution and to remind one of the fact that different monomers can associate to a dimer. Two dimers then associate in a half-staggered manner to anti-parallel tetramers (III). Head to tail addition of further tetramers causes elongation (IV). (c) Interactions of microtubules and intermediate filaments in axons. In axonal extensions of neurons, microtubules and neurofilaments (type IV intermediate filaments) interact. The C-terminal of neurofilaments (blue ellipsoids) can link to microtubules, as can plakin (brown bar), which binds to neurofilaments (modified from Lariviere and Julien (2003)).

Characteristics

Microtubules (MTs)

The globular polypeptides α - and β -tubulin have molecular weights of about 50 kD. Whereas both tubulins bind GTP, only the GTP bound by β -tubulin is exchangeable. Higher eukaryotes have up to 7 isoforms of α - and β -tubulin, respectively, which are encoded by

different genes. Via non-covalent bonds, α - and β -tubulins associate to heterodimers, which polymerize head to tail to form protofilaments. Thirteen protofilaments make up a hollow tube with a diameter of 25 nm, a microtubule (MT) (Fig. 1a). These polarized structures have β -tubulins at the plus ends and α -tubulins at the minus ends. Within the MTs, non-covalent bonds between α - and β -subunits cause lateral stabilization. When the β -tubulins at the plus end of the protofilament are in the GTP binding state, addition of further heterodimers is possible, if the concentration of free heterodimers is higher than a certain critical concentration (Fig. 1). GTP bound β -tubulins at the plus end protect the MT from degradation. They form a GTP cap. When the β -tubulins at the plus end of the protofilament are in the GDP binding state however, depolymerization is more likely. The transition from growing to pause and shortening observed at the plus end is called “**dynamic instability**”. In contrast, “**treadmilling**” is a net addition of tubulin heterodimers at the plus end and the balanced net loss from the minus end, so that the total length of the MT remains constant. In most cells there are at least 2 pools of MTs, i.e. dynamic MTs, which show dynamic instability with a half-life of 10 min and stabilized MTs, which can persist for several hours. De-tyrosination of the C-terminal of the α -subunit exposes a Glu residue and stabilizes MTs. Dimers cannot be added or removed (5).

MTs are stabilized at their minus ends by the **centrosome** (also called the MT organizing center, MTOC). Centrosomes are protein complexes containing among other proteins 2 centrioles (ring-like structures) and γ -tubulin. Centrosomes serve as nucleation points for microtubular polymerization and constrain the lattice structure of an MT to 13 protofilaments. Centrosome-independent nucleation is possible and may be induced by DNA or proteins such as cadherin.

Intermediate Filaments

The proteins of the intermediate filament superfamily form polymeric filaments with a diameter of 10 nm, which serve as scaffolds and stabilize cells. Proteins of the lamin family are found only in cell nuclei, whereas the filaments of the other families occur cytosolically. The various subunits have in common a central rod domain with 4 highly conserved alpha-helical domains (1a, 1b, 2a, 2b), which are separated by linkers (Fig. 1b). The alpha-helical domains consist of heptad repeats with apolar residues in positions 1 and 4 at the outside of the helix, which are essential for dimerization. Gene specific amino-terminal head domains and a carboxy-terminal tail domain flank the rod domain. Hydrophobic interactions cause the rod domains of 2 subunits to wrap around each other, thus generating a coiled-coil (Fig. 1b). Vimentin or desmin

filaments form homodimers, while keratin filaments or neurofilaments form heterodimers. Two dimers then associate in a half-staggered manner to antiparallel tetramers (Fig. 1b). Longitudinal growth is due to head to tail addition of further tetramers. Eight tetramers packed together side-by-side in a helical array form a filament with a diameter of 10 nm.

Six Types of Intermediate Filaments Are Known

Type I and II intermediate filaments comprise the acidic and basic keratins. Type I keratins (subunits K9 to K20, IRSa1-Irsa3, Ha1-Ha9) are acidic, whereas type II keratins (subunits K1-K8, Hb1-Hb6) are neutral to basic. Both types form heterodimers. Keratins are present in endothelia, epithelia and epidermis as well as in hair follicles and nails. Keratins have also been found in hepatocytes. Expression of the various subunits is tissue-specific and, in the ►epidermis, layer-specific (6).

Type III filaments are also called vimentin-like filaments and form homodimers. Vimentin is present in fibroblasts and embryonic tissues, desmin in muscle cells, glial fibrillary acidic protein in non-myelinated Schwann cells and astroglial cells and peripherin in developing neurons. Nestin is found in early neuroectodermal cells.

Type IV filaments comprise the neurofilaments and α -internexin. Neurofilaments consist of the subunits NF-H (115 kD), NF-M (90 kD) and NF-L (61 kD). They form obligatory heterodimers, which consist of NF-L together with NF-M or NF-H. Neurofilaments are present in developing and mature neurons. α -internexin (also called NF66; 64–66 kD) is considered to be a neurofilament protein. It can form homopolymers, although co-assembly with NF subunits is possible. It is present in neurons during early development.

Type V: nuclear lamins. Intermediate filaments of the lamin family (lamins A/C and B) line the inner nuclear membrane.

Type VI intermediate filaments are filensin and phakinin. Little is known about the function of these recent additions to the group of intermediate filament proteins.

Molecular Interactions

Microtubules (MTs)

►Microtubule associated proteins (MAPs) bind to MTs *in vivo* and play a role in their nucleation, growth, stabilization (by cross-linking) and motion. Of the numerous MAPs (MAP1, MAP2 and tau in neurons; MAP4 in non-neuronal cells), the tau family proteins have received special attention, since they have been implicated in neurodegeneration (see below). MAPs are negatively regulated in their activity by phosphorylation.

►Motor proteins move along MTs in an ATP-dependent manner. Members of the superfamily of

kinesin motors move only to the plus and dynein motors only to the minus end. The respective motor domains are linked to their cargoes *via* adaptor proteins. The binding of the motors to MTs is regulated by kinases and phosphatases. When motors are immobilized at their cargo binding area, they can move MTs.

MTs extend from the centrosome throughout the cytoplasm to the plasma membrane, where they are stabilized by caps. Sliding along the MTs, kinesin and dynein motors transport their cargoes between the center and the periphery of the cell. MTs present in the axons of neurons are extended not only by addition of heterodimers to the plus ends but also by use of short MTs that have been initiated in the centrosome. Their axonal transport is mediated by dynein motors, which are passively moved along actin filaments. Once formed in the axon, MTs serve as tracks for fast axonal transport, i.e., the movement of membranous organelles and membrane proteins to the nerve ending.

Some specialized eukaryotic cells have cilia, which show a whip-like motion. Sperm cells move with one flagellum, which is much longer than a cilium but has a nearly identical internal structure called an axoneme. It is composed of nine doublet MTs which form a ring around a pair of single MTs. Numerous proteins bind to the MTs. Ciliary dynein motors generate the force by which MTs slide along each other to cause the bending of the axoneme necessary for motion.

Intermediate Filaments

Since the regulatory mechanisms of the intermediate filaments are little known, their molecular interactions and defects resulting from mutations will be briefly summarized.

Intermediate Filaments of Type I and II (Keratins)

In some cell types, e.g. in epidermis, hair follicle or nail bed, keratins of type I and II make up a large percentage of the total protein content. Keratin filaments are essential for cellular resistance to mechanical strain. Mutations in keratins can thus result in the rupture of epidermal layers or of epithelial coherence. Numerous disorders related to keratin mutations have been reported (6). They include various epidermolytic diseases, corneal dystrophy, cryptogenic liver disease and disturbances of hair follicles and nails.

Intermediate Filaments of Type III (Vimentin-like Filaments)

Vimentin filaments are bundled by plectin, i.e. a member of the phakin protein family. Mutations resulting in diseases have not been reported.

Glial fibrillary acidic protein plays an important role in the function of astroglial cells. There is increasing evidence that heterozygous missense mutations in the

gene are connected to Alexander's disease, which is characterized by leukodystrophy, mental retardation and developmental delay (4).

Missense mutations and deletions of the desmin gene (located on chromosome 2q35) can lead to myopathies including cardiomyopathy (desmin-related myopathy) with autosomal-dominant or autosomal-recessive inheritance (7).

Peripherin can form polymers with neurofilament subunits. In the exploration of neurodegeneration, the role of peripherin currently elicits great interest (2).

Intermediate Filaments of Type IV (Neurofilaments)

Links of neurofilaments to cytoskeletal structures, such as MTs, are possible *via* the nonhelical C-terminal domains of the polymerized subunits, as exemplified by the tails of NF-M and NF-H in neuronal axons (Fig. 1C). Although neurofilaments are not necessary for axon elongation in neurons, deletions of the subunits NF-M, NF-H and NF-L cause loss of motor axons. In addition, deletions of NF-M or NF-L reduce the axonal diameter by >50%. Since the caliber of the axon determines its conductance, neurofilaments seem to play an important role in neuronal function (2). It is noteworthy that decreased levels of NF-L have been found in some patients with Alzheimer disease or **▶amyotrophic lateral sclerosis** (ALS). In addition, mutations in the NF-H gene and in the NF-L gene have been found in subpopulations of patients with ALS and **▶Charcot-Marie-Tooth disease**, respectively. Of the proteins interacting with neurofilaments those of the plakin family (such as plectins, BPAG1, ACF7, desmoplakin, envoplakin or periplakin) have been studied most. Plakins link neurofilaments to other cytoskeletal structures (Fig. 1). Deletion of BPAG1 causes neurodegeneration characterized by disorganization of MTs and neurofilaments (2).

Intermediate Filaments of Type V (Lamins B and A/C)

The lamins form a meshwork, which lines the inner surface of the nuclear membrane. Since heterochromatin can attach to the lamin lining, mutations in lamins A/C may alter the organization of heterochromatin and thus of gene expression. Mutations in lamin A/C may also cause mechanical strain within a cell nucleus, which results in changes in gene expression. These hypotheses are currently being studied to explain why mutations in lamins A/C, which are present in all nuclei of an organism, can cause tissue-specific defects, such as myopathy in striated muscle, partial lipodystrophy syndromes and peripheral neuropathy as well as premature aging syndromes (8).

Extensive information on intermediate filaments, i.e. their structures, gene sequences, mutations and connected diseases, is available at <http://www.interfil.org>.

Regulatory Mechanisms

Microtubules (MTs)

The membrane tubules and lamellae of the endoplasmic reticulum (ER) are extended in the cell with the use of MTs and actin filaments. Kinesin motors are required for stretching out the ER, whereas depolymerization of MTs causes the retraction of the ER to the cell centre in an actin-dependent manner. Dynein motors move proteins synthesized in the ER along MTs to the Golgi complex (GC), where they are modified and packaged. The resulting vesicles move along the MTs to the cell periphery transported by kinesin motors. MTs also determine the shape and the position of the GC. Their depolymerization causes the fragmentation and dispersal of the GC. Dynein motors are required to rebuild the GC.

When cells enter **▶mitosis**, the interphase array of MTs is dismantled. The centrosome duplicates, and the daughters move to opposite poles of the nucleus. After disassembly of the nuclear envelope, MTs emanating from both centrosomes show pronounced dynamic instability. They grow and shorten and thus probe the cytoplasm until they find attachment sites at the condensed chromosomes, the **▶kinetochores**. Dynein motors are involved in the attachment and in the subsequent segregation of the two sets of chromosomes. In some cells, spindles are formed in the absence of centrosomes. Here, DNA itself initiates microtubular nucleation at the kinetochores with the help of the GTPase Ran and/or the MT-destabilizing factor stathmin. Kinesin and dynein motors then bundle and focus the MTs at their minus ends (1).

MTs are present in neuronal axons and dendrites. In axons, they are linked to neurofilaments (Fig. 1C). They extend and stabilize axons and facilitate their internal transport. Tau protein binds to and stabilizes MTs, unless it is hyper-phosphorylated. There is accumulating evidence that tau is involved in several neurodegenerative diseases (so-called **▶tauopathies**), such as Alzheimer disease, frontotemporal dementia, progressive supranuclear palsy and corticobasal degeneration. These diseases have in common neurofibrillary tangles, i.e. filamentous cytosolic inclusions containing tau, which can lead to cell death. The cytotoxic effect of the inclusions is currently being investigated. They may either sequester tau or other proteins, which then cannot perform their normal functions, or may have direct toxic effects

Therapeutic Use of Agents with Action on Microtubules

The alkaloids **▶vinblastine** and vincristine extracted from the plant *Catharantus roseus* (also called *Vinca rosea*) and the semisynthetic derivatives vindesine, vinorelbine and vinflunine can block mitosis and are used clinically for cancer treatment. When used at

micromolar concentrations, the alkaloids bind with 1:1 stoichiometry to β -tubulin in a region between residues 175 and 213. The bound disassembled tubulin heterodimers can no longer polymerize to MTs and form paracrystalline tubulin-*Vinca* alkaloid arrays. Pre-existent MTs are subsequently depolymerized. Chromosomes are no longer separated but dispersed or clumped in the cytoplasm. The cells undergo apoptosis. When used at nanomolar concentrations, *Vinca* alkaloids may block mitosis at the metaphase-anaphase transition by inhibiting the dynamic instability and treadmilling of MTs, without affecting microtubular mass. The cytostatic agent estramustin has similar effects. It blocks heterodimer polymerization at micromolar concentrations and inhibits dynamic instability at nanomolar concentrations.

In contrast, ►**paclitaxel** (from the bark of the Western yew tree) and its more potent analogue docetaxel bind to MTs, when used at nanomolar concentrations. By attaching to β -tubulin between amino acids 239 and 254, the agents block the release of heterodimers from MTs and prevent the dynamic instability necessary for the capturing of chromosomes without affecting microtubular mass. Both taxols are used for cancer treatment. Labeled paclitaxel is also used for the immunocytochemical localization of MTs.

The alkaloid ►**colchicine** (from *Colchicum autumnale*) blocks tubulin polymerization by binding to heterodimeric β -tubulin between amino acids 239 and 254. Since it inhibits the MT-dependent migration of granulocytes into areas of inflammation and their MT-dependent release of pro-inflammatory agents, it is used to treat attacks of gout. Its antimitotic effect in the gastrointestinal system induces diarrhea. Nocodazole competes for the same binding site as colchicine and has similar effects on heterodimeric β -tubulin.

References

1. Karsenti E, Vernos I (2001) The Mitotic Spindle: a Self-Made Machine. *Science* 294:543–547
2. Lariviere RC, Julien JP (2004) Functions of Intermediate Filaments in Neuronal Development and Disease. *J Neurobiol* 58:131–148
3. Lee VM, Giasson BI, Trojanowski JQ (2004) More Than Just Two Peas in a Pod: Common Amyloidogenic Properties of Tau and Alpha-Synuclein in Neurodegenerative Diseases. *Trends Neurosci* 27:129–134
4. Messing A, Brenner M (2003) GFAP: Functional Implications Gleaned From Studies of Genetically Engineered Mice. *Glia* 43:87–90
5. Palazzo AF, Cook TA, Alberts AS et al (2001) mDia Mediates Rho-Regulated Formation and Orientation of Stable Microtubules. *Nat Cell Biol* 3:723–729
6. Porter RM, Lane EB (2003) Phenotypes, Genotypes and Their Contribution to Understanding Keratin Function. *Trends Genet* 19:278–285
7. Taratuto AL (2002) Congenital Myopathies and Related Disorders. *Curr Opin Neurol* 15:553–561
8. Worman HJ, Courvalin JC (2004) How Do Mutations in Lamins A and C Cause Disease? *J Clin Invest* 113:349–351

Cytosol

Definition

Cytosol describes a fraction of the cell that contains most of the extranuclear material.

►**Morpholino Oligonucleotides, Functional Genomics by Gene ‘Knock-down’**

DA

► Dopamine (DA)

daf

Definition

daf stands for dauer formation defective mutant of *C. elegans* that inappropriately regulates entry into the *dauer diapause*, a long lived alternate third larval stage.

► [C. Elegans as a Model Organism for Functional Genomics](#)

Dalton

Definition

In molecular biology and biochemistry the term dalton, denoted Da, is used for the unified atomic mass unit (amu). The unit honours the English chemist John Dalton (1766–1844), who proposed the atomic theory of matter in 1803.

► [Mass Spectrometry: MALDI](#)

Dapper (Dpr)/Frodo

Definition

Dapper (Dpr)/Frodo describes a family of vertebrate proteins that bind and modify ► [Dishevelled](#) (Dvl) activity.

► [Wnt/Beta-Catenin Signaling Pathway](#)

Data Availability

Definition

Data availability is a prerequisite to inform and to treat patients as well as to counsel carriers of minor or major predispositions for lifestyle, genetic or environmental health risk; data availability should be dealt with within the wider framework of legal and moral rights to privacy and data protection.

► [Ethical Issues in Medical Genetics](#)

Data Normalisation

Definition

During analysis of microarrays, data normalisation is a data transformation during which a set of spot quantitation matrices is transformed into a gene expression data matrix, by removing systematic noise, scaling and other nontrivial data processing steps.

► [Microarray Data Analysis](#)

Data-Mining in Biology, “How to Find a Needle in a Haystack?”

KARYN MÉGY

Sygen International plc, Department of Pathology,
University of Cambridge, Cambridge, UK
km369@cam.ac.uk

Definition

The recent increase in the amount of biological data raises the questions of storage and accessing, modelling and computing, description and understanding of these

data ... and makes data mining indispensable. Fortunately, in the same period, new and cheaper computers with a larger storage capacity and faster processing have been produced, enabling storage and computing of this information (= cheap data). New machine learning methods based on logic programming have been developed. Associated with the traditional statistical tools, they have made the modelling and analysis of these data possible and easier. Storage in databases allows the linkage of this information and assignation of functions. Here we have reached an apparently illogical point; we have first produced and collected more and more data, and now we try to find a use for them, a question which they could answer, whereas the classical scientific approach is first to have a question and then to collect data to try to answer it. In fact, the amount of data may be voluminous, but it is of low value, as no direct use can be made of it. It is the information hidden in this large volume of raw data that is useful. The idea is that it is possible to strike gold in unexpected places by extracting information not obviously discernable (or so obvious that no-one noticed it before!). Data mining is the process of automatically extracting such hidden knowledge. It can be like searching for a needle in a haystack ... and it often is!

The following review describes data mining in general and then focuses on data mining in biology with the example of ►Expressed Sequence Tag (EST) mining.

Characteristics

General Method or “How to Mine Data”

Data processing can be divided into several steps (Fig. 1). Scientists do not really agree on a common definition; some argue that it is the extraction of data, others that it is the whole process of data processing or a part only of this process. Let us consider in this review three general stages in data mining.

Exploration

This stage usually starts with data preparation, which involves data cleaning, data transformation, data selection and feature selection if data are characterized by a large number of features.

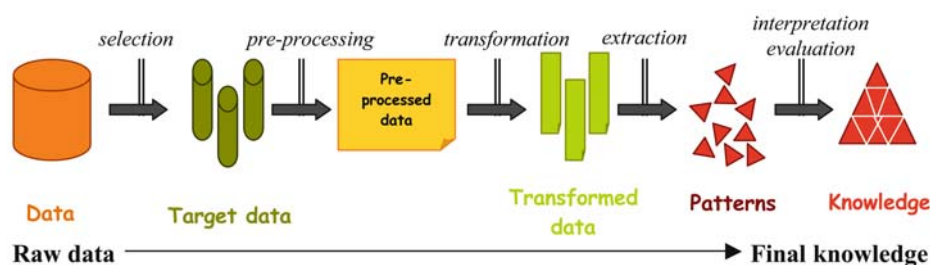
The data-cleaning step has to be done carefully. The old saying “garbage-in, garbage out” is particularly applicable to data mining projects where large sets of automatically collected data serve as input. The method of gathering data is often not well controlled and data might contain out of range values (e.g. sequence size: – 100) or impossible combinations (e.g. newborn – age: 30 years), producing highly misleading results.

If the dataset is too large, it has to be reduced to a more manageable size by selecting specific data or reducing the feature number. Data can be selected according to some criteria (e.g. sequences >500 bp) or reduced using aggregation (e.g. group together sequences >500 bp) or ►clustering. The selection of features consists of discarding those that are not useful for the actual model-building phase.

Optimal Representation of the data structure and Validation

This stage involves considering various models and choosing the best one based on its predictive performances, i.e. in explaining the variability in question and producing stable results across samples. This may sound like a simple operation but it sometimes involves a very elaborate process. There are a variety of techniques developed to achieve this goal; many of them are based on the “competitive evaluation model”. This consists of applying different models, based on different methods, to the same set of data and comparing their performances to choose the best. Classical methods are based on ►nearest neighbour, clustering, ►decision tree, ►neural networks and ►genetic algorithms. The choice of one or the other method depends on the data gathered, on the problem to be solved and on the available computing tools.

Data verification is a primordial step often neglected to save time. Exploration of data is only the first stage of the analysis and the results cannot be considered as reliable as long as they are not confirmed. The validation of the results can be achieved by applying the model to another data set - or independent subsets - and checking for the predictions or experimentally confirming the results.



Data-Mining in Biology, “How to Find a Needle in a Haystack?”. Figure 1 The pipeline of data processing.

Application of the model to New Data in order to generate predictions

The final stage involves using the model selected as best in the previous stage and applying it to new data in order to generate predictions or estimate the predicted outcome.

EST Mining

Applications of data mining in biology are countless, sequence mining (6), pattern identification (5, 9), pathway discovery (7, 12), gene expression analysis (micro arrays, ESTs, ►SAGE...) to the whole genome or to specific tissues (4, 8, 10, 13) and much more.

►ESTs are tags representing expressed genes, thus they are often used in genome annotation or in pseudogene discovery. Because the frequency of ESTs representing a gene gives information on its abundance, ESTs are widely used in gene expression analysis as well. Moreover, the large amount of EST data allows analysis of a large number of genes from a given organism(s) and in a given condition(s). Countless resource centres based in almost every country provide EST data and/or tools to help in EST mining. The main ones are the NCBI (National Centre for Biotechnology Information) and the EBI (European Bioinformatics Center). Besides storing all kinds of data, they provide tools for their analysis such as sequence comparison, expression analysis and genome comparison (Table 1).

The analysis of EST generally involves the following steps.

Exploration

• Data Selection

dbEST lists all the ESTs and libraries publicly available. In April 2004, dbEST contained 20,624,959 entries (release 040204) in more than 700 organisms. But not all of them are useful in a given analysis. Thus the first step in any EST analysis is to select the useful ESTs and libraries. Some ESTs are too short to be of any use and must be removed. Others contain repeats, vectors or low quality sequences that have to be masked (using RepeatMasker – A.Smit and P.Green, unpublished data). Some libraries containing too few ESTs are not considered reliable enough for analysis of the gene expression level and have to be discarded. Libraries generated from the same organ can be pooled because the protocols used to generate them are more or less the same. Such virtual libraries are often used instead of the individual libraries. For each new study, this step of data selection has to be redone because the criteria for EST or library usefulness may vary.

Data-Mining in Biology, “How to Find a Needle in a Haystack?”. Table 1 URL of some bioinformatics resources

EBI	► http://www.ebi.ac.uk/
Expression Profiler	► http://www.ebi.ac.uk/expressionprofiler/
Array Express	► http://www.ebi.ac.uk/Databases/microarray.html
NCBI	► http://www.ncbi.nlm.nih.gov
dbEST	► http://www.ncbi.nlm.nih.gov/dbEST/
Geo	► http://www.ncbi.nlm.nih.gov/geo/
UniGene	► http://www.ncbi.nlm.nih.gov/UniGene/
Cap3	► http://genome.cs.mtu.edu/cap/cap3.html
ENSEMBL	► http://www.ensembl.org
GenEST	► http://genenest.molgen.mpg.de/

• EST Clustering

Because a gene can be represented by several ESTs, the next step is often an EST clustering. It consists of grouping ESTs corresponding to a single gene. Depending on the stringency of the clustering, a gene can be represented by one cluster or by as many clusters as it has transcripts. Some clusterings already exist. UniGene for example, contains GenBank sequences partitioned into a non-redundant set of gene-oriented clusters. Each UniGene cluster contains sequences representing a unique gene and related information, such as the tissue types in which the gene is expressed and the genomic location. But some prefer to do their own clustering, choosing parameters more suitable to their study. For example, ESTs can be pairwise compared using the well known program ►BLAST (1) and overlapping ESTs are grouped in the same clusters. This solution is time consuming but generally gives better results because the clustering fits the study requirements exactly.

• EST Contiguing

In some cases, a last step of EST contiguing is done. EST from a cluster are concatenated in one or several sequences, depending on the stringency of the clustering and the concatenation, representing a gene and its transcripts. “Ready-to-use” contigs were generated by some groups. In GeneNest, for example, each gene is represented by a single

cluster of ESTs and/or mRNAs. A cluster might be sub-divided into several contigs because of alternative splicing, genomic sequences or artefacts like chimeric sequences. Such gene indices are available for a few organisms, including human. Again, some prefer to do their own concatenation. Using CAP3 from the STADEN package, for example, ESTs from the same cluster are aligned and a consensus sequence representing the alignment is generated. A putative identity can be assigned to each gene by querying it against the GenBank database.

Representation and Validation

Several pieces of information can be obtained for the genes represented by these clusters and/or contigs.

- **Genomic Location**

The consensus sequence of a gene generated through this method could be the basis for mapping to the genome. Because the consensus sequences represent expressed genes, it helps to discard genes (expressed) to pseudogenes (non expressed). ENSEMBL provides genomic maps of genes for various species. If available, the EST information is shown on the maps.

- **Expression Profiles and Differential Expression**

The expression profile of a gene can be derived from the frequency of the cognate ESTs in the various libraries. This method is called Northern electronic as opposed to the classical Northern (inappropriate for large scale analysis) or to micro-arrays and is cheaper and easier to develop. However, more and more micro-array analyses are now available on the Internet. The Gene Expression Omnibus from the NCBI is a repository for gene expression and molecular abundance data and contains micro-array data. Similarly, the EBI provides access to a public repository for micro-array based gene expression data through ArrayExpress. The probability of differential expression of a gene can be derived from its expression profile. Several tests were developed, according to the initial data set and the question (2). Most biologists do not really like using statistics; fortunately several "ready-to-use" tests exist. Expression Profiler developed at the EBI, is a set of tools for clustering, analysis and visualization of gene expression and other genomic data. It even allows integration of a specific test or method.

- **Co-regulation? Co-expression?**

By comparing expression profiles of several genes, co-expressed genes can be identified. Such genes can be involved in the same metabolic pathway and/or be co-regulated. Gene correlation is a powerful tool to determine the function of unknown genes or the regulation pattern.

If a gene of unknown function is co-expressed with genes of known function, it may have the same function or be involved in the same pathway. By comparison of the promoter regions of co-expressed genes, common motifs can be identified. Such motifs are potentially involved in the regulation of genes.

The last step of data mining previously described, "Application to new data", is not represented here. However, the same protocol can be applied to any other tissue and/or condition, given that there are enough EST data.

Conclusion

Everybody can do data mining; provided that the data set is large enough. The questions we must ask are "Are the results reliable?", "Did the user take all the precautions required?". Always remember that data mining has limitations (noisy, dynamic and missing data). It is not always possible to eliminate, or at least to minimize these problems, but being aware of these limitations enables us to better interpret the results.

Clinical Relevance

EST mining is not a final analysis *en-soi*. The expression profiles generated for hundreds of genes suggest more questions. Are genes co-expressed? Are they co-regulated? Are they specific to a tissue? These questions are more relevant to the medical field. By combining the results of EST mining with *in situ* hybridisation techniques, Gitton et al. (3) have associated a list of genes with a potential biological role and suggested candidates for Down syndrome. But the biological role of these candidates has to be confirmed before any clinical use. If EST mining, and data mining in general, raise some questions, they do not ultimately answer any of them, they only give predictions that have to be experimentally validated.

References

1. Altschul SF, Gish W et al (1990) Basic local alignment search tool. *J Mol Biol* 215: 403–410
2. Claverie JM (1999) Computational methods for the identification of differential and coordinated gene expression. *Human Molecular Genetics* 8: 1821–1832
3. Gitton Y, Dahmane N et al (2002) A gene expression map of human chromosome 21 orthologues in the mouse. *Nature* 420: 586–590
4. Jiang Z, Zhang M et al (2003) Census of genes expressed in porcine embryos and reproductive tissues by mining an expressed sequence tag database based on human genes. *Biol Reprod* 69: 1177–1182
5. Jonassen I, Eidhammer I et al (2002) Structure motif discovery and mining the PDB. *Bioinformatics* 18: 362–367

6. Lawson D (1999) Data mining parasite genomes: haystack searching with a computer. *Parasitology* 118: 15–18
7. Le Bras M, Bensaad K et al (2003) Data mining the p53 pathway in the Fugu genome: evidence for strong conservation of the apoptotic pathway. *Oncogene* 22: 5082–5090
8. Megy K, Audic S et al (2002) Heart-specific genes revealed by expressed sequence tags (ESTs) sampling. *Genome Biology* 3:research 0074.1–0074.11
9. Narasimhan G, Bu C et al (2002) Mining protein sequences for motifs. *J Comput Biol* 9: 707–720
10. Pires Martins R, Leach R et al (2001) Whole-body gene expression by data mining. *Genomics* 72: 34–42
11. Pregibon D (1997) Data mining. *Statistical Computing and Graphics*, pp 7–8
12. Ramirez-Parra E, Frundt C et al (2003) A genome-wide identification of E2F-regulated genes in *Arabidopsis*. *Plant J* 33: 801–811
13. Xu D, Olman V et al (2003) EXCAVATOR: a computer program for efficiently mining gene expression data. *Nucleic Acids Res* 31: 5582–5589

DCM

► [Dilated Cardiomyopathy](#)

DCp

► [Heat Capacity Change](#)

DCP1/DCP2

Definition

DCP1/DCP2 stands for a heterodimer that removes the 5' cap from mRNAs.

► [RNA Stability](#)

DcpS

Definition

DcpS is an abbreviation that stands for a “scavenger” 3',5' exonuclease. This enzyme degrades mRNAs

containing poly (A) tails that have been shortened to less than 10 nucleotides by ► [deadenylation](#).

► [RNA Stability](#)

ddNTP

2',3'–Dideoxynucleosid–5'–Triphosphate

► [SNP Detection and Mass Spectrometry](#)

De Novo (Mutation)

Definition

De novo (mutation) refers to a newly occurring defect at the genomic, chromosomal, or gene level caused by physical, chemical, or biological agents, which is not present in the somatic cells of the parents. ► [Recurrence risk](#) is very low but above zero, because ► [germline mosaicism](#) is present in a small subset of parents.

► [Heritable Skin Disorders](#)

► [Microdeletion Syndromes](#)

► [Spinal Muscular Atrophy](#)

De Novo Pathway

Definition

De novo pathway designates a sequence of enzyme-catalyzed reactions, which builds up the complex heterocyclic structure of purine or pyrimidine nucleotides from simple precursors, for example, amino acids.

► [Nucleotide Biosynthesis](#)

Deacylation

Definition

Deacylation refers to the removal of an acyl group from a chemical compound.

► [Proteases and Inhibitors](#)

Deadenylation

Definition

Deadenylation describes the process of removing the poly (A) tail from the 3' end of mRNAs.

►RNA Stability

Deamination

Definition

Deamination is the removal of an amine group from a molecule. In DNA, deamination refers to the conversion of 5-methyl-cytosine to thymine. This spontaneously occurring process is believed to cause the loss of CpG dinucleotides from the genome.

►CpG Islands

►SNP Detection and Mass Spectrometry

Death Domain Fold Protein Superfamily

Definition

The death domain (DD), death effector domain (DED), and caspase recruitment domain (CARD) comprise of families of proteins that share a conserved fold characterized by a bundle of six helices. The proteins of these families are therefore grouped together in the death domain-fold super family. Based on computational analysis, the recently discovered PYRIN domain has also been predicted to accomplish a death domain-fold like structure. The death domain-fold represents a functional unit mediating protein-protein interactions found in many proteins with a role in signaling events that regulate apoptosis and inflammation. Typically, the subfamilies of the death domain-fold proteins undergo homophilic interactions. Thus, a DD interacts with a DD, a DED with a DED and a CARD with a CARD.

►NFκB Pathway

►TNF Receptor/Fas Signalling Pathways

Death Inducing Signaling Complex

Definition

Death Inducing Signaling Complex (DISC) describes a complex of apoptosis-inducing proteins. It was

originally used to describe the ►Fas signaling complex, but now no longer refers to a special structure within the cell.

►Apoptosis, Regulation and Clinical Implications
Degradation signals

Decision Tree

Definition

Decision tree is a simple representation to a finite number of classes. A tree is composed of nodes (object name), edges (possible values for the object) and leaves (different classes). Objects are classified by following a path down the tree, by taking the edges, corresponding to the values of an object.

►EST Mining for Expression Analysis

Defective Protein Folding Disorders

CLAUDIO SOTO

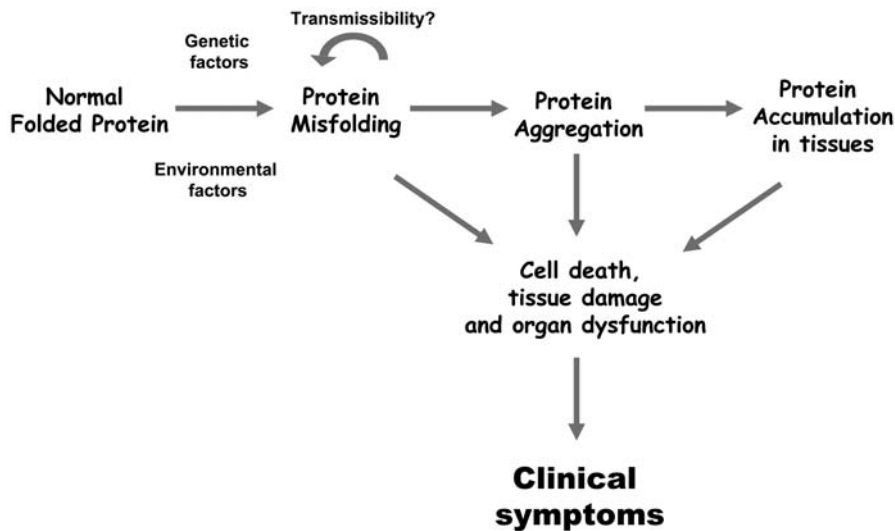
Department of Neurology, University of Texas Medical Branch, Texas, USA
clsoto@utmb.edu

Synonyms

Amyloidosis; protein conformational diseases; protein misfolding disorders; proteopathies

Definition

Defective ►protein folding disorders (DPFDs) are a group of diverse neurological and systemic diseases in which the hallmark pathological event is the misfolding, aggregation and accumulation of a protein in different organs, inducing cellular ►apoptosis, tissue damage and organ dysfunction (Fig. 1). This group includes ►Alzheimer's disease, transmissible spongiform encephalopathies, serpin-deficiency disorders, sickle cell anemia, ►Huntington's disease, diabetes type II, amyotrophic lateral sclerosis, ►Parkinson's disease, dialysis-related amyloidosis, spinocerebellar ataxias, secondary or reactive amyloidosis and more than 15 other less well-known diseases (Table 1).



Defective Protein Folding Disorders. Figure 1 A working hypothesis for the molecular mechanism of defective protein folding disorders. Misfolding, aggregation and accumulation in tissues leading to cell death and organ failure might be the common cause of several etiologically diverse human diseases.

Characteristics

DPFDs include diseases with very different clinical manifestations and symptoms. However, these disorders share some common features, namely most of them have both sporadic and inherited origins, all of them appear later in life (usually after the fourth or fifth decade) and their pathology is characterized by cellular loss and protein deposits in the dysfunctional organ. Several studies from different disciplines and of distinct diseases strongly support the hypothesis that the central event in DPFDs is a change in the secondary and/or tertiary structure of a normal protein, which is not always linked to alterations in the primary structure (Fig. 1). There is no evident sequence or structural homology among the proteins implicated in DPFDs. However, the striking feature of these proteins is their ability to fold into a stable alternative conformation, which in most cases is rich in β -pleated sheet structure and has a high tendency to aggregate and accumulate as intra- or extra-cellular fibrillar deposits known as **►amyloid**. The correlation and co-localization of protein aggregates with degenerating tissue and disease symptoms is a strong indication of the involvement of protein misfolding and aggregation in the pathogenesis of DPFDs. Moreover, protein deposits have become a typical signature of these diseases and their presence is used for definitive diagnosis. However, it is still a matter of controversy as to whether the deposits of aggregated protein are the cause of the disease or an inseparable epiphenomenon.

Mechanism of Protein Misfolding and Aggregation

Structural studies have shown a large conformational re-arrangement of the polypeptide chain during the process of misfolding and aggregation. The starting point in DPFDs is the natural protein folded in the native and active conformation, which is usually a mixture of α -helical and random structure and the end point is the same protein aggregated and adopting a β -pleated sheet conformation (Fig. 1). However, it is unknown whether the misfolding triggers protein aggregation or protein oligomerization induces the conformational changes. On the basis of the available evidence, it is likely that slight conformational changes result in the formation of a misfolded intermediate, which is unstable in an aqueous environment because of exposure of hydrophobic segments to the solvent. This unstable intermediate is stabilized by intermolecular interactions with other molecules, forming small β -sheet oligomers, which by further growth produce amyloid fibrils. In this model the conversion of the folded protein into the pathological form is triggered by structural changes, but complete misfolding depends upon protein oligomerization.

Several environmental factors have been proposed to promote protein misfolding and aggregation, including metal ions, pathological **►chaperone proteins**, pH, **►oxidative stress**, macromolecular crowding and an increase in the concentration of the misfolded protein. Many of these alterations are associated with aging, consistent with the late onset of most DPFDs. Kinetic

Defective Protein Folding Disorders. Table 1 A list of some Defective Protein Folding Disorders and the protein implicated

Diseases	Protein involved
Alzheimer's disease	Amyloid- β protein, Tau
Parkinson's disease	α -synuclein
Huntington's disease	Huntingtin
Spinocerebral ataxias	Ataxins
Transmissible spongiform encephalopathies (Creutzfeldt-Jakob disease, Gerstmann-Straussler, etc)	Prion protein
Amyotrophic lateral sclerosis	Superoxide dismutase
Type II diabetes	Islet amyloid polypeptide
Primary amyloidosis (implicated in multiple myeloma, β -cell dyscrasias)	Immunoglobulin light chain
Secondary or reactive amyloidosis (implicated in familial Mediterranean fever, rheumatoid arthritis, etc)	Amyloid-A
Hemodialysis-related amyloidosis	β 2-microglobulin
Senile systemic amyloidosis, familial amyloid polyneuropathy	Transthyretin
Hereditary cerebral hemorrhage with amyloidosis Icelandic-type	Cystatin C
Familial amyloidosis, Finnish-type	Gelsolin
Sickle cell anemia, inclusion-body hemolysis	Hemoglobin
Familial amyloid polyneuropathy	Apolipoprotein A-I
Amyloidosis in senescence	Apolipoprotein A-II
Hereditary non-neuropathic systemic amyloidosis, familial visceral amyloidosis	Lysozyme
Serpin deficiency disorders (cirrhosis, thromboembolic disease, angioedema)	Serpins (α 1-antitrypsin, α 1-antichymotrypsin, etc)

studies have shown that protein aggregation follows a seeding/nucleation mechanism, which resembles a crystallization process. The critical event is the formation of protein oligomers that act as a nucleus to direct further growth of aggregates. Nucleation-dependent polymerization is characterized by a slow lag phase in which a series of unfavorable interactions forms an oligomeric nucleus, which then rapidly grows to form larger polymers. The lag phase can be minimized or removed by addition of pre-formed nuclei or seeds.

Mechanism of Tissue Damage by Misfolded Proteins

The conformational change may promote the disease by either the toxic activity of the misfolded protein or

by the lack of the biological function of the natively folded protein, which is depleted during misfolding and aggregation. Some of the proteins associated with DPFs have known biological functions. However, knock out animals lacking these proteins have in general failed to show a phenotype similar to the disease. The most widely accepted theory of tissue degeneration in DPFs proposes that misfolding and aggregation result in the acquisition of a toxic function by the misfolded protein. This concept is based on direct induction of cellular apoptosis *in vitro* by aggregates of several misfolded proteins (or fragments of the proteins). Additional support for this hypothesis comes from experiments with ►transgenic animals, where incorporation of the human mutated gene for the

misfolded protein has been shown to be enough to trigger protein accumulation and tissue degeneration. There is also evidence that β -sheet oligomerization of non-disease-related proteins result in highly cytotoxic structures, indicating that misfolding and aggregation of any protein may lead to inherent toxicity. This concept provides a unifying mechanism of cell death and tissue degeneration in protein misfolding disorders. Several mechanisms have been proposed for the neurotoxic activity of misfolded aggregates, and it is likely that different pathways operate, depending on whether the proteins accumulate intra- or extracellularly. Extracellular aggregates might activate a signal transduction pathway leading to apoptosis by interacting with specific cellular receptors. Intracellular aggregates might damage cells by recruiting factors essential for cell viability into the fibrillar aggregates. Another proposed mechanism is membrane disruption and depolarization mediated by incorporation of the misfolded aggregates into the membrane inducing the formation of ion channels, resulting in alteration of ion homeostasis and deregulation of cellular signal transduction leading to cell death. Finally, protein aggregates could induce oxidative stress by producing free radical species, resulting in protein and lipid oxidation, elevation of intracellular calcium and mitochondrial dysfunction.

Genetics

Most DPDFs have both inherited and sporadic origins. The percentage of hereditary cases varies depending on the particular disease. Interestingly, mutations in the genes encoding the protein component of fibrillar aggregates are genetically associated with inherited forms of the disease. The familial forms usually have an earlier onset and higher severity than sporadic cases. In the familial cases, a mutation may destabilize the normal protein folding, favoring the misfolding and aggregation of the protein. Mutations in the respective fibrillar proteins have been associated with familial forms of many diseases, including Alzheimer's, Parkinson's, Huntington disease and related polyglutamine disorders, amyloid polyneuropathy, cardiac amyloidosis, visceral amyloidosis, cerebral hemorrhage with amyloidosis of the Dutch and Icelandic types, cerebral amyloidosis of the British and Danish types, thromboembolic disease, angioedema, emphysema, sickle cell anemia, amyotrophic lateral sclerosis, etc.

► **Transmissible spongiform encephalopathies** (TSEs) are unique among DPDFs in that, in addition to having genetic and sporadic origins, they can also be transmitted by infection. TSEs, also known as prion disorders, include Creutzfeldt-Jakob disease, fatal familial insomnia, Gerstmann-Straussler syndrome in humans and bovine spongiform encephalopathy,

chronic wasting disease and scrapie in animals. The TSE infectious agent, known as a prion, is composed mainly (or exclusively) of misfolded prion protein, which is called PrP^{Sc} to differentiate it from the normally folded protein, termed PrP^{C} . The fact that the full pathological and clinical characteristics of prion diseases can be propagated from individual to individual by serial passage of pure PrP^{Sc} is a good indication that the misfolding of the prion protein is the most likely cause of the disease. Transmission depends on the conversion of host PrP^{C} into the misfolded protein, which is induced by infectious PrP^{Sc} . None of the other protein conformational disorders has been convincingly shown to be transmissible. However, in animal models of Alzheimer's disease and systemic conformational disorders associated with deposition of amyloid-A in spleen and apolipoprotein A-II in several organs, the disease pathology can be accelerated by injection of tissue homogenate enriched in the misfolded protein. Although not associated with a disease, other examples of propagation of phenotypic changes by transmission of protein misfolding have been found in yeast and other fungi. In these organisms different proteins can adopt alternative conformations with different activities, and the misfolded protein aggregates can convert the monomeric folded protein *in vitro* and *in vivo*.

Molecular Diagnostics and Therapy

At present there is no accurate molecular diagnosis for DPDFs. The diagnosis is done mostly by observation of the clinical symptoms and by exclusion of other similar pathologies. The accuracy of clinical diagnosis is variable depending on the specific disease of the group. Since many of these diseases have an inherited origin, genetic testing for mutations associated to proteins involved in misfolding is very useful for diagnosis. However, in most DPDFs only a fraction (sometimes less than 1%) of the cases are inherited. Therefore, genetic testing cannot be used as the only diagnostic tool in these diseases. In many cases the definitive diagnosis is established by post-mortem examination and detection of misfolded aggregates accumulated in the tissue. In the diseases where the protein is deposited in organs in which biopsy is possible, this helps to establish a definitive diagnosis. No imaging technology or biochemical markers have been validated for diagnosis of diseases associated with protein misfolding and aggregation. However, recent efforts have resulted in the identification of small chemical compounds, amyloid binding proteins and peptides that can bind specifically to the aggregates. Labeling of these molecules with a reagent that allows imaging may lead to the design of non-invasive diagnostic tools.

Despite dramatic progress in understanding the contribution of defective protein folding to the pathogenesis of DFPDs, none of these diseases has, as yet, a treatment aimed at inhibiting and correcting protein misfolding. Many of the illnesses of the group have no efficient treatment, some have a therapy directed to alleviation of the clinical symptoms, but there is certainly an important need for treatments focused on altering the cause of the disease. At least four approaches have been proposed to attack protein misfolding and aggregation, stabilization of the native protein conformation, inhibition and reversal of defective protein folding, competitive inhibition of protein oligomerization and increased clearance of the misfolded protein. Recent exciting research has shown *in vitro* and in animal models the validity of these strategies and some compounds are already in clinical trials to evaluate their efficacy in humans. The results of these tests might provide not only novel and more efficient therapies, but might also definitively prove the involvement of protein misfolding and aggregation in the pathogenesis of a variety of diseases.

References

1. Soto C (2003) Unfolding the role of Protein Misfolding in Neurodegenerative Diseases. *Nat Rev Neurosci* 4:49–60
2. Lomas D, Carrell RW (2002) Serpinopathies and the Conformational Dementias. *Nat Rev Genetics* 3:759–768
3. Dobson CM (1999) Protein Misfolding, Evolution and Disease. *Trends Biochem Sci* 24:329–332
4. Kelly JF (1998) The alternative conformations of amyloidogenic proteins and their multi-step assembly pathways. *Curr Opin Struct Biol* 8:101–106
5. Thompson AJ, Barrow CJ (2002) Protein conformational misfolding and amyloid formation: characteristics of a new class of disorders that includes Alzheimer's and prion diseases. *Curr Med Chem* 9:1751–1762

Degradation Signals

Definition

In early experiments, the Varshawsky group searched for special destruction signals for the ubiquitin system. One of the most prominent signals was deg-1, which represents the N-terminal part of the yeast Mat α 2-repressor. The fusion protein between deg-1 and other proteins is rapidly degraded by the ubiquitin-proteasome system. Two other destruction signals are the Destruction-box (D-box) and the KEN-box. Both are located in the N-terminus of cyclins and CDC20, which are substrates for the APC.

► Ubiquitination

Dejerine-Sottas Neuropathy

► Hereditary Neuropathies, Motor and/or Sensory

Deletion

Definition

Deletion is a particular kind of mutation characterized by a loss of a piece of DNA from a chromosome. Deletion of a gene or part of a gene can lead to a disease or abnormality.

► COPD and Asthma Genetics

Deletion Loop

► Insertion/Deletion Loop

Delta I-Like (DII) 1,3,4

Definition

Delta I-like (DII) 1,3,4 are the three orthologues of the *Drosophila* Delta gene in the mouse.

► Notch Pathway

Dementia Praecox

► Schizophrenia, Genetics

Demyelination

Definition

Demyelination is a degenerative defect with myelin as the primary target. Demyelination can be an acquired (e.g. multiple sclerosis) or a genetic disease (e.g. storage diseases).

► Glial Cells and Myelination

Denaturing High-Performance Liquid Chromatography

Definition

Denaturing high performance liquid chromatography (DHPLC) is a liquid chromatographic method that is very sensitive for the detection of DNA sequence variation. The selected temperature and buffer gradient conditions can resolve mutations in various genes.

- ▶ Hereditary Neuropathies, Motor and/or Sensor
- ▶ Spinal Muscular Atrophy

Dendrites

Definition

Dendrites are long, branching postsynaptic extensions from the cell body (soma) of a neuron that conduct electrical impulses received from other nerve cells from the synapses to the cell body. They are regarded as the primary input station of a neurone, where input summation and synaptic integration occurs. In some neuronal cell types, dendrites may also sustain Na^+ and Ca^{2+} action potentials.

- ▶ Cell Polarity
- ▶ Fragile X Syndrome
- ▶ Neurons

Dendritic Extensions

Definition

Dendritic extensions are branching postsynaptic extensions of the nerve cell used to conduct electric impulses received from other nerve cells.

- ▶ Cell Polarity

Dendritic Protein Synthesis

Definition

Dendritic protein synthesis describes protein synthesis by translation of messenger RNA, at ribosomes located

near dendritic synapses distant from the neuronal cell body, where most protein synthesis takes place.

- ▶ Fragile X Syndrome

Dentatorubral Pallidoluysian Atrophy

Definition

Dentatorubral pallidoluysian atrophy (DRPLA) is a rare neurodegenerative disease, primarily reported to occur in Japan, with a spectrum of multiple system degenerations resembling ▶ Huntington's disease and spinocerebellar atrophy.

- ▶ Repeat Expansion Diseases
- ▶ Polyglutamine Disease, the Emerging Role of Transcription Interference

Deoxyhexose

Definition

Deoxyhexose is a hexose monosaccharide in which an alcoholic hydroxyl group is replaced by a hydrogen atom.

- ▶ Glycosylation of Proteins

Deoxyribophosphate Lyase

Definition

A deoxyribophosphate lyase is an enzyme that is able to cleave the sugar phosphate backbone of deoxyribose phosphate at the 5'-prime side of an abasic site. The enzyme uses a β -elimination reaction.

- ▶ DNA Polymerases

Dependence

Definition

Dependence describes the physical and/or psychological reliance on drugs. Psychological dependence and physical dependence have to be strictly separated. With physical dependence, the body has adapted to the presence of the drug and withdrawal symptoms may occur if use is reduced or stopped.

- ▶ Addiction, Molecular Biology

Depolarization

Definition

Depolarization describes the change of the membrane potential towards positive potentials and it constitutes the initiation phase of the action potential.

- ▶ Ion Channels/Excitable Membranes
- ▶ Neurons

Depurination

Definition

Depurination defines loss of purines in nucleic acids due to energy absorption.

- ▶ SNP Detection and Mass Spectrometry

Dermal Papilla

Definition

Dermal papilla describes a ball-shaped group of cells of mesodermal origin which is located in the center of the hair bulb, at the base of the hair follicle. The dermal papilla cells are the source of inductive signals required for growth and differentiation of keratinocytes of the hair matrix and the hair bulge.

- ▶ Skin and Hair

Dermis

Definition

Dermis is the part of the skin that is below the outermost layer, the epidermis. The dermis assumes the important functions of thermoregulation and supports the vascular network to supply the avascular epidermis with nutrients. All skin appendages, e.g. hair follicles, sweat glands, and sebaceous glands, are located in the dermis. The dermis contains mostly fibroblasts, which are responsible for the synthesis of connective tissue constituents that give the support and elasticity of the skin. Also present are immune cells that are involved in

the defence against foreign invaders passing through the epidermis.

- ▶ Wound Healing

Dermomyotome

Definition

Dermomyotome is an epithelial cell layer constituting of the dorsal part of the somite lying under the ectoderm. As its name implies, it will give rise to dorsal dermis and to the skeletal muscle of the myotome, as well as the precursor cells of other skeletal muscles.

- ▶ Somitogenesis

Des Species

Definition

Des species refers to a disulfide-bonded intermediate of a protein with all but one of its native disulfide bonds.

- ▶ Protein Disulfide Bonds

Desferrioxamine

Definition

Desferrioxamine is a small molecule produced by bacteria to chelate iron. Desferrioxamine has an extraordinarily high affinity for the metal. It is used medically to remove unwanted iron from the body.

- ▶ Hemochromatosis

Desmal Ossification

Definition

Desmal ossification describes the direct formation of bones through differentiation of mesenchymal progenitor cells in osteoblasts and subsequent production of bone matrix.

- ▶ Bone Disease and Skeletal Disorders, Genetics

Desmosomes

AMANDA E. BASS, KATHLEEN J. GREEN
Departments of Pathology, Dermatology, and the
Robert H. Lurie Cancer Center, Northwestern
University Feinberg School of Medicine, Chicago,
Illinois, USA
bass@northwestern.edu,
kgreen@northwestern.edu

Definition

The desmosome is a ►cytoskeleton-linked attachment junction connecting cells to each other in epithelia and some muscle cell tissues. Desmosomes form structural connections between keratin networks from one cell to another. Keratins bind to desmoplakin proteins in the cytoplasmic plaques of these junctions and special cadherins, the desmogleins and desmocollins, span the membrane and interact across the intercellular space with desmosomal cadherins of the neighbouring cell, forming electron-dense threads and the midline overlap of the junction. Other components provide stability and regulation of the junctions.

Characteristics

Desmosomes, also known as maculae adherens, are specialized cellular structures required for strong ►cell-cell adhesion. One of several types of ►intercellular junctions, desmosomes mediate strong adhesion by anchoring the ►intermediate filament (IF) cytoskeleton to molecular plaques, or “spot welds” between adjacent cells. Desmosomes are found in several tissues including the epidermis and within the myocardium. Due to the fact that these tissues are subject to frequent mechanical stress, strong intercellular adhesion is necessary to

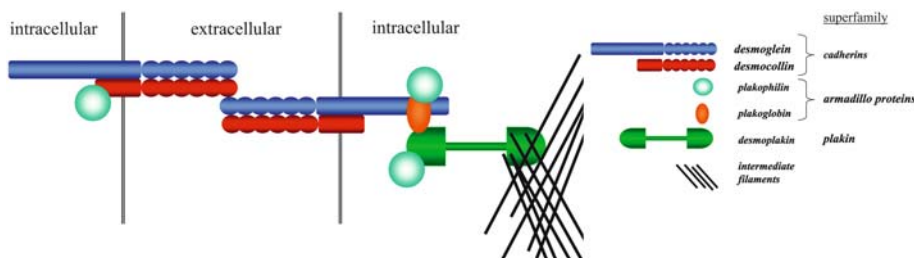
maintain their integrity (1). Desmosomes have also been described in non-stress bearing tissues such as the arachnoid meningeal layer and dendritic cells in lymphoid follicles (2, 3).

The desmosomal plaque is composed of three major protein families, ►cadherins, ►Armadillo proteins, and plakins (2). The desmosome is similar to the ►adherens junction in that it forms cadherin mediated cell-cell attachments by anchoring cytoskeletal filaments into a juxtamembrane plaque. Whereas adherens junctions contain classical E-cadherin and are associated with the ►actin cytoskeleton, desmosomes have a more heterogeneous collection of desmosomal cadherins and are coupled to the IF cytoskeleton through different adaptor proteins (Fig. 1). While the plaque components differ between the adherens junction and the desmosome, several desmosomal plaque proteins have been observed to associate with adherens junction components. During processes such as wound healing in which migrating cells are brought into contact, adherens junctions form earlier in the process than desmosomes and the formation of adherens junctions is thought to be required for the ensuing formation of desmosomes (2, 4).

Recently, additional roles for desmosomal adhesion have been elucidated. Desmosome components are important in inhibiting the motility and invasive activity of tumor cells, regulating tissue development and morphogenesis and wound healing and have been implicated in non-adhesive functions including modulation of intracellular signaling pathways, transcriptional gene activation, cellular response to growth factor stimulation and ►apoptosis.

Ultrastructural Appearance

In the electron microscope, the desmosome appears as an electron-dense structure at cell-cell appositions.



Desmosomes. Figure 1 Desmosome and Adherens Junction Schematic Diagram. This simplified diagram demonstrates cadherin heterodimer-mediated adhesion between adjacent cells. The desmosomal cadherins desmoglein and desmocollin span the plasma membrane and interact with other desmosomal cadherins through conserved residues in the amino-terminal domains in the extracellular space. Armadillo proteins plakoglobin and plakophilin interact with the cytoplasmic domains of the cadherins and mediate the attachment of intermediate filament-binding plakin protein desmoplakin to the cadherin complex.

This structure has been divided ultrastructurally into three domains, the extracellular domain, the outer dense plaque and the inner dense plaque. Flanked by the adjacent plasma membranes is a 30 nm-wide extracellular domain with a dense midline resembling a zipper. This extracellular domain is composed of the ectodomains of the desmosomal cadherins desmoglein and desmocollin. Immediately adjacent to the plasma membrane in the cytoplasm one finds the outer dense plaque (ODP). The outer dense plaque is a 15–20 nm-wide electron-dense band that includes the intracellular domain of the cadherins, the Armadillo proteins plakoglobin and plakophilin, and the amino-terminal domain of desmoplakin. Intracellular to the ODP is a region called the inner dense plaque, which is less electron-dense than the ODP and contains the desmoplakins. IFs can be observed to insert directly into the IDP (Fig. 2) (1, 5).

By immunofluorescence light microscopy, desmosomes appear at the cell-cell borders in epithelial colonies in cell culture. Whereas adherens junctions stained by fluorophore-conjugated antibodies have a more continuous, linear appearance around the cell-cell borders, desmosomes appear in a discontinuous “railroad track” pattern. Viewed *en-face*, desmosomes display a punctate, “spot-weld”-like appearance. When the IF cytoskeleton is counterstained, the filaments appear to project almost to the plasma membrane where they co-localize with the desmosomal components. While desmosomes are present at the lateral borders of cells in a monolayer or in simple epithelial tissues, desmosomes in a

▶**stratified epithelium** are found on all cell-cell interfaces except at the basement membrane.

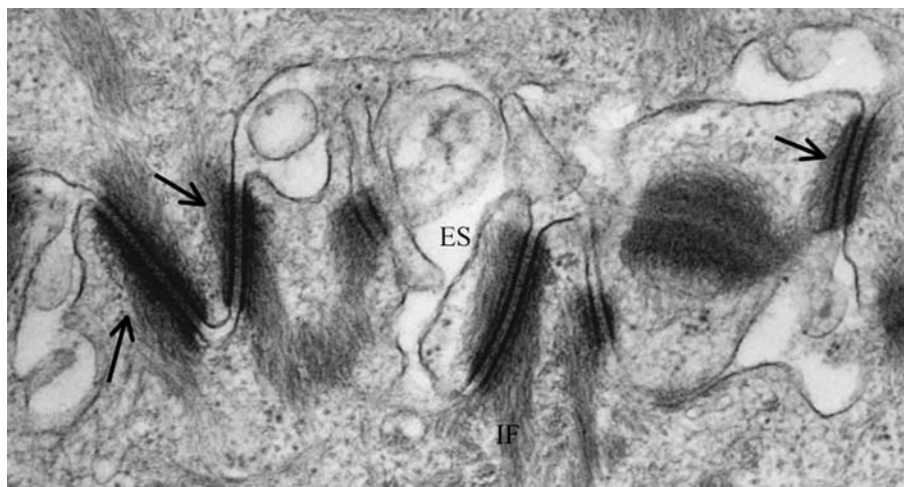
Physiological Significance

The importance of desmosomes in maintaining tissue integrity is demonstrated by the phenotypes of the ▶**bullous** skin diseases. ▶**Autoimmune** diseases ▶**pemphigus** vulgaris and pemphigus foliaceus are two such skin diseases in which desmosome components are the specific targets of autoantibodies. Additionally, bacterial toxins can target desmosome components in the skin and achieve similar skin pathologies. Certain genetic mutations in desmosome components result in congenital diseases involving multiple organ systems including the skin and its appendages and the cardiovascular system. Mouse models of these genetic diseases display severe skin defects or even lethality during embryogenesis. Finally, loss of desmosomal intercellular adhesion occurs in some types of tumor cells, conferring to these transformed cells a more metastatic tumor profile.

Molecular Interactions

Desmosomal Cadherins

Desmosomal cadherins are calcium dependent ▶**type I transmembrane glycoproteins** that mediate direct intercellular attachment most likely through heterotypic cadherin-cadherin interactions. There are two subfamilies of desmosomal cadherins, the desmogleins (Dsgs), and the desmocollins (Dscs).



Desmosomes. Figure 2 Electron Micrograph Image of Bovine Tongue Epithelium. Desmosomes appear as electron-dense “spot weld” structures between adjacent epithelial cells (arrows). Regions of the plasma membrane that contain desmosomes are closely apposed whereas extracellular spaces are observed in regions without desmosomes. Intermediate filaments project from the cytoplasm into the plaque, where they act to anchor the junction and strengthen intercellular adherence. ES, extracellular space; IF, intermediate filaments.

There are four Dsg isoforms in humans, Dsg 1–4, encoded by separate genes on chromosome 18. Dsgs contain four conserved extracellular ectodomains important in calcium binding and cadherin hetero- and homophilic interactions. Flanking the membrane-spanning region are two more variable anchor domains, the extracellular and intracellular anchors. The cytoplasmic region of the Dsgs contains an intracellular cadherin-type segment (ICS), which is highly conserved among the cadherins, and a proline rich linker (IPL) region. At the carboxy-terminal, the Dsgs have a long intracellular tail encoding a repeating unit domain (RUD), and a Dsg terminal domain (DTD). The three Dsc isoforms are similar in structure to the Dsgs except for a truncated intracellular region consisting only of the intracellular cadherin segment followed by eleven amino acids specific to one of the Dsc isoforms (2, 3, 6). In stratified epithelia such as the skin, [▶keratinocytes](#) originate at the basal layer and move superficially as they become terminally differentiated. The different Dsg and Dsc isoforms are expressed at distinct stages of epithelial differentiation. The Dsg 1/Dsc 1 pair is concentrated in the superficial layers of the epidermis whereas the Dsg 3/Dsc 3 pair is expressed abundantly in the less differentiated basal and suprabasal layers. Dsg 2 and Dsc 2 are expressed basally with Dsc 2 tapering off more superficially than Dsg 2 (2). Variations in this differential expression pattern are associated with different types of complex epithelia. For example, the ratio of Dsg 3 to Dsg 1 is high in the basal layers of skin epidermis, but it drops off sharply in the superficial layers where Dsc 1 is concentrated. On the other hand, the Dsg 3/Dsg 1 ratio is high throughout all layers of oral mucous membrane, where Dsg 3 is expressed more uniformly compared with epidermis (7). Furthermore, desmosomal cadherin differential expression is considered to be a possible mechanism for the regulation of epithelial development and differentiation.

Cadherin Related Diseases

Desmosomal cadherins are the targets of both autoimmune and bacterial pathogenesis. Autoimmune blistering diseases pemphigus vulgaris (PV) and pemphigus foliaceus (PF) are caused by autoantibody production against Dsg proteins in the epidermis. Inactivation of desmosomal cadherin interactions then leads to [▶acantholysis](#), which is the formation of intercellular spaces within the epidermis between keratinocytes. This loss of adhesion leads to weakened skin integrity and the development of large blisters develop upon mechanical insult.

In mucosal PV, production of autoimmune antibodies against Dsg 3 leads to disruption of desmosomal junctions between cells in the deep basal layers of the

epidermis resulting in erosions in the oral mucosa. In this form of PV, the skin is less affected by anti-Dsg 3 antibodies than the mucosa because Dsg 1 expression compensates for the loss of Dsg 3. However, in the mucocutaneous form of PV, autoantibodies against Dsg 3 and Dsg 1 result in oral and epidermal blistering. In PF, autoantibodies target Dsg 1, leading to less severe blistering due to the more superficial expression of Dsg 1 relative to Dsg 3. PF-like blisters can be observed in *Staphylococcus aureus* infections such as [▶staphylococcal scalded skin syndrome](#) (SSSS) and [▶bullous impetigo](#). Rather than an autoantibody attack on Dsgs, the secreted bacterial exfoliative toxin targets and cleaves Dsg 1 resulting in loss of adhesion in the epidermis (7).

Mutations in the *DSG* genes manifest in an array of dermatologic and cardiac diseases. Loss of Dsg 1 expression results in a skin disorder called striate palmoplantar keratoderma (SPPK), in which the patient's palms and soles exhibit [▶hyperkeratosis](#) or epidermal thickening in a striped or striated pattern and superficial epidermal blisters. Mutations in Dsg 4 result in defective hair follicle differentiation. Similar epidermal fragility phenotypes were observed in *DSG/DSC* knockout mouse models (3).

Armadillo Proteins

The presence of desmosomal cadherins alone is not sufficient to form desmosomal cell-cell adhesion. Adaptor proteins, which mediate the attachment of the IF cytoskeleton to the intracellular plaque are required for the formation of mature desmosomes. One group of such adaptor proteins comprises Armadillo family members that directly interact with the cadherins. Both desmosomes and adherens junctions contain Armadillo protein family members that link IF- or actin-binding proteins to the cadherin cytoplasmic tail. Armadillo, a *Drosophila* protein, activates gene transcription when it accumulates in the cytoplasm in response to [▶Wnt/Wg](#) signaling and is important for determining segment polarity during *Drosophila* morphogenesis. Armadillo related proteins contain three major domains, an amino-terminal head domain followed by a central domain consisting of arm repeats and a carboxy-terminal tail. The central arm domain is highly conserved between the different family members and comprises twelve imperfect 42-amino acid repeats. This region is important in mediating protein-protein interactions and is responsible for binding to cadherins and other proteins. In the desmosome, plakoglobin (γ -catenin) and the plakophilins are the major Armadillo proteins whereas [▶ \$\beta\$ -catenin](#) and p120-catenin are the major Armadillo proteins present in the adherens junction.

Plakoglobin

In the desmosome, the Armadillo family member plakoglobin acts as a linker between the cadherin cytoplasmic tail and the IF-binding protein desmoplakin. While it is required for keratinocyte cell-cell adhesion and desmosome formation, plakoglobin, like β -catenin can be found in the nucleus and is implicated in the modulation of \blacktriangleright TCF/Lef transcriptional activation and in regulating desmosome size. Plakoglobin binds directly to desmosomal or classical cadherins, IF- or actin-binding proteins and plakophilins (see below). In contrast to β -catenin, whose junctional localization is limited to the adherens junction, plakoglobin can associate with both desmosomes and adherens junctions. These junctional interactions are mutually exclusive, meaning that plakoglobin bound to desmosomal components cannot also bind to adherens junction components, and *vice versa* (2). The ability of plakoglobin to associate with both types of adhesion junctions suggests a possible role in mediating cross talk between junctions.

The importance of plakoglobin in maintaining the desmosomal plaque is underscored by the diseases in which plakoglobin is not present. Plakoglobin truncation mutations in humans result in Naxos disease, which is a congenital syndrome consisting of woolly hair, SPPK and arrhythmogenic right ventricular cardiomyopathy. Some plakoglobin-null mice die in the embryonic stage due to heart failure, and some die later due to epidermal defects (3).

Outside of its desmosomal role, plakoglobin has been implicated in a number of signaling pathways in certain cell types. In addition to interacting with LEF/Tcf signaling molecules, plakoglobin may regulate intracellular signaling by binding to and acting as a substrate for tyrosine kinases like epidermal growth factor receptor (EGFR) and Src and some phosphatases (2). Plakoglobin may also play a role in neoplastic transformation and tumor progression. This idea is supported by its ability to induce proliferation when over-expressed in tumor cells, and to stimulate the anti-apoptotic pathway by induction of Bcl-2 (8). However, plakoglobin can also act as a tumor suppressor by suppressing hyperproliferation in other cell types. Moreover, loss of plakoglobin expression has been exhibited in some types of ovarian, lung and breast cancers.

Plakophilin

The desmosomal adaptor protein plakophilin is another Armadillo family member that plays structural and possible regulatory roles in the desmosome and in mediating transcriptional activation within the nucleus. Plakophilin is closely related to the Armadillo protein p120catenin, which is present in the adherens junction and in the nucleus and plays similar structural and

possibly signaling roles. Plakophilins have been suggested to mediate lateral junctional interactions, possibly by stabilizing junctions and increasing adhesive strength. There are three plakophilin isoforms, plakophilins 1–3, which are expressed in a differential tissue distribution. Junctional plakophilin 1 is expressed suprabasally in stratified epidermis. Plakophilin 2 is the most widely expressed plakophilin, found in all layers of epithelia, myocardium and in lymph nodes. Plakophilin 3 is found in most simple and stratified epithelia. Only one human plakophilin-related disease has been described to date. Congenitally inherited loss of plakophilin 1 expression leads to ectodermal dysplasia or skin fragility syndrome in which patients exhibit cutaneous fragility, and epidermal thickening of the palms of the hands and soles of the feet (2, 3).

The role of plakophilin is not as well characterized as that of other Armadillo proteins; however, there is biochemical evidence demonstrating that the plakophilins bind to the desmosomal cadherins, keratin (an IF expressed in keratinocytes and other epithelial cells), plakoglobin, and the adherens junction protein β -catenin. The plakophilins also bind directly to desmoplakin and serve in its recruitment to cell-cell borders. It is thought that plakophilin and p120catenin are also important in regulating the actin cytoskeleton because over-expression of p120catenin and some plakophilins has been shown to induce the formation of filopodial-like cytoplasmic protrusions in several different cell types. Other signaling functions of the plakophilins are demonstrated by their ability regulate cytoplasmic β -catenin signaling and by the localization of plakophilin 2 to the nucleus where it binds to the RNA polymerase III complex (2, 3, 6). However, plakophilin's specific role in these signaling pathways has yet to be elucidated.

Desmoplakin

Desmoplakin is the major mediator of IF attachment to desmosomal cadherins and is essential for desmosome formation. Desmoplakin is a member of the plakin family, which also includes, but is not limited to, plectin, envoplakin and periplakin. Desmoplakin mRNA is alternatively spliced to yield two isoforms, desmoplakin I and desmoplakin II. Desmoplakin is a large dumbbell-shaped molecule with globular amino- and carboxy-terminal domains and a coiled coil central rod domain. In desmosomes, desmoplakin interacts with cadherins and Armadillo proteins through its amino-terminal domain and with the IF cytoskeleton through its carboxy-terminal domain (3).

Haploinsufficiency or loss of desmoplakin results in SPPK. Patients with SPPK exhibit striations on the palms and soles, woolly hair and can also develop arrhythmogenic right ventricular cardiomyopathy

(ARVC). Desmoplakin is critical for the maintenance of desmosome strength through its attachments to the IF cytoskeleton (2, 3). Desmoplakin-null mouse models die very early in embryogenesis, however chimeric desmoplakin knockouts in which only extra-embryonic tissues express desmoplakin die later in embryogenesis due to abnormalities in the vascular, epidermal-, cardiac- and neuroepithelial-tissue precursors. Conditional desmoplakin-null mice in which desmoplakin is knocked out in the epidermis display abnormal desmosomes and severe skin fragility (3).

Regulatory Mechanisms

Regulation of desmosome assembly and disassembly is the focus of current studies. It is known that desmosomal cadherin interactions are calcium sensitive. Early, immature desmosomal junctions disengage in the presence of low calcium and reform when calcium levels are restored to higher levels. However, when desmosomes mature they become less susceptible to disengagement due to alterations in calcium concentrations. It is thought that desmosomes are formed by the movement of desmosomal precursors containing some of the desmosomal components to the plasma membrane (2, 4).

Regulation of cadherin turnover and desmosome stability might also occur through phosphorylation of desmosomal components. Growth factor receptor signaling pathways lead to the phosphorylation of Dsgs and plakoglobin, resulting in disruption of the interactions between desmosomal components and the eventual dissolution of the junction. Evidence from some tumor systems reveals the over-expression or over-activation of Epidermal Growth Factor Receptor (EGFR) and a more motile, non-adherent tumor cell phenotype. Activation of such intracellular signaling pathways results in negative regulation of cell-cell adhesion (2, 6).

Once considered to be static complexes acting solely as anchors for cell adhesion, desmosomes and their components are becoming understood to display complex and dynamic behaviors as they relate to cell adhesion, signaling, growth, shape, motility and development.

Acknowledgements

We would like to thank Spiro Getsios for critical review of this essay and the members of the Green laboratory for valuable discussions. This work was supported by NIH grants R01AR41836, R01AR43380, and project #4 of P01 DE12328.

►Intermediate Filaments

References

1. Garrod DR, Merritt AJ, Nie Z (2002) Desmosomal adhesion: structural basis, molecular mechanism and regulation. *Mol Membr Biol* 19:81–94
2. Getsios S, Huen AC, Green KJ (2004) Working out the strength and flexibility of Desmosomes. *Nat Rev Mol Cell Biol* 5:1–11
3. Godsel LM, Getsios S, Huen AC et al (2004) The Molecular Composition and Function of Desmosomes. In: Behrens J, Nelson WJ (eds) *Handbook of Experimental Pharmacology* 165:137–193
4. Garrod D, Chidgey M, North A (1996) Desmosomes: differentiation, development, dynamics and disease. *Curr Opin Cell Biol* 8:670–678
5. North AJ, Bardsley WG, Hyam J et al (1999) Molecular map of the desmosomal plaque. *J Cell Sci* 112:4325–4336
6. Huber O (2003) Structure and function of desmosomal proteins and their role in development and disease. *Cell Mol Life Sci* 60:1872–1890
7. Amagai M (2003) Desmoglein as a target in autoimmunity and infection. *J Am Acad Dermatol* 48:244–252
8. Van Aken E, De Wever O, Correia da Rocha AS et al (2001) Defective E-cadherin/catenin complexes in human cancer. *Virchows Arch* 439:725–751

Desorption

Definition

Desorption refers to the uptake of laser energy by matrix molecules in MALDI-TOF-mass spectrometry; desorption leads to sudden phase transition.

►MALDI

►Mass Spectrometry

►SNP Detection and Mass Spectrometry

Destruction Complex

Definition

Destruction complex refers to a complex of proteins that are required for degradation of β -catenin which includes the core components ►APC, ►Axin, and GSK3.

►Wnt/Beta-Catenin Signaling Pathway

Detergent

Definition

Detergent is an amphiphilic molecule with limited water solubility as a monomer, but highly soluble in the micellar form.

►Two-dimensional Crystalization of Membrane Proteins

Developmental Timing

Definition

Developmental timing describes the global orchestration of how developmental events or stages temporally unfold, specified in strict sequence by a gene regulatory network.

► *C. Elegans* as a Model Organism for Functional Genomics

Devic's Disease

Definition

Devic's disease is a demyelinating disease, which is restricted to the optic nerves and spinal cord; most usually seen in Japanese and Africans.

► Genetic Predisposition to Multiple Sclerosis

DG

► Free Energy Change (DG)

DG/VCFS

► DiGeorge/Velocardiofacial Syndrome

DH

► Enthalpy/Enthalpy Change (DH)

DHF

Definition

Dihydrofolate, the oxidized form of THF (Tetrahydrofolate), is a metabolite in the Folate and B12 Metabolism.

► Nucleotide Biosynthesis

DHPLC

► Denaturing High-Performance Liquid Chromatography

Di- and Trihydroxycholestanoic Acid

Definition

Di- and trihydroxycholestanoic acid are intermediates in the formation of the primary bile-acids ► **chenodeoxycholic acid** and ► **cholic acid**, respectively, from cholesterol in the liver.

► Peroxisomal Disorders

Diabetes

► Diabetes Mellitus, Genetics

► *C. Elegans* as a Model Organism for Functional Genomics

Diabetes Insipidus

Definition

Diabetes insipidus refers to the disease caused by vasopressin (► **ADH**) deficiency (termed-central at level of the hypothalamus or pituitary), or dysfunction (termed-nephrogenic at the level of the kidney), which results in excessive urination (polyuria) with secondary excessive drinking (polydipsia).

► Hypothalamic and Pituitary Diseases Genetics

Diabetes Insipidus, a Water Homeostasis Disease

MARIEL BIRNBAUMER

Receptor Regulation Section, Laboratory of Signal Transduction, National Institute of Environmental Health Sciences, Research Triangle Park, N.C., USA
birnbau2@niehs.nih.gov

Definition

Diabetes insipidus (DI) is a water ►homeostasis disorder characterized by the partial or total inability of the patient kidneys to produce concentrated urine. The most dramatic symptom of this condition is the passage of large volumes of dilute urine. The urine of these patients contains low concentrations of salts and metabolic products and as a consequence it is less colored. The English rendition of the term diabetes insipidus is “tasteless ►polyuria” as opposed to diabetes mellitus (honey flavored polyuria). The names emphasize the distinction between the polyuria associated with poor water reabsorption and the one caused by urinary excretion of large quantities of glucose. The impairment in water recovery characteristic of DI increases the sodium concentration in blood (hypernatremia), triggering a physiological increase in thirst. If water intake is restricted, hypernatremia can damage the central nervous system leading to mental retardation and death. Individuals suffering any form of DI are very susceptible to severe dehydration.

Characteristics

Water Homeostasis

Production of urine by the mammalian kidney is a consequence of filtration, re-absorption and secretion. The filtered plasma (►glomerular filtrate) is processed by the kidney; amino acids, salts, and glucose are reabsorbed whereas urea, inorganic acids and other products of dietary metabolism are secreted and excreted. Re-absorption is aimed at recovering nutrients and water necessary to maintain the composition of body fluids. The selective process reabsorbs close to 99% of the nutrients and glucose but only 95% of the water. Filtered metabolic waste products and secreted substances remain in the filtrate and become the solutes found in urine. It is estimated that a healthy adult produces approximately 180 l of glomerular filtrate in 24 h. Approximately 170 l/day of filtrate are reabsorbed together with salts, glucose and nutrients, whereas most of the remaining volume (approximately 8 l/day) is re-absorbed by a specific water recovery mechanism. In normal individuals water recovery is hormonally regulated by arginine vasopressin (AVP, also known as antidiuretic hormone or ►ADH), and results in around 2 l of urine/day, a process that is defective in DI patients.

AVP is secreted from the posterior pituitary when special brain receptors detect increases in the plasma concentration of salts. The process is very sensitive, since increases of 1% are sufficient to promote AVP release and stimulate water re-absorption in the kidney collecting ducts. Once the reabsorbed water brings the plasma salt concentration back to normal values (around 300 mOsmol), hormone secretion is inhibited.

AVP controls re-absorption by regulating water permeability in the last portion of the kidney tubules that process the glomerular filtrate.

The V2 vasopressin receptor (►V2R) mediates AVP regulation of collecting duct water permeability. Specialized V2R-containing cells of the renal collecting duct termed ►principal cells are responsible for regulated water recovery. Activation of the V2R by AVP stimulates the production of 3'5'-cyclic adenosine monophosphate (cAMP) by the enzyme ►adenylyl cyclase. The elevated intracellular levels of cAMP activate the enzyme protein kinase A that in turn promotes the movement of the aquaporin2 water channel from the intracellular vesicles where it resides to the apical membrane of the cells that form the tubule. Tubular permeability to water augments due to the increased abundance of aquaporin2 in the membrane and re-absorption is facilitated by the high concentration of salt and urea in the tissue surrounding the collecting tubules. Thus, AVP adjusts the volume of urine excreted to match water intake, producing concentrated urine when the intake is restricted and allowing the excretion of dilute urine when fluid intake is abundant.

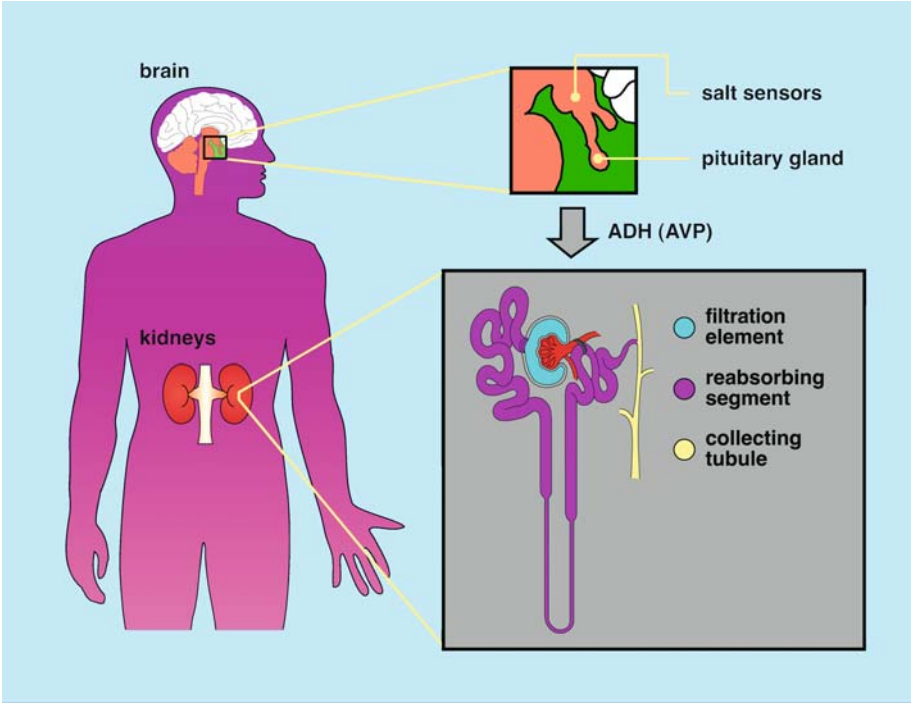
Cellular and Molecular Regulation

The Genetics of Diabetes Insipidus

A combination of traditional genetic linkage studies and new molecular biology methods applied to individuals suffering DI identified three genes that if altered can produce this disease. These genes encode the hormone arginine vasopressin (AVP), the V2 vasopressin receptor and the water channel aquaporin2. A mutation in any of these genes can cause DI. The characteristics of the different genotypes are summarized in Table 1; all are rare diseases.

Familial neurogenic or neurohypophyseal DI is caused by lack of the hormone AVP. Its manifestation is autosomal dominant, the gene does not reside in a sex chromosome and a single mutant allele produces the phenotype. Nephrogenic DI is caused by a lack of active V2 receptors or functional aquaporin2 channels. Alterations of the V2 receptor cause approximately 95% of the cases of nephrogenic DI (NDI), while aquaporin2 changes account for the remaining 5%. Both proteins are expressed only in the principal cells of the collecting ducts and are required for AVP regulation of water re-absorption. These diseases have distinct inheritance patterns determined by the chromosomal location of the genes, autosomal for aquaporin2 or sex-linked for the V2 receptor.

Some medications and toxic agents can alter water re-absorption in the kidney mimicking diabetes insipidus; whether their target is specific or not, remains speculative. This summary focuses on the genetic causes of the disease.



Diabetes Insipidus, a Water Homeostasis Disease. Figure 1 The diagram summarizes the anatomical location of the organs involved in maintaining water balance, the pituitary gland located at the base of the brain that releases AVP and the nephron, the functional unit of the kidney. The segments of the nephron where filtration and re-absorption take place are color-coded. The segment of the nephron where water permeability depends on the blood level of AVP (collecting tubule) is identified by the yellow color.

Diabetes Insipidus, a Water Homeostasis Disease. Table 1 Genetics of Diabetes Insipidus: a water homeostasis disease

Gene	Phenotype	Defect	Inheritance	Chromosomal Locus
AVP precursor	Neurogenic DI	No hormone	Autosomal dominant	20p13
AVP V2 receptor	Nephrogenic DI	No cell signaling	X-linked recessive	Xq28-ter
Aquaporin2 channel	Nephrogenic DI	No water channel	Autosomal recessive	12q13

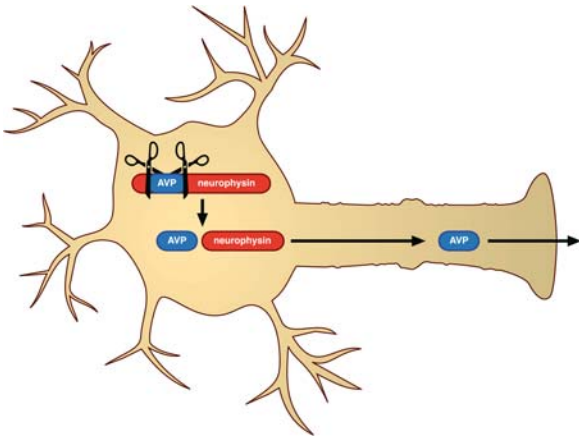
Lack of AVP - Familial Neurohypophyseal DI

The hormone AVP is a peptide of nine amino acids synthesized in the neuronal cells of the supraoptic and magnocellular nuclei of the hypothalamus as a precursor protein; this precursor protein is processed and cleaved to produce neurophysin and AVP (see Fig. 2). Mutations in the gene encoding the precursor protein change its amino acid composition and interfere with the proteolytic processing, thereby blocking hormone production. The accumulation of unprocessed precursor leads to degeneration of the neuronal cells. The presence of one mutant allele that disrupts AVP production is sufficient to damage or kill the neurons and prevent expression of the normal allele. As a

consequence one mutant allele often has a dominant phenotype. The devastating consequence of the accumulation of defective protein in the neuron is analogous to the retinal degeneration leading to blindness caused by one mutant rhodopsin allele in a heterozygous patient suffering retinitis pigmentosa. More than 40 mutations of the AVP-encoding gene have been described with prevalence of the dominant presentation.

Defective V2R - X-linked Recessive NDI

The V2 receptor gene was identified as the one responsible for this disease almost one hundred years after the symptoms and inheritance pattern were

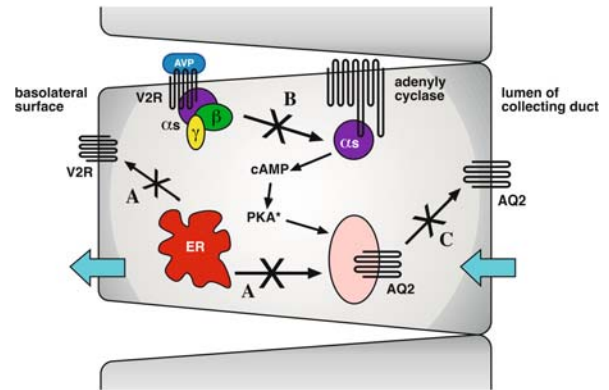


Diabetes Insipidus, a Water Homeostasis Disease.

Figure 2 The diagram represents an AVP producing neuron from the hypothalamus. The newly synthesized precursor protein is cleaved by special enzymes, represented here by the scissors, producing the nine amino acid hormone AVP and the larger accompanying protein neurophysin. Both are transported along the axon to reach the posterior pituitary where they are secreted in response to increased sodium concentration in the blood.

described. More than 180 mutations of the V2R gene have been identified. Most mutations disrupt protein synthesis by promoting alterations in receptor structure that interfere with the normal placing of the protein in the plasma membrane of the principal cell. As a consequence, the majority of the mutant receptors accumulate inside the cell, and the small percentage that reaches the cell surface is usually functionally defective. Mutations that cause the appearance of premature stop codons produce truncated proteins that are trapped inside the cell and never reach the surface. A few mutations with less dramatic effects on protein synthesis lower the ability of the receptor to promote intracellular signaling, thereby reducing the increase of intracellular levels of cAMP. These faulty receptors mobilize water channels to the apical surface of the cell only when AVP blood levels are very high. Hence, these mutations cause a partial phenotype of the disease; the kidneys produce concentrated urine only when circulating AVP reach the peak levels induced by dehydration. When blood salt concentration returns to normal values causing a drop in AVP levels, the symptoms reappear.

The gene for the V2 receptor resides in the human X-chromosome and, as is the case with most X-linked recessive disorders, the heterozygous female carriers are asymptomatic whilst the sons that inherit the mutant allele have the disease. The existence of heterozygous females with NDI symptoms puzzled researchers for



Diabetes Insipidus, a Water Homeostasis Disease.

Figure 3 Diagram of a *principal cell* of the collecting duct where water re-absorption takes place. The vasopressin V2 receptor (V2R) is present in the basolateral surface where AVP arrives by diffusing through the intracellular fluid. All surfaces of the cell are water permeable except for the apical surface, which only allows water passage when the aquaporin2 (AQ2) water channels have translocated from the intracellular vesicles to the cell surface. Mutations that interfere with the normal traffic of the V2R or AQ2 from the endoplasmic reticulum (ER) where they are synthesized to the cell surface or to the aggregophore vesicles, respectively, produce the blockade identified as A. Mutations that interfere with the ability of the V2R to stimulate cAMP production by adenylyl cyclase are identified as B. Finally, mutations of the AQ2 gene interfering with water channel activity or its insertion into the apical surface are identified as C. The green arrows identify the direction of water movement when re-absorption takes place.

some time. Recently, evidence of altered lyonization of the X chromosome has been reported in those symptomatic female carriers. The term lyonization describes the inactivation of one of the alleles of every gene present in the X chromosome. The inactivation or silencing of the genes is a consequence of a chemical change (DNA methylation) that takes place in cells bearing two X-chromosomes. Gene silencing works in a random manner inactivating one or the other allele. The consequence for the locus heterozygous for a mutation is that 50% of the cells will express a normal gene and 50% will express the mutant one. An alteration of the randomness of lyonization has been found to be a genetic trait in some families. This results in a skewed inactivation process that favors one allele over another. If silencing of the normal allele is favored, the abnormal protein will predominate and the heterozygous female will suffer NDI.

Defective Aquaporin2 - Autosomal NDI

Mutations identified in the aquaporin2 gene affect the activity and location of the water channel in the principal cells. Depending on the phase of protein synthesis and processing that is altered, the phenotype of this form of DI can be autosomal dominant or autosomal recessive.

In most cases, the disease is recessive. Mutation of a single allele does not interfere with the synthesis of the water channel from the normal allele and sufficient molecules are made to allow effective AVP regulation of water permeability without interference by the abnormal protein. The mutant channel protein is often unable to mature properly and remains trapped inside the ►[endoplasmic reticulum](#) where it is synthesized. Other mutant AQP2 channels may not be translocated properly from the intracellular vesicles to the cell surface without interfering with the movement of AQP2 normal proteins. Patients bearing such mutations will manifest the disease only if they have mutations in both alleles.

Recently, a dominant form of autosomal NDI has been characterized. In this form the presence of a single mutant gene results in a dominant phenotype. Water channels are homo-tetramers of identical subunits as seen by high-resolution electron microscopy and confirmed in biochemical experiments. The newly identified mutant subunit moves from the ER to the trans Golgi network where it mixes with the normal subunit produced by the normal allele giving rise to inactive heterogeneous channels. Simultaneous expression of normal and mutant aquaporin2 proteins in the same cells (*Xenopus* oocytes), confirmed that the ►[trans-Golgi network](#) is the formation site of the tetrameric channels and that mutant subunits can produce inactive heterogeneous water channels leading to a dominant phenotype.

References

1. Knepper MA (1997) Molecular physiology of urinary concentrating mechanism: regulation of aquaporin water channels by vasopressin. *Am J Physiol* 272:F3–F12
2. Gainer H (1998) Cell-specific gene expression in oxytocin and vasopressin magnocellular neurones. In: Zingg HH et al (eds) *Vasopressin and Oxytocin*. Plenum Press pp 15–27
3. Rutishauser J, Kopp P, Gaskill MB et al (2002) Clinical and molecular analysis of three families with autosomal dominant neurohypophyseal diabetes insipidus associated with a novel and recurrent mutations in the vasopressin-neurophysin II gene. *Eur J Endocrinol* 146:649–656
4. Birnbaumer M (1999) Vasopressin Receptor Mutations and Nephrogenic Diabetes Insipidus. In: Conn PM, Ulloa A (eds) *G protein-coupled receptors and G-proteins in Endocrine Disease Archives of Medical Research* 30: 465–474

5. Marr N, Bichet DG, Lonergan M et al (2002) Hetero-ligomerization of an Aquaporin-2 mutant with wild-type Aquaporin-2 and their misrouting to late endosomes/lysosomes explains dominant nephrogenic diabetes insipidus. *Hum Mol Genet* 11:779–789

Diabetes Mellitus

Definition

Diabetes mellitus is a group of metabolic disorders with one common manifestation: hyperglycemia. If untreated, hyperglycemia causes damage to the eyes, kidneys, nerves, heart and blood vessels. There are two major forms of diabetes mellitus: Type 1 diabetes mellitus (formerly called Type I, IDDM or juvenile diabetes) is characterized by destruction of the pancreatic beta cells caused by an autoimmune process, usually leading to absolute insulin deficiency. Type 2 diabetes mellitus (formerly called NIDDM, Type II or adult-onset) is characterized by insulin resistance in peripheral tissue, and an insulin secretory defect of the pancreatic beta cells. The latter is the most common form of diabetes mellitus, and is highly associated with a family history of diabetes, older age, obesity and lack of exercise.

►[Affinity Chromatography and In Vitro Binding \(Beads\)](#)

►[Diabetes Mellitus, Genetics](#)

►[Hypothalamic and Pituitary Diseases Genetics](#)

►[Repeat Expansion Diseases](#)

Diabetes Mellitus, Genetics

DONALD W. BOWDEN

Center for Human Genomics, Wake Forest University
School of Medicine, Winston-Salem, NC, USA
dbowden@wfubmc.edu

Synonyms

Type 1 diabetes mellitus: type 1 diabetes; juvenile diabetes; insulin dependent diabetes

Type 2 diabetes mellitus: type 2 diabetes; maturity onset diabetes (MODY); noninsulin dependent diabetes

Definition

►[Diabetes mellitus](#) is one of the most common chronic diseases in the world. Diabetes mellitus is characterized

by the inability to metabolize glucose efficiently, resulting in abnormally high levels of glucose (►**hyperglycemia**) in the blood. Continued hyperglycemia over time can lead to serious medical complications of diabetes, including heart disease (►**cardiovascular disease**), kidney disease (►**nephropathy**), damage to vision (►**retinopathy**), and nerve damage (►**neuropathy**) that contribute to early mortality and disability. There are two common forms of diabetes mellitus: ►**type 1 diabetes**, characterized by early age of onset and the inability to produce ►**insulin**, the key hormone for glucose metabolism, due to ►**autoimmune** destruction of insulin producing β -cells of the pancreas and ►**type 2 diabetes**, characterized by late age of onset and inability to utilize insulin properly.

It is widely accepted that susceptibility to type 1 diabetes, type 2 diabetes, and diabetes complications results from genes that individuals inherit acting in combination with their lifestyle and environment.

Characteristics

Evidence for Genetic Contributions to Diabetes

Multiple studies suggest that genes contribute to diabetes. In twin studies of either type 1 or type 2 diabetes, ►**monozygotic twins** have a higher concordance for diabetes than ►**dizygotic twins**. Concordance rates for monozygotic type 1 diabetes twins are 30–40% compared to 6% for dizygotic twins. Concordance rates for monozygotic type 2 diabetes twins are approximately 75% compared to 35% for dizygotic twins. ►**Familial aggregation** studies compare the frequency of diabetes within families to the frequency of diabetes in the general population. If a child with type 1 diabetes has a sibling, that sibling is 10–15 times more likely to develop type 1 diabetes than an unrelated child. For type 2 diabetes this sibling risk is 2–3.5 higher for a sibling of a type 2 diabetes affected individual. The incidences of type 1 and type 2 diabetes vary greatly between different ethnic and racial groups, which has been interpreted as being due to both genetic and lifestyle/environment differences between peoples.

Approaches for Locating and Identifying Genes for Diabetes and Diabetic Complications

Many researchers have sought to identify and characterize the genes contributing to diabetes and associated complications. There are two commonly followed molecular strategies, ►**candidate gene** analysis and whole ►**genome scans**, and two analytical approaches, ►**association studies** and ►**linkage** studies. Association studies compare the frequency of specific alleles of a genetic marker between different populations, conventionally case:control populations. Linkage studies evaluate the inheritance of a genetic locus in

families. Association and linkage studies have different strengths and limitations. Association studies can detect small (e.g., approximately 5% or less of the risk) genetic contributions. The limitation of this type of study is that results are frequently difficult to replicate independently. Technically association studies are “short-sighted”, they are able to look at only very limited lengths of DNA in the genome (approximately 50,000 bp). Linkage analysis in family sets is more robust, but less sensitive. Linkage analysis has the power to detect only significant (e.g., 25% of the risk or more) genetic contributions. Linkage analysis in family sets is “far-sighted” with an ability to detect genetic factors at a distance of 5 million base pairs or more, but has little ability to locate genes precisely. The genome scan is based on linkage analysis. In the genome scan a systematic survey of the entire human genome is carried out by collecting inheritance data on 400 or more genetic guideposts spaced at uniform intervals across all of the human chromosomes in a collection of appropriate families. These are major undertakings, but have the ability to located previously unknown genes contributing to diabetes or its complications.

Type 1 Diabetes

The human leukocyte antigen (*HLA*) gene complex was recognized as the major genetic contributor to type 1 diabetes over 30 years ago. The *HLA* complex of genes encompasses approximately 3.5 million base pairs of DNA on chromosome 6 and contains approximately 150 genes, many of which are directly involved in regulating the immune system. Imperfect regulation of the immune system is believed to lead to the autoimmune destruction of insulin producing β -cells in the pancreas. Detailed investigation of the *HLA* gene complex has lead to finer and finer mapping of the genes that contribute most directly to type 1 diabetes. A current focus is on the *HLA-DQ* ►**locus** made up of 2 genes, *HLA-DQA1* and *HLA-DQB1*. These genes present foreign antigens to helper T-cells and help mediate the immune response. The *HLA-DQ* genes are highly ►**polymorphic**, meaning there are multiple different forms of the DNA sequence, each of which is called an ►**allele**, for the same locus in the population. The alleles designated *DQA1*0301* and *DQB1*0302* are 3–4-fold more common in type 1 diabetes patients than in non-diabetic subjects. Other alleles show less dramatic differences between diabetic and non-diabetic individuals. The “risk” alleles are associated with a higher likelihood of developing diabetes, but are also common in the general population and only a small percentage of individuals with the risk alleles will develop diabetes. Combinations of *DQA1* and *DQB1* alleles, called ►**haplotypes**, may contribute greater risk than individual alleles. In addition, the risk can be modified by other *HLA* genes and other genes

in the genome. Due to these complexities, prediction of type 1 diabetes by DNA analysis is still not possible.

It has been estimated that *HLA* contributes approximately 42% of the genetic risk that a person will develop type 1 diabetes, with the rest of the risk being due to other genes and additional risk being contributed by environmental factors. At least 17 additional regions of the human chromosomes have been implicated in the development of type 1 diabetes, but the impact of these other putative genes is modest compared to *HLA*. Of these genes, the gene coding for insulin has the strongest impact, contributing an estimated 10% of the genetic risk of type 1 diabetes susceptibility.

Environmental factors undoubtedly play a key role, in concert with genetic factors, in type 1 diabetes susceptibility. These environmental factors are poorly understood but race, geography, viruses and nutrition have been suggested as possible factors. Clinical trials (1) are currently underway to assess whether early intervention can prevent or delay the onset of type 1 diabetes in individuals at high genetic risk for developing type 1 diabetes.

Type 2 Diabetes

Type 2 diabetes risk is also thought to be due to a combination of genetic, lifestyle and environmental factors. In contrast to type 1 diabetes, the major lifestyle and environment risk factors, obesity, poor nutrition, sedentary lifestyle and tobacco use, contributing to type 2 diabetes have been identified and extensively documented for decades. The genetic contributors to type 2 diabetes are, in contrast, poorly understood. A current hypothesis is that two metabolic defects, resistance to the action of insulin in peripheral tissues (such as skeletal muscle) and defects in the β -cell production of insulin are necessary for the development of type 2 diabetes. These defects are believed to have a genetic basis that is modified by lifestyle and environment factors such as obesity. In spite of 20 years or more of research, no major genes for type 2 diabetes susceptibility have been identified, which suggests that genetic risk for type 2 diabetes is due to the combination of many genes, each contributing a relatively small amount of risk.

Over 100 different genes have been genetically tested for contributing to type 2 diabetes. These candidate genes have been implicated in type 2 diabetes due to their known or hypothesized function in metabolism. A characteristic of these candidate gene studies is that multiple studies of the same gene often result in conflicting results as to whether the gene is a true contributor to type 2 diabetes. Evidence is now clear, however, that the most obvious candidate genes, insulin, insulin receptor and the common glucose transporter genes do not contribute significantly to type 2 diabetes in the general population. Only rarely have

genetic mutations been found in such genes and individuals affected by such mutations often have dramatic clinical presentations markedly different from type 1 or type 2 diabetes.

Some of the strongest evidence that a gene contributes to type 2 diabetes susceptibility is for the [▶peroxisome proliferator-activated receptor](#) (PPAR γ [▶ gene](#)), which is a member of the nuclear hormone receptor subfamily of transcription factors. PPAR γ regulates transcription of many genes and is believed to be involved in adipocyte differentiation. A common genetic polymorphism that changes the coding sequence of PPAR γ from proline to alanine at amino acid 12 of the coding sequence (Pro12Ala) has been evaluated in a large number of studies. Altshuler et al. (2) combined the results of 16 studies encompassing over 3000 subjects and carried out a [▶meta analysis](#) of the genetic data. They found that the common proline coding allele contributes a 1.25-fold increased risk for developing type 2 diabetes that was statistically significant and which could contribute to as much as 25% of the risk to develop type 2 diabetes in the general population.

A second gene for type 2 diabetes is the [▶calpain 10](#) gene identified in Hispanic American families (3). The *calpain 10* gene codes for a member of the calpain (a calcium activated protease)-like cysteine protease family. This gene was identified through its position on the long arm of human chromosome 5, which was inherited with type 2 diabetes in Hispanic American families in a genome scan. This gene was previously unknown and the mechanism by which a calcium activated protease could contribute to type 2 diabetes is unclear. Molecular genetic analysis suggested that combinations of alleles in different polymorphisms within the gene, called haplotypes, are the likely source of diabetes susceptibility. Individuals with the "risk" haplotypes are at 3–7 times greater risk of developing type 2 diabetes than other Hispanic Americans. Most surprisingly, the DNA sequences that make up the risk alleles are in non-coding intronic sequences within the gene, i.e. they do not directly affect the protein sequence. These non-coding sequences play a role in processing the RNA transcribed from the *calpain 10* gene and ultimately could affect levels of *calpain 10* protein levels. These results have puzzled researchers, but similar observations have now been made with other genes identified in common diseases such as Crohn's disease, asthma and schizophrenia. The identification and analysis of *calpain 10* could thus represent a window into previously unknown mechanisms for regulation of cellular metabolism.

Over 20 genome scans have been carried out in collections of type 2 diabetes families in an effort to locate type 2 diabetes genes. These studies have genetically analyzed many different populations and ethnic groups including Caucasian Americans,

Caucasian Europeans (French, Finnish, English), Hispanic Americans, Native Americans and Han Chinese. Three locations on the human chromosomes, the long arms of chromosomes 1, 10, and 20, have been consistently implicated as probable locations for genes that contribute to type 2 diabetes. The gene (or genes) within these regions have not, as yet, been identified

Maturity Onset Diabetes of the Young (MODY)

Maturity onset diabetes of the young (►MODY) (4) is a distinct subtype of type 2 diabetes, which constitutes 1–3% of type 2 diabetes. In contrast to type 2 diabetes in the general population, characterized by complex patterns of inheritance and late age of onset, MODY is recognized as a clearly defined pattern of ►autosomal dominant inheritance and early age of onset (childhood to early adulthood). Clinically MODY-affected individuals are ►hypoinsulinemic; they produce reduced amounts of insulin. Great progress has been made in understanding the molecular basis of MODY with the identification of mutations in at least 8 genes. Seven of these genes code for transcription factors that are involved in a wide variety of regulatory pathways for gene expression. For example the MODY1 form of diabetes is caused by defects in the ►HNF4A gene, which codes for ►hepatocyte nuclear factor 4 α protein. This protein interacts with regulatory elements in the promoters and enhancers of more than 40 target genes involved in glucose, cholesterol, and fatty acid metabolism. Mutations in the *insulin promoter factor 1* gene, *IPF1*, are responsible for the MODY4 form of diabetes. The *IPF1* coded transcription factor binds to specific sequences in the promoter of the insulin and other genes. Mutations in *IPF1* lead to reduced expression of insulin and abnormalities in pancreas development. Distinct from the transcription factor related diseases, MODY2 is caused by mutations in the glucokinase gene leading to dysfunctional glucokinase, which performs the rate-limiting step in glycolysis. Despite the many insights that MODY genes have provided into the molecular basis of glucose metabolism, none of these genes have been implicated as contributors to type 2 diabetes in the general population.

Genetics of Diabetes Complications

Like the disease entities of type 1 and type 2 diabetes, there is considerable evidence that the complications associated with diabetes have a genetic component. It has been widely observed that complications appear at different times and progress at different rates in diabetes-affected patients, even when they have similar glucose control and have had diabetes for similar durations. These observations have led to efforts to quantify the magnitude of the genetic components of

complications and identify the genes that contribute to susceptibility to developing complications.

A primary hypothesis is that genes that contribute to diabetic complications are different from the genes that contribute directly to diabetes. Susceptibility to cardiovascular (CVD) or renal disease, for example, is thought to be amplified by the environment of diabetes.

Evidence for Genetic Contributions to Diabetic Complications

Renal disease can be evaluated clinically by relatively inexpensive methods such as measuring urinary protein and serum creatinine. This has facilitated accumulation of extensive evidence that renal disease in diabetes has a genetic component. These conclusions are based on the results of familial aggregation studies and the observation of significantly different incidence rates in disparate racial and ethnic groups. For example, relatives of an African American with type 2 diabetes and nephropathy are at an 8-fold greater risk of developing diabetic nephropathy than an individual without nephropathic family members. These qualitative assessments have been complemented by mathematical modeling studies of renal disease called ►segregation analysis that quantitatively estimate the genetic contribution to diabetic renal disease. In such a study (5) it was calculated that 27% of a measure of renal health (albumin/creatinine ratio) in Caucasian type 2 diabetes families is contributed by genetic factors.

The genetic component of CVD in diabetes is less well documented since few studies have focused on purely diabetes affected populations and evaluation of CVD is more technically challenging than renal disease. Using a mathematical modeling approach called variance components analysis that calculates an estimate of the genetic and environmental contributions to a trait, subclinical cardiovascular disease has been estimated to be due 30–40% to environmental factors and 40–50% to genetic factors (6).

There is also evidence that genetics plays a role in retinopathy. In the Diabetes Control and Complications Trial family members of type 1 diabetes patients with severe retinopathy were at a 3-fold higher risk of developing retinopathy than families of retinopathy negative patients. In a study of racial/ethnic differences, African Americans are 2.5-fold as likely to develop retinopathy as a Caucasian American control group. It is likely that other diabetes complications, such as neuropathy also have genetic contributors, though definitive studies have not been completed.

The Search for Diabetes Complications Genes

Searches for diabetes complications genes have been fewer in number and are less advanced than searches for diabetes genes, though hundreds of reports of

analyses of candidate genes have been published. Of note are studies of the angiotensin converting enzyme > (►*ACE*) gene, which have been so extensive that meta analyses, are possible. In meta analyses of up to 50,000 subjects, the *ACE* gene is associated with both nephropathy and cardiovascular disease, but not retinopathy. Given that these effects are most clearly seen in a very large sample suggests the actual contribution of *ACE* to these traits is significant, but of small magnitude. Multiple other genes from the renin-angiotensin system, cytokines and growth factors, inflammatory and vascular factors, glycosylation genes, nitric oxide synthases, and genes of metabolic regulation have varying degrees of evidence for association with specific complications.

Genome scans targeted at locating complication genes are few in number. A study of Pima Indian families for both nephropathy and retinopathy revealed evidence for nephropathy genes on the long arm of chromosome 7. Additional segments of chromosomes 3, 9, and 20 showed some evidence of linkage to nephropathy and chromosomes 3 and 9 showed evidence of linkage to retinopathy. Other genome scans have been recently completed or are under way. A genome scan of Turkish kindreds with multiple diabetic nephropathy-affected individuals found strong evidence for linkage to chromosome 18. Evaluation of this location in the Pima Indian dataset showed evidence of confirmation.

Cellular and Molecular Regulation

Type 1 diabetes, type 2 diabetes, and diabetic complications have complex patterns of inheritance presumably due to the effects of multiple genes interacting with environmental and lifestyle factors. In only a few cases, e.g. *PPAR γ* in type 2 diabetes and the transcription factors identified in MODY, is the actual mechanism of action of these gene products understood. Even in these cases where the protein action is understood at a molecular level, extrapolation to prediction of physiological impact on the whole organism is difficult. The mechanisms by which the *HLA* genes in type 1 diabetes contribute to autoimmune destruction of the β -cells remain poorly understood. Knowledge of genes that contribute to the complications of diabetes remains limited, but these genes are likely to be different from the diabetes susceptibility genes themselves.

Clinical Relevance

It is currently possible to identify individuals that are at increased risk for developing type 1 diabetes due to the presence of higher risk *HLA* haplotypes. The presence of these haplotypes is in no way predictive however. If desired, confirmation of MODY can now frequently be determined through DNA analysis, and this same

technology could be used to predict this autosomal dominant form of diabetes. The primary genetic components of type 2 diabetes in the general population remain to be identified. Though at an early stage, efforts to identify genes that contribute to complications may have a significant impact on diabetes care. Identification of individuals at higher risk for developing specific complications early in the course of their diabetes may provide a window of opportunity to prevent or delay complications.

References

1. Schatz DA, Bingley PJ (2001) Update on major trials for the prevention of type 1 diabetes mellitus: the American Diabetes Prevention Trial (DPT-1) and the European Nicotinamide Diabetes Intervention Trial (ENDIT). *J Pediatr Endocrinol Metab* 14:619–622
2. Altshuler D, Hirschhorn JN, Klannemark M et al (2000) The common *PPAR γ* Pro12Ala polymorphism is associated with decreased risk of type 2 diabetes. *Nat Genet* 26:76–80
3. Horikawa Y, Oda N, Cox NJ et al (2000) Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. *Nat Genet* 26:163–175
4. Fajans SS, Bell GI, Polonsky KS (2001) Molecular mechanisms and clinical pathophysiology of maturity-onset diabetes of the young. *N Engl J Med* 345:971–980
5. Fogarty DG, Hanna LS, Wantman M et al (2002) Segregation analysis of urinary albumin excretion in families with type 2 diabetes. *Diabetes* 49:1057–1063
6. Wagenknecht LE, Bowden DW, Carr JJ et al (2001) Familial aggregation of coronary artery calcium in families with type 2 diabetes. *Diabetes* 50:861–866

Diabetes Polyuria

Definition

Diabetes polyuria stands for any disorder that leads to the production of large volumes of urine.

►Diabetes Insipidus, a Water Homeostasis Disease

Diagnostic Gene Testing

Definition

Genetic testing is performed on a person suffering from a disease to determine whether they have a mutation in a gene known to cause that disease.

►Hereditary Spastic Paraplegias

Diaphragm

Definition

Diaphragm describes a muscle separating the thoracic and abdominal cavities necessary to support breathing.

► [Muscle Development](#)

Diaphysis

Definition

Diaphysis is the shaft of the long bone.

► [Bone and Cartilage](#)

Diastole

Definition

Diastole designates the relaxation and dilatation phase of the heart chambers, during which they fill with blood.

► [Cardiac Signalling: Cellular, Molecular and Clinical Aspects](#)

Diastrophic Dysplasia

Definition

Diastrophic dysplasia refers to a severe recessive chondrodysplasia with abnormalities of the joints (contractures), scoliosis and abnormal ears. It is caused by mutations in a sulfate transporter *DTSDT* gene.

► [Bone Disease and Skeletal Disorders, Genetics](#)

Dicentric Chromosomes

Definition

Dicentric chromosomes arise by end-to-end fusion of two broken chromosomes and have two centromeres.

► [DNA Amplification](#)

Dichromatic Beamsplitter

Definition

Dichromatic beamsplitter is an optical filter, which reflects light of wavelengths shorter than the specified wavelength, and transmits light of longer wavelengths.

► [Fluorescence Microscopy: Spectral Imaging vs. Filter-Based Imaging](#)

Dickkopf

Definition

Dickkopf (Dkk) is a head inducer secreted from the vertebrate head organizer and induces anterior development. Dkk is a high-affinity ligand for binding ► [low density lipoprotein receptor-related protein 6 \(LRP6\)](#) and inhibits Wnt signaling.

► [Wnt/Beta-Catenin Signaling Pathway](#)

DI-CMT

Definition

DI-CMT stands for dominant-intermediate ► [Charcot-Marie-Tooth disease](#).

► [Hereditary Neuropathies, Motor and/or Sensor](#)

Dictyostelium Discoideum

Definition

Dictyostelium discoideum is a haploid, unicellular social amoeba. During starvation the unicellular amoebae polarises and aggregates by chemotaxis towards a cAMP source forming multicellular structures. The migrating cells show numerous similarities to leukocytes, providing an experimental system to investigate ► [chemotaxis](#) and ► [cell motility](#). Cell development, cell differentiation, cell adhesion, molecular genetics and a completed genome analysis are major research areas of this model organism.

► [Actin Cytoskeleton](#)

Differential Display

Definition

Differential display refers to a comparative analysis where total gene or protein contents from different sources are displayed and compared to each other to identify differences. Commonly used display methods include RT-PCR, Northern/Western blotting and Genechip(r)/Proteinchip(r) etc.

► [Proteomics in Human-Pathogen Interactions](#)

Differential in-Gel Electrophoresis

Definition

Samples of proteins from 2 or 3 experimental conditions are labeled with 2 or 3 respective fluorescent dyes, mixed together and run on a single 2D-gel electrophoresis, reducing variability. Confocal scanning and computer softwares then allow the quantification of the spots belonging to the respective samples.

► [Proteomics in Ageing](#)

Differential Scanning Calorimetry

T. REINISCH¹, J. RÖSGEN², H.-J. HINZ¹

¹Institute for Physical Chemistry, WWU Münster, Münster, Germany

²Institute of Human Biological Chemistry and Genetics, University of Texas Medical Branch, Galveston, TX, USA

hinz@uni-muenster.de

Definition

Differential Scanning Calorimetry is an extremely versatile technique to characterize quantitatively protein stability and conformational equilibria. It probes temperature inducible reactions by monitoring the apparent heat capacity of a sample solution relative to that of a reference solution as a function of temperature. The biologically interesting temperature range from 0°C to 130°C can be studied. The differential arrangement of two, practically identical measuring

units, which are filled with the sample and reference solution, respectively, allows for the compensation of the large background heat capacity of water. This constitutes an enormous improvement of the signal to noise ratio of the measurement. ► **DSC** is most powerful in the analysis of mechanisms of reversible reactions, however, for simple models irreversible processes can also be treated.

Description

Isothermal Titration Calorimetry measures at constant temperature the enthalpy change resulting from a reaction. In contrast Differential Scanning Calorimetry monitors the changes in apparent heat capacity of a sample in the course of a constant-rate temperature scan.

Application in Life Sciences

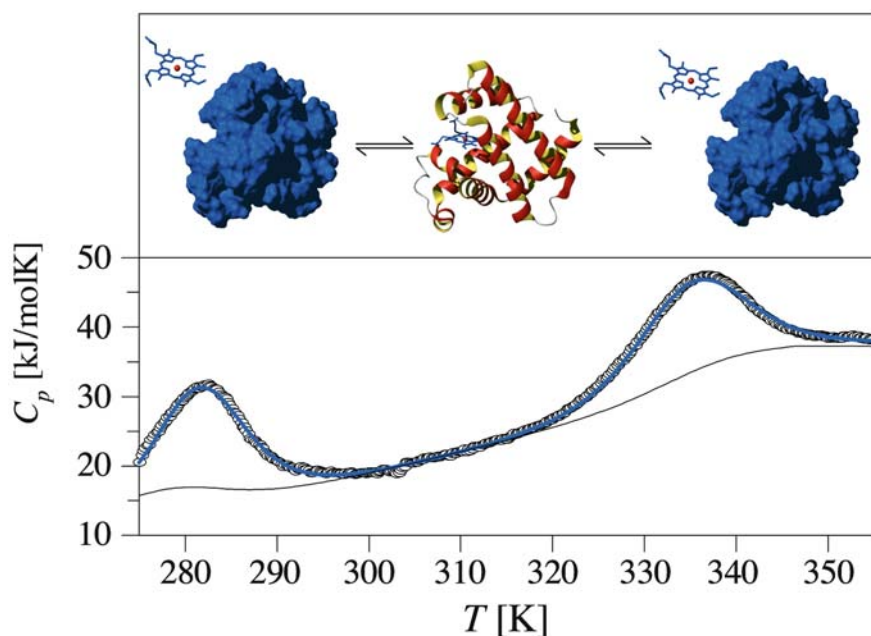
DSC is ideally suited to the thermodynamic characterization of all temperature inducible processes such as folding-unfolding transitions of proteins, helix to coil transitions of nucleic-acids, phase changes in membranes and changes in the degree of oligomerization. This implies that characterisation of stability of proteins, nucleic acids and membranes as well as of their complexes is a major domain of DSC.

Instrumental Aspects and Principles of Operation

To maintain equilibrium conditions during the temperature scan heating rates must be slower than the relaxation rates of the reaction observed. This can lead to very slow scan rates (1/1000 of a degree per min) for example in case of some lipid phase transitions. However, generally protein or nucleic acid solutions can be scanned at rates between 0.1 and 2 K/min. As in ► **ITC** modern DSC instruments use twin cells that operate in differential mode. Thus Fig. 1 of the article on ITC is also representative for a DSC instrument with the exception of the stirrer. The reference cell – filled with the identical equilibrium dialysis buffer solution as the sample cell (excluding the sample) – compensates for the 99.9% heat capacity signal of the buffer in a 1mg/ml solution. This arrangement improves the signal to noise ratio significantly and results in a differential heat capacity signal as a function of temperature, $\Delta c_{p,app}(T)$, that originates solely from the replacement of buffer solution by sample.

$$\Delta c_{p,app}(T) = c_{p,buffer} \cdot m_{buffer} - c_{p,sample}(T) \cdot m_{sample} \quad (1)$$

m_{buffer} is the mass of buffer solution replaced by the mass of the sample, m_{sample} . As sample and reference cell have the same volume, V_{cell} , one can express the masses by the partial specific volumes, $\bar{v} \left[\frac{cm^3}{g} \right]$.



Differential Scanning Calorimetry . Figure 1 Heat capacity profile and fit (blue) of Myoglobin ($c = 10$ g/l, 50 mM NaAc pH = 4.3), low temperature peak: $r = 0.2$ K/min, high temperature peak: $r = 0.5$ K/min.

$$m_{\text{buffer}} = \frac{V_{\text{cell}}}{\bar{V}_{\text{buffer}}}, m_{\text{sample}} = \frac{V_{\text{cell}}}{\bar{V}_{\text{sample}}}$$

Introducing these expression into eq.(1) and solving for the heat capacity of the sample one obtains the working equation of DSC

$$c_{p,\text{sample}}(T) = c_{p,\text{buffer}}(T) \frac{\bar{V}_{\text{sample}}(T)}{\bar{V}_{\text{buffer}}(T)} - \frac{\Delta c_{p,\text{app}}(T)}{m_{\text{sample}}} \quad (2)$$

Unless strong intermolecular interactions occur, the partial specific volumes do not significantly differ from the specific volumes at the low concentrations employed usually. However, for precise c_p data the partial specific volume should be used. $c_{p,\text{buffer}}$ is usually assumed to be equal to $c_{p,\text{water}}$. This is only acceptable for very low buffer concentrations. To obtain accurate absolute heat capacities for the biopolymer systems, $c_{p,\text{buffer}}$ must be determined by a reference run with one cell filled with buffer and the other one filled with water.

$$T_{1/2} = 333.4 \text{ K}, \Delta H^0(T_{1/2}) = 212 \text{ kJ/mol}, \Delta C_p^0(T_{1/2}) = 6.8 \text{ J/molK}$$

Figure 1 shows a typical DSC study on myoglobin at pH = 4.3. At this pH myoglobin is unfolded at low temperature, refolds as indicated by the low temperature peak to form the native structure between 2 and 16°C, and finally unfolds again at high temperature. The enthalpy changes accompanying the low temperature renaturation and the high temperature denaturation

can be obtained without a model assumption by integration of the excess heat capacity, which is defined as the difference between the observed heat capacity $c_p(T)$ and the baseline $c_{p,\text{baseline}}$.

$$c_{p,\text{excess}}(T) = c_p(T) - c_{p,\text{baseline}}(T) \quad (3)$$

$$\Delta H^0(T_{1/2}) \equiv \int_{T_1}^{T_2} c_{p,\text{excess}} dT; \quad (4)$$

$$\Delta S^0(T_{1/2}) = \frac{\Delta H^0(T_{1/2}) - \Delta G^0(T_{1/2})}{T_{1/2}}$$

T_1 and T_2 refer to the temperatures of the start and end, respectively, of the corresponding low or high temperature heat capacity peak. Although the transition extends over the temperature range $T_2 - T_1$, both the transition enthalpy and entropy are assigned by definition to the temperature of half conversion, $T_{1/2}$. For an N to D transition $\Delta G^0(T_{1/2})$ vanishes at $T_{1/2}$ and

$$\Delta S^0(T_{1/2}) \text{ equals } \frac{\Delta H^0(T_{1/2})}{T_{1/2}}.$$

The exact values of the so-called transition temperatures, $T_{1/2}$, which are defined as those temperatures at which 50% conversion of the sample has occurred, are given in the figure legend. The heat capacity curve of myoglobin is unique compared to that of the large majority of other proteins in that the so-called cold

denaturation is observed in addition to heat denaturation in the accessible temperature range of about 0°C–130°C. The reason is the large positive heat capacity change ΔC_p associated with unfolding and the relatively small ΔH^0 value. This combination renders possible a fast change in the fractions of both native, $\alpha_N(T)$, and unfolded proteins, $\alpha_D(T)$, as a function of temperature. This becomes evident from the equations for a simple two state transition

$$\begin{aligned} N &\rightleftharpoons D; \\ K &= \frac{[D]}{[N]}; \\ \alpha_N(T) &= \frac{[N]}{[N] + [D]} = \frac{1}{1 + K}; \\ \alpha_D(T) &= \frac{[D]}{[N] + [D]} = \frac{K}{1 + K}; \end{aligned} \quad (5)$$

The equilibrium constant is related to the standard Gibbs energy change, ΔG^0 , via the relation

$$\begin{aligned} K(T) &= e^{-\frac{\Delta G^0(T)}{RT}} \\ \Delta G^0(T) &= \Delta H^0(T) - T\Delta S^0(T) \\ &= \Delta H^0(T_{1/2}) + \Delta C_p(T - T_{1/2}) \\ &\quad - T \left\{ \frac{\Delta H^0(T_{1/2})}{T_{1/2}} + \Delta C_p \ln \frac{T}{T_{1/2}} \right\} \end{aligned} \quad (6)$$

Equation (6) implies a linear temperature dependence of the enthalpy as described by $\Delta H^0(T) = \Delta H^0(T_{1/2}) + C_p(T - T_{1/2})$ and the corresponding relation for the entropy change: $\Delta S^0(T) = \Delta S^0(T_{1/2}) + \Delta C_p \ln \frac{T}{T_{1/2}}$.

For the N (D) transition ΔG^0 equals 0 at $T_{1/2}$. A comparison with eq. (24) in the chapter on ITC shows the identity of the two equations. Here the reference temperature is $T_{1/2}$, the temperature of half conversion, while there it is any convenient experimental reference temperature T_R . Inspection of eq.(6) reveals that for a vanishing ΔC_p there will be no cold unfolding, since ΔG^0 does not turn negative with decreasing temperature. However, in general unfolding of proteins is associated with a significantly positive value of ΔC_p and this can lead in combination with a moderate enthalpy change at $T_{1/2}$ to a change in sign of ΔG^0 above 0°C, as seen in Fig. 1 for myoglobin.

Fitting of Heat Capacity Curves to Different Transition Models Using Partition Functions

Insight into the mechanism of the reaction underlying the heat capacity curve can be obtained by fitting of the heat capacity data to analytical equations describing various transition models. An elegant approach is based

again on the relative partition function Z . One of the prominent models used in abundance for protein unfolding is the so-called **two-state** model. Within the framework of this model it is assumed that only two populations of protein molecules exist, native (N) and unfolded (D) molecules and that the relative fractions α_N and α_D are shifted by temperature.

The relative partition function Z for this model is defined as

$$Z = \frac{[N] + [D]}{[N]} = 1 + K \quad (7)$$

The enthalpy relative to the native state N can be calculated from eq.(8)

$$H(T) - H_N(T) = RT^2 \frac{\partial \ln Z}{\partial T} = \frac{K}{1 + K} \Delta H^0(T) = \alpha_D(T) \cdot \Delta H^0(T) \quad (8)$$

The heat capacity $c_p(T)$ is the variation of the enthalpy with temperature and is described by the following equation

$$\begin{aligned} C_p(T) &= C_{p,N}(T) + \frac{\partial(H - H_N)}{\partial T} \\ &= C_{p,N}(T) + \Delta C_p(T) \frac{K}{1 + K} \\ &\quad + \frac{[\Delta H^0(T)]^2}{RT^2} \cdot \frac{K}{1 + K} \cdot \frac{1}{1 + K} \\ &= C_{p,N}(T) + \Delta C_p(T) \cdot \alpha_D \\ &\quad + \frac{[\Delta H^0(T)]^2}{RT^2} \cdot \alpha_N \alpha_D \end{aligned} \quad (9)$$

$C_{p,N}(T)$ is the heat capacity function of the native state. For small globular proteins it can be represented by a linear function, having values of approximately $1 - 2.5 \text{ JK}^{-1}\text{g}^{-1}$ near 25°C and a slope of $(6 - 8) \cdot 10^{-3} \text{ JK}^{-2}\text{g}^{-1}$. Deviations from these average values have been interpreted as indicating loose structural packing with concomitant large conformational fluctuations. $\Delta C_p(T) = C_{p,D}(T) - C_{p,N}(T)$ is the difference between the extrapolated heat capacity functions of the denatured and native states at temperature T .

$C_{p,N}(T) + \Delta C_p(T) \cdot \alpha_D(T)$ is the equation for the baseline and the third term in eq.(9) represents the excess heat capacity function resulting from the transition.

There is another transition mechanism which applies to a large number of protein- and DNA reactions. It is the dissociation of a dimer with concomitant loss of native structure. The corresponding equilibrium constants are

$$N_2 \rightleftharpoons 2D \quad K = \frac{[D]^2}{[N_2]} \quad (10)$$

$$AB \rightleftharpoons A + B \quad K = \frac{[A][B]}{[AB]} \quad (11)$$

Equation (10) refers for example to unfolding of a homodimeric protein, while eq.(11) describes melting of double-stranded DNA. The relative partition function Z for the $N_2 = 2D$ equilibrium is

$$Z = \frac{[N] + [D]}{[N]} = \frac{2[N_2] + [D]}{2[N_2]} = 1 + \frac{2}{\sqrt{1 + \frac{8c}{K}} - 1} \quad (12)$$

Z has been formulated in terms of subunit monomer concentrations with $c = 2[N_2] + [D] = \text{const.}$ being the total monomer concentration.

Proper derivation of the relative partition function with respect to temperature provides the enthalpy relative to the native state.

$$H - H_N = RT^2 \frac{\partial \ln Z}{\partial T} = \Delta H^0 \frac{\alpha_D}{1 + \alpha_D} \quad (13)$$

$\alpha_D(T) = \frac{K}{1+K}$ is the degree of unfolding, $K(T) = e^{-\frac{\Delta G^0(T)}{RT}}$ is the equilibrium constant and ΔH^0 is the transition enthalpy.

The heat capacity function is the variation with temperature of the enthalpy.

$$\begin{aligned} C_p - C_{p,N} &= \frac{\partial(H - H_N)}{\partial T} \\ &= \frac{\partial}{\partial T} \left\{ RT^2 \frac{\partial \ln Z}{\partial T} \right\} \\ &= C_{p,N} + \Delta_N^D c_p \frac{\alpha_D}{1 + \alpha_D} \\ &\quad + \frac{(\Delta H^0)^2}{RT^2} \cdot \frac{2\alpha_N \alpha_D}{(1 + \alpha_D)^3} \end{aligned} \quad (14)$$

It is important to note that the weighting factors for both ΔH^0 in eq.(13) and (ΔH^0) in eq.(14) are different from those reported previously in the literature.

An analogous treatment of the reaction $AB \rightleftharpoons A + B$ yields the following expressions:

$$Z = \frac{[AB] + [A] + [B]}{[AB]} \quad (15)$$

The total monomer concentration is constant due to mass conservation

$$c = 2[AB] + [A] + [B] = 2[AB] + 2[A] \quad (16)$$

$$Z = 1 + \frac{2[A]}{[AB]} = 1 + \frac{2[A]}{\frac{[A]^2}{K}} = 1 + \frac{2K}{[A]} \quad (17)$$

The enthalpy relative to the native AB complex is given by the following equation

$$H - H_N = \Delta H^0 \cdot \frac{Q - 1}{Q + 1} \quad (18)$$

The expression for Q is:

$$Q = 1 + \frac{2}{\left(\sqrt{1 + \frac{2c}{K}} - 1\right)} \quad (18a)$$

The heat capacity function is again obtained by the derivation of the enthalpy function with respect to temperature:

$$C_p - C_{p,N} = \Delta C_p^0 \frac{Q - 1}{Q + 1} + \frac{(\Delta H^0)^2}{RT^2} \cdot \frac{2Q(Q - 1)}{(Q + 1)^3} \quad (18b)$$

The differences in the formulae for the two dimer-to-monomer equilibria are the result of the non-identity of subunits in the $[AB] = [A] + [B]$ case.

Cooperativity Parameters

A variety of further unfolding mechanisms is elaborated in reference [1]. The enthalpy values obtained by the fitting routines from the heat capacity curves are so-called van't Hoff enthalpies, $\Delta H^0 = \Delta H_{vH}$, as they are derived ultimately from the variation with temperature of the equilibrium constant.

The ratio between the model free calorimetric enthalpy, ΔH_{cal} , obtained by integration of the transition peak and the ΔH_{vH}^0 value derived from the fit, is a rigid criterion for the size of the cooperative unit defined by the ratio $(\Delta H_{cal} / \Delta H_{vH})$. A ratio of $\frac{\Delta H_{cal}}{\Delta H_{vH}^0} = 1$ indicates unambiguously the occurrence of a cooperative two-state transition involving the complete protein or DNA macromolecule. Ratios larger than one signify that probably intermediate states are populated in the course of the transition. Intermediate transitions are likely to require less than the maximal enthalpy for the transformation between the native state and the intermediate conformation.

A ratio of $\frac{\Delta H_{cal}}{\Delta H_{vH}^0}$ is obtained, if intermolecular co-operation exists as, for example, in phase transitions of lipid vesicles. Such a ratio smaller than 1 can, however, also point to an irreversible process coupled to unfolding, as for example the aggregation of unfolded proteins during a heating scan.

Stoichiometry Tests

DSC provides an excellent test of the stoichiometry of a reaction. For the $N_n \rightleftharpoons nD$ transition the equilibrium constant at the transition temperature $T_{1/2}$ can be expressed as

$$K(T_{1/2}) = n\left(\frac{c}{2}\right)^{n-1} \quad (19)$$

with $c = n[N] + [D]$ being the constant, analytical monomer concentration.

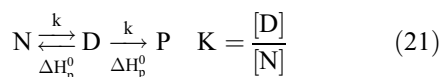
By performing several DSC studies at different concentrations c the variation of the transition temperature $T_{1/2}$ with $\ln c$ can be analyzed. To estimate the stoichiometry n of the reaction the following equation is used:

$$\frac{1}{T_{1/2}} = \frac{R \ln \left[n \left(\frac{c}{2} \right)^{n-1} \right]}{\Delta H^0(T_{1/2})} - \frac{\Delta S^0(T_{1/2})}{\Delta H^0(T_{1/2})} \quad (20)$$

$\Delta H^0(T_{1/2})$ and $\Delta S^0(T_{1/2})$ are the respective standard transition enthalpies and entropies at $T_{1/2}$.

Treatment of Irreversible Reactions

Several protein unfolding reactions are associated with irreversible behaviour of the unfolded species. If the reaction can be described by a so-called Lumry-Eyring- or, equivalently, Michaelis-Menten type mechanism, both equilibrium and kinetic parameters can be deduced from the heat capacity curve. The following equations apply to this mechanism:



If the rate constant k is small compared to the rate constants governing the equilibrium, the equilibrium constant K is to a good approximation represented by the ratio of the fractions of denatured (α_D) and native protein (α_N)

$$K = \frac{(\alpha_D)}{(\alpha_N)} \quad (22)$$

Considering mass conservation by the condition $1 = \alpha_D + \alpha_N + \alpha_P$ and introducing the rate of product formation by

$$\frac{\partial \alpha_P}{\partial t} = k \cdot \alpha_D = k \cdot \alpha_D^{\text{eq}} (1 - \alpha_P) \quad (23)$$

The fraction of irreversible product, α_P , at temperature T can be obtained by integration

$$\alpha_P = 1 - \exp \left(- \int_{T_0}^T \frac{k \alpha_D^{\text{eq}}}{r} dT \right) \quad (24)$$

The time t can be transformed to temperature T by using the definition of heating rate $r = \frac{dT}{dt}$.

The enthalpy relative to the native state results from the following equation

$$\begin{aligned} H - H_N &= \Delta H_N^0 (\alpha_D + \alpha_F) + \Delta H_P^0 \alpha_P \\ &= \Delta H^0 [\alpha_D^{\text{eq}} (1 - \alpha_P) + \alpha_P] + \Delta H_P^0 \alpha_P \end{aligned} \quad (25)$$

ΔH^0 is the enthalpy change associated with the $N \rightleftharpoons D$ equilibrium, ΔH_P^0 that accompanying the irreversible reaction forming P . α_D^{eq} is the equilibrium fraction of unfolded protein defined by $\alpha_D^{\text{eq}} = \frac{K}{1+K}$.

The final fitting equation is obtained from the temperature derivative of the enthalpy

$$\begin{aligned} C_p &= \frac{\partial}{\partial T} (H - H_N) \\ &= C_{p,N} + \Delta_N^D C_p (\alpha_D + \alpha_F) + \Delta_D^P C_p \alpha_P \\ &\quad + \frac{(\Delta_N^D H^0)^2}{RT^2} - \alpha_D^{\text{eq}} \alpha_D^{\text{eq}} \exp \left(- \int_{T_0}^T \frac{k \alpha_D^{\text{eq}}}{r} dT \right) \\ &\quad + \{ (1 - \alpha_D^{\text{eq}}) \Delta_N^D H^0 + \Delta_D^P H^0 \} \frac{k \alpha_D^{\text{eq}}}{r} \\ &\quad \times \exp \left(- \int_{T_0}^T \frac{k \alpha_D^{\text{eq}}}{r} dT \right) \end{aligned} \quad (26)$$

As ancillary condition one usually assumes Arrhenius behaviour for the rate constant $k(T)$

$$k(T) = k(T_{1/2}) \exp \left[- \frac{E_a}{R} \left(\frac{1}{T} - \frac{1}{T_{1/2}} \right) \right] \quad (27)$$

E_a is the activation energy of the irreversible $D \rightarrow P$ process. R is the gas constant and $T_{1/2}$ is the temperature of 50% transition.

Estimation of the Average Degree of Ligandation

DSC is favourably used for quantifying the stabilization of proteins by ligand binding. Particularly the average degree of ligand binding, \bar{X} , can be conveniently estimated from the variation of the transition temperature $T_{1/2}$ with the logarithm of the ligand concentration, x , using the equation

$$\frac{\partial \left(\frac{1}{T_{1/2}} \right)}{\partial \ln x} = -R \frac{\bar{X}}{\Delta H^0(T_{1/2})} \quad (28)$$

The slope of a plot of the reciprocal transition temperature versus $\ln x$ gives the average number of ligands bound per mole of protein provided the overall transition enthalpy in the presence of ligands, $\Delta H^0 \times (T_{1/2})$, at $T_{1/2}$ is known. This first estimate of \bar{X} forms a convenient basis for quantitative numerical fitting strategies described in [1].

Summary

This concise description of DSC neglected the wealth of information available for the molecular interpretation of the thermodynamic parameters. Instead it has emphasized the quantitative aspects of the method. DSC is the only technique that provides a direct access to all thermodynamic parameters relevant to protein- or in general biopolymer stability. These are the standard Gibbs energy change, ΔG^0 , the transition enthalpy, ΔH^0 , the heat capacity change, ΔC_p , that determines the dependence on temperature of ΔH^0 . As these parameters are usually obtained by DSC at the upper biologically significant temperature range compared to values from ITC, the two techniques are favourably complementary.

There is a growing awareness that, specifically for rational drug design, for the quantification of mutational effects on the stability of proteins, for the characterization of conformational changes and for ligand interactions, structural information alone is insufficient for a deeper understanding of the forces responsible for biomolecular interactions. However, the combination of thermodynamic and structural knowledge has opened new strategies with increased predictive power for a successful route in drug design (4, 5).

In this context both ITC and DSC play a dominant role which will even become more prominent on the basis of new high-throughput instruments currently under construction.

► [Thermodynamic Properties of DNA](#)

References

1. Rös gen J, Hinz H-J (1999) Theory and Practice of DSC Measurements on Proteins. In: Kemp RB (ed) Handbook of Thermal Analysis and Calorimetry. From Macromolecules to Man, vol 4
2. Rös gen J, Hinz H-J (2001) Folding Energetics of Ligand Binding Proteins: I. Theoretical Model. J Mol Biol 306:807–824
3. Rös gen J, Hinz H-J (1999) Statistical thermodynamic treatment of conformational transitions of monomeric and oligomeric proteins. Phys Chem Chem Phys 9:2327–2334
4. Todd MJ, Semo N, Freire E (1998) The Structural Stability of the HIV-1 Protease. J Mol Biol 283:475–488
5. Velazquez- Campoy A, Freire E (2001) Incorporating Target heterogeneity in drug design. J Cell Biochem Suppl 32:82–88

Differentially Methylated Region

Definition

Differentially methylated region (DMR) designates the DNA segments that are methylated differentially on the

two parental alleles. Such elements usually play a role in the process of Genomic Imprinting.

► [Genomic Imprinting](#)

Differentiation

Definition

Differentiation describes the transformation, as a result of altered gene expression, of a ► [precursor cell](#) into an individual cell type with a characteristic structure and function. Differentiation creates the diversity of cells that arise from a fertilized egg. Differentiation also occurs in microorganisms.

► [Biochemical Engineering of Glycoproteins](#)

► [Cancer Invasion and Metastasis: Cellular, Molecular and Clinical Aspects](#)

► [Lung](#)

► [Muscle Development](#)

Diffraction Intensities

Definition

The square of the amplitude of the diffracted wave; these are the experimentally measured quantities for which there is no direct phase information.

► [X-Ray Crystallography—Basic Principles](#)

Diffusion

Definition

Diffusion refers to the process by which molecules in the gaseous or liquid phase randomly move as a result of their thermal movement, and usually diffuse along a concentration gradient. In cell biology, diffusion is described as a form of “passive transport” by which substances cross membranes.

► [FRAP](#)

► [Proteomics in Microfluidic Systems](#)

DIGE

Definition

The DIGE (difference gel electrophoresis) system is based on the labelling of proteins on primary amines using fluorescent dyes (Cy3 and Cy5). Two protein

extracts (e.g. from two cell populations) are labelled with a dye and then mixed before 2D-gel electrophoresis. The visualization of the proteins in the gel is realized by fluorescence. The system allows the distinction of the proteins differentially expressed between the two extracts in a single gel and their quantification.

► [Proteomics in Cancer](#)

DiGeorge/Velocardiofacial Syndrome

Definition

DiGeorge/velocardiofacial syndrome (DG/VCFS), alias 22q11 deletion syndrome, alias CATCH phenotype (formerly known as CATCH22), including DiGeorge syndrome, including velocardiofacial syndrome (alias Shprintzen syndrome, alias Shprintzen-Sedláčková syndrome), including conotruncal anomalies face syndrome, is the second most common (1:4,000 newborns) autosomal disorder in man after ► [Down syndrome](#) (trisomy 21). It is characterized by extremely variable defects of the third and fourth branchial arches (conotruncal heart defects, facial anomalies, thymic hypo- or aplasia, cleft lip/palate, ► [hypoparathyroidism](#)). Intelligence is normal or borderline, but young adults have a 30% risk of schizoid psychosis. DG/VCFS is caused (>98%) by a chromosome 22q11 deletion, which is detected by ► [FISH](#), or (<1%) by deletion of the *GATA3*-D10S585 area at chromosome 10p14.

► [Microdeletion Syndromes](#)

Dihydrofolate Reductase

Definition

Dihydrofolate reductase is the enzyme which catalyzes the reduction of dihydrofolate (DHF) to tetrahydrofolate, and which is the target for the anticancer drug, methotrexate (MTX).

► [Nucleotide Biosynthesis](#)

Dilated Cardiomyopathy

Definition

Dilated Cardiomyopathy (DCM) is a disease of the heart muscle that causes the heart to become enlarged, and to pump less strongly.

► [Familial Dilated Cardiomyopathy](#)

Dimer

Definition

Dimer refers to a complex consisting of two independent molecules (mostly proteins). If both components are identical, the complex is called homodimer; if the components are of different nature, it is called heterodimer.

► [Neutrophilic Factors](#)

Dioxygenases, Monooxygenases and Oxidases

Definition

Dioxygenases, Monooxygenases and oxidases are enzymes that metabolize molecular oxygen (dioxygen). They are defined as dioxygenases when both atoms of oxygen are incorporated into their products. Monooxygenases incorporate one atom into a substrate and the other is reduced to water. Oxidases catalysed reactions reducing both oxygen atoms to water.

► [Enzyme Catalyzed Post-translational Hydroxylation of Proteins](#)

Diploid

Definition

Diploid defines a chromosomal status in a cell with two copies of each chromosome. It represents the normal status for most human and zebrafish somatic cells, except haploid cells of the gonads, which have only a single copy of each chromosome.

► [Mutagenesis Approaches in the Zebrafish](#)

► [SNP Detection and Mass Spectrometry](#)

Diploid Genome

Definition

Diploid genome refers to a genome that contains a balanced set of chromosomes derived equally from maternal and paternal sources.

► [Centromeres](#)

DISC

► Death Inducing Signaling Complex

Discordance / Discordant

Definition

Discordant means that only one member of a twin pair shows the trait.

► Schizophrenia Genetics

Discovery Approach

Definition

A discovery approach is a hypothesis-free view on a biological situation. For example, cell structures may be viewed by a microscope, protein species compositions by 2-DE.

► Two-Dimensional Gel Electrophoresis

Disease Models

Definition

Disease models are conditions identified in humans and/or animals that are of value for studying biological phenomena and pathological mechanisms, and are useful in studying disease in general.

► *C. Elegans* as a Model Organism for Functional Genomics

► Repeat Expansion Diseases

Disease-Modifying Anti-Rheumatic Drugs

Definition

Disease-modifying anti-rheumatic drugs are therapies for rheumatoid arthritis with the capability to retard the rate of progression of structural damage to joints.

► Rheumatoid Arthritis

Dishevelled

Definition

Dishevelled (Dvl) describes a modular protein that is required for Wnt/ β -catenin signalling, and that also functions in other signalling pathways such as the Planar Cell Polarity (PCP) pathway in *Drosophila*, which controls the polarized orientation of cells, and thereby the formation of functional organs.

► Wnt/Beta-Catenin Signaling Pathway

Dispersive Difference

Definition

The dispersive difference is the difference of equal reflections hkl at different wavelengths. Optimal wavelengths are λ_{infl} at the inflection point of the absorption curve, and λ_{native} at a shorter wavelength with low absorption.

$$\Delta F_{\text{disp}} = |F_{\text{obs}}(\lambda_i, hkl)| - |F_{\text{obs}}(\lambda_j, hkl)|$$

► MAD Phasing

Dissolution Point

Definition

Dissolution point characterizes the point in the solubilization process where a sufficient number of detergent molecules are available to solubilize all components of a biological membrane.

► Two-dimensional Crystallization of Membrane Proteins

Distributive and Processive DNA Synthesis

Definition

A DNA Polymerase is distributive if it dissociates from the primer-template after each incorporation of a single

nucleotide. In contrast, incorporation of several nucleotides of a DNA polymerase without dissociation from the primer-template is named processive DNA synthesis. The processivity of an enzyme describes the average number of nucleotides it incorporates without dissociation from the primer-template.

► [DNA Polymerases](#)

are as well-protected as their disulfide bonds. Structural fluctuations (usually global unfolding) that expose the thiols are likely to expose the disulfide bonds as well. Under typical oxidative folding conditions, disulfide-insecure species preferentially reshuffle to unstructured disulfide species.

► [Protein Disulfide Bonds](#)

Disulfide Regeneration

Definition

Disulfide regeneration designates a process by which an unfolded, reduced cysteine-containing protein recovers its native disulfide bonds.

► [Protein Disulfide Bonds](#)

Disulfide-Protected Species

Definition

Disulfide-protected species are molecules containing disulfide bonds whose stable tertiary structure inhibits the reduction or reshuffling of its disulfide bonds.

► [Protein Disulfide Bonds](#)

Disulfide Reshuffling

Definition

Disulfide reshuffling describes a thiol-disulfide exchange reaction during which a protein thiolate attacks a protein disulfide to form a new protein disulfide bond, and release a thiolate that was involved in the former protein disulfide bond.

► [Protein Disulfide Bonds](#)

Disulfide-Secure Species

Definition

Disulfide-secure species denote a disulfide-protected intermediate in oxidative folding whose thiol groups can be exposed by a local unfolding, which maintains the burial (and protection) of the disulfide bonds. Under typical oxidative folding condition, disulfide-secure species preferentially oxidize to the native protein.

► [Protein Disulfide Bonds](#)

Disulfide Species

Definition

Disulfide species designates a protein with a specific pairing of its cysteines that form a disulfide bond(s).

► [Protein Disulfide Bonds](#)

Disulfide-Insecure Species

Definition

Disulfide-insecure species refers to a disulfide-protected intermediate in oxidative folding, whose thiols

Diuretics

Definition

Diuretics promote the urinary excretion of sodium (Na^+), by inhibiting the absorption of filtered fluid across the renal tubular epithelium. The ensuing reduction in Na^+ reabsorption reduces the Na^+ content of the body, the critical determinant of extracellular and plasma fluid volumes. Thus, the use of diuretics is primarily indicated in the treatment of oedematous diseases and arterial hypertension.

► [Mendelian Forms of Human Hypertension and Mechanisms of Disease](#)

Dizygotic

Definition

Dizygotic means fraternal or unidentical twins developed from separate zygotes.

► [Schizophrenia Genetics](#)

Dkk

► [Dickkopf](#)

DM

► [Diabetes Mellitus](#)

DM Domain

Definition

The DM domain has a zinc finger-like DNA-binding motif and was originally described as a DNA-binding motif shared between doublesex (dsx) in *Drosophila* and mab-3 in *C. elegans*. Since the initial characterizations of dsx and mab-3, DM(dsx mab)-related genes have been identified from virtually all species examined.

► [Mutagenesis Approaches in Medaka](#)

DM1

► [Myotonic Dystrophy Type 1](#)

DM2

► [Myotonic Dystrophy Type 2](#)

DMPK

► [Myotonic Dystrophy Protein Kinase](#)

DMR

► [Differentially Methylated Region](#)

DMs

Definition

DMs are double minute chromosomes, i.e. small circular chromosomes. They have an origin of replication, so that they are multiplied but lack a centromere. Their distribution between mother and daughter cell during mitosis occurs randomly.

► [DNA Amplification](#)

DNA

Definition

DNA stands for Deoxyribose nucleic acid. It consists of a specific sequence of 2'-deoxynucleoside 5'-monophosphates (deoxyribonucleotides) specifying the genetic code.

► [DNA Structure](#)

► [Fragile X Syndrome](#)

► [Nucleotide Biosynthesis](#)

► [Shotgun Libraries](#)

DNA Amplification

ERHARD WINTERSBERGER

Institute of Medical Biochemistry, Division of Molecular Biology, Medical University of Vienna, Vienna, Austria

erhard.wintersberger@univie.ac.at

Definition

DNA amplification is a process whereby a limited part of the genome is increased in copy number. It can comprise DNA sequences of the length of ►oligonucleotides up to large parts of the genome. Usually the definition of DNA amplification does not include polyploidization, where the total genome (i.e. the number of chromosomes) increases by an integral number. In some organisms, amplification of specific genes takes place in particular cells and is necessary for normal development. In other cases DNA amplification gives cells a growth advantage by partially releasing them from normal regulation or allowing them to survive under otherwise unfavorable conditions (2).

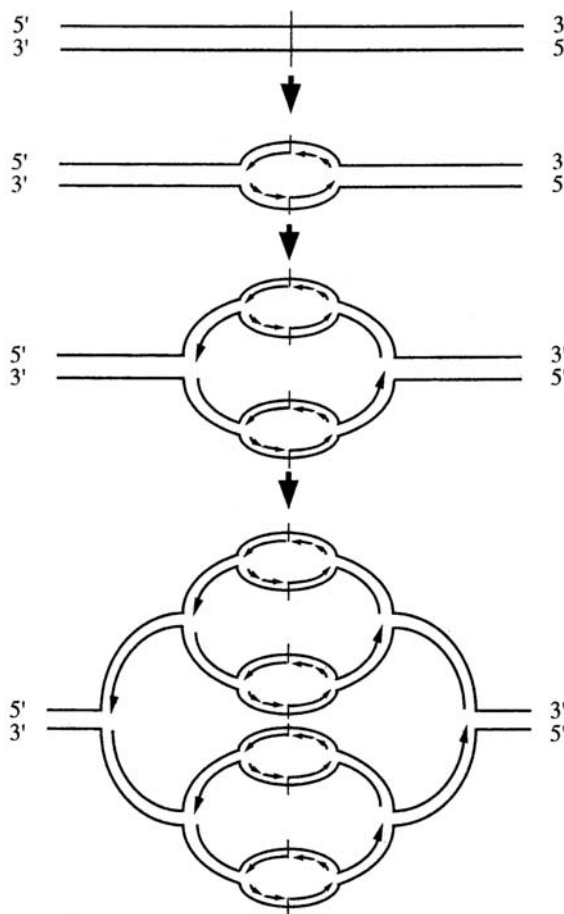
Finally, DNA amplification also means a technique by which, with the help of specific oligonucleotides and the polymerase chain reaction, the presence of a

particular DNA sequence can be determined qualitatively and quantitatively. This procedure is widely used for diagnostic purposes. Examples for specific applications of this technique are described in numerous reviews and original papers. The subject will not be dealt with any further here.

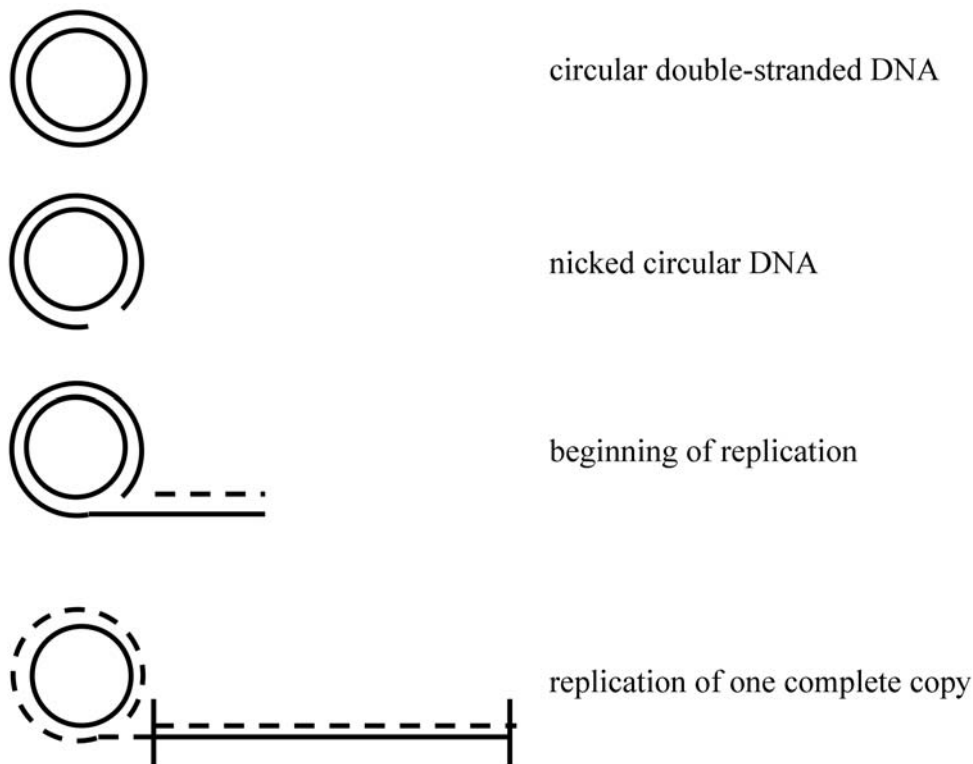
Characteristics

Among the developmentally regulated DNA amplifications, the almost complete overreplication of the genome during the formation of ►polytene chromosomes in some organisms other than mammals (e.g. in the salivary glands of *Diptera*) is the extreme case. Less extensive amplification events serve the purpose of providing particular cells with supernumerary copies of selected genes to cope with a temporary and exceptionally high demand for transcripts of these genes. One well-studied example is the amplification of the genes for ribosomal RNA in the macronucleus of *Tetrahymena* and during amphibian oogenesis, which leads to a more than thousand-fold increase in rDNA. Genes for ribosomal RNA are present in many copies in the genome in nearly all organisms. This is necessary to cope with the need for heavy transcription of the gene for a sufficient supply of ribosomal RNA in all cells. During oogenesis an even larger stock of ribosomal RNA has to be produced because transcription does not take place in fertilized eggs for several cell-cycles. Hence, in order to avoid oogenesis taking too long, genes for ribosomal RNA are heavily overreplicated. The overreplicated DNA is present extrachromosomally. The programmed amplification of the chorion genes in follicular cells of the ovary of *Drosophila* is also well known. In this case the extra copies of the gene remain side by side in the genome, like the DNA puffs of *Sciara*.

In contrast to developmentally regulated amplification processes, many aberrant cases are known that occur rarely in cells and that are noticed if they involve genes carrying selectable markers such as those conferring drug resistance. Such amplification events are frequently seen in cells of tumor patients receiving chemotherapy, where amplification gives cells a growth advantage and selects against the effects of the drug. Also relevant in this context are amplifications which are again frequently observed in tumor cells and which involve genes whose products play a role in growth control such as cyclins or transcription factors like c-myc. Normal human cells grown in culture are incapable of amplification (3). This contrasts with the majority of immortal cell lines and tumor cells. There are two types of cytological manifestation of amplified DNA in such cells namely homogeneously staining regions (►HSRs) and double minute chromosomes (►DMs). The preferred



DNA Amplification. Figure 1 DNA amplification by overreplication (regulated or unscheduled) from the same origin of replication. This gives rise to structures resembling onionskins.



DNA Amplification. Figure 2 Rolling circle mechanism of DNA amplification. Replication starts from circular double stranded DNA. This mechanism is used in the replication of the DNA of some bacteriophages and in the amplification of DNA sequences coding for ribosomal RNA in amphibia.

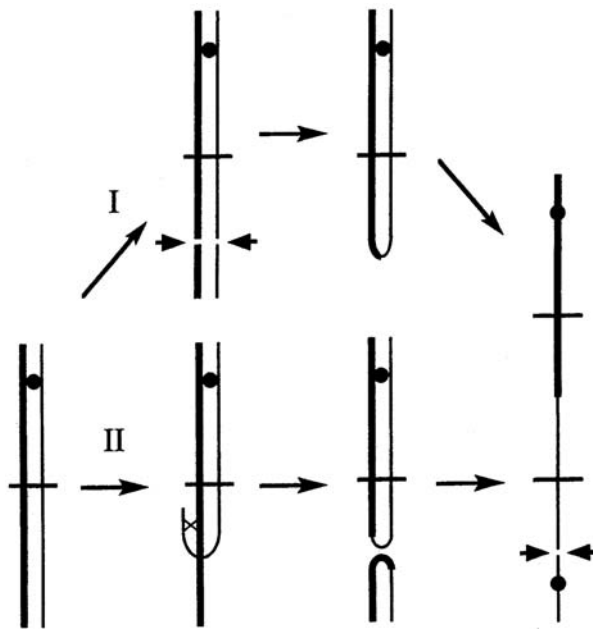
cytological characteristic of amplified DNA is in part cell type specific and in part depends on whether cells are grown in culture or isolated from solid tumors. Cultured mouse cells, for instance, display mostly DMs, rat and hamster cells mostly HSRs while human cells have either HSRs or DMs. Very rarely are the two forms seen side by side within one cell. It is interesting that human tumor cell lines in culture display mainly HSRs, while the presence of amplified DNA in tumor cells *in vivo* is manifested by the occurrence of DMs. In some cases one form can be converted into the other, i.e. when cell lines are established from tumors exhibiting amplified DNA in the form of DMs, these disappear and are converted into HSRs and *vice versa*. In a well established experimental system, the Chinese hamster ovary cells selected for ▶methotrexate resistance and carrying the amplified gene coding for dihydrofolate reductase in the form of HSRs, it was found that even the earliest detectable HSRs already contain several extra copies of the gene plus sequences surrounding it (4). This is of relevance when mechanisms of DNA amplification are considered.

Amplification events involving short tandemly repeated sequences can include trinucleotide repeats,

minisatellites and microsatellites. Amplification of trinucleotide motifs is the cause of a variety of human diseases (repeat expansion diseases) and may result in long chains of the same amino acid (often glutamine, more rarely other amino acids like alanine).

Mechanisms of DNA Amplification

The mechanisms of DNA amplification vary between different amplification events. The simplest way to amplify specific parts of the genome is a violation of the rule that the total genome is replicated once and only once during one cell-cycle. This mechanism could be regulated or accidental. Regulated rereplication appears to be the mechanism by which polytene chromosomes are formed in flies. Here large parts of the genome are replicated many times during one S phase in particular cells. Some sequences however are saved from this process and remain unamplified (or amplified to a significantly lower degree than the rest of the genome). Hence this process does not lead to polyploidization. Regulated rereplication from a specific origin of DNA replication was also proven in case of the amplification of chorion genes during oogenesis in *Drosophila*. In this case this process leads to the



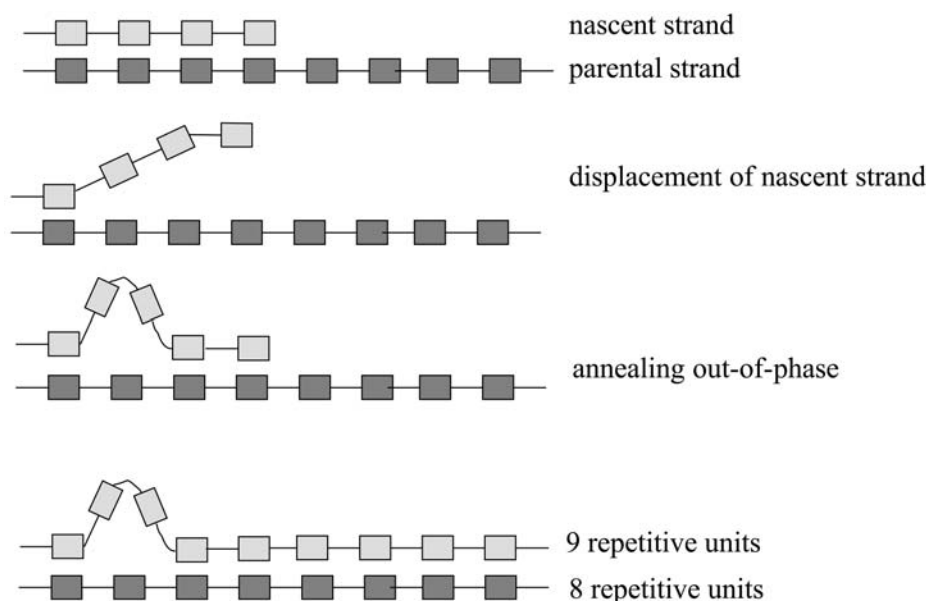
DNA Amplification. Figure 3 Initiation of B-F-B (breakage-fusion-bridge) cycles occurs by chromosome breakage, loss of telomeres (as in I) or by unequal sister chromatid exchange, an example of which is shown in II. In both cases dicentric chromosomes arise which are again broken in the next mitosis. The two lines of different thickness represent sister chromatids, the filled circle is the centromere and the horizontal bar the gene that will be amplified.

appearance of structures resembling onionskins, hence this mechanism was named ►onionskin mechanism (Fig. 1). Interestingly, in another case of developmentally regulated amplification, that of the genes coding for ribosomal RNA during oogenesis in frogs, a different mechanism is used. Here rereplication is carried out by the so-called ►rolling circle mechanism (Fig. 2). Since the rolling circle mechanism to be started requires that the genes are present in circular form, it is likely that this amplification process starts with a homologous recombination process in which one or a few of the repeating units of rRNA genes present in the genome are excised. Such excision of parts of repetitive genomic sequences are well known. Unregulated DNA amplification events which may accidentally take place in some cells and which are manifest by the emergence of DMs or HSRs occur by a different mechanism. The two forms can interchange, but the question is which form, DMs or HSRs, emanates first. During the selection of methotrexate-resistant hamster cells with amplified genes for dihydrofolate reductase HSRs were observed early. This

was interpreted to indicate that breakage-fusion-bridge (B-F-B) cycles rather than overreplication are responsible for their formation. B-F-B cycles could be initiated by double strand breaks (Fig. 3). Erosion of telomeres leads to chromosome ends which resemble double strand breaks, which may in addition arise elsewhere in the chromosome when replication forks are stalled for prolonged periods of time or the replication apparatus did not remain at the stalled fork until the cause for the replication block is eliminated. Such double strand breaks may be repaired by fusion of two broken chromosomes. Alternatively B-F-B cycles could be started by unequal sister chromatid exchange (Fig. 3). Either case leads to ►dicentric chromosomes that are likely to be broken during the next mitosis. This leads to more broken chromosomes that are healed by fusion, again giving rise to dicentrics continuing the B-F-B cycle. Normal cells are unlikely to initiate such a reaction chain because intact check point controls would drive cells with double-strand breaks which are not properly repaired, or those with too short or lost telomeres into apoptosis. This is probably the reason why methotrexate-resistant cells arise at a frequency below the detection limit in normal human cells (3). In other words, cells have to have previously acquired mutations in control systems to allow DNA amplification, a situation that is characteristic of tumor cells. Among the different DNA amplifications thus arising, those that provide a growth advantage for the cell will be selected. The very same reactions are very likely responsible for the emergence of drug resistance in patients receiving chemotherapy.

While the mechanism leading to HSRs was proven for the formation of methotrexate resistant hamster cells, it is clear that other possibilities for the initiation of DNA amplification exist. DMs could arise by unscheduled DNA synthesis starting from normal replication origins (just as it is the case in developmentally regulated overreplication). Alternatively they may form by recombination events in which portions of a chromosome are removed in circular form (5). Such circular structures will persist in the cells if they carry a functional origin of replication. Because DMs lack centromeres, their segregation to daughter cells occurs at random with the result that one of the daughter cells may receive more DMs (and therefore has a higher grade of amplification) than the other one. DMs may fuse to larger circles and reintegrate into the genome to give rise to HSRs. It is very likely that all these possibilities occur and that different cells differ in the way DNA amplification is initiated.

Amplification of short tandemly repeated DNA sequences present in satellite DNA or in trinucleotide repeats most probably occurs by slippage of the replication apparatus during DNA replication (Fig. 4).



DNA Amplification. Figure 4 Replication slippage. During the replication of repetitive sequences (indicated by the blocks) hybridisation of nascent single stranded DNA can occur with nearby copies of the repetitive units. Continuation of the replication process then leads to an extension of the units in the replicated DNA.

Amplification of trinucleotide repeats can happen within 5' untranslated regions of genes or within genes themselves. In the first case such amplifications can interfere with the regulation of gene expression (as it may be the case in the fragile X syndrome). They may also give rise to polyglutamine repeats within the protein coded by the gene (as in the case of Huntington chorea). Such polyglutamine regions were recently found to interact with specific components of the transcription apparatus, which again may result in a deterioration of gene expression.

Clinical Relevance

Amplification of oncogenes, growth promoting transcription factors or G1 cyclins like cyclin D is often observed in tumor cells (1). The important question is whether these amplifications have a role in the initiation of tumor formation or whether they are secondary and late incidents. The evidence from *in vitro* studies in cell culture indicates that intact control systems in normal cells do not allow the survival of cells with chromosomal abnormalities as they turn up in the course of unscheduled DNA amplification. This would favor the idea that DNA amplification is an event which takes place later in the evolution of tumor cells when changes in check point controls have already occurred and cells are defective in mechanisms which control the intactness of the genome and the

karyotype. Mutational inactivation of p53 in particular appears to assist DNA amplification. Such a situation seems also to be responsible for the frequently observed emergence of tumor cells that resist chemotherapy. In this case the DNA amplification events, such as those involving dihydrofolate reductase (resistance to methotrexate) or multi drug resistance (MDR, resistance to inhibitors of mitosis) result in a decline in the effectiveness of chemotherapy. Amplifications of trinucleotide repeats can lead to severe, frequently slowly developing neurodegenerative diseases such as Huntington chorea, muscular dystrophy or the fragile X syndrome.

References

1. Schwab M (1999) Oncogene amplification in solid tumors. *Seminars in Cancer Biology* 9:319–325
2. Stark GR (1993) Regulation and mechanisms of mammalian gene amplification. *Adv Cancer Res* 53:447–491
3. Tlsty TD (1990) Normal diploid human and rodent cells lack a detectable frequency of gene amplification. *Proc Natl Acad Sci USA* 87:3132–3136
4. Trask BJ, Hamlin JL (1989) Early dihydrofolate reductase gene amplification events in CHO cells usually occur on the same chromosome arm as the original locus. *Genes Dev* 3:1913–1925
5. Wahl GM (1989) The importance of circular DNA in mammalian gene amplification. *Cancer Res* 49:1333–1340

DNA Arrays

► DNA Microarrays/DNA Arrays

DNA Binding Domains

Definition

The recognition and binding of specific DNA sequences is achieved by protein domains structured as basic domains next to helix-loop-helix or leucine zipper structures, or as zinc finger, helix-turn-helix, or less well defined motifs.

► Transcription Factors and Regulation of Gene Expression

DNA Catenanes

Definition

DNA catenanes are topoisomers of circular DNA molecules consisting of two or more circular DNA molecules, which are held together noncovalently by one DNA circle encircling the DNA strand of another like links of a chain. Catenanes are products of replication of circular DNA when Type II topoisomerases are absent. In order to segregate into daughter cells, DNA catenanes must be resolved by the reaction catalyzed by Type II topoisomerase.

► DNA Topoisomerases

DNA Chips

► DNA Microarrays

DNA Chip Technology

► Biochip Technology
 ► DNA Chip Technology
 ► Gene Chip Technology
 ► Microarray Technology

DNA Conformations

Definition

DNA conformations are different shapes of DNA double helices.

► Repeat Expansion Diseases

DNA Damaging Agents

BERND KAINA, GERHARD FRITZ

Division of Applied Toxicology, Institute of Toxicology, University of Mainz, Mainz, Germany
 kaina@uni-mainz.de

Definition

Each human diploid cell contains in the pre-replicating state 6×10^9 base pairs in the DNA, which corresponds to a total amount of 12×10^{22} nucleotides in the human body (given that the body consists of 10^{13} cells). This huge amount of DNA, packed into 46 chromosomes (diploid), is under constant attack by endogenous and exogenous reactive molecules that causes damage in DNA bases or its sugar phosphate backbone. Furthermore, some DNA damaging agents (DDA) may cause the formation of covalent bonds between DNA and proteins (DNA protein cross-links) thus affecting DNA-histone and DNA-transcription factor interactions. This has impact on DNA packing (in the case of histones), cell division, replication and transcription of DNA.

Characteristics

Most of the DNA lesions that are formed spontaneously or by the attack of endogenous and exogenous reactive molecules are repaired, i.e. removed in an error-free manner (for review see ref. 1). Some of the lesions, although they block transcription and replication, are tolerated because of the existence of tolerance mechanisms, which include low-fidelity DNA polymerases and recombination processes. The occurrence of lesion tolerance mediated by low-fidelity DNA polymerases leads to an increase in mutation frequency. Lesion tolerance, by recombination bypass, is supposed to be error-free (2), although it may also result in incorrect recombination, which will lead to chromosomal aberrations. The final outcome of ► DNA damage induced in the genome depends on the structure of the DNA lesion, the cellular repair and replication

capacity, and the signaling that a given lesion evokes. Therefore, specific lesions can be considered as pre-cytotoxic (O^6 -methylguanine, pyrimidine dimers, most bulky DNA lesions), pre-mutagenic (O^6 -methylguanine, 8-oxo-guanine, O^4 -methylthymine, (6-4) photoproducts), and pre-clastogenic (apurinic sites, double-strand breaks, CPDs). Here, the various classes of DNA damaging agents will be reviewed briefly. This review will include their action on DNA and the lesions they induce. The mechanisms of repair of these lesions will be reviewed separately in this monograph.

Classification of DNA Damaging Agents (DDA)

In Table 1, several classes of DDA are summarized. The most prominent source for “spontaneous” DNA base modifications are reactive oxygen species (ROS) that are formed inside the cell during normal metabolism (respiratory function), or outside of the cells, e.g. during inflammatory processes. Another source of endogenous damage results from (cellular) methylating species. Thus, the intracellular methyl group donor S-adenosylmethionine has been shown to

be able to methylate DNA (3). Another important cause of “spontaneous” DNA damage is the inherent instability of DNA that results in $\sim 10^4$ base depurinations/cell/day, base deaminations and DNA strand breaks. The most important exogenous DDA are radiation and chemical mutagens (Table 1).

Ionizing Radiation

Humans are exposed to ionizing radiation (IR) every day due to cosmic radiation, terrestrial γ -radiation e.g. by K-40, Uran-238 (external irradiation) and due to ingestion and inhalation (internal irradiation). The natural radiation load is (2.4 mSv/year from cosmic and terrestrial sources and in the range of 4–6 mSv/year for man-made sources. Several sub-populations, such as nuclear plant workers, flight personal and smokers, are exposed to additional IR doses of 6 mSv/year. Tobacco smoke contains natural radioisotopes (Pb210 and Po210) causing an effective dose of 1.2 μ Sv/cigarette and an accumulative load of 8.8 mSv/year (lung dose: 106 mSv/year). This is important to note in view of the discussion of human exposure due to emission from nuclear plants, which is low in

D

DNA Damaging Agents. Table 1 DNA damaging agents

Type of agent	Example	Type of reaction with DNA
Radiation		
Ionizing radiation	X-ray, K-40, Pb210, Po210	a) DNA single and double strand breaks b) Oxidative DNA damage, e.g. Formamidopyrimidine Guanine 8-Hydroxyguanine 5-Hydroxyhydantoin Thymine glycol Cyclobutane pyrimidine dimer (CPD) 6-4-photoproduct Oxidative DNA damage
Ultraviolet light	UV-A, UV-B, UV-C light	
Chemical mutagens		
Aromatic hydrocarbons (PAH)	Benzo(a)pyren 7,12-Dimethylbenzanthrazen	N2 of guanine N2 of guanine
Halogenated hydrocarbons	Vinylchloride	N1 of adenine
Aromatic amines	N-acetyl-aminofluorene (AAF)	C8 and N2 of guanine
Alkylating agents	Dimethylnitrosamine (DMNA) Alkyl nitrosamides Mustards Mono- and bifunctional)	N1, N3, N7 of adenine; N3, N7, O6 of guanine and phosphate ester Inter- and intrastrand cross-links
Natural compounds	Aflatoxine Saflrol	Adduct at O6 and N7 of guanine Adduct at N1 and N3 of adenine Adduct at O6 of guanine

comparison. IR causes the formation of DNA single-strand breaks (SSB), double-strand breaks (DSB), oxidation of bases, DNA protein cross-links and local denaturation of DNA due to base damage. Since SSB are repaired quickly, they are usually not harmful for the cell. More dangerous are DSBs (induction ratio 25:1 for SSB and DSB) which, unless repaired, causes the generation of chromosomal breaks, sister-chromatid exchanges, gene inactivation and cell death (4). Other critical lesions are purine and pyrimidine modifications (Fig. 1a). A replication-blocking lesion induced by IR is thymine glycol that might account for replication-dependent cytotoxic effects. The most critical mutagenic lesion after IR exposure is 8-oxo-guanine (the isomeric form is 8-hydroxyguanine). 8-Oxo-G mispairs with adenine giving rise to GC → TA transversion ► [mutations](#). 8-Oxo-G is also formed spontaneously at detectable levels (estimations range from 100–500 lesions/cell/day). It is repaired by the enzyme 8-oxoguanine-DNA glycosylase, which facilitates the repair of the lesion via the base excision repair (BER) pathway. Whether gene mutations and malignant transformation induced by IR is due to non-repaired base damage or DSBs, which form chromosomal aberrations, is still a matter of debate.

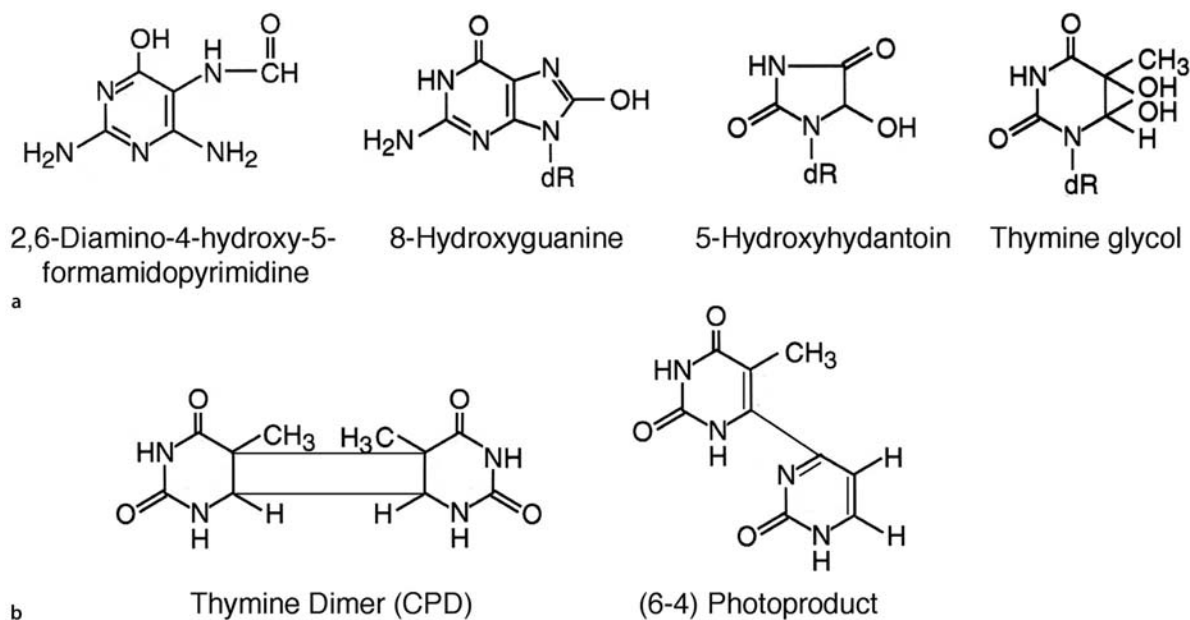
Ultraviolet Light (UV)

UV-light is classified as UV-C (220–280 nm), UV-B (280–320 nm) and UV-A (320–400 nm) according to differences in wavelength. The atmosphere absorbs UV-C, whereas UV-B and UV-A reaches the earth and

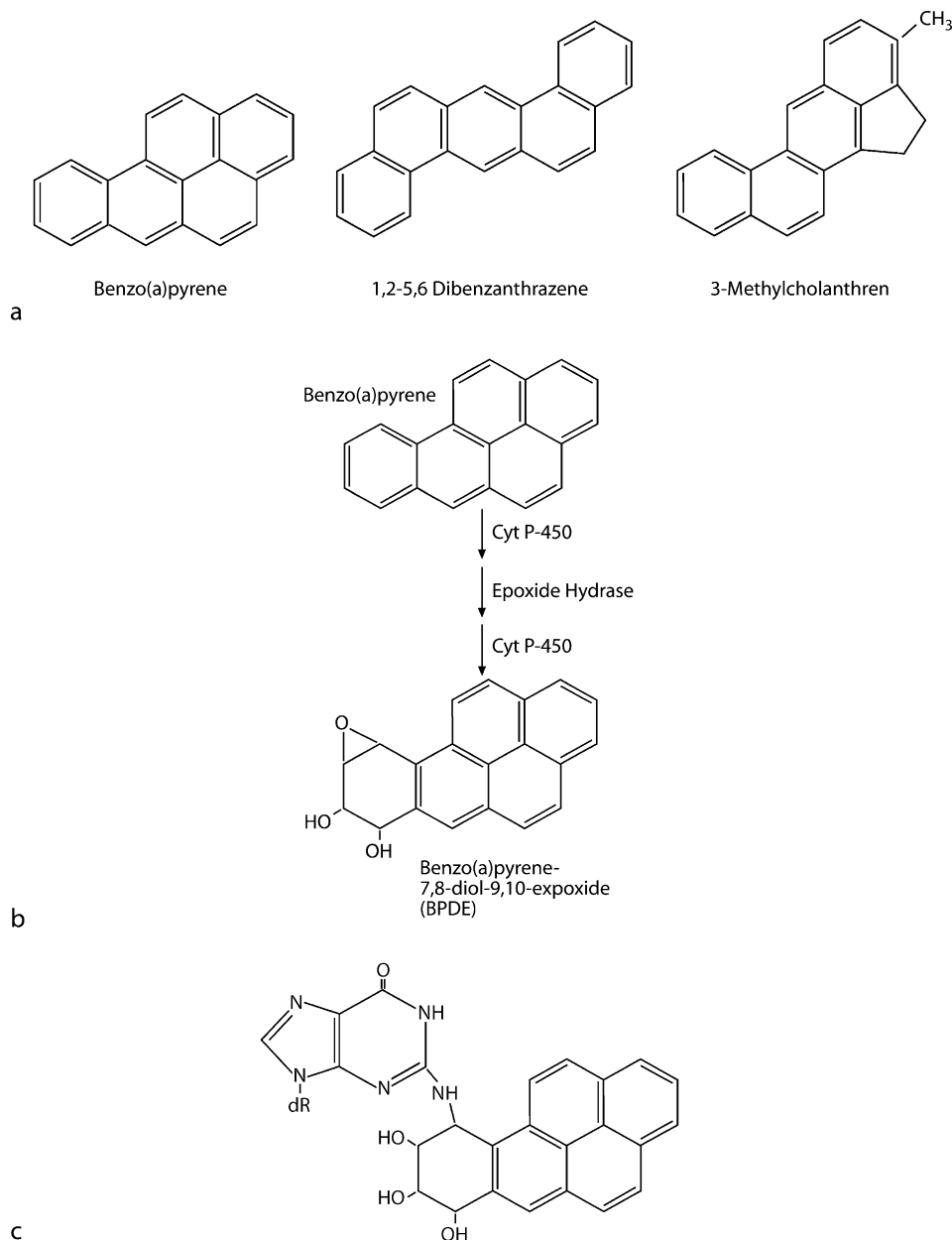
accounts for 10 and 90% of the natural UV exposure, respectively. UV-A light is considered to be responsible for sun-based effects on skin. It causes, upon chronic exposure, aging of skin and skin ► [cancer](#) (among them malignant melanomas). UV-A and UV-B irradiation induces cyclobutane pyrimidine dimers (CPDs) and (6-4) photoproducts (6-4PPs) (Fig. 1B). CPDs are induced at a three times higher frequency than 6-4PPs. UV-light also induces oxidative stress, which is true notably for long-wave UV-light and, therefore, causes the formation of 8-oxo-G and DNA-protein cross-links. There is evidence that CPDs and 6-4PPs account for most of the harmful effects of UV-light, including cell death, mutation and cancer formation. The lesions cause severe blockage in DNA replication and, thereby, provoke cell death via apoptosis (5). They also cause, in cells that survived exposure, the formation of chromosomal aberrations and gene mutations. The latter is due to error-prone (SOS) polymerases that allow cells to tolerate pre-toxic DNA lesions at the cost of mutation accumulation. The importance of ► [DNA repair](#) for defense against UV-light is strikingly illustrated by various DNA repair deficient disorders that lead to hypersensitivity to sunlight and also various metabolic defects (1). These hereditary disorders comprise xeroderma pigmentosum, Cockayne's syndrome, trichothiodystrophy, and xeroderma pigmentosum variant.

Chemical DDA

Some chemical DDA (DNA damaging agents) are listed in Table 1. Most of the agents are powerful



DNA Damaging Agents. Figure 1 Base damage in DNA induced by (a) ionizing radiation and (b) ultraviolet light.



DNA Damaging Agents. Figure 2 Polycyclic aromatic hydrocarbons. (a) Structure of various polycyclic aromatic hydrocarbons, (b) its metabolism and (c) site of attack in DNA of benzo(a)pyrene forming N²-guanine adduct.

mutagens and carcinogens. Many chemical DDA are stable. They do not react with DNA unless they are activated by metabolic conversion (e.g. many polycyclic hydrocarbons and nitrosamines). Metabolic activation occurs notably by cytochrome P450, converting the substances into reactive, ultimate carcinogens and mutagens. Various other chemical DDA are inherently active or undergo spontaneous hydrolysis

(e.g. nitrosamides) thereby forming reactive metabolites without the need for enzymatic metabolization.

Aromatic Hydrocarbons (PAH)

PAHs such as benzo(a)pyrene, dibenzanthracene, dimethylbenzanthracene, and methylcholanthrene are formed during incomplete combustion of coal, oil, gas, or tobacco. In order to become carcinogenic, they

need metabolic conversion, which is accomplished by cytochrome P450 mono-oxygenases. The structure of various PAHs is illustrated in Fig. 2a. As an paradigmatic example, the metabolic conversion and reaction with DNA of benzo(a)pyrene, a major carcinogen in cigarette smoke, is outlined in Fig. 2b. Upon metabolism by Cyp1A1, a diol epoxide is formed that reacts with DNA at the exocyclic amino group of deoxyguanosine, forming an N²-BP-adduct (Fig. 2C). This adduct provokes severe steric alteration of DNA that impairs DNA-dependent metabolic functions, including DNA replication and transcription. These adducts can be removed from DNA by nucleotide excision repair (NER), which is the main cellular defense mechanism against these types of lesions. A more distal defense mechanism is mediated by detoxifying (phase II) enzymes, which includes epoxide hydrolases and glutathione transferases.

Halogenated Hydrocarbons (HHC)

HHC such as chloroform, halothan, CCl₄, trichlorethan or trichlorethylen are narcotic and highly cytotoxic for liver and/or kidney, because of the formation of radicals that lead to lipid peroxidation. Some of the HHC's, such as vinylchloride, dibromoethane or epichlorhydrine, are carcinogenic. Vinylchloride, which is extensively used for PVC production, is enzymatically converted into an epoxide, which in turn leads to a non-enzymatic formation of chloroacetaldehyde. Both metabolites attack DNA at N1 position of adenosine. Rearrangements of the molecule gives finally rise to an ethano-adenosine-adduct (Fig. 3).

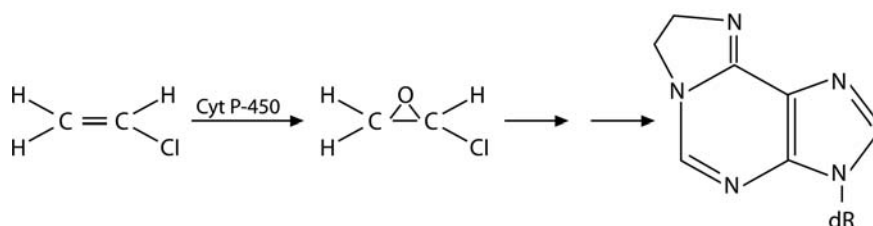
Aromatic Amines (AA)

AA such as benzidine, naphtylamine, toluidine, aminoanthracene or acetylaminofluorene (AAF) (Fig. 4a) are widely distributed industrial compounds used e.g. for the synthesis of colored dyes, pestizides and manufacturing of plastics. They can also be formed upon high heating of food. Upon metabolic activation (i.e. N-hydroxylation) by cytochrome P450 1A2 a reactive metabolite is generated which is responsible

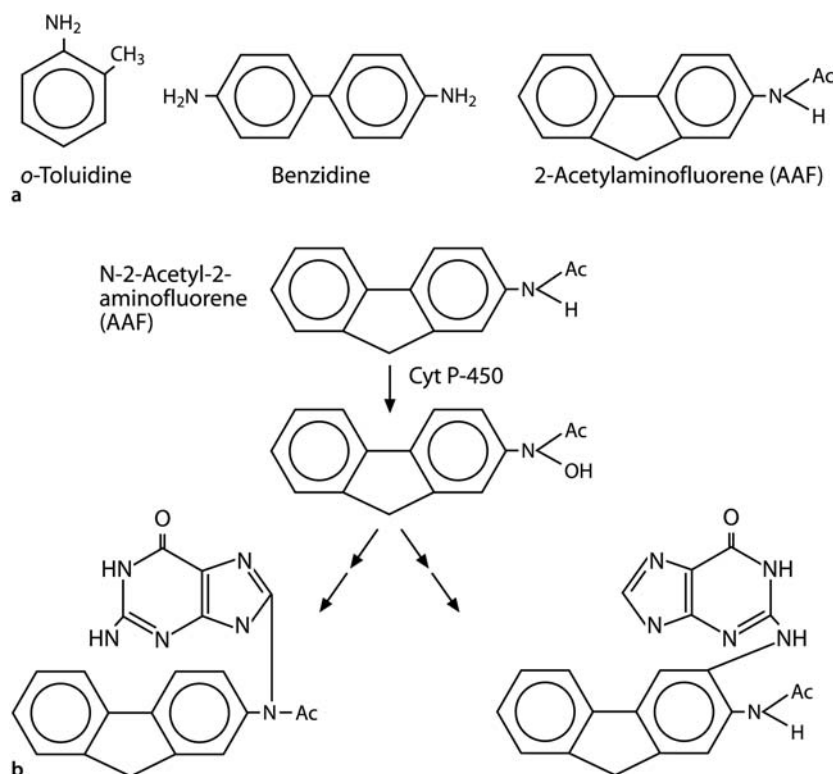
for both cytotoxic effects (i.e. formation of methemoglobin) and DNA adduct formation which occurs mostly on position C8 and N2 of guanine (Fig. 4b). One main detoxification reaction consists of N-acetylation catalyzed by polymorphic N-acetyltransferases. In human, the main target organ for AA-induced cancer is the bladder.

Alkylating Agents: N-Nitrosamines and N-Nitrosamides

Perhaps the most powerful DNA reactive agents and, at the same time, mutagens, clastogens and carcinogens are alkylating agents. Most relevant, both as environmental carcinogens and endogenously formed alkylating agents, are the N-nitrosamines (Table 1 and Fig. 5a). N-nitrosamines, e.g. dimethylnitrosamine (DMNA), are stable compounds that need metabolic activation in order to turn into the ultimate reactive agents, which are highly reactive electrophilic species (Fig. 5b). In contrast, N-nitrosamides, e.g. N-methyl-N-nitrosourea, are quite unstable, undergoing spontaneous hydrolysis in aqueous solution that yields again electrophils (e.g. carbenium ions). These electrophilic species react with DNA at 12 different sites (Table 2). The extent of the reaction with these DNA alkylation sites is dependent on the electrophilicity of the agents (i.e. their S_N1 or S_N2 character). The major alkylation site for all S_N1 and S_N2 agents is N7 of guanine, followed by the phosphate in the DNA backbone and the N3 of adenine. Irrespective of this, the most biologically important attack occurs at a minor alkylation site, the O⁶ of guanine. O⁶-alkylguanine encompasses 0.3–8% of total DNA alkylation. The mutagenic and carcinogenic potency of simple alkylating agents correlates with their O-alkylating potency. Thus, methyl methanesulfonate (MMS) is a very weak carcinogen and induces point mutations at low frequency, whereas N-methyl-N-nitrosourea and, even better, N-ethyl-N-nitrosourea are highly carcinogenic and induce point mutations at high frequency. The underlying reason for this has been clarified: O⁶-methylguanine and O⁶-ethylguanine mispair with



DNA Damaging Agents. Figure 3 Metabolic activation and mode of action of vinylchloride forming an etheno adenosine adduct.



DNA Damaging Agents. Figure 4 Aromatic amines. (a) Structure and (b) metabolism of AAF inducing guanine adducts.

thymine, which gives rise to GC \rightarrow AT transition mutations. O⁶-alkylguanine is repaired by the repair protein MGMT (1).

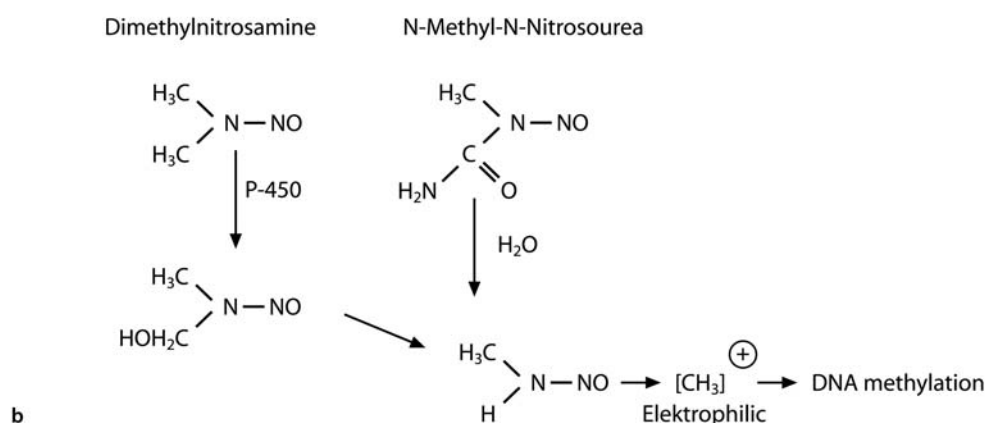
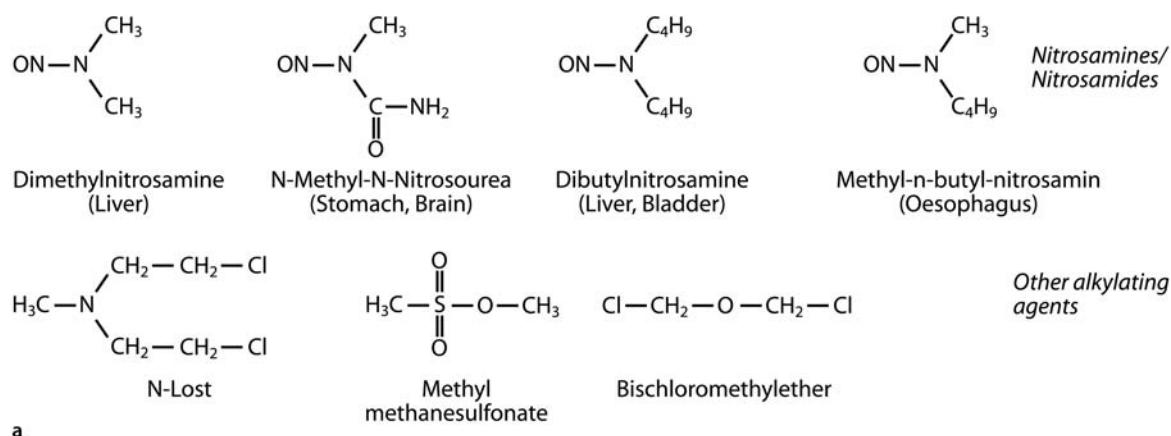
Simple alkylating agents exert a marked organotropy in tumor formation. Thus, ENU is extremely powerful in inducing brain tumors in rats; DMNA induces liver tumors and MNU thymic lymphomas. Brain tumor formation was shown to be related to the lack of MGMT expression at various stages of development of the rat brain, whereas liver tumor formation upon chronic exposure of rats with DMNA is likely due to cytochrome P450 metabolism of this agent in the liver.

Simple alkylating agents, notably the N-nitrosamines, are found widely in the environment (6). They are also formed endogenously, e.g. in the stomach from nitrite and secondary amines and also, presumably, by microbial metabolism in the stomach and intestine. It should also be noted that tobacco smoke contains high amounts of dimethylnitrosamine and at least 5 tobacco smoke-specific carcinogens, which are degradation products of nicotine (see below). Nitrosamines can also be formed during the preparation of food, e.g. roasted meat and alcoholic beverages such as Whisky.

Various alkylating drugs exhibit crosslinking activity, because they are either bifunctional agents or monofunctional mutagens forming adducts that undergo secondary rearrangement in the DNA. Paradigmatic examples of bifunctional alkylating agents are N-Lost and S-Lost compounds (Fig. 5a). N-Lost (LOST = Lommel and Steinkopf, according to the German chemists who discovered compounds to be used as chemical weapons in world war I) alkylates guanines, without metabolic activation, at N7 of guanine, forming an interstrand N7-G-N7-G crosslink. These crosslinks are considered to be highly cytotoxic. This principle has been utilized in tumor therapy: a derivative of N-Lost is cyclophosphamide, which is most often applied in cancer treatment. Monofunctional crosslinking agents are chloroethylating nitrosoureas (BCNU, ACNU, CCNU) that are used in the therapy of glioblastomas, malignant melanomas and various hematopoietic tumors.

DDA in Tobacco Smoke

Tobacco smoke contains more than 3200 different compounds. These compounds include 47 well-known



DNA Damaging Agents. Figure 5 Alkylating agents. (a) Structure of different mono- and bifunctional alkylating agents and (b) conversion into electrophilic species.

DNA Damaging Agents. Table 2 Sites of attack of alkylating agents

DNA adduct	Dimethylnitrosamine N-methyl-N-nitrosourea1, 2-dimethylhydrazine	Methyl methanesulfonate	Diethylnitrosamine N-ethyl-N-nitrosourea
N1-Alkyladenine	0.7	1.2	0.3
N3-Alkyladenine	8.0	11.0	4.0
N7-Alkyladenine	1.5	1.9	0.4
N3-Alkylguanine	0.8	0.7	0.6
N7-Alkylguanine	68.0	83.0	12.0
O6-Alkylguanine	7.5	0.3	8.0
N3-Alkylcytosine	0.5		0.2
O2-Alkylcytosine	0.1		3.0
N3-Alkylthymine	0.3		0.8
O2-Alkylthymine	0.1		7.0
O4-Alkylthymine	0.7		1.0–4.0
Alkylphosphate	12.0	1.0	53.0

carcinogens. The major DDAs in tobacco smoke are benzo(a)pyrene, benzanthrazene, dimethylnitrosamine, the tobacco-smoke-specific nitrosamine N-nitroso-nornicotine, cadmium, nickel, anilin, acrolein and formaldehyde. Tobacco smoke also contains ROS that cause oxidative DNA damage. Overall, tobacco smoke is a main source of exogenous man-made DNA damage. Because of the diversity of the DDA it contains, it is difficult to assess the component that is most dangerous and that is also responsible for tobacco smoke-induced tumor formation, such as lung cancer.

DDA in the Natural Environment

Microorganisms, fungi, and plants produce various highly reactive DNA damaging agents. They are considered as a serious hazard for man, even contributing to some types of cancer. Most important are aflatoxins such as aflatoxin B₁ (AFT-B₁). It is a mycotoxin produced by different aspergillus species. Food contaminated with aspergillus (rice, soy, peanuts) contains high amounts of aflatoxins. AFT-B₁ is a powerful hepatocarcinogen. Chronic exposure causes liver cirrhosis. AFT-B₁ is metabolically activated in the liver by Cyp P450, forming a reactive epoxide that reacts with DNA preferably at the N7 position of guanine.

Another DNA damaging agent that causes liver cancer in animals is saffrol. The compound was found in pepper, anise, celery, and the tree *Sassafras albidum*. There are various other compounds produced in plants that have been described to be mutagenic and carcinogenic. These compounds include cycasin, arecolin, coumarin, limonene, aristolochic acid and reserpine. For some of them the DNA reaction products still need to be defined.

Some DDAs produced by microorganisms are in use as well-established anticancer drugs. Examples are mitomycin D and bleomycin. Mitomycin C causes DNA interstrand crosslinks that are cytotoxic. Bleomycin is an example of a chemical that induces radiomimetic effects, notably DNA strand breaks.

Heavy Metals

Various heavy metal ions, such as Ni, Cd, and As, are mutagenic and carcinogenic. They cause DNA strand breaks, which could be due to direct attack of DNA or intracellular radical formation. There are data available indicating that heavy metals inhibit specific DNA repair enzymes of the BER and NER pathway and, therefore, cause accumulation of DNA damage.

Topoisomerase Inhibitors

Various agents that play an important role in tumor therapy cause DNA damage indirectly, for example by

inhibition of topoisomerase I or II. Topo I and II are DNA helicases that cleave DNA, thereby facilitating their unwinding. Topo I causes single-strand breaks, topo II double-strand breaks. The breaks they induce in DNA are transient and become religated by a second catalytic function of the topoisomerases themselves. Camptothecin and etoposide inhibit this religation function of topo I and II, respectively, thereby causing an accumulation of DNA strand breaks due to the inability of the enzyme to reseal the formed breaks. Topoisomerase inhibitors are examples of agents that cause DNA damage by inhibiting cellular enzymatic functions involved in DNA metabolism.

► Base Excision Repair

References

1. Christmann M, Tomicic-Christmann M, Roos W et al (2003) Mechanisms of DNA repair – an update. *Toxicology* 193:3–34
2. Kaina B (2004) Mechanisms and consequences of methylating agent-induced SCEs and chromosomal aberrations: a long road traveled and still a far way to go. *Cytogen Genome Res* 104:77–86
3. Rydberg B, Lindahl T (1982) nonenzymatic methylation of DNA by the intracellular methyl group donor S-adenosyl-L-methionine is a potentially mutagenic reaction. *EMBO J* 1:211–216
4. Friedberg EC, Walker GC, Siede W (1995) *DNA Repair*, ASM Press, Washington
5. Kaina B (2003) DNA damage-triggered apoptosis: critical role of DNA repair, replication, double-strand breaks and signaling. *Biochem Pharmacol* 66:1547–1554
6. Koehl W, Eisenbrand G (1999) N-nitroso compounds. In: Marquardt H, Schäfer SG, McClellan RO (eds) *Toxicology*, Academic Press, pp 743–753

DNA Double-Strand Break Repair

► Double-Strand Break Repair

DNA Gyrase

Definition

DNA gyrase is a bacterial Type IIA topoisomerase that introduces negative supercoiling into circular DNA. The reaction mechanism involves the wrapping of DNA around the gyrase molecule in a positive superhelical turn, which is then converted into a negative one by strand passage. The negative supercoiling of bacterial DNA produced by gyrase serves to monitor DNA integrity, as any single or double strand DNA break would result in relaxation of supercoiling.

► [DNA Topoisomerases](#)

DNA Helicases

NAYEF A. MAZLOUM¹, NATHAN A. ELLIS²

¹Weill College of Medicine, Cornell University, NY, USA

²Department of Medicine, University of Chicago, Chicago, IL, USA

nellis@medicine.bsd.uchicago.edu

Definition

The unwinding of double-stranded DNA is an important cellular function essential to a variety of nucleic acid transactions such as ► [DNA replication](#), ► [DNA repair](#), ► [DNA recombination](#) and mRNA ► [transcription](#). Unwinding can be catalyzed by a group of ubiquitous enzymes referred to as DNA helicases (reviewed in 1, 2, 3, 4, 5, 6, 7). DNA helicases are molecular motors that disrupt the hydrogen bonds holding together the two strands of the DNA duplex to produce single-stranded DNA. The reaction is driven forward by the hydrolysis of nucleoside 5'-triphosphate (most often ATP), owing to an intrinsic ► [NTPase](#) activity. Unwinding provides access to the single-stranded DNA that other proteins require to carry out their functions (1, 2, 3, 4, 5, 6, 7). Another class of structurally similar enzymes, the RNA helicases, bind and unwind RNA as substrate. A small number of enzymes can unwind both RNA and DNA (6).

Characteristics

The DNA helicases have been studied at the biochemical, molecular, cellular and biophysical levels. Classification as a DNA helicase requires the biochemical demonstration of unwinding activity (2). However,

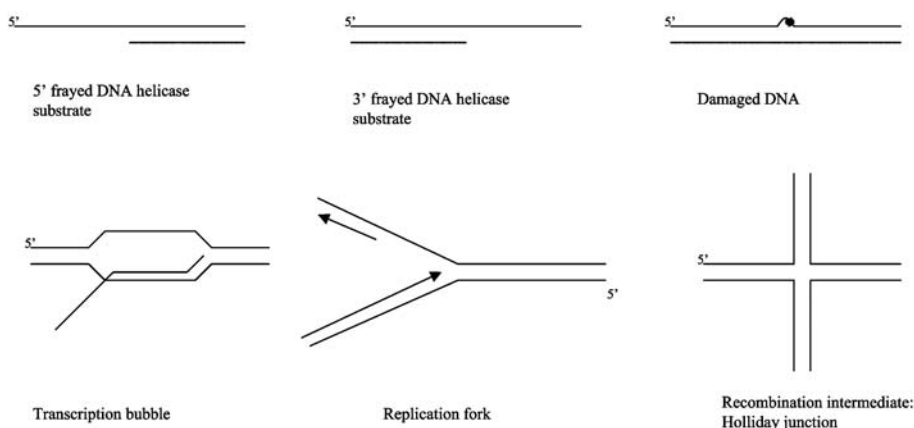
large numbers of similar molecules have been identified by bioinformatic criteria as putative helicases or, at the very least, putative molecular motors whose activities and functions require further characterization. For example, by informatic analyses it is estimated that 2% of *Saccharomyces cerevisiae* genes are helicase-related proteins (4).

Biochemical Characteristics

The biochemical activities shared by DNA helicases include DNA binding, specificity for the NTP or dNTP bound and hydrolyzed, DNA-dependent or DNA-stimulated NTP (or dNTP) hydrolysis and NTP (or dNTP) hydrolysis-dependent unwinding of double-stranded DNA. Mg²⁺ is invariably a cofactor required in these reactions. Helicases are characterized based on their biochemical properties such as substrate specificity, affinity for single-stranded as compared to double-stranded DNA, the directionality of the enzyme's movement along single-stranded DNA, i.e., movement in a 5' to 3' direction *versus* in a 3' to 5' direction, processivity, efficiency, ► [oligomerization](#) and other proteins with which they interact (1, 2, 3, 4, 5, 6, 7).

The substrate specificity of a DNA helicase is defined by the kinds of nucleic acids they unwind (Fig. 1). Substrates are usually formed experimentally by annealing a labeled oligonucleotide with a complementary oligonucleotide(s) or purified viral or plasmid DNAs. Experimental substrates are constructed to resemble substrates that might be encountered in the cell, such as a frayed broken DNA end, a ► [telomere](#), a replication fork, a recombination intermediate, etc. Most DNA helicases bind efficiently to and translocate along a single strand of DNA. Some can unwind a DNA-RNA hybrid, but most often they do so by translocating along the DNA strand and displacing the complementary RNA strand as they would the complementary DNA strand (6).

The processivity of a given helicase is defined as the probability that the enzyme will move forward one step from a distinct position along the nucleic acid substrate divided by the probability that the enzyme will dissociate from that position. The efficiency is defined by the number of nucleotides unwound per single ATP molecule hydrolyzed by the helicase (also known as the step size). Both processivity and efficiency can be affected by other proteins with which a helicase might interact (7). For example, the hexameric helicase DnaB of *Escherichia coli*, which unwinds DNA ahead of the replication fork, interacts with ► [DNA polymerase III](#); this interaction enhances its processivity and efficiency >10 fold. DNA helicases are typically present as part of a specific macromolecular complex (es), where activity is coupled to other nucleic acid transactions (2) (Table 1).



DNA Helicases. Figure 1 Schematic drawing of selected cellular DNA helicase substrates. DNA helicases often require some single-stranded DNA to begin unwinding. For example, single-stranded DNA can be exposed at the end of DNA duplex by an exonuclease. Cyclopurine dimers generated by UV irradiation are recognized by XPC protein or RNA polymerase. These enzymes expose single-stranded DNA and recruit the TFIIH helicase to the lesion-containing site. Similarly, TFIIH is recruited to the promoter regions of genes by other transcription factors that bind DNA and presumably expose single-stranded DNA by distorting the DNA helix. Replicative DNA helicases that unwind DNA helix ahead of the replication fork are recruited to the origin of replication by specific replication proteins. Structure-specific DNA helicases recognize and resolve recombination intermediates such as Holliday junctions. These proteins stabilize a form of the Holliday junction in which the four strands entering the junction are unbasepaired.

Bioinformatic Characteristics

Sequence homology analyses of many DNA and RNA helicases from a wide spectrum of prokaryotic and eukaryotic organisms and viruses have identified the presence of seven highly conserved motifs (I, Ia, II, III, IV, V, and VI), which led to the classification of many known and putative helicases into two superfamilies—superfamily 1 (SF1) and superfamily 2 (SF2). From the amount of sequence homology between the motifs, the predicted secondary structures of SF1 and SF2 members are similar (3, 4). Because motifs I and II are homologous to Walker boxes A and B, they are believed to take part in NTP and Mg^{2+} binding. These two motifs are shared among many ATPases as well. The remaining motifs participate in binding the substrate and in coordinating the enzyme's movements during catalysis (6).

Superfamily 3 (SF3) consists of a smaller group of proteins whose members are found in only RNA and DNA viruses, and they have only three conserved motifs (A, B, and C). Only one member of this family—SV40 T antigen—has been shown to possess helicase activity. Family 4 members, also small in number, are present in bacteria and **bacteriophages**, share five conserved motifs (1, 1a, 2, 3, and 4) and are presumed to act in DNA replication, because the enzymes tested interact with **DNA primases** (e.g., *E. coli* DnaB is in family 4) (3, 4).

Other helicases that do not belong to any of these families include the hexameric Rho transcription factor and the **MCM** (minichromosome maintenance) protein complex (3, 4). The MCM complex belongs to the AAA family of ATPases (ATPases associated with various cellular activities, such as DNA replication, protein folding and degradation, etc.). This family of ATPases carries out chaperone-like functions to assist in the assembly and disassembly or the efficient operation of various protein complexes. Members of this family share a conserved region of about 220 amino acids that contains an ATP-binding site. It is thought that AAA domains act as ATP-dependent protein clamps. Along with the conserved N-terminal Walker motifs A and B, they contain another highly conserved region in the central part of the ATPase domain. The utilization of the AAA domain in DNA unwinding activity is not currently understood.

The presence of the seven SF1 or SF2 helicase motifs in a protein does not guarantee the presence of helicase activity *in vitro*. An example is eukaryotic SWI2/SNF2 that has a DNA-dependent ATPase activity but no helicase activity. SWI2/SNF2 is part of a large protein complex that is thought to re-model chromatin as part of its function in transcriptional regulation, DNA repair and recombination (6). The SWI2/SNF2 protein appears to act as a molecular motor that changes the disposition of histones on the DNA.

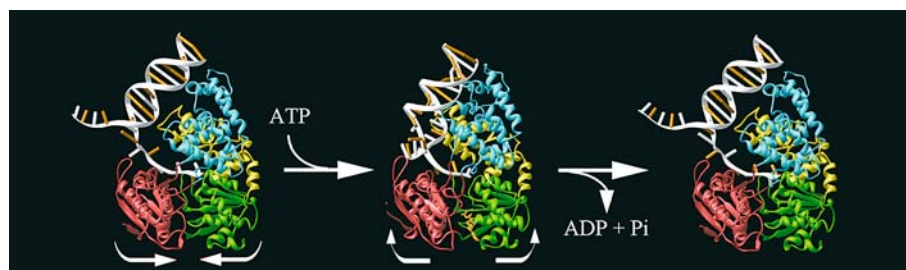
DNA Helicases. Table 1 Selected DNA helicases that participate in various essential functions: macromolecular complexes in which their activities are coupled to other cellular activities

Function	Helicase	Associated Proteins ^a	Source
Replication	DnaB	DNA polymerase III SSB DnaG (primase)	<i>E. coli</i>
	MCM 2, 3, 4, 5, 6, 7 complex	DNA polymerase α CDC45 ORC	Eukaryotes
DNA Repair	UvrD	DNA polymerase I SSB UvrA	<i>E. coli</i>
	XPB XPD	TFIIH	Eukaryotes
mRNA transcription	RNA polymerase core ^b (α_1 , α_2 , β_1 , β_2)	σ factor	<i>E. coli</i>
	XPB, XPD	TFIIH	Eukaryotes
Recombination	RecBCD ^c	RecA SSB	<i>E. coli</i>
	BLM	TOPIII RAD51 RPA MLH1 FANC proteins	Eukaryotes
	WRN	DNA polymerase δ RPA Ku70/Ku80 TRF2	Eukaryotes

^a Abbreviations for the associated proteins are as follows: SSB, single-stranded DNA binding protein; CDC45, an essential protein required for the initiation of DNA replication; ORC, origin recognition complex, consisting of six subunits ORC (1-6), that recognize the origin of replication and load MCM's onto the origin; UvrA, a base excision repair protein, functioning in the recognition of DNA damage; TFIIH, a transcription factor, consisting of six core peptide chains XPB, XPD, p44, p62, p52, and p34 and three interacting proteins Mat1, cdk7, and cyclin H; σ factor, a transcription factor, that binds to the RNA polymerase core enzyme and enhances promoter specificity; RecA, an essential prokaryotic DNA recombination gene, that facilitates strand invasion during homologous recombination; TOPIII, DNA topoisomerase III; RAD51, an essential recombination protein physically and functionally homologous to RecA protein; RPA, replication protein A, that binds single-stranded DNA and is comprised of three subunits p70, p34, and p17; MLH1, a DNA mismatch repair protein, of which mutation leads to colon cancer predisposition; FANC proteins are encoded by the gene mutated in the rare genetic disease Fanconi anemia; Ku70/Ku80, a DNA double-strand break repair protein, that binds DNA-dependent protein kinase (DNA-PK) and to the broken DNA end; TRF2, telomeric repeat binding factor 2, that protects telomeres against end-to-end fusion of chromosomes.

^b Though RNA polymerase is not biochemically classified as DNA helicase, it is capable of unwinding DNA during transcription.

^c RecBCD consists of three peptide chains RecB, RecC and RecD and it has exonuclease activities.



DNA Helicases. Figure 2 Proposed mechanism of action for PcrA helicase. ATP binding and hydrolysis mediates conformational changes that allow for DNA translocation and duplex destabilization. PcrA is comprised of four domains 1A (green), 1B (yellow), 2A (red), and 2B (blue). The binding of domains 1A and 2A to ssDNA induces swiveling of domain 2B, which in turn prepares the potential to bind duplex DNA. Together domains 1A and 2A form the RecA-like domains, which when bound to ATP, act as a crosslinking bridge leading to a cleft closure that bring these two domains closer. This conformational change within the recA like domains spreads a larger conformational change across the surface of domains 1B and 2B, which results in direct interaction with the duplex. Domain 2B 'grabs' the duplex at 12-13 bases away from the ss-dsDNA junction (distal contact) and pulls it toward the protein. Meanwhile, domain 1B forms a contact with the duplex closer to the junction (proximal contact) leading to destabilization in the duplex for half a helical turn as well as strand separation at the junction. Upon ATP hydrolysis, the bridge formed between domains 1A and 2A is broken, which is translated in a reversal of the conformational changes in domains 1B and 2B. While domain 1B releases the duplex, 2B pushes the DNA away from the protein while maintaining contact. This process allows the duplex to translocate across the surface of the protein. Reprinted with permission from (5).

DNA Helicases. Table 2 Human helicase genes and the clinical entities in which they are mutated

Genes mutated	Syndrome	Selected clinical features ^a	Cancer predisposition	Selected cellular phenotypes	Biochemistry/function
<i>XPB</i> and <i>XPD</i>	Xeroderma pigmentosum ^b (XP)	Skin abnormalities Photosensitivity Cataracts Neurological abnormalities ^c	Premalignant actinic keratoses Basal cell and squamous cell carcinomas of the skin Melanomas Ocular cancers	Unscheduled DNA synthesis (UDS) Increased UV sensitivity Increased mutation rate post-irradiation Decreased repair of cyclopurimidine dimers (CPDs) and 6-4 photoproducts	DNA-dependent ATPase activity XPB: 3' → 5' helicase XPD: 5' → 3' helicase Components of TFIIH transcription factor Nucleotide excision repair mRNA transcription initiation
	Cockayne syndrome (CS)	Growth deficiency Neurological abnormalities Skeletal abnormalities Progeria Photosensitivity Mental deficiency Cataracts Hearing loss Dental caries	None ^d	Increased UV sensitivity Decreased repair of DNA lesions of the transcribed DNA strand Decreased recovery of DNA and RNA synthesis post-irradiation Normal UDS	
	Trichothiodystrophy (TTD)	Photosensitivity Intellectual impairment Sulfur-deficient brittle hair Decreased fertility Short stature	None ^d	Increased UV sensitivity Decreased repair of CPDs and 6-4 photoproducts Decreased recovery of RNA synthesis post-irradiation	
<i>BLM</i>	Bloom syndrome (BS)	Proportional dwarfism Sun-sensitive facial erythema Telangiectasia Moderate immunodeficiency Infertility/subfertility Predisposition to diabetes	Leukemias Lymphomas Carcinomas	Chromosome breakage Hyper-recombination High-sister chromatid exchange Sensitivity to alkylating agents	DNA-dependent ATPase 3' → 5' strand displacement Substrates: recombination intermediates and G4 DNA
<i>WRN</i>	Werner syndrome (WS)	Premature graying and alopecia Atrophy of extremities Atherosclerosis Arteriosclerosis Osteoporosis Diabetes Subfertility	Soft tissue sarcoma Osteosarcoma Melanoma Meningioma Hematological disorders	Reduced cell-division potential Variegated chromosomal mosaicism Increased telomere shortening Sensitivity to 4-NQ and etoposide	DNA-dependent ATPase 3' → 5' strand displacement Substrates: recombination intermediates and G4 DNA 3' → 5' exonuclease

DNA Helicases. Table 2 Human helicase genes and the clinical entities in which they are mutated (Continued)

Genes mutated	Syndrome	Selected clinical features ^a	Cancer predisposition	Selected cellular phenotypes	Biochemistry/function
<i>RTS</i>	Rothmund-Thomson syndrome (RTS)	Short stature Skin atrophy Telangiectasia Poikiloderma Skeletal abnormalities Hair thinning and loss	Soft tissue sarcoma Osteosarcoma Skin cancer	Increased mosaicism	Unknown

^a There is variable expressivity of many of these clinical features.

^b There are seven major complementation groups in XP that have been identified (A, B, C, D, E, F, and G). All of these components are essential for the nucleotide excision repair pathway. XP variant (XPV) syndrome shows typical XP features, but an absence of a defect in NER. XPV cells have a defect in post-replication repair.

^c Only in complementation groups XPA, XPB, XPD, and XPG.

^d Average lifespan of these persons is low, but a lack of cancer predisposition is suggested by the clinical data.

Structural Studies

Recent crystal structures have provided new insights into the mechanism of catalysis by DNA helicases (Fig. 2). The helicases studied in this way consist of 4 domains (1A, 1B, 2A, and 2B) with a central lateral groove along which single-stranded DNA binds. Domains 1A and 2A are structurally similar to the ►*RecA* protein. ATP binds in the groove between 1A and 2A. Translocation along single-stranded DNA follows an inchworm model. The various states of ATP/ADP binding are associated with dynamic changes in the enzyme's conformation and changes in the interactions between conserved amino acid residues in the central lateral groove and the single-stranded DNA. Hydrolysis of ATP accompanies a power stroke, in which the leading edge of the enzyme (domains 2A and 2B) pumps forward on the DNA while the lagging edge (domains 1A and 1B) holds onto the DNA (5).

Clinical Relevance

Helicases and Human Disease

Mutational defects in several DNA helicase genes can cause distinctive genetic disorders. Six human genetic disorders have been reported thus far, including xeroderma pigmentosum (XP), Cockayne syndrome (CS), trichothiodystrophy (TTD), Bloom syndrome (BS), Werner syndrome (WS) and Rothmund-Thomson syndrome (RTS) (Table 2). These syndromes represent a subset of disorders in which DNA repair deficiency, ►*genomic instability* and cancer predisposition are common features.

Persons with XP exhibit a defect in nucleotide excision repair (NER). NER is an important DNA repair mechanism that removes cyclopurine dimers and other lesions generated by UV irradiation and bulky DNA

adducts, such as those generated by cis-platinum treatment. A defective NER pathway results in increased mutagenesis particularly in the skin from failure to repair sunlight-generated lesions, which explains the development of skin cancers in XP. Two DNA helicases function in the NER pathway, XPB and XPD, which have opposite directionalities on single-stranded DNA. The XPB and XPD helicases are part of the transcription factor complex, TFIIH, which contains up to 9 proteins. In NER, the XPB and XPD subunits of TFIIH provide helicase activities that open up the DNA duplex so that the lesion-containing DNA can be cleaved from the duplex by specific endonucleases. In transcription, TFIIH opens up the DNA duplex so that RNA polymerase can bind to the transcribed DNA strand. Thus, the role of TFIIH in each process is similar. Some ►*mutations* of *XPB* and *XPD* result in XP, whereas other mutations in these genes result in either of the clinically distinct syndromes CS and TTD. The defects of CS and TTD cells are also distinct from those of XP cells, and they are believed to be caused by a deficiency of TFIIH in transcription but not NER. A small number of persons with *XPB* or *XPD* mutations have XP and CS or TTD, and presumably these mutations affect both NER and transcription. CS is a progeroid syndrome, and mouse models of these syndromes have shown that these genes are important in mechanisms of aging (6).

The DNA helicases that are defective in persons with BS, WS, and RTS belong to a family of DNA helicases referred to as the RecQ helicases, which is conserved from bacteria to humans. The corresponding genes that are mutated in these syndromes are *BLM*, *WRN*, and *RECQL4* (also referred to as *RTS*), respectively. The syndromes are clinically distinct (Table 2): BS is characterized by proportional dwarfism and a sun-sensitive facial rash, WS by progeroid features such

as alopecia, atherosclerosis and osteoporosis and RTS by small size, poikiloderma and alopecia. A common feature of cells that contain mutations in one of the RecQ genes is a disturbance in recombination. For example, BS cells exhibit excessive recombination between homologous chromosomes and between sister chromatids, as well as chromosome breakage and WS cells exhibit increased chromosome translocations. The genomic instability that is prominent in these syndromes is associated with susceptibility to a wide spectrum of cancers. The RecQ enzymes are unusual DNA helicases in that they are most efficient in binding and unwinding substrates that resemble recombination intermediates, including Holliday junctions, D-loops, and [▶G-quartet DNA](#). Although the function of the RecQ helicases is elusive, a number of functions have been proposed, including resolution of Holliday junctions, suppression of illegitimate recombination events that might occur between imperfectly homologous DNA sequences and participation in one or more of the processes involved in restarting a stalled replication fork, which can occur when the replication machinery (i.e. the polymerase) encounters DNA damage (6). Mutations in RecQ helicases have uncovered the important connections between DNA recombination, genomic stability, cancer predisposition and mechanisms of aging

[▶Bloom Syndrome](#)

[▶Replication Fork](#)

References

1. Caruthers JM, McKay DB (2002) Helicase structure and mechanism. *Curr Opin Struct Biol* 1:123–33
2. Delagoutte E, von Hippel PH (2003) Helicase mechanisms and the coupling of helicases within macromolecular machines. Part II: Integration of helicases into cellular processes. *Q Rev Biophys* 36:1–69
3. Hall MC, Matson SW (1999) Helicase motifs: the engine that powers DNA unwinding. *Mol Microbiol* 34:867–877
4. Singleton MR, Wigley DB (2002) Modularity and specialization in superfamily 1 and 2 helicases. *J Bacteriol* 184:1819–1826
5. Soultanas P, Wigley DB (2001) Unwinding the ‘Gordian knot’ of helicase action. *Trends Biochem Sci* 26:47–54
6. Van Brabant AJ, Stan R, Ellis NA (2000) DNA helicases, genomic instability, and human genetic disorders. *Annu. Rev. Genomics Human Genet* 1:409–459
7. von Hippel PH, Delagoutte E (2001) A general model for nucleic acid helicases and their “coupling” within macromolecular machines. *Cell* 104:177–90

DNA Lesion

Definition

DNA lesion refers to a section of a DNA molecule containing a primary damaged site i.e. a base alteration,

a base deletion, a sugar alteration or a strand break. Replication before repair, or inefficient repair, can result in the fixation of a primary lesion as a permanent mutation.

[▶Chromosomal Instability Syndromes](#)

DNA Ligases

ALESSANDRA MONTECUCCO, ROSSELLA ROSSI
Istituto di Genetica Molecolare, CNR, Pavia, Italy
montecucco@igm.cnr.it

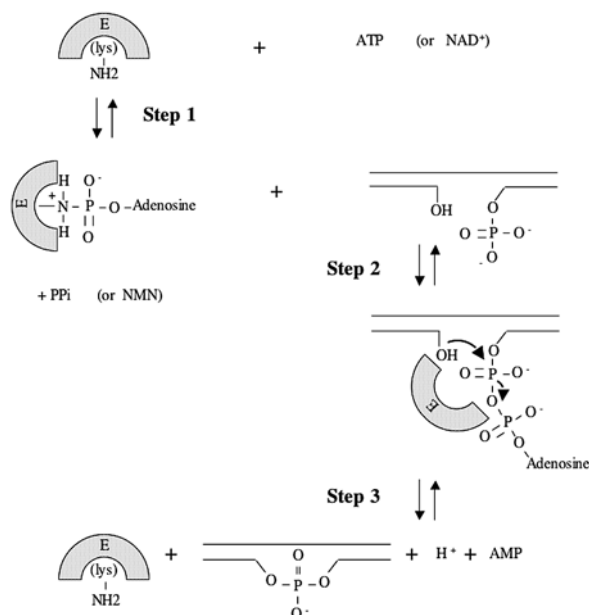
Definition

DNA ligases are a large family of evolutionarily related enzymes that catalyze the joining of single-strand or double-strand breaks between adjacent 3'-hydroxyl and 5'-phosphate termini in the phosphodiester backbone of double-strand DNA. For this reason they play a vital role in DNA metabolism. DNA strand breaks can occur as a result either of the direct action of [▶DNA damaging agents](#) or as reaction intermediates during [▶DNA replication](#), repair and recombination; therefore the sealing of these breaks by DNA ligase is critical for maintaining genome integrity.

Characteristics

Reaction Mechanism

The joining reaction is coupled to the hydrolysis of a high-energy bond in the cofactor molecule. On the basis of the required cofactor DNA ligases are separated into two major groups, [▶ATP-dependent](#) and NAD^+ ([▶NADPH+H+NAD⁺](#))-dependent enzymes. DNA ligases from most organisms break down ATP to [▶AMP](#) and pyrophosphate (PPi). Eubacterial enzymes, instead, utilize NAD^+ with the production of AMP and nicotinamide mononucleotide (NMN). One ATP (or NAD^+) molecule is consumed per nick-sealing event. The DNA joining reaction involves three successive steps. The first is a nucleophilic attack on the α -phosphate of ATP (or the adenyl-phosphate of NAD^+) by the lysine residue in the active site of the enzyme resulting in a covalent enzyme-AMP intermediate (Fig. 1). Certain cofactor analogs, such as dATP, can be used, but the anomalous enzyme-nucleotide intermediates appear to function poorly in the successive steps. During the second step, the AMP group is transferred to the phosphorylated 5'-end of a single-strand break to produce an inverted (5')-(5') pyrophosphate bridge structure. In the final step the enzyme catalyzes the sealing of the nick with the displacement of AMP.



DNA Ligases. Figure 1 DNA joining reaction catalyzed by DNA ligase. Step 1: DNA ligase (E) interacts with ATP or NAD⁺ depending on the class of the enzyme, to form an enzyme-AMP intermediate in which the AMP group is covalently linked to the lysine (lys) of the active site. Step 2: the AMP is transferred from the enzyme to the 5'-phosphate terminus of a nick in duplex DNA to form a DNA-adenylate intermediate. Step 3: the interaction of non-adenylated DNA ligase with the DNA-adenylate complex results in phosphodiester bond formation and the release of AMP.

In the so-called ‘reverse reaction’ the AMP group can be used as a cofactor for the ligase-directed removal of superhelical turns from a covalently closed **▶supercoiled DNA**. This occurs through a nicking/closing mechanism mimicking the DNA topoisomerase activity. The lysine of the active site is required also for the AMP-dependent **▶DNA relaxation**.

Some DNA ligases are also able to catalyze blunt-end ligation. Blunt-end joining proceeds less efficiently than ligation of nicks or of cohesive ends and is stimulated by polyethylene glycol.

Structure

The three-dimensional structures of both an ATP-dependent and a NAD⁺-dependent DNA ligase have been determined by X-ray crystallography (1). Immediately apparent from the structures is the conservation of a common ‘core’ domain between the two enzymes. Alignment of DNA ligase sequences from different sources has identified a series of co-linear conserved motifs (motif I, III, IIIa, IV, V, VI) that define a family of related nucleotidyl transferases including eukaryotic GTP-dependent mRNA-capping enzymes.

With the exception of motif VI, these lie in the core domain. Motif I corresponds to the active site and contains the consensus sequence ‘lys-x-asp-gly’ shared with RNA ligases. The lysine in motif I is engaged in the covalent binding to AMP. The similarity of structural features between DNA ligases and mRNA capping enzymes suggests a similar mechanism of catalysis and the evolution of these two classes of enzymes from a common ancestral nucleotidyl transferase. The polynucleotide specificity of the enzyme (RNA or DNA) is dictated by additional protein ‘modules’ around the catalytic conserved core.

NAD⁺-Dependent DNA Ligases

Eubacterial DNA ligases appear to be almost unique in utilizing NAD⁺ as high-energy cofactor. The prototype of this class is the *Escherichia coli* (*E. coli*) enzyme encoded by the essential *ligA* gene. NAD⁺-dependent DNA ligases are also found in *Bacillus subtilis* and *Staphylococcus aureus* and probably are essential genes in all eubacteria. The enzymes from different bacteria have now been cloned and sequenced. They show a high degree of sequence similarity and a similar size (about 75 kD). The fact that NAD⁺-dependent DNA ligases are specific to eubacteria and are essential for cell survival makes these proteins potential targets for novel antibacterial strategies.

ATP-Dependent DNA Ligases

Most eukaryotic DNA ligases together with archeal and bacteriophage enzymes use ATP as a cofactor. In eukaryotic cells there are three classes of ATP-dependent DNA ligases that are related in sequence and structure. These enzymes are built around a common catalytic core that is similar to the smaller ATP-dependent DNA ligase of the T7-bacteriophage. Numerous pieces of evidence indicate that the three classes of DNA ligases are targeted to different pathways in DNA replication, repair and recombination and show a limited, if any, ability to substitute for each other. The main functional differences between these DNA ligases seem to be mediated by large regions located outside the catalytic core that can establish specific protein-protein interactions and govern the sub-cellular distribution.

The three mammalian enzymes are called DNA ligase I, III and IV (Table 1).

DNA ligase I is required for chromosomal DNA replication as well as for several DNA-repair pathways. Orthologs have been identified in organisms as diverse as yeast and mammals. The human *lig1* gene has been mapped on chromosome 19q13.2-13.3 close to other DNA repair genes. This proximity is conserved in mouse. During chromosome replication DNA ligase I is responsible for the joining of the **▶Okazaki fragments**. The enzyme acts as part of a replication complex called

DNA Ligases. Table 1

Human enzyme	Gene	Chromosomal location	Molecular mass (kDa)	Amino acids encoded by cDNA	Yeast homolog	Cellular localization	Function <i>in vivo</i>	DNA ligase-interacting proteins
DNA ligase I	<i>lig1</i>	19q13.2-13.3	102	919	<i>cdc 9</i>	Nucleus Replication foci	DNA replication BER NER	PCNA DNA polymerase β CyclinA/ cdk2
DNA ligase III α	<i>lig3</i>	17q11.2-12	103	922		Nucleus Mitotic chromosomes	BER	XRCC1
DNA ligase III β	<i>lig3</i>	17q11.2-12	96	862		Nucleus Mitotic chromosomes	Meiotic recombination	
Mitochondrial DNA ligase	<i>lig3</i>	17q11.2-12				Mitochondria	Mitochondrial DNA repair	
DNA ligase IV	<i>lig4</i>	13q33-34	96	844	<i>dnl 4</i>	Nucleus Mitotic chromosomes	NHEJ V(D)J recombination	XRCC4

the replisome formed by several different replication proteins. DNA ligase I is divided into two clearly distinct regions, a highly conserved C-terminal region, containing the active site and a less conserved hydrophilic N-terminal portion (residues 1–216), that functions as a regulatory domain. The N-terminal domain controls the sub-cellular distribution of the enzyme and mediates the interaction with other replicative proteins. Human DNA ligase I is phosphorylated *in vivo* on four serine residues (at positions 51, 66, 76, 91) and is a substrate for both casein kinase II (CKII) and cyclin-dependent kinases (CDK). Consistently with its role in DNA replication, the DNA ligase I gene is more expressed in proliferating than in resting cells. DNA ligase I activity during the cell cycle is controlled by both phosphorylation and sub-nuclear compartmentalization. The enzyme associates with [▶replication foci](#) during S-phase while displaying a diffuse nucleoplasmic distribution in non-S phase nuclei. Its function requires interaction with the proliferating cell nuclear antigen (PCNA) mediated by a p21-type PCNA-binding motif in the first 20 amino acids of the protein (2). The PCNA binding site is required for recruitment to the replication foci and for the efficient joining of the Okazaki fragments during DNA replication.

Similarly to other enzymes involved in the maturation of the Okazaki fragments, DNA ligase I functions in

long-patch base excision repair ([▶BER](#)). In addition DNA ligase I interacts with DNA polymerase β that catalyzes DNA synthesis in short-patch BER. Finally, *in vitro* reconstitution assays have implicated DNA ligase I in nucleotide excision repair (NER).

A case of DNA ligase I-deficiency in humans has been described which is responsible for the 46BR syndrome, characterized by growth retardation, abnormal development, sun sensitivity and immunodeficiency (3). Different point mutations have been found in each of the *lig1* alleles resulting in extremely reduced enzyme activity. Skin fibroblasts from this patient show hypersensitivity to DNA damaging agents, replication defects and genetic instability. Transgenic mice carrying the same mutation display an increased incidence of spontaneous epithelial tumors (4). *Lig1* knock out mice have been recently obtained. DNA ligase I null mutants die during embryogenesis and fibroblasts from these embryos show accumulation of DNA replication intermediates and increased genome instability.

DNA ligase III is unique to vertebrates, and functions both in the nucleus and in mitochondria. It is involved mainly in nuclear and mitochondrial DNA repair. Alternative splicing of *lig3* transcripts produces two isoforms called DNA ligases III α and β with different carboxyl-terminal sequences (5). DNA ligase III α is ubiquitously distributed and has a role in the

short-patch BER. DNA ligase III β is expressed only in the testes where it is believed to play a role in homologous recombination during meiosis. DNA ligase III α but not the β form, has a C-terminal **►BRCT domain** through which it interacts with the repair protein XRCC1 (X-ray cross complementing factor 1). The heterodimeric complex formed by XRCC1 and DNA ligase III α functions *in vivo* in short-patch BER. At the N-terminus of both DNA ligase III α and β there is a consensus sequence for a **►zinc finger** DNA binding domain that could trigger the binding of the enzyme and of associated proteins to sites of DNA damage. Both isoforms can be targeted to mitochondria thanks to a mitochondrial targeting sequence (MTS) located upstream of the first amino acid of the mature protein. This raises the possibility that both enzymes play a part in the repair of mitochondrial DNA. Upon import into the mitochondria, the MTS is cleaved by a mitochondrial peptidase. DNA ligase IV is conserved in all eukaryotes. Orthologs have been identified in yeast, higher plants and vertebrates. It functions in non-homologous end joining (NHEJ), which is the main repair pathway for double strand breaks caused by ionizing radiation or certain classes of chemical mutagens. This pathway is also required for V(D)J recombination, the process by which immunoglobulin and T-cell receptor genes are rearranged to generate antibody diversity. *In vivo* DNA ligase IV is strongly associated with XRCC4, a nuclear phosphoprotein. XRCC4 stabilizes DNA ligase IV and stimulates its activity by targeting the protein to sites of DNA double strand breaks. DNA ligase IV is characterized by a C-terminal region comprising two BRCT domains separated by a short linker that contains the binding site for XRCC4. DNA ligase IV is a nuclear protein and is associated with chromosomes during mitosis. Recent evidence indicates that DNA ligase IV co-localizes with a human condensin subunit, a member of the structural maintenance of chromosome (SMC) super-family proteins. Inactivation of the *lig4* gene in mouse leads to late embryonic lethality associated with a block of lymphopoiesis, V(D)J recombination failure and extensive apoptotic cell death in the embryonic central nervous system (6). DNA ligase IV-deficient embryonic fibroblasts also show marked sensitivity to ionizing radiation, growth defects and premature senescence. All of these phenotypic characteristics, except embryonic lethality, resemble those associated with deficiencies in other proteins involved in NHEJ. To date five patients have been described with a defect in DNA ligase IV. One patient was homozygous for a missense mutation that reduced the adenylation and ligation activities of the enzyme. The patient did not display any overt immunodeficiency but he developed leukemia at the age of 14 and showed pronounced radio-sensitivity.

The other four patients had mutations that disrupted the catalytic domain or impaired the interaction with XRCC4. All of them showed common features including immunodeficiency, developmental and growth delay and chromosomal instability. Their clinical phenotype closely resembled the DNA damage response disorder, Nijmegen breakage syndrome (NBS). Unlike NBS, *lig4* cells showed a normal cell-cycle checkpoint response but impaired DNA double strand break rejoining. These findings have unveiled the existence of an inherited disorder, designated *lig4* syndrome that arises from a specific DNA repair defect and is associated with primary human immunodeficiency, genomic instability and developmental delay (7).

Clinical Relevance

Similarly to other replicative enzymes, DNA ligase I is a marker of cell proliferation and its sub-nuclear distribution is regulated during the cell cycle. It is therefore possible to use anti-DNA ligase I antibodies to recognize S-phase cells by immunohistochemistry and to determine the labeling index of a cell population.

DNA Ligase Inhibitors with Anticancer Activity

Given its role in DNA replication, DNA ligase I is a target for a number of anticancer drugs. Antitumor anthracyclines and distamycins are strong inhibitors of human DNA ligase I. Anthracyclines block the joining reaction without affecting enzyme adenylation. This inhibition is mediated by the interaction of anthracyclines with the DNA substrate. A potent and specific inhibitor of human DNA ligase I is tallimustine, a distamycin derivative with antitumor activity. Unlike the anthracyclines, this drug inhibits the adenylation of human DNA ligase I, preventing the successive steps of the joining reaction. Etoposide is an anticancer drug belonging to the topoisomerase poisons. Although unable to inhibit the DNA joining reaction *in vitro*, this drug has a dramatic effect on the phosphorylation and sub-nuclear distribution of DNA ligase I.

DNA Ligase Inhibitors with Antibiotic Activity

Multiple drug resistance among bacterial pathogens has made many of the currently available antibiotics ineffective. As a consequence the search for novel antibacterial classes with innovative mechanisms of action is crucial. NAD⁺-dependent DNA ligases are promising target candidates because they are indispensable for many fundamental aspects of DNA metabolism. The different cofactor requirements raise the chance of finding specific inhibitors directed exclusively against the eubacterial enzymes. Derivatives of the anti-malaria drug chloroquine inhibit the *E. coli* DNA ligase. Human DNA ligase I is also affected but at a 10-fold higher concentration. More recently

pyridochromanones that possess substantial antibacterial activity have been identified as a novel class of potent DNA ligase inhibitors. The compounds specifically inhibited the purified NAD⁺-dependent enzymes from both *E. coli* and *Streptococcus pneumoniae* in the nanomolar concentration range.

Mouse Models

Both DNA ligase I null mouse cells and a mouse model for the 46BR syndrome have been recently produced. These mutants represent useful models to investigate the relationship between DNA replication failure, genome instability and cancer susceptibility. New insights into the role of DNA ligase IV have come from the targeted disruption of the gene encoding DNA ligase IV in mice. Unlike other partners in NHEJ, DNA ligase IV has an essential function in early mammalian development.

► [Base Excision Repair](#)

References

1. Timson DG, Singleton MR, Wigley DB (2000) DNA ligases in the repair and replication of DNA. *Mutation Res* 460:301–318
2. Montecucco A, Rossi R, Levin DS et al (1998) DNA ligase I is recruited to sites of DNA replication by an interaction with proliferating cell nuclear antigen: identification of a common targeting mechanism for the assembly of replication factories. *EMBO J* 17:3786–3795
3. Barnes DE, Tomkinson AE, Lehman AR et al (1992) Mutations in the DNA ligase I gene in an individual with immunodeficiencies and cellular hypersensitivity to DNA-damaging agents. *Cell* 69:495–504
4. Harrison C, Ketchen A-M, Redhead NJ et al (2002) Replication failure, genome instability and increased cancer susceptibility in mice with a point mutation in the DNA ligase I gene. *Cancer Res* 62:4065–4074
5. Mackey ZB, Ramos W, Levin DS et al (1997) An alternative splicing event which occurs in mouse pachytene spermatocytes generates a form of DNA ligase III with distinct biochemical properties that may function in meiotic recombination. *Mol Cell Biol* 17:989–998
6. Frank KM, Sekiguchi JM, Seidl KJ et al (1998) Late embryonic lethality and impaired V(D)J recombination in mice lacking DNA ligase IV. *Nature* 369:173–177
7. O'Driscoll M, Cerosaletti KM, Girard P-M et al (2001) DNA ligase IV mutations identified in patients exhibiting developmental delay and immunodeficiency. *Mol Cell* 8:1175–1185

DNA Marker

Definition

DNA marker defines any unique DNA sequence in the genome that can be used in DNA hybridization, PCR or

restriction mapping experiments as a landmark for genetic analysis of phenotypes such as a disease.

► [SNP Detection and Mass Spectrometry](#)

DNA Methylation

Definition

DNA methylation is a covalent modification of the DNA. In mammals, the most frequent modification is the addition of a methyl group to cytosine usually found in CG doublets. This type of modification does not change the nucleotide sequence. Methylation can change the ability of some proteins to bind to a sequence that contains methylated CGs, and by this means it can influence the chromatin structure and activity, and consequently, gene expression. Therefore, DNA methylation is an epigenetic mark. The modification is catalyzed by methyltransferases, a family of enzymes with several identified members.

► [CpG Islands](#)

► [Fragile X Syndrome](#)

► [Genomic Imprinting](#)

DNA Microarrays/DNA Arrays

Definition

DNA arrays/DNA microarrays (also referred to as ► [DNA chips](#)) denote an analytical tool to probe RNA expression levels in a cell. DNA probes or oligonucleotides are immobilized at high density onto a microcarrier. The underlying principle of DNA microarray technology is nucleobase-pairing. A single array may represent parts of a genome or the entire set of genes of a given organism/cell. The most common gene expression microarrays fall into two categories: Affymetrix Genechips (a proprietary commercial array using 25 base oligonucleotides probe sets), and spotted DNA microarrays in which long cDNA probes or oligos are deposited and bound to a substrate, usually a coated glass microscope slide.

► [C. Elegans as a Model Organism for Functional Genomics](#)

► [DNA Chips](#)

► [High-Throughput Approaches to the Analysis of Gene Function in Mammalian Cells](#)

► [Microarray Data Analysis](#)

- ▶ Multifactorial or Common Diseases
- ▶ PNA chips
- ▶ Protein Microarrays as Tool for Protein Profiling and Identification of Novel Protein-Protein Interactions
- ▶ Proteomics in Human-Pathogen Interactions
- ▶ Rheumatism Related Genes, Identification
- ▶ RNA Interference in Mammalian Cells
- ▶ Thermodynamic Properties of DNA
- ▶ *Xenopus* as a Model Organism for Functional Genomics

DNA Polymerase Switch

Definition

DNA polymerase switch refers to a replication factor C dependent process, which provides the transition from initiation to elongation of DNA replication by substitution of DNA polymerase α /primase by DNA polymerase δ .

- ▶ DNA Polymerases
- ▶ Replication Fork

DNA Polymerase δ Holoenzyme

Definition

DNA polymerase δ holoenzyme designates a replicative DNA polymerase complex that consists of DNA polymerase δ , PCNA (proliferating cell nuclear antigen) and replication factor C.

- ▶ DNA Polymerases
- ▶ Replication Fork

DNA Polymerases

HEINZ-PETER NASHEUER¹, HELMUT POSPIECH^{1,2}, JUHANI SYVÄÖJA^{1,3}

¹Department of Biochemistry, National University of Ireland, Galway, Ireland

²Biocenter Oulu and Department of Biochemistry, University of Oulu, Oulu, Finland

³Department of Biology, University of Joensuu, Joensuu, Finland
h.nasheuer@nuigalway.ie

Definition

The balance between stability and variability in the human genome influences the development of inherited diseases such as cancer and modulates adaptability to new environments and the immunoresponse. DNA polymerases are at the heart of human DNA metabolism and act as DNA synthesis machines. Thus, they play a central role in the mechanisms of genome stability as well as its instability (for additional discussion see ▶ [Replication Fork](#)). In the human cell, four extended protein families A, B, X and Y have evolved to cope with the different tasks of sustaining the genome as well as inducing specific mutations in certain regions or 'rescuing' the genome after specific stress situation e.g. DNA damage (Table 1). This orchestra of human DNA polymerases requires tight regulation of the various DNA synthetic tasks, which has been only partially understood (Table 2).

DNA polymerases catalyse DNA synthesis in a template-dependent fashion. They require a divalent cation, preferentially Mg^{2+} , to incorporate the deoxyribonucleoside monophosphate (dNMP) in a 5'-3' direction. DNA polymerases are not able to start DNA synthesis *de novo* since they depend completely on the presence of a free 3'-OH group. The reaction is driven energetically by the fact that the incoming dNTP reacts chemically with the 3'-OH of the primer to produce the corresponding dNMP and pyrophosphate, which is hydrolysed to inorganic phosphate. During this reaction, the dNMP is covalently linked to the primer, which is extended by one nucleotide, and a free 3'-OH is 'created' again. The reaction occurs with high fidelity as regards correspondence of the incorporated nucleotide with the template. The error rates in formation of A-T and G-C pairs are in the range of 10^{-3} to 10^{-5} depending on the DNA polymerase. The accuracy of the DNA synthesis reaction is further enhanced by 3'-5' proofreading ▶ [exonucleases](#), which are often an intrinsic activity of DNA polymerase polypeptides, although the two activities are usually located on different domains (Table 2).

Characteristics

DNA Polymerase α -Primase

Since DNA polymerases cannot start a DNA reaction *de novo*, a special activity – the primase – has been evolved which associates with one specific DNA polymerase and synthesises a short oligonucleotide – the primer – to start the DNA synthesis reaction. The heterotetrameric DNA polymerase α -primase complex (Pol α -prim) initiates eukaryotic DNA replication at the

DNA Polymerases. Table 1 Distribution of DNA polymerases in Archaea, Bacteria, Eukaryotes and Viruses

DNA polymerase family						
Domain	A	B	C	D	X	Y
Archaea	–	Pol BI, Pol BII ^a , (Pol BIII) ^b	–	Pol DI ^c	+	DinB
Bacteria	Pol I	(Pol II)	Pol III	–	+	Pol IV/DinBPol V/UmuC
Eukaryotes	Pols γ and θ	Pols α, δ, ε and ζ, (Pol φ) ^d	–	–	Pols β, λ, μ and σ	Pols η, ι and κ, REV1
Viruses	+	+	–	–	(+)	–

DNA polymerases are listed by name, or, if not available, their presence is indicated by a '+'

^a Only in Crenarchaeota

^b Types of polymerases that are only found sporadically in a domain are indicated by brackets

^c Only in Euryarchaeota

^d Only in yeasts

DNA Polymerases. Table 2 Nomenclature and properties of mammalian DNA polymerases.

Name [synonyms]	Family	Subunit composition	Associated activities	Functions
Pol α ^a	B	165,67,58,48 ^b	primase	replication, repair
Pol β	X	39	dRP lyase ^c	repair
Pol γ	A	125,35	3'→5' exonuclease	mitochondrial replication & repair
Pol δ	B	125,66,50,12	3'→5' exonuclease	replication, repair
Pol ε	B	261,59,17,12	3'→5' exonuclease	replication, repair
Pol ζ [Rev3]	B	344		translesion synthesis, somatic hypermutation
Pol η [Rad30A, Y XPV]		78		translesion synthesis, somatic hypermutation, sister chromatid cohesion
Pol θ [Mus308]	A	198	helicase ^d	repair
Pol ι [Rad30B]	Y	80		translesion synthesis, somatic hypermutation
Pol κ [DinB, Pol θ]	Y	99		translesion synthesis, somatic hypermutation
Pol λ [Pol β2]	X	66	dRP lyase	translesion synthesis, repair, recombination
Pol μ	X	55		somatic hypermutation, repair
Pol σ [Trf4/5, Pol κ] ^e	X	42/55		sister chromatid cohesion
Rev1 ^f	Y	138		translesion synthesis

^a By convention, eukaryotic DNA polymerases are designated by the lowercase letters of the Greek alphabet in order of their description

^b Molecular masses in kD

^c Deoxyribophosphate lyase

^d Inferred from the sequence

^e In different organisms, various numbers of Pol σ orthologues are present

^f REV1 is a DNA-dependent deoxycytidyl transferase, since it only incorporates dCMP, but no other nucleotides into DNA

origin of DNA replication to start leading strand replication, synthesises RNA primer on the lagging strand and elongates them to produce the 5'-end of Okazaki fragments (for review see 1, 2; see Replication Fork). These activities are also necessary in DNA repair processes, especially homologous DNA recombination and DNA checkpoint signalling. Of the constituent four subunits the largest, p180, has DNA polymerase activity whereas the smallest subunit, p48, is a primase (1, 2). The remaining subunits, called p58 and B subunit or p67/p70, have no known enzymatic activities but evidence exists that they fulfil regulatory roles (2).

The largest subunit p180 contains at least seven sequence motifs conserved between eukaryotic Pol α homologues (2). The central part of p180 shows the highest sequence conservation and is sufficient to support the catalysis of DNA synthesis since it contains phosphoryl transfer and Mg^{2+} , DNA and dNTP binding activities (2). The C-terminus is required for the assembly of the heterotetrameric complex (3). Although Pol α -prim lacks exonuclease activity, it interacts with the tumour suppressor protein p53, which has an intrinsic exonuclease activity and which can co-operate with Pol α -prim to exchange a mispaired 3'-primer end on a template (4). In addition, the protein binds various cellular and viral proteins such as replication protein A (RPA), poly(ADP-ribose) polymerase, Simian Virus 40 (SV40) T antigen and HPV protein E2. Pol α -prim is also involved in the maintenance of telomere length probably *via* its interaction with the telomere-associated protein Cdc13 (2).

The second largest subunit has an apparent molecular mass of about 66–70 kD depending on its posttranslational modifications (1,2). This polypeptide, also known as the B subunit, shows the highest sequence variation of the four Pol-prim subunits. Interestingly, each of the eukaryotic multisubunit replicase complexes – Pol α -prim, Pol δ and Pol ϵ – contains a distantly related B subunit (5). Recently it was reported that the B subunit of mouse Pol α -prim is required for the nuclear import of p180 (2). In addition, the human B subunit interacts with the initiator protein Cdc45 and various viral initiator proteins such as SV40 T antigen, HPV11 and HPV18 E1 (2). These associations with replication initiation factors suggest that the B subunit of Pol α -prim might be involved in formation and regulation of the initiation complex at DNA replication origins. Its cell- cycle-dependent phosphorylation by cyclin-dependent kinases (Cdks) supports the view that the B subunit carries out regulatory functions in DNA replication. Recent results suggest that other subunits of Pol α -prim are targets of protein kinases. p180 is a substrate of Cdks, Dbf4-Cdc7 and casein kinase II in a cell cycle-dependent manner (2).

Its ability to synthesise RNA primer is the central function of Pol α -prim. p48 carries the catalytic centre for the RNA polymerase (primase) (2, 6). The p48 subunit is highly conserved in eukaryotic primases and even has homologies to archaeal primases (2). Three conserved aspartates (D109, D111 and D306) are involved in metal ion binding. In addition, a conserved arginine at position 304 probably contacts the synthesised primer as occurs in yeast RNA polymerase II (7). In the N-terminus, the p48 polypeptide contains a motif with 4 conserved cysteines, which probably bind Zn. The p58 subunit stabilises the DNA primase activity of the p48 subunit and fully functional DNA primase activity requires both subunits' activities, e.g. p58 is required for the nuclear import of p48, for the binding of p48 to p180 and for control of the primase product lengths (6). Each primase subunit contains sequences with homology to Pol β , which is reflected in the three-dimensional structures of Pol β and an archaeal primase p40 (2).

DNA Polymerase β

The 3-D structure and catalytic mechanism of Pol β have been thoroughly studied (1, 8). It is a single polypeptide of 39 kD, composed of two domains, an 8 kD N-terminal domain and a 31 kD C-terminal domain, connected by a protease sensitive hinge region. The N-terminal domain has a template binding and 5'-►**deoxyribophosphate lyase** (5'-dRP lyase, see glossary) activity that removes an abasic 5'-deoxyribose phosphate from the 5'-primed side of a single-strand break, while the C-terminal domain carries the polymerase activity.

Consistent with its ability to deal with ►**abasic sites** of DNA, known intermediates of base excision repair (BER), Pol β has a role in DNA repair (8). The enzyme is well suited for a small gap-filling function in BER. Besides overwhelming evidence on the importance of this enzyme in BER, there is some indication that Pol β may have additional roles associated with DNA nucleotide excision repair (NER) and meiotic recombination. The repair role of Pol β is further supported by the fact that mutated forms of Pol β occur in numerous cancers. Targeted deletion of the mouse Pol β gene indicates that the enzyme is not essential for mammalian cell viability. However, homozygous mice die immediately after birth showing respiratory failure and defective neurogenesis.

DNA Polymerase γ

Pol γ is the principal DNA replicase and also the sole repair DNA polymerase in human mitochondria (1). Since aberrations in the mitochondrial DNA have been associated with tissue degeneration and aging, the functions and structure of Pol γ are of great clinical interest in understanding the evolution of mutations

in mitochondrial DNA. The enzyme complex consists of two subunits with molecular masses of 140 and 55 kD (p140 and p55) and synthesises DNA with high processivity. The large catalytic subunit has DNA polymerase, 3'-5' and 5'-3' exonuclease activities and can perform proofreading functions. In addition, the polypeptide shows 5'-dRP lyase activity. The small subunit binds DNA on its own and is necessary for the high processivity of the enzyme complex. Moreover, p55 stimulates the exonuclease activity of p140.

DNA Polymerase δ

Mammalian Pol δ was identified more than 20 years ago. Pol δ is implicated as one of the major DNA polymerases required for DNA replication (6, 8; see ►[Replication Fork](#)). This is mainly based on studies of SV40 DNA replication, which has served as a model for the process in mammalian cells. Consistently, the catalytic and the second largest B subunits have been found to be essential for cell viability, both in budding and in fission yeast. Mutants have a cell-cycle arrest phenotype, but surprisingly, a major amount of DNA is synthesised during replication before arrest in these mutants. Apart from its major function in DNA replication, Pol δ appears to be involved in several DNA transactions, including mismatch repair, nucleotide and base excision repair and bypass of DNA damage, as well as recombinational processes (see ►[DNA Recombination](#)).

Pol δ possesses a 3'→5' exonuclease activity in addition to its polymerase activity. The enzyme is therefore able to remove nucleotides immediately if incorporated erroneously. This proofreading greatly increases the fidelity of DNA synthesis and provides an important mechanism for the eukaryotic cell to avoid mutations. Proofreading is a feature of replicative DNA polymerases to achieve the superior accuracy in DNA synthesis that is required to duplicate large genomes. A targeted mutation in the mouse Pol δ gene inactivating proofreading causes an elevated mutation rate and mice with the proofreading defect have an increased risk of developing cancer (9). Mutations in Pol δ have also been identified in non-polyposis colon cancer cells. The clinical relevance of this finding remains unclear since the affected cell lines carry additional mutations in DNA mismatch repair factors that commonly underlie this disease (10).

Pol δ appears to be ubiquitous in all eukaryotic organisms. Apart from the catalytic subunit of approximately 125 kD, Pol δ isolated from various sources contains two or more additional smaller, structurally conserved subunits. Although these subunits appear to lack any enzymatic activity, they carry out crucial roles in maintaining the integrity of the enzyme and in mediating interactions with other proteins.

To fully understand the function of Pol δ , it has to be seen in the context of the two factors RFC and PCNA. These proteins provide the moving platform that mediates the dynamic functions of the enzyme (1; see Replication Fork). PCNA is a eukaryotic sliding clamp that is structurally and functionally related to the β subunit of *E. coli* Pol III holoenzyme. Three PCNA molecules form a closed ring able to encircle duplex DNA. Once loaded onto DNA by RFC in an ATP-dependent manner, PCNA cannot dissociate but can freely translocate on double-stranded DNA. PCNA then associates with Pol δ at the primer-template junction and supports ►[processive DNA synthesis](#) by the polymerase whilst at the same time preventing non-productive binding of Pol δ to single-stranded DNA. Pol δ in conjunction with PCNA and RFC is therefore generally referred to as the Pol δ holoenzyme.

DNA Polymerase ϵ

Pol ϵ is one of the three major replicative DNA polymerases in eukaryotic cells. It has been isolated and cloned from several species from yeast to human (11). The human enzyme consists of a 261 kD catalytic subunit and three smaller subunits of 59, 17 and 12 kD. The catalytic subunit contains DNA polymerase and proofreading exonuclease activities. Notably, it possesses a large C-terminal domain of unclear function that accounts for half of the molecular mass. Yeast homologues of the catalytic subunit and 59 kD subunit are both essential for viability. No catalytic activity has been assigned to the 59 kD subunit and its role has remained obscure. As shown with the yeast model, the two smallest subunits are not essential for viability. They possess a histone-fold and the 17 kD subunit is also an integral part of the chromatin remodelling factor CHRAC that regulates chromatin accessibility and nucleosome spacing, a process associated both with DNA replication and transcription.

Simian Virus 40 (SV40) DNA replication *in vitro* as a model of mammalian DNA replication has allowed identification and elucidation of the roles of cellular factors involved in replication, including Pols α and δ (11; see Replication Fork). SV40 DNA replication can not be utilised to study the role of Pol ϵ since it is not involved in this system. However, studies on cellular model systems have revealed that it is an essential component of cellular DNA replication machinery *in vivo*, being already involved in the initiation phase of replication (12). During elongation it seems to be supported by the same partner proteins as Pol δ , including at least RFC and PCNA. In most models presented, Pols δ and ϵ are believed to share their labour between leading and lagging strand synthesis or between elongation and maturation of Okazaki fragments. A third possibility is that the two DNA polymerases act in different regions of complex

eukaryotic chromatin, in both cases having Pol α as their sole DNA polymerase partner. Besides DNA replication, there is evidence that Pol ϵ is involved in recombinative repair, nucleotide excision repair and base excision repair. It may also be among the players in cellular response to DNA damage.

Other Eukaryotic DNA Polymerases

Apart from Pols α , β , γ , δ and ϵ , which are considered the classical cellular DNA polymerases, it became apparent that a number of additional DNA polymerases are present in many organisms (Table 2). These novel DNA polymerases can be best considered in their functional context (13, 11, 1).

One of the most prevalent roles for several of these enzymes is replicative DNA translesion DNA synthesis. Pol ζ is the first eukaryotic enzyme found capable of synthesising past a UV-induced thymidine cyclobutane dimer. This enzyme is the fourth widely distributed eukaryotic DNA polymerase related to Pol α . **Translesion DNA synthesis** by Pol ζ is error-prone and this enzyme is responsible for most of the damage-induced mutagenesis in *Saccharomyces cerevisiae*. More recent work indicates that the major importance of Pol ζ lies in its unusual ability to extend from poorly paired primer ends and to use distorted templates. One of several DNA polymerases may incorporate a nucleotide opposite to various DNA lesions. Pol ζ then uses this inserted nucleotide for limited extension beyond the lesion where the normal replication apparatus can take over again. Inactivation of the mouse gene for the catalytic subunit of Pol ζ results in early embryonic lethality. This indicates that DNA translesion synthesis may be of great importance in higher eukaryotes with large genomes.

It is probably mainly the DNA polymerases of family Y that incorporate nucleotides opposite to damaged DNA (13). Budding yeast Rev1 protein was the first enzyme of this family shown to possess DNA polymerase-related activity. Rev1 incorporates dCMP opposite to an abasic site, but appears to possess also a second, poorly understood function that is important for DNA translesion synthesis in general.

Human Pol η has been identified as the activity that prevents the sensitivity to UV light seen in cells derived from patients with the rare autosomal hereditary disorder xeroderma pigmentosum variant (XPV). The protein is encoded by the *XPV* gene that is affected in those patients. Pol η synthesises efficiently past a thymidine-thymidine *cis-syn* cyclobutane dimer, the principal damage inflicted by UV light on DNA. What is more, this translesion synthesis is apparently error-free.

Higher eukaryotes possess a second Pol η homologue, designated Pol ι . It shares several features with Pol η ,

including highly **distributive DNA synthesis** and the ability to bypass certain DNA lesions, but in contrast to Pol η , Pol ι is highly inaccurate during DNA synthesis. Translesion synthesis by Pols η and ι appears to be tightly linked to DNA replication. A fourth human DNA polymerase of family Y was recognised and designated Pol κ . As expected, Pol κ is capable of mutagenic DNA lesion bypass, but differs from other family members in that it introduces -1 frameshifts into the sequence and efficiently extends from a mispaired primer end. Crystal structures of family Y DNA polymerases reveal a conventional right hand-like architecture with domains assigned as palm, finger and thumb. Compared to other DNA polymerases, the finger and thumb domains are small, resulting in an open spacious active site that can accommodate DNA lesions and mispaired primer ends.

A process that is related to DNA translesion synthesis is **somatic hypermutation** (13). Here, mutations are intentionally introduced into certain immunoglobulin gene segments in a controlled manner. In addition to the DNA translesion polymerases, the novel Pol μ of family X is indicated for a role in this process, since it is specifically expressed in lymphoid tissues (14, 13).

Human Pol λ is an enzyme of family X related to Pol β that is involved in DNA repair. The latest, but still controversial family X DNA polymerase member is Pol σ . Eukaryotic organisms seem to possess multiple related genes for Pol σ proteins that are at least in part redundant. Pol σ enzymes execute an essential function in the sister chromatid adhesion that links the newly synthesised sister DNA strands until mitosis, but yeast as well as human Pol σ proteins apparently also have divergent functions as poly-A polymerases in the cytoplasm (13, 2, 15). Pol θ of family A is found in higher eukaryotes and is unique in that it contains a putative **helicase**, a 3'→5' exonuclease and a DNA polymerase domain in one polypeptide chain. Pol θ is probably involved in the repair of DNA interstrand cross-links.

Clinical Relevance

Human DNA replication and repair processes are highly co-ordinated and tightly controlled to maintain the stability of the genetic information. To interrupt cell-cycle progression after damage to the genome or replication stress is essential for survival of eukaryotic cells. Failure to induce a transient cell cycle arrest leads to accumulation of mutations that ultimately induce cell death or cancer. A variety of DNA lesions can damage the cellular genetic information and induce a complex set of mechanisms to repair them. The activation of these processes is the ultimate goal of the checkpoint systems. Once the integrity of the DNA is again

established, signals are sent to continue cell-cycle progression. In recent years it has become clearer that Pol α -prim and Pol ϵ appear to play a key role in co-ordinating DNA replication, DNA repair and cell-cycle checkpoints in addition to their enzyme activities during DNA replication. These replicative DNA polymerases appear to serve as sensors for the functional integrity of the replication apparatus (11, 1). In recent years, mutations of DNA polymerases such as Pols α , δ and η have been found in cancer tissues and cell lines suggesting that interference with DNA polymerase functions might cause genetic instability and then cancer. The increased cancer incidence in transgenic mice lacking the proofreading exonuclease of Pol δ and the identification of Pol η mutants as the underlying defect in xeroderma pigmentosum variant support this view (9, 13). Many DNA viruses target cellular DNA polymerases and other components of the replication apparatus to ensure viral replication. Pol α -prim, which interacts with replication factors of multiple viruses, is the major species-specific factor for polyoma virus DNA replication.

The posttranslational modifications of the central DNA replication enzyme are highly complex and only partially understood. However, the understanding of these signalling pathways will most probably allow establishment of new targets for finding new cancer drugs.

The contribution of different DNA polymerases to somatic hypermutation is not yet well understood. Mouse and cell models suggest that Pol μ and, to a lesser extent, Pol η are required for comprehensive immunoglobulin maturation (14, 13).

►Base Excision Repair

►DNA Helicases

►Nucleotide Biosynthesis

►Replication Fork

►Shotgun Libraries

►SNP Detection and Mass Spectrometry

References

- Hübscher U, Maga G, Spadari S (2002) Eukaryotic DNA polymerases. *Annu Rev Biochem* 71:133–163
- Nasheuer HP, Smith R, Bauerschmidt C et al (2002) Initiation of eukaryotic DNA replication: regulation and mechanisms. *Prog Nucleic Acid Res Mol Biol* 72:41–94
- Smith RW, Steffen C, Grosse F et al (2002) Species specificity of simian virus 40 DNA Replication *in vitro* requires multiple functions of human DNA polymerase α . *J Biol Chem* 277:20541–20548
- Melle C, Nasheuer HP (2002) Physical and functional interactions of the tumor suppressor protein p53 and DNA polymerase α -primase. *Nucleic Acids Res* 30:1493–1499
- Mäkinen M, Pospiech H, Kilpeläinen S et al (1999) A novel family of DNA-polymerase-associated B subunits. *Trends Biochem Sci* 24:14–16
- Zerbe LK, Kuchta RD (2002) The p58 subunit of human DNA primase is important for primer initiation, elongation, and counting. *Biochemistry* 41:4891–4900
- Cramer P, Bushnell DA, Kornberg RD (2001) Structural basis of transcription: RNA polymerase II at 2.8 angstrom resolution. *Science* 292:1863–1876
- Memisoglu A, Samson L (2000) Base excision repair in yeast and mammals. *Mutat. Res.* 451:39–51
- Goldsby RE, Lawrence NA, Hays LE et al (2001) Defective DNA polymerase- δ proofreading causes cancer susceptibility in mice. *Nat Med* 7:638–639
- Flohr T, Dai JC, Buttner J et al (1999) Detection of mutations in the DNA polymerase delta gene of human sporadic colorectal cancers and colon cancer cell lines. *Int J Cancer* 80:919–929
- Hübscher U, Nasheuer H-P, Syväoja JE (2000) Eukaryotic DNA polymerases, a growing family. *Trends Biochem Sci* 25:143–147
- Pospiech H, Syväoja J (2003) DNA Polymerase ϵ – More than a Polymerase. *The Scientific World Journal* 3:87–104
- Goodman MF (2002) Error-prone repair DNA polymerases in prokaryotes and eukaryotes. *Annu Rev Biochem* 71:17–50
- Bertocci B, De Smet A, Berek C et al (2003) Immunoglobulin kappa light chain gene rearrangement is impaired in mice deficient for DNA polymerase μ . *Immunity* 19:203–211
- Read RL, Martinho RG Wang SW et al (2002) Cytoplasmic poly(A) polymerases mediate cellular responses to S phase arrest. *Proc Natl Acad Sci USA* 99:12079–12084

DNA Primase

Definition

DNA primase is a ►DNA polymerase that requires a primer in order to begin incorporation of deoxyribonucleotides into the nascent DNA strand. A DNA primase primes DNA synthesis by incorporating ribonucleotides complementary to the DNA template to generate an RNA primer. This enzyme then extends the RNA primer by incorporating several deoxyribonucleotides. Later in the replication process, other proteins remove the RNA primer and replace it with deoxyribonucleotides.

►DNA Helicases

►Replication Fork

DNA Recombination

JEAN-MARIE BUERSTEDDE

Institute of Molecular Radiobiology, GSF,
Neuherberg, Germany
buersted@gsf.de

Definition

DNA recombination is a process in which specialized proteins interact with DNA to create molecules with altered base sequence content. Depending on the details of the reaction, the outcomes are deletions, duplications or simply a new order of allelic variation.

Recombination can be classified as site-specific or homologous. Site-specific recombinases recognize particular signal sequences within the target DNA molecules. In contrast, homologous recombination requires inter- or intra-molecular sequence homology to start the reaction.

DNA recombination is employed by all forms of life to change and reshuffle the genome. Whereas site-specific recombination is primarily used by viruses and **transposons** to enter and exit their host's genomes, homologous recombination enables meiotic exchanges between chromosomes and the segregation of beneficial and deleterious mutations. Most types of recombination reactions are related to **DNA repair** and require the participation of general repair factors. In return, homologous recombination helps to patch spontaneous and induced DNA damage by copying information from intact homologous **sister chromatids** or chromosomes.

Characteristics

As recombination revolves around the venerable DNA double helix, it comes as no surprise that it is itself highly conserved during evolution. Therefore, bacterial and simple eukaryotic organisms have been ideal models to explore the molecular basis of recombination reactions.

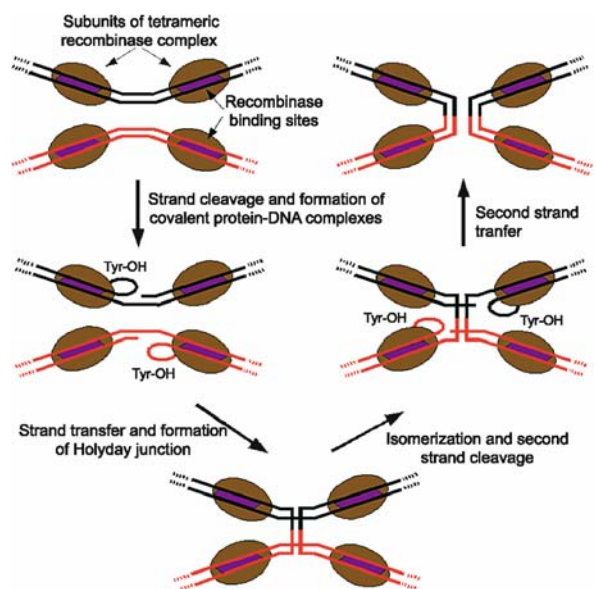
Site-specific recombination is initiated by the assembly of a multimeric protein complex around recognition sites in the target DNA, which then tightly control the timing and the directionality of the DNA **strand transfers**. The two best understood forms of site-specific recombination are **integration** and **transposition** catalyzed by tyrosine recombinases and transposases respectively.

Tyrosine recombinases are defined by a conserved catalytic domain and although their biological missions vary, the mechanism of action is unique for the whole family (1). At the beginning, the recombinase scans the target DNA for recognition sites, which consist of two

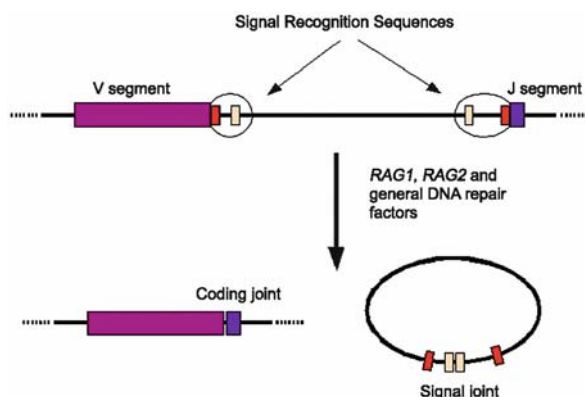
inverted recombinase-binding sites separated by a spacer. If a pair of recognition sites is detected, a tetrameric recombinase complex is assembled in which each of the recombinase subunits interacts with one of the binding sites. In the next step, a duo of coordinated strand transfers first produces a **Holiday junction** and then a complete helical exchange within the aligned spacer regions (Fig. 1). Intermediates in these tightly coordinated strand exchanges are recombinase-DNA phosphotyrosine linkages which preserve the energy of the cleaved DNA phosphodiester bonds.

Another paradigm of site-specific recombination is transposition (2). Like integration, the reaction is initiated by the assembly of a multimeric complex at the transposon DNA ends. The transposase within the complex then cleaves the 3' DNA ends of the transposon and induces a nucleophilic attack by the generated OH end groups against the target DNA. Some transposition reactions include the formation and cleavage of a hairpin at the transposon end as an intermediate. Interestingly, retroviral integration into the host chromosomes also occurs by a transposition-like mechanism.

Another variant of a transposition reaction is **antigen receptor gene rearrangement** (**V(D)J recombination**) in the immune system (3). The receptor genes are not stored in a functional form in the germline, but are assembled from families of receptor coding segments during lymphocyte development (Fig. 2). The boundaries of the rearranging gene segments are defined by conserved signal recognition sequences (SRS). Two lymphoid specific proteins, recombination activating



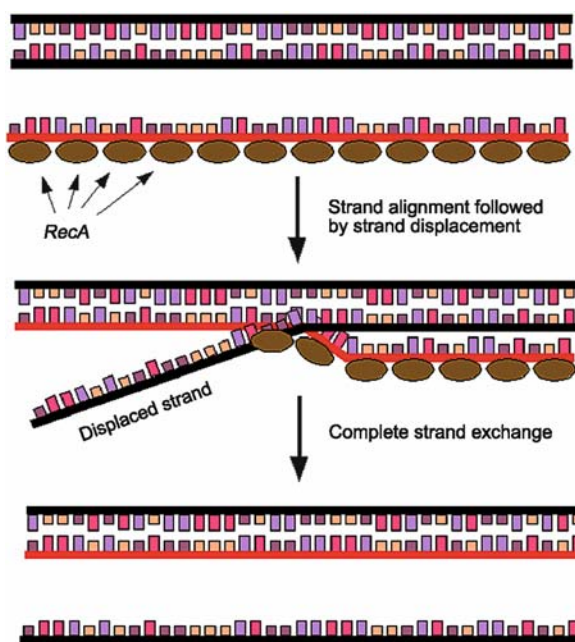
DNA Recombination. Figure 1 Site-specific integration catalysed by tyrosine recombinases.



DNA Recombination. Figure 2 V(D)J recombination at the murine immunoglobulin light chain locus.

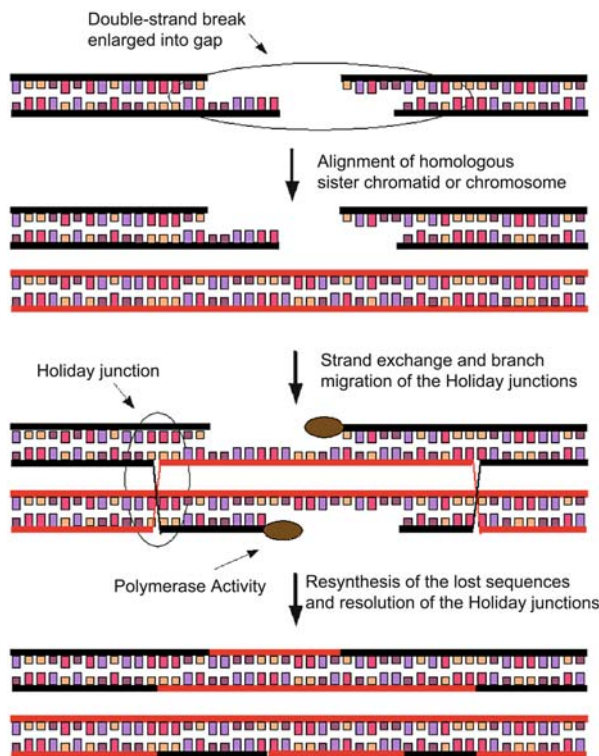
gene 1 (*RAG1*) and recombination activating gene 2 (*RAG2*) recognize the signal recognition sequences and introduce a nick precisely between the coding segment and the SRS. The liberated 3' OH group of the nicked strand then attacks and hydrolyzes the phosphodiester bond of the opposite strand resulting in a double strand break at the SRS and a ▶hairpin end at the coding segment. These intermediates are further processed with the assistance of general DNA repair factors like the DNA-protein kinase (DNA-PK) and the Artemis protein, which nick the hairpins of the coding segments. The end products are coding joints between the receptor coding segments and signal joints between the removed or inverted signal recognition sequences. V(D) J recombination assures a single receptor specificity for each lymphocyte, because as a rule only one of the two light and heavy chain receptor loci successfully rearranges. A second function is to contribute diversity to the antigen receptor repertoire by the combination of segments randomly chosen from gene segment families.

First insights into the mechanism of homologous recombination (4) were gained from studies of the bacterial *RECA* gene whose mutants are severely suppressed in most forms of homologous recombination. Although *RecA* is a versatile protein with a number of functions, the best characterized activity related to recombination is a strand exchange reaction in which a single stranded DNA sequence aligns with a homologous double-stranded molecule and replaces the identical strand of the duplex DNA (Fig. 3). As long as new base pairs are formed during the strand invasion, the reaction is energetically neutral and ATP hydrolysis is only needed to proceed through areas leading to ▶mismatches. Structural analysis indicates that *RecA* coats single stranded DNA with a regularity of 3 bases per monomer and facilitates homology



DNA Recombination. Figure 3 Strand exchange catalysed by the *RecA* protein.

search and strand exchange by stretching the molecule. If two homologous double helices interact, strand exchange is often reciprocal between the aligned sequences leading to Holiday junctions. At least in theory, Holiday junctions between homologous sequences can slide back and forth with minimal energy requirements, as base pairs are broken and formed at equal rates. Nevertheless, the RuvA and RuvB proteins are recruited to Holiday junctions and most probably force the branch migration into a certain direction. In the last stage of the reaction, the RuvC protein cuts the Holiday junctions and the recombined DNA molecules are relieved of their close encounter. Studies of homologous recombination in eukaryotic models offer insight into meiotic cell division and can account for the effects of chromatin structure. Mutant screens in the yeast *Sacharomyces cerevisiae* uncovered the so-called *RAD52* genetic pathway whose mutants are characterized by high radiation sensitivity, altered recombination frequencies and a block in meiosis. The mutant phenotypes in the *RAD52* pathway can be explained by a defect in the repair of double strand breaks that initiate mitotic and meiotic recombination in yeast. The observation that the *RAD52* pathway can repair even gaps in transfected genes using endogenous genes as templates led to the influential double-strand break repair model of recombination (5) which postulates that the ends of the broken strands invade the homologous partner



DNA Recombination. Figure 4 The double-strand break repair by homologous recombination model.

sequence leading to reciprocal strand displacement, branch migration and ultimately the formation of patch and slice recombinants (Fig. 4). Although the biochemistry of homologous recombination in eukaryotes still needs to be explored in greater detail, several genes in the *RAD52* pathway encode structural and functional homologues of the bacterial RecA protein and thus probably facilitate the initial strand exchange between the donor and the target sequences.

Structural homologues of the *RAD52* pathway in the forms of *RAD51*, *RAD52* and *RAD54* genes have also been found in vertebrates and at least the *RAD54* mutant shows the expected phenotype of radiation sensitivity and homologous recombination deficiency. However, the roles of the *RAD51* and *RAD52* genes have evolved in vertebrate cells as the *RAD51* gene is essential and disruption of the *RAD52* gene which produces the most severe recombination defect in *S. cerevisiae* has only a mild effect

Clinical Relevance

Homozygous null mutations in the *RAG1*, *RAG2* and *Artemis* (5) lead to severe combined immunodeficiency (SCID) due to lack of mature T and B cells whereas *RAG* mutations which only impair the protein

function manifest themselves with the milder Omenn immunodeficiency syndrome.

Defects in the homologous recombination pathway might manifest themselves by infertility due to a block in meiosis or by cancer predisposition due to DNA repair deficiency. However, these diseases have not been associated with mutations of the *RAD52* pathway and viable *RAD54* and *RAD52* ▶knockout mice are neither infertile nor do they develop cancer. There is however evidence that the human *Rad51* protein interacts with the proteins encoded by breast cancer genes *BRCA1* and *BRCA2* suggesting that an alteration in homologous recombination causes hypermutation or chromosomal instability in breast cancer cells of affected patients.

Whereas the jury is still out as to whether human diseases directly relate to homologous recombination defects, experimental medicine has taken a lesson from Mother Nature and adapted recombination to redesign the genome of living cells. Artificial gene constructs are introduced into cells and then clones are isolated which have integrated the construct by homologous recombination. This powerful approach is now combined with embryonic stem cell technology to study gene function by disruption and to create animal models for human disease. Remarkably, site-specific recombination has also found a place in the toolbox of molecular biologists to regulate gene expression and to simulate chromosomal translocations (6). Homologous recombination offers great promise for gene therapy, because it can correct genetic defects by the targeted integration of wild type constructs. All current strategies of stable gene therapy integrate transgenes at random chromosomal positions and suffer from problems of insertional mutagenesis, variable gene expression and possible dominant effects of the causative mutation. Although the frequency of targeted integration events is still too low for *in vivo* applications, it might be possible to increase it or to expand selected stem cell transfectants for the repopulation of the patient.

▶Chromosomal Instability Syndromes

▶DNA Helicases

References

1. Subramaniam S, Tewari AK, Nunes-Duby SE et al (2003) Dynamics and DNA substrate recognition by the catalytic domain of lambda integrase. *J Mol Biol* 329:423–439
2. Kennedy AK, Haniford DB, Mizuuchi K (2000) Single active site catalysis of the successive phosphoryl transfer steps by DNA transposases: insights from phosphorothioate stereoselectivity. *Cell* 101:295–305
3. Gellert M (2002) V(D)J recombination: RAG proteins, repair factors, and regulation. *Annu Rev Biochem* 71:101–132

4. West SC (2003) Molecular views of recombination proteins and their control. *Nat Rev Mol Cell Biol* 4:435–445
5. Moshous D, Callebaut I, de Chasseval R et al (2001) Artemis, a novel DNA double-strand break repair/V(D)J recombination protein, is mutated in human severe combined immune deficiency. *Cell* 105:177–186
6. Yu Y, Bradley A (2001) Engineering chromosomal rearrangements in mice. *Nat Rev Genet* 2:780–90

DNA Relaxation

Definition

DNA relaxation designates the removal of negative or positive supercoils from DNA.

►DNA Ligases

DNA Repair

Definition

DNA repair refers to the process of identification and repair of spontaneous or induced DNA damage by enzymatic reactions. DNA repair is essential to protect the integrity of the genome, and prevents somatic mutation that can lead to cancer.

►Base Excision Repair
 ►DNA Helicases
 ►DNA Recombination
 ►Double-Strand Break Repair
 ►Repeat Expansion Diseases
 ►DNA Repair Mechanisms

DNA Repair Mechanisms

WOLFGANG GOEDECKE
 Institute for Genetics, University of Essen,
 Essen, Germany
 wolfgang.goedecke@uni-essen.de

Definition

Many chemical and physical agents are capable of altering the chemistry and structure of DNA (1). DNA damaging processes occur in every cell at all times and

do not only result from environmental stress but mainly from intrinsic metabolic activities. If such lesions in DNA are not repaired, mutations may arise. In general, DNA repair reactions comprise three steps. In the first step, the damaged site is recognised and marked. The second step involves the elimination of the lesion, and during the third step, the original sequence is restored by DNA synthesis and subsequent single-strand ligation. Cells with deficient DNA-repair systems show increased levels of ►genome instability, leading to accumulations of various types of mutations (mutator phenotype). Depending on the repair mechanism affected, the mutational spectra of these cells may exhibit significant increase in spontaneous base pair substitutions, insertions and deletions, or even chromosomal rearrangements. In many cases, acquired genomic instability observed in tumour cells results from defects in the DNA repair systems, which illustrates the ►caretaker function of DNA repair genes for the maintenance of genome stability (2). These caretaker functions are also consistent with the observed phenotypes of human syndromes known to be associated with DNA repair deficiencies (see below) (3).

Characteristics

As long as only one DNA strand is damaged, the remaining undamaged strand can be used as a template for new DNA repair synthesis. Repair mechanisms that take advantage of an undamaged template include postreplicative ►mismatch repair (MMR), ►base excision repair (BER), and ►nucleotide excision repair (NER). The repair of damage that affects both strands of the DNA molecule, e.g. a double strand break (DSB) or an interstrand cross link (ICL), is more difficult. In such cases, the original sequence information can be restored by ►homologous recombination repair (HRR). If HRR cannot be used, the two sugar phosphate backbones of the double helix can be covalently closed by a mutagenic process called ►nonhomologous end joining (NHEJ) (4).

MMR

The misincorporation of deoxyribonucleotide triphosphates during DNA replication leads to mismatched base pairs. Other frequent replication errors are loop structures induced by slippage of the newly synthesised DNA strand relative to the template strand on sequences containing short repeat units in tandem (microsatellites). If not repaired, such lesions will generate either base pair substitutions (in the case of base pair mismatches) or insertions or deletions (in the case of strand slippage) in one of the daughter strands after another round of DNA replication and sister chromatid segregation. Postreplicative MMR provides a tool to remove such replication errors, because it

ensures the removal of the mispaired nucleotides exclusively from the newly synthesised strand while the parental “correct” strand remains intact and serves as a template for repair synthesis.

BER

Mutations may also arise as a consequence of chemical modification or elimination of the normal DNA bases by spontaneous deamination ($\sim 100 \text{ cell}^{-1} \text{ day}^{-1}$; e.g. cytosine \rightarrow uracil; adenine \rightarrow hypoxanthine), oxidation ($\sim 20,000 \text{ cell}^{-1} \text{ day}^{-1}$; e.g. guanine \rightarrow 8-oxo-guanine), or depurination ($\sim 10,000 \text{ cell}^{-1} \text{ day}^{-1}$; loss of adenine or guanine). In this way, base derivatives with altered base pairing properties or apurinic sites are formed. If not repaired, these types of lesions may lead to base pair substitutions in the next S-phase. Chemically modified bases are excised by DNA glycosylases leaving behind apurinic or apyrimidinic (AP-) sites. Several different DNA glycosylases exist, each of which can recognise one specific type of lesion. In the subsequent repair steps, the AP-site is recognised by an AP-endonuclease and a deoxyribosephosphodiesterase, which opens the phosphate backbone and removes the baseless sugar moiety to generate a new 3'-OH group that can prime the DNA polymerase-catalysed incorporation of one or a few deoxyribonucleotide triphosphates.

NER

Lesions in which chemical base modifications are combined with a distortion of the double helix are called bulky adducts. A typical example of this type of lesion is UV-light-induced covalent crosslinks between adjacent pyrimidines (e.g. thymine-dimers). Such bulky adducts represent major obstacles to DNA replication and transcription. They are recognised by distinct multi-protein complexes that are, depending on the state of transcription of the DNA, composed of several different polypeptides (products of the XP-, CS-, and TTD-gene-groups, see below). Lesions in the overall chromatin are repaired at a slower rate and reduced efficiency (global genome repair; GGR) as compared to the transcribed chromatin (transcription coupled repair; TCR). This ensures that active genes are repaired with higher priority than inactive genes or non-coding DNA. Both pathways share common downstream activities, characterised by incision steps on the 5'- and 3'-site of the lesion, removing a single-stranded oligonucleotide of 24 up to 32 nucleotides that contains the lesion. The remaining single-strand gap is subsequently filled in and ligated.

HRR

In cases where both DNA strands are damaged (e.g. DSB), HRR can use the undamaged sister chromatid or homologous chromosome as template and is, therefore,

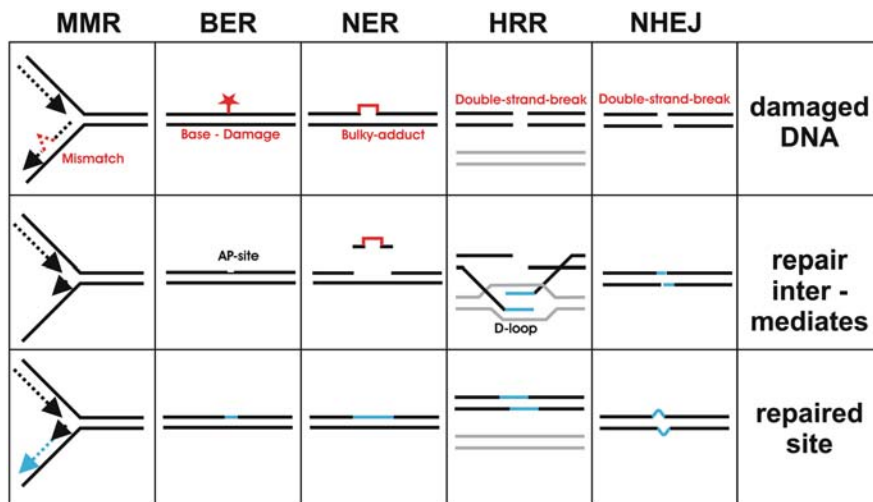
capable of accurate restoration of the original sequence information. In the mitotic cell cycle, this pathway is predominantly used during the S- and G2-phase, when the sister chromatid can serve as a template for repair. The homologous chromosome is usually not used during the normal mitotic cell cycle, but becomes a template for HRR during prophase I of meiosis. HRR is initiated by the invasion of the single-stranded 3'-ends of the broken DNA molecule (recipient) in the homologous sequences of the undamaged double-strand (donor), which is catalysed by homologous pairing proteins (Rad52 and Rad51). The 3'-ends in the resulting displacement (D-) loop serve as primers for DNA synthesis on the intact DNA molecule. Downstream activities include branch migration of the resulting Holliday junction (HJ), DNA synthesis, ligation, and HJ-resolution by endonucleases.

NHEJ

Since DSB are potentially lethal DNA lesions, it is obvious, that there is a need for repair mechanisms during G1 or G0 where no sister chromatids are available to stimulate HRR. In this stage of the cell cycle, DSB are rejoined by NHEJ, an activity that can operate in the absence of sequence homology but preferentially uses sequence homology of fortuitously complementary bases whenever available. The NHEJ mechanism includes the alignment of the broken DNA ends by a protein complex and their conversion into a ligatable structure. It is obvious that such a mechanism is prone to introduce mutations, because trimming of the DNA ends often includes degradation or insertion of a few nucleotides before the remaining nicks are sealed by a DNA ligase. It is worth mentioning that the same NHEJ mechanism is responsible for the generation of the diversity of the immunoglobulins by a process called V(D)J-recombination which occurs during B- and T-cell development in the mammalian immune system.

Genetics

As outlined above, DNA repair genes have a caretaker function to maintain genome stability. A reduction in DNA repair capacity leads to an increase in the mutation rate in the affected cell; the cell develops a so-called mutator phenotype. The known accumulation of mutations in tumour cells may be explained by a two-step mechanism. First, a cell becomes deficient in a certain DNA repair function. As a result of the increased mutation rate other genes, the so-called **gatekeepers**, may be affected. Gatekeepers are represented by oncogenes and tumour suppressor genes that regulate cell division (cell cycle checkpoints), cell differentiation, and cell death (apoptosis). If mutated, these in combination with other genes promote the further development into a cancer cell. This may



DNA Repair Mechanisms. Figure 1 The figure shows repair mechanisms as discussed in the text. Double stranded DNA molecules are given as a black double-line. The Y-shaped structure symbolises the replication fork and the grey double line at the HRR mechanism indicates the homologous DNA molecule used as template for repair synthesis. In the first row, different types of damaged DNA are indicated in red. The second shows typical intermediate structures occurring during the repair process itself. The bottom line shows the repaired site with newly synthesised DNA due to repair synthesis given in blue. The bent blue line in the NHEJ mechanism indicates that this pathway is mutagenic due to loss of base pairs and mispaired bases remaining left at the junction.

explain why a common feature of many, albeit not all, human syndromes associated with defective DNA repair is a predisposition to cancer.

Information about human DNA repair genes comes on the one hand from the study of cell lines deficient in DNA repair which are derived from patients with spontaneous or inherited tumour diseases. The other approach uses the information resulting from the human genome project, to compare DNA sequences with DNA repair genes already known from simpler organisms like bacteria and yeast. Using this latter approach, about 125 genes have been identified, which are directly involved in DNA repair (5, 6).

Clinical Relevance

In humans, no disease is known so far which affects genes involved in BER. This may reflect the fact, that deficiencies in the BER system are not compatible with life. Since DNA bases are permanently chemically damaged in the normal cellular metabolism, deficiencies in BER probably cause lethality.

Deficiencies in DNA repair pathways in humans are genetically characterised by inherited syndromes with more or less distinct phenotypes. Deficiencies in the postreplicative MMR were found to be associated with **hereditary nonpolyposis colon cancer (HNPCC)**. Patients suffering from this disease are heterozygous for genes involved in the recognition of mismatches or loop structures (MLH1, MLH6, MLH1 and PMS2) generated during DNA replication. Tumours from these

patients lack MMR activity, indicating that they became homozygous due to loss of heterozygosity or a second mutation in the homologous gene. In agreement with the MMR mechanism, tumours from HNPCC patients are further characterised by microsatellite instabilities, which can be explained by the lack of repair of single-stranded loops generated during replication.

Three syndromes are associated with deficiencies in NER function, **xeroderma pigmentosum (XP)** with seven complementation groups (XP-A to XP-G), **Cockayne syndrome (CS)** with two complementation groups (CS-A and CS-B) and **trichothiodystrophy (TTD)** with three complementation groups (TTD, XP-B or XP-D). The fact, that mutations in XP-B or XP-D may cause the XP- or TTD-phenotype explains the importance of identifying the nature of the mutation causally connected with the disease. Surprisingly, cancer predisposition due to exposure to UV-light is associated with the XP, but not with CS or TTD. Another interesting point is the existence of combined phenotypes of XP and CS, caused by certain mutations in XP-B, XP-D and XP-G.

There is strong evidence, that the genes associated with familiar forms of **breast cancer BRCA1 and BRCA2** are involved in HRR. Vertebrate cells deficient in HRR have severe problems in traversing S-phase and reinitiating replication at sites where replication forks have collapsed. The mechanistic relation of these two genes with HRR is not entirely understood but may be

due to the binding of both proteins to RAD51. Like its bacterial counterpart *recA*, RAD51 is involved in D-loop formation, a key step in HRR (see above). One of the genes causing the autosomal recessive **Fanconi anaemia** (FA) (FANC-D2) was recently shown to interact with BRCA1 suggesting problems in HRR in patients with FA as well. Overall, eight FA complementation groups are known (FA-A to FA-H). FA patients are predisposed to acute myeloid leukaemia and squamous cell carcinomas.

Mutations in LIG4, an NHEJ-specific DNA ligase, are found in cells from patients with delay in growth and development combined with immunodeficiency. As well as their involvement in DNA repair, the components of the NHEJ machinery also play an important role in V(D)J-recombination, which explains the complex phenotype of these patients.

Ataxia telangiectasia (AT) is a disease caused by mutation in the ATM gene, which is involved in the sensing and signalling of DSB. It is not entirely clear, if this mutation has consequences for HRR, NHEJ, or both pathways. Interestingly, there exists another disease with a related phenotype, the **ataxia telangiectasia-like disease** (AT-LD), which is caused by a truncated MRE11 gene product. This protein is supposed to act as a nuclease in HRR and NHEJ.

Chromosome Instability Syndromes

References

1. Ames BN, Gold LS, Willett WC (1995) The Causes and Prevention of Cancer. *PNAS* 92:5258–5265
2. Loeb KR, Loeb LA (2000) Significance of multiple mutations in cancer. *Carcinogenesis* 21:379–385
3. Cahill DP, Kinzler KW, Vogelstein B et al (1999) Genetic instability and Darwinian selection in tumours. *Trends Cell Biol* 9:M57–M60
4. Hoeijmakers JH (2001) Genome maintenance mechanisms for preventing cancer. *Nature* 411:366–374
5. Ronen A, Glickman BW (2001) Human DNA repair genes. *Environ Mol Mutagen* 37:241–283
6. Wood RD, Mitchell M, Sgouros J et al (2001) Human DNA repair genes. *Science* 291:1284–1289

DNA Replication

Definition

DNA replication refers to the duplication process by which an accurate copy of the DNA is made. The complementarity of base pairing between the two strands of the double-helix forms the basis for DNA replication. Most often, replication precedes cell division by mitosis. Notable exceptions include the DNA replication that precedes meiosis, and the

replication that occurs in cells undergoing endoreplication as a way to amplify gene copy number in a cell.

- DNA Helicases
- DNA Replication Initiation
- Replication Fork
- Chromosome Remodeling

DNA Replication Initiation

MASATOSHI FUJITA

Virology Division, National Cancer Center Research Institute, Chuohku, Tokyo, Japan
mafujita@gan2.res.ncc.go.jp

Definition

Genomic DNA, which is distributed into multiple chromosomes in eukaryotes, has to be replicated completely and only once during a single cell-cycle to maintain integrity. The necessary cell-cycle regulation is mainly executed at the initiation rather than the elongation step of DNA replication. Our knowledge of mechanisms responsible in eukaryotes largely depends on findings with budding yeast, since identification of specific DNA sequences acting as replication origins on plasmids, autonomously replicating sequences (**ARSSs**), has allowed extensive analysis of the initiation step. Several factors essential for regulation of initiation have been identified, including **ORC** (origin recognition complex), CDC6, Cdt1 and **MCM**. These players in DNA replication initiation are strictly controlled by **Cdks** (cyclin-dependent kinases). Subsequent work has suggested that the fundamental machinery for DNA replication is conserved in metazoans including mammalian cells. In this essay, current concepts of cell-cycle regulation of DNA replication in mammalian cells and their clinical relevance are summarized. For further reading, see reviews cited in the references (1, 2, 3).

Characteristics

Pre-Replication Complex Formation During G1 Phase

To initiate genome duplication, a multi-protein complex, termed the pre-replication complex (pre-RC), is generated during the G1 phase based on ORC binding to chromosomal DNA. ORC was initially identified in budding yeast as a protein complex binding sequence-specifically to ARSSs, and later found to be conserved in mammalian cells. Unlike yeast however, the interaction of mammalian ORC with chromosomal DNA is not simply determined by the primary DNA sequence and ORC localization may be influenced by high-order chromatin/nuclear structures associated with transcription (2). It is also suggested that chromatin regions

affixed to the nuclear matrix could more easily access the ORC, since functional ORC is also associated with the matrix (4). These imply a complex architecture of mammalian replication origins.

CDC6 and Cdt1 proteins are recruited to the origin possibly by physical interaction with ORC. The resultant machinery, consisting of ORC, CDC6 and Cdt1 proteins, may function as a loader for the MCM heterohexameric complex, which could function as a helicase in the initiation as well as the elongation steps. In an *in vitro* DNA replication system with *Xenopus* egg extracts, once MCMs are loaded, ORC/CDC6 is no longer required for DNA replication initiation. Therefore, replication initiates at sites onto which MCM complexes are loaded. In the case of budding yeast, these appear to be near ORCs bound to ARSs. As a consequence, the ARS acts as an initiation site as well as a *cis* element required for assembly of the initiation machinery. In mammalian cells, it appears that multiple MCM complexes are loaded beyond each ORC site (1, 3, 4). Therefore, it is possible that initiation sites are not always close to an ORC binding site. While such a broad distribution of MCMs on pre-replication chromatin seems appropriate for explaining the initiation zone observed in mammalian cells, the issue also makes the replication origin problem complicated. Finally, it is worth noting that transcription of ORC1, CDC6 and all MCM subunits is driven by the ►E2F transcription factor.

Firing of the Pre-Replication Complex During S phase

When cells enter S phase, multiple other proteins or protein complexes need to be recruited onto the pre-RCs to initiate DNA synthesis. This step is dependent on two kinases, cyclin-dependent kinase (cyclin E/Cdk2 and cyclin A/Cdk2 in mammals) and CDC7 kinase. Cdk activity is absolutely required for promoting DNA replication, yet the proteins that must be phosphorylated remain largely unknown. Cdks physically interact with and phosphorylate ORC and CDC6. However, considering that ORC and CDC6 are no longer necessary for initiation after loading MCMs, such phosphorylation is unlikely to be involved in promotion of replication. Rather, it may contribute to prohibition of re-replication (see below). Certain MCM subunits in the hexameric complexes are also phosphorylated by Cdk2 kinase, but the outcome is not clear at present (also see below).

Prior to the DNA unwinding step, CDC45, an essential replication protein, is loaded onto chromatin, probably *via* physical interaction with the MCM complex. Since the loading is dependent on both Cdk and CDC7 kinase activities and both these kinases phosphorylate certain subunits of the MCM complex, one might conclude that phosphorylation of the latter leads to stable physical association with CDC45. However, there is

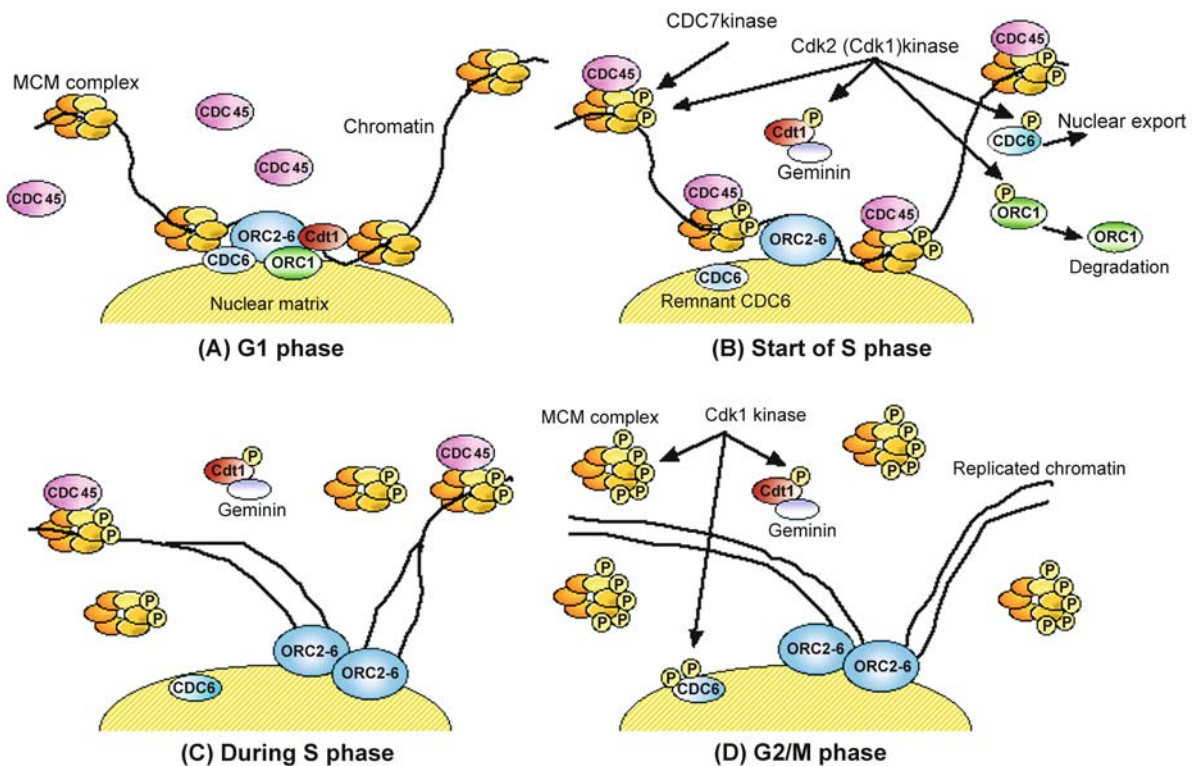
no evidence of this at present. It has been reported that the MCM complex bound to CDC45 shows stronger helicase activity than when lacking CDC45. On the other hand, an *in vitro* study has demonstrated that Cdk modification of the MCM complex rather reduces its helicase activity. Whatever the case, it seems clear that Cdk and CDC7 kinase activity-dependent loading of CDC45 is a prerequisite for the unwinding of DNA. Once CDC45 is loaded, the DNA is unwound with the help of RPA, a eukaryotic single-strand DNA binding protein complex. Subsequently, many components of the DNA synthetic machinery, such as ►DNA polymerases α , δ and ϵ , are assembled onto the unwound DNA, starting DNA replication. DNA polymerase ϵ complex consists of many subunits in addition to the catalytic subunit. Very recently, it was shown in budding yeast that Cdk-phosphorylation of Sld2 protein, a subunit of the DNA polymerase ϵ complex, is essential for chromosomal DNA replication (5). This is the first protein to be clearly demonstrated to promote replication by Cdk phosphorylation. It is clearly of interest to examine whether such machinery is also conserved in human cells.

Prevention of Re-Replication by Cdks

Genomic DNA is tightly controlled so that re-replication does not occur. After functioning in DNA replication, the MCM complexes are displaced from chromatin through an unknown mechanism(s). Therefore, re-establishment of pre-RC, in other words re-binding of the MCM should be suppressed during S, G2 and M phases of the cell-cycle. Recent studies suggest that Cdks play a central role in this context. Thus, Cdk activity has a bipartite function in the cell-cycle regulation of eukaryotic DNA replication.

In mammals, cyclin E/Cdk2 and cyclin A/Cdk2 are the S phase promoting Cdks, as already mentioned, while cyclin B/Cdk1 is the mitotic Cdk. In addition, cyclin A/Cdk1 appears active during S, G2 and M phases, although the biological consequences have not been clarified. Direct evidence is lacking, but Cdk2 kinase has been suggested to prevent formation of pre-RC. The importance of Cdk1 kinase is clearly demonstrated by the fact that its inactivation results in re-binding of MCM proteins and subsequent re-replication (1). Since Cdk2 kinase activity is still maintained in cells lacking Cdk1, the latter presumably plays the primary role in prevention of re-replication. If this is the case, the main contribution of cyclin A/Cdk1 would be inhibition of re-replication in S and G2 phases.

How do Cdks prevent re-establishment of pre-RC? One means is phosphorylation of CDC6, leading to degradation in yeast or nuclear export in mammalian cells. In the latter situation, cyclin A/Cdk2 may be a responsible kinase. Cyclin B/Cdk1 can phosphorylate CDC6 as



DNA Replication Initiation. Figure 1 A model for the state of pre-replication chromatin and cell-cycle regulation in mammalian cells. Note that the phosphorylation states in this model are figurative. (A) In the G1 phase with low Cdk2/Cdk1 activity, ORC and CDC6 are associated with the matrix, together with Cdt1 and load MCM complexes onto chromatin beyond ORC binding sites. (B) and (C) When cells enter S phase, dependent on Cdk2 and CDC7 kinase activity, CDC45 and some other proteins are recruited around MCM and they unwind DNA. Then, DNA synthetic proteins are assembled on single-stranded DNA. The latter step is omitted in this model. Activated MCM plays an essential role in DNA replication, possibly as a DNA helicase and is simultaneously displaced from chromatin through an unknown mechanism. In S phase, reloading of dissociated MCM may be suppressed by Cdk2 kinase, at least through CDC6 nuclear export and ORC1 degradation (and possibly inhibition of Cdt1 function). Alternatively, cyclin A/Cdk1 kinase may be responsible for the inhibition. Geminin also prevents the MCM rebinding by sequestering Cdt1. (D) During the G2/M phase, Cdk1 kinase hyperphosphorylates CDC6 and MCM complexes (and possibly Cdt1 and ORC1), thereby prohibiting the reloading of MCM. During exit from mitosis, Cdk activity is restrained and the pre-replication chromatin structure is again constructed.

well. ORC is another target of Cdk modification. In human cells, ORC1 is degraded when cells enter into S phase. As it appears to be phosphorylated by cyclin A/Cdk2, it could be that this is the trigger leading to its degradation. In budding yeast, ORC2 is also the target of Cdk phosphorylation. Cdk modification controls the function of the MCM proteins and it has been shown that MCM2 and MCM4 in the complex are phosphorylated by Cdk1 kinase, inactivation resulting in MCM re-binding to chromatin. During exit from mitosis, Cdk activity is restrained through destruction of the cyclin by ▶APC (anaphase promoting complex) and the pre-replication chromatin structure is again constructed. As described above, Cdks appear to employ at least three redundant mechanisms, that is inhibition of ORC, CDC6 and MCM activity, to prevent re-replication. Is it necessary to block all three pathways for induction of

re-replication without inhibiting Cdk activity? In budding yeast, the answer is clearly “yes” (3). Also in mammalian cells, alteration of the regulation of the individual components, for example over-expression of the phosphorylation-deficient CDC6 mutant, fails to induce re-replication, consistent with the findings in budding yeast. Although over-expression of fission yeast CDC6 causes re-replication, it could be a consequence of inactivation of Cdk by CDC6 over-expression.

Prevention of Re-Replication by Geminin: Is It Independent of Cdk Activity?

In metazoans, ▶geminin has been identified as another inhibitor of pre-RC formation. This protein prevents the loading of MCM proteins onto chromatin by binding to and inhibiting Cdt1, appears after cells enter S phase and is destroyed during exit from mitosis in an

APC-dependent manner. Thus, geminin may contribute to prevention of re-replication during S, G2 and M phases. It has been suggested that its action is not mediated by elevated Cdk activity, which would appear favorable in some situations; even when Cdk activity is down-regulated by the checkpoint mechanism in cells undergoing DNA damage, re-formation of pre-RC could be prevented. On the other hand, it has been clearly demonstrated that inactivation of Cdk1 activity results in re-binding of MCM proteins and subsequent re-replication. Therefore, it remains possible that geminin functions are also under the control of Cdks. This point need to be further addressed in future.

Clinical Relevance

It is easily imaginable that studies of DNA replication initiation proteins will have a major impact on an understanding of cancer cell biology. However, to date, data are limited compared to those for upstream regulators such as cyclin/Cdks and E2F, whose involvement in cancer has been well established. In general, initiation proteins controlled by the E2F transcription factor like ORC1, CDC6 and MCMs, are over-expressed in tumor cells. Our group has recently found that Cdt1 protein levels are also increased in tumor cell lines. However, the biological significance of over-expression of these proteins has yet to be well addressed. Interestingly, it has now been proposed that over-expression of Cdt1 can enforce re-replication in certain tumor cell lines, suggesting an involvement in genomic instability (6 and our unpublished data). Cell responses to CDC6 over-expression are still controversial, and thus further experimentation is required.

As described above, geminin plays a negative role in regulation of DNA replication. Thus, it is conceivable that its enforced expression would inhibit tumor cell growth and a recent report provided evidence that this is indeed the case. It is also conceivable that the function of geminin might be impaired, for example by mutation of its gene, in some tumors, a possibility that should undoubtedly be addressed in future. If this were the case, geminin would be a new tumor suppressor gene.

Whatever the importance to carcinogenesis, over-expression of initiation proteins could have potential for cancer diagnostics, as shown by our group for MCM7 (7). Several subsequent studies by others support this notion. Since Cdt1 over-expression might play a role in carcinogenesis, its detection might be more significant for cancer diagnostics.

References

1. Fujita M (1999) Cell cycle regulation of DNA replication initiation proteins in mammalian cells. *Front Biosci* 4:816–823
2. Tsurumi T, Fujita M (2002) Molecular basis for Epstein-Barr Virus DNA replication. *Curr Top Virol* 2:85–100
3. Bell SP, Dutta A (2002) DNA replication in eukaryotic cells. *Annu Rev Biochem* 71:333–374
4. Fujita M, Ishimi Y, Nakamura H et al (2002) Nuclear organization of DNA replication initiation proteins in mammalian cells. *J Biol Chem* 277:10354–10361
5. Masumoto H, Muramatsu S, Kamimura Y et al (2002) S-Cdk-dependent phosphorylation of Sld2 essential for chromosomal DNA replication in budding yeast. *Nature* 415:651–655
6. Vaziri C, Saxena S, Jeon Y et al (2003) A p53-dependent checkpoint pathway prevents rereplication. *Mol Cell* 11:997–1008
7. Hiraiwa A, Fujita M, Nagasaka T et al (1997) Immunolocalization of hCDC47 protein in normal and neoplastic human tissues and its relation to growth. *Int J Cancer* 74:180–184

DNA Structure

Definition

DNA, desoxyribonucleic acid, consists of a long, unbranched polymer composed of four types of subunits, the deoxyribonucleotides. Each “monomer” consists of a 5-carbon sugar (deoxyribose), a nitrogen containing base attached to the sugar, and a phosphate group. The bases can be either the ►purin bases adenine (A), and guanine (G), or the ►pyrimidine bases cytosine (C), and thymine (T). The corresponding complexes with sugar residues are called ►nucleosides (adenosine, guanosine, cytidine, thymidine). The nucleotides are linked together by covalent phosphodiester bonds, that connect the 5' carbon of the deoxyribose group to the 3' carbon of the next nucleotide. In the 1950s, James Watson and Francis Crick proposed and described an hypothetical structure for DNA - the DNA double helix: while the alternating ribose sugar molecules and phosphate groups form the outer backbone of the polynucleotide strand, nitrogenous bases are “inside” like rungs on a ladder. Two of those strands of DNA wrap around each other in an antiparallel fashion and form a right-handed double helix. Adenine on one side pairs with thymine on the other by hydrogen bonding, and cytosine pairs with guanine (►complementary base-pairing). The C-G pair is stabilized by three hydrogen bonds while the A-T pair has only two hydrogen bonds, which avoids wrong pairing. The double helix has a 2-fold axis of symmetry. There is a wide (major) and a narrow (minor) groove between the backbones on opposite strands. Purine and pyrimidine rings are spaced 0.34 nm apart, and have intermolecular base-pairs in the center. The bases are stacked one on top of the other which is very important for stability.

The DNA structure model, proposed by Watson and Crick in 1953, has been verified in many different ways and its identification is recognized as one of the most significant breakthroughs in modern science, because it explains how the chain of nucleotides can encode instructions for an organism; the genetic information is carried in the linear sequence of the nucleotides and its variability along the molecule can account for the variety in the genetics code. The model also explains how the genetic information can be copied from one generation (of cells) to the next. Duplication of the genetic information occurs by the polymerization of a new complementary strand onto each of the old strands of the double helix during ►DNA replication.

One important feature of the DNA double helix regarding its function is that all the information in it is redundant. If one base is lost, the complementary base on the opposite strand still contains the information. This is the basis of one strand DNA repair. In this process a damaged or missing base is replaced using the information on the opposite strand.

►DNA replication

►DNA repair mechanisms

DNA Topoisomerases

JOHN L. NITISS, ALEXANDER STEPANOV

Department of Molecular Pharmacology,
St. Jude Children's Research Hospital, Memphis,
TN, USA

john.nitiss@stjude.org

Synonyms

Nick-closing enzymes; ω protein (specifically for *E. coli* DNA topoisomerase I); DNA gyrase (refers to a sub-family only)

Definitions

DNA topoisomerases are a class of enzymes that modulate DNA topology by the transient introduction of DNA strand breaks. Double-stranded DNA provides considerable advantages for organisms. Double stranded DNA provides critical redundancy that can serve as a template to repair damaged DNA. The finding by Meselson and Stahl that DNA replication is semi-conservative, i.e. that each daughter cell receives one parental and one newly replicated strand indicated that cells must manage to completely separate the DNA strands during the process of DNA replication and chromosome segregation. The problem of separating DNA strands after replication was better appreciated

following the discovery by Cairns that the *E. coli* chromosome is circular. While it could be imagined that the two strands of a linear double stranded DNA molecule could separate by rotation, this is impossible with a circular double stranded molecule. Although linear double stranded DNA molecules are not formally subject to topological constraints, in practice, a long (chromosome sized) DNA molecule poses the same topological problems as a circular molecule. The solution to this topological problem posed by double stranded DNA is a class of enzymes termed DNA topoisomerases. These enzymes resolve different DNA structures by the introduction of transient DNA strand breaks, and then carry out strand passage reactions that result in changes in DNA topology. This article provides a brief overview of the different topoisomerases found in prokaryotic and eukaryotic cells, their enzyme mechanisms and some of their important biological roles. Recent reviews describe enzyme mechanisms, topoisomerase protein structure and the biological roles of topoisomerases (1, 2).

Characteristics

The problem of separating replicated DNA strands was originally termed the swivel problem, since the problem was first envisioned as requiring the unwinding of parental DNA strands during replication. James C. Wang identified a protein from *E. coli* that had the ability to nick DNA but also was capable of resealing the break following a change in DNA topology. These enzymes were termed DNA topoisomerases because their principal reactions are to change DNA topology and the first discovered enzyme was named *E. coli* topoisomerase I. DNA topoisomerases were first identified in eukaryotic cells by Dulbecco and Champoux. Early studies showed that the eukaryotic enzymes had substantially different properties from the *E. coli* topoisomerase I, notably, the eukaryotic enzyme could efficiently relax both positively and negatively supercoiled DNA, while the prokaryotic enzyme was unable to relax positively supercoiled DNA. However, both the prokaryotic and eukaryotic enzymes introduced single strand breaks in DNA and are therefore termed type I topoisomerases. Since their structures and biochemical mechanisms are distinct, enzymes similar to *E. coli* topoisomerase I are termed type IA enzymes while the eukaryotic enzyme is termed a ►type IB topoisomerase. Both type IA and type IB enzymes are typically monomeric and carry out their reactions without a high-energy cofactor. Although originally found in eukaryotic cells, subsequent experiments have shown that the type IB enzyme is also encoded by some viruses such as pox viruses and is also found in many bacterial species (although not *E. coli*).

Gellert and co-workers identified a second topoisomerase from *E. coli* cells. Unlike *E. coli* topoisomerase I,

the enzyme discovered by Gellert could introduce negative supercoiling, hence the enzyme was named DNA gyrase. DNA gyrase makes double strand breaks in DNA, and is termed a type II enzyme. Type II enzymes are typically multimeric, with DNA gyrase consisting of a tetramer of two pairs of subunits termed *gyrA* and *gyrB*. The eukaryotic type II enzymes are dimers of identical subunits. Also unlike type I topoisomerases, type II enzymes require a high-energy co-factor such as ATP. Type II enzymes are also found in some viruses, including the T-even phages and are found in all eukaryotes. The type II enzymes from eukaryotes and T-even phages also make double strand breaks in DNA, but are unable to introduce negative supercoils into DNA.

More recent work has shown that eukaryotic cells also possess ▶[type IA topoisomerases](#). Hence, type IA, type IB and ▶[type IIA topoisomerases](#) are spread throughout all taxonomic kingdoms. Additionally, some unusual topoisomerases have been found in archaeobacteria, including reverse gyrase, a type I enzyme that can introduce positive supercoils in DNA and a novel type II enzyme termed Top VI. Since this enzyme has substantial structural differences from other type II topoisomerases, this class of enzymes has been termed type IIB topoisomerases. Type IIB topoisomerases have been found in plant somatic cells, and clearly play important roles. One of the subunits of the type IIB enzyme also appears in eukaryotic cells under a somewhat different guise. The ▶[Spo11](#) protein, which is required for introducing double strand breaks to initiate meiotic recombination is homologous to archaeobacterial type IIB enzymes and is thought to break DNA using a topoisomerase-like mechanism.

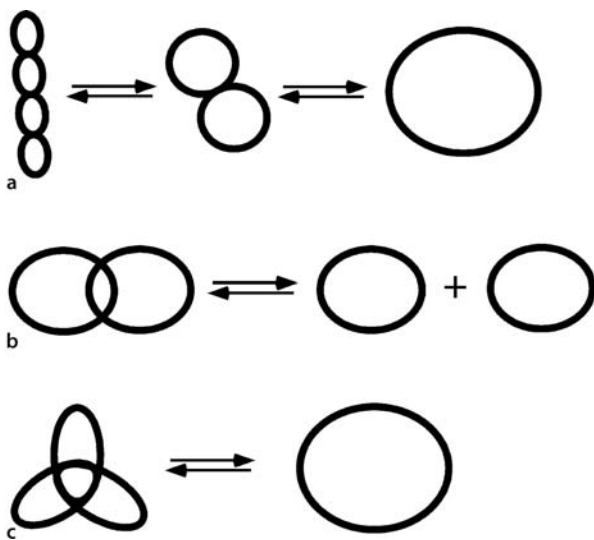
Reactions Carried Out by Topoisomerases

DNA topoisomerases carry out the reactions that change DNA topology by making a transient break in one or both DNA strands. DNA strand breaks are introduced by the reaction of tyrosine residue(s) of the topoisomerase with the phosphodiester backbone of DNA, resulting in the formation of a covalent phosphotyrosine bond between the enzyme and DNA, a transesterification reaction. An important aspect of cleavage *via* transesterification is its reversibility. After strand passage, a second transesterification reaction reforms the phosphodiester backbone of DNA and releases the active site tyrosine residue. Although all topoisomerases carry out a transesterification reaction with an active site tyrosine to cleave DNA, there are also clear differences between different classes of topoisomerases. The type IA and type II enzymes form 5' phosphotyrosyl linkages with DNA,

while the type IB enzymes form a 3' phosphotyrosyl linkage. Type II topoisomerases absolutely require a divalent cation in order to cleave DNA, while a divalent cation is completely dispensable for type IB enzymes. Interestingly, divalent cations are required for type IA enzyme strand passage, but are not required for DNA cleavage. Taken together, these considerations suggest important differences in the chemistry involved in DNA cleavage by different classes of DNA topoisomerases.

The introduction of DNA breaks by topoisomerases allows the enzymes to catalyze changes in DNA topology as illustrated in Fig. 1. The canonical reaction of topoisomerases, relaxation of supercoiled DNA is carried out by both type I and type II topoisomerases, although the efficiency of relaxation varies with the enzyme and the substrate. Type IA topoisomerases normally only relax negatively supercoiled DNA and are inactive against positive supercoiled DNA, while type IB enzymes are highly active against both positively and negatively supercoiled DNA. Recent elegant single molecule studies by Cozzarelli and colleagues have shown that *E. coli* top IV has much greater activity relaxing positively supercoiled DNA (3). Two enzymes have been shown to be able to introduce supercoils into DNA. As alluded to above, bacterial DNA gyrase can introduce negative supercoils in DNA. The unique ability of gyrase to introduce negative supercoils into a DNA molecule depends on the unusual DNA binding properties of this molecule. About 140 base pairs of DNA are wrapped around gyrase in right-handed orientation to produce one turn of positive superhelical turn, which is then converted into a negative superhelical turn by the strand passage reaction catalyzed by DNA gyrase. Some hyperthermic archaeobacteria carry an enzyme termed reverse gyrase, that is able to introduce positive supercoiling into DNA. Interestingly, reverse gyrase is a type IA topoisomerase, which introduces positive supercoiling by the concerted action of a DNA helicase and the type IA topoisomerase activity.

In addition to relaxation of supercoiled DNA, topoisomerases also carry out other topological changes in DNA. A reaction of critical biological importance is the decatenation of catenated DNA molecules (Fig. 1B). As described below in the section concerning DNA replication, near the end of replication, separation of replicated molecules requires passage of DNA double strands, i.e. a type II topoisomerase activity. In the absence of a type II topoisomerase, the product of replication of a circular molecule is a multiply intercatenated DNA molecule. Not surprisingly, a failure to separate such molecules before cell division prevents their proper segregation. Type I topoisomerases can also carry out decatenation reactions



DNA Topoisomerases. Figure 1 Reaction catalyzed by topoisomerases. Figure 1 shows a schematic representation of reactions catalyzed by topoisomerases. In all three panels, the solid lines represent double stranded DNA. Panel 1A shows relaxation of supercoiled DNA, a reaction catalyzed by both type I and type II topoisomerases. As indicated in the text, the reverse reaction, introduction of DNA supercoiling is carried out only by DNA gyrases. Panel B diagrams a decatenation reaction. Decatenation of intact double strand DNA can be carried out by type II topoisomerases and also by type III topoisomerases in collaboration with a *recQ* helicase (e.g., *E. coli* top III + *recQ*). Decatenation of double stranded DNA can also be carried out by type I topoisomerases provided that there is a gap in one of the two strands. Type II topoisomerases also can knot and unknot DNA (panel C).

with single strand DNA and, provided that there is a gap in one strand, type I topoisomerases can also decatenate double stranded DNA. It has recently been shown that a type IA topoisomerase in association with specific helicases of the *recQ* family can also carry out decatenation of intact double stranded DNA. An intermediate of this reaction is thought to be a hemi-catenane, a molecule in which one DNA strand of a DNA duplex is wrapped around another double strand. A reaction that is closely related to decatenation is unknotting of DNA. A knot can be considered to be an intramolecular catenane. While unknotting is useful as an assay for type II topoisomerases, knotted DNA arises only rarely in normal DNA metabolism.

Topoisomerase Strand Passage Mechanisms

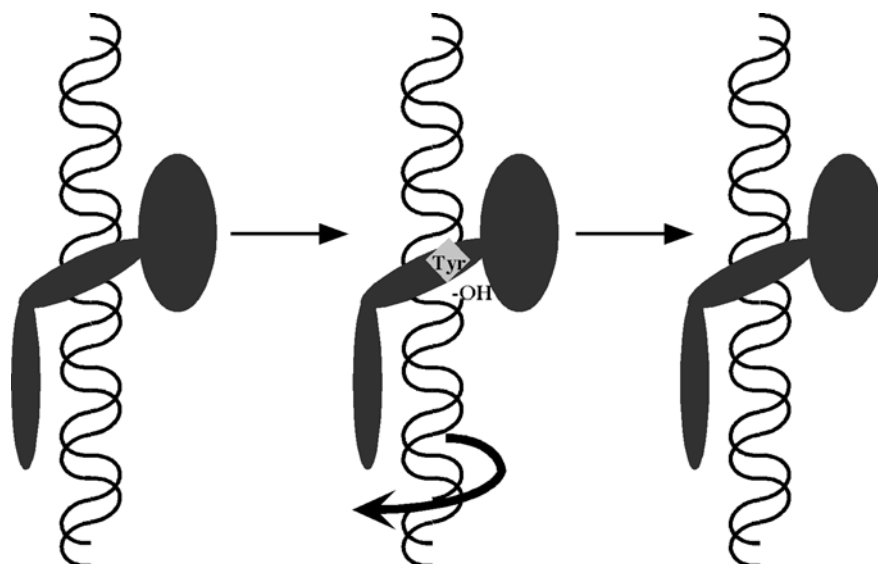
There are two major mechanisms that topoisomerases use to move DNA strands during the enzyme reaction.

Type IB enzymes use a rotation, or swiveling mechanism, whereas both type IA and type II enzymes use an enzyme bridging or strand passage mechanism. Only type IB and type II enzyme mechanisms will be considered here; several recent reviews comprehensively discuss all three types of enzyme mechanisms (4, 5, 6, 7).

Type IB enzymes introduce a single strand break in DNA, through a 3' phosphotyrosine intermediate. The enzyme does not maintain tight interaction with the other end of the cleaved DNA (with a 5' -OH, see Fig. 2). The free DNA end has the potential to rotate; in general the rotation will be in the direction that results in relaxation of the supercoiled DNA. The rotation may be somewhat constrained by the enzyme, but the enzyme can relax multiple superhelical turns per cleavage cycle. After rotation, the enzyme reseals the break and can then dissociate or initiate another round of catalysis.

The reaction cycle of topoisomerase II is illustrated in Fig. 3, and is considerably more complex than the reaction cycle for type IB topoisomerases. There are several important features distinguishing the reaction mechanisms. The type II topoisomerases use a high-energy cofactor, typically ATP, however the high-energy cofactor does not directly contribute to breakage or rejoining of DNA strands, rather ATP binding and hydrolysis controls the conformational states of the enzyme. The enzyme bridging mechanism restricts the enzyme to the passage of one DNA (double) strand per breakage/reunion cycle. The reaction cycle is illustrated in Fig. 3.

Topoisomerase II interacts with two different double stranded DNA duplexes during its reaction. The enzyme generates a DNA "gate" through which a second strand can be passed. The cleaved strand is therefore referred to as the G-segment, while the duplex transferred through the break is called the T-segment. DNA cleavage occurs by the formation of phosphotyrosine bonds between a catalytic tyrosine of each enzyme subunit and the phosphodiester backbone of each strand (of the G-segment). In the case of DNA gyrase, the *gyrA* subunits carry the tyrosine involved in the transesterification reaction. DNA cleavage by type IIA enzymes results in a four base pair staggered double strand break in the G-segment, while for type IIB enzymes, a two base pair stagger occurs. In the presence of ATP, topoisomerase II undergoes dimerization of the N-terminal domains of the subunits. The enzyme forms a "closed clamp" that can trap a second DNA duplex, the T-segment. The T-segment passes through the double strand break in the G-segment, which exits the enzyme complex through the carboxy terminal portion of the enzyme (the bottom of the enzyme as illustrated in Fig. 3). Following T-segment transport, the enzyme reverses the cleavage reaction,



DNA Topoisomerases. Figure 2 Strand passage by strand rotation. Type IB topoisomerases relax DNA by a strand rotation mechanism. After DNA cleavage, the free end of the cleaved strand can rotate and alter the supercoiling of DNA. There are limited contacts between the enzyme and the cleaved strand, although recent work suggests that the rotation of the cleaved strand may be somewhat constrained. Nonetheless, multiple superhelical turns can be removed per cleavage/rejoining cycle.

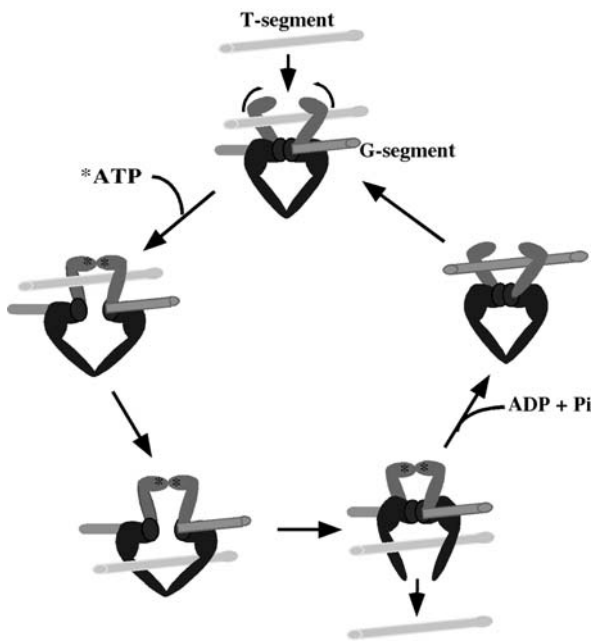
and reforms the phosphodiester bonds of the G-segment. After religation of the cleaved G-segment, hydrolysis and release of ATP is required to dissociate the N-terminal domains of the subunits and restore the enzyme to a conformation where the enzyme can either dissociate from the G-segment of DNA or initiate a new catalytic cycle. For eukaryotic enzymes, the capture of a T-segment is random and will be dictated by the overall three dimensional distribution of DNA molecules.

The Complement of Topoisomerases in Cells: Genetic Studies on the Physiological Roles of Topoisomerases

The complete DNA sequencing of multiple prokaryotic and eukaryotic genomes has provided a powerful complement to the initial biochemical and genetic studies that identified many of the topoisomerases in prokaryotes and eukaryotes. There is only very limited homology among the different major classes of topoisomerases, although there are some key similarities between type IA and type II topoisomerases (8). Within each family of topoisomerases, there is significant amino acid homology. For example, determination of the DNA sequence of the gene encoding yeast Top2 revealed substantial homology with the genes encoding *E. coli* DNA gyrase. The identified homology made it possible to identify a set of residues

that are invariant within each class, making identification of topoisomerase genes relatively straightforward. All bacterial cells have at least one type IA and one type II topoisomerase. Some archaeobacteria that carry a type IIB topoisomerase do not have an obvious type IIA enzyme. Recently, type IB topoisomerases have also been found in eubacteria, however, neither *E. coli* nor *B. subtilis* carry this type of enzyme. *E. coli* cells carry two type IA enzymes (Top I and Top III) and two type IIA enzymes (DNA gyrase and Top IV). Of these four enzymes, all but Top III are essential for viability. It should be noted that in *S. typhimurum*, Top I is not essential for viability, although mutants lacking this enzyme have pleiotropic defects in gene expression and DNA repair. *E. coli* cells lacking Top I are viable provided that they also carry compensatory mutations, most commonly mutations that reduce the activity of DNA gyrase. *E. coli* cells lacking both Top I and Top III have not been identified, suggesting that the compensatory mutations do not alleviate a requirement for at least one type IA topoisomerase. These results indicate that type IA topoisomerases and type II enzymes have indispensable roles in DNA metabolism in bacterial cells. As described below, both DNA gyrase and Top IV play key roles in DNA replication and chromosome segregation.

The yeasts *Saccharomyces cerevisiae* and *Schizosaccharomyces pombe* have served as important model systems for understanding the biological functions of



DNA Topoisomerases. Figure 3 Catalytic cycle of type II topoisomerases. Figure 3 illustrates the catalytic cycle of type II topoisomerases. The enzyme first interacts with the G-segment, the strand that the enzyme will cleave. ATP binding results in dimerization of the N-terminal domain (the top of the enzyme as drawn). The dimerization of the N-terminal domain can lead to the capture of a second DNA segment termed the T-segment. A double strand break in the G-segment creates a DNA gate that the T-segment can pass through. The T-segment exits the enzyme through a C-terminal protein gate. The dynamics of the exiting of the T-segment are poorly understood. After the T-segment transits through the double strand break in the G-segment, the G-segment is religated, and the ATP is hydrolyzed. Release of ADP and Pi leads to dissociation of the N-terminal dimerization and resets the enzyme for a new catalytic cycle.

DNA topoisomerases in eukaryotic cells. Both yeasts have three topoisomerases, termed Top1, Top2, and Top3, in addition to a Spo11 (type IIB) meiotic enzyme. In both yeasts, *top1* (a type IB enzyme) null mutants are viable, indicating the enzyme is not essential for viability. *top1* mutants have relatively modest phenotypes, most notably changes in gene expression and alterations in genome stability (discussed below). The yeasts are the only known eukaryotes that do not require a type IB enzyme for viability. By contrast, null mutations in *top2* (a type IIA enzyme) are lethal. Conditional *top2* mutants have normal DNA replication and lose viability specifically at mitosis, suggesting a defect in chromosome segregation. However, the combination of a *top1* null mutation with a *top2*

mutation results in much more severe phenotypes, including defects in replication and transcription. Current models suggest that either Top1 or Top2 can perform many functions requiring a topoisomerase. *S. cerevisiae* cells lacking *top3* (a type IA enzyme) are viable, but grow extremely slowly and accumulate compensatory mutations in *SGS1*, a helicase that interacts with Top3 protein. Interestingly, *S. pombe* cells lacking *top3* are unviable unless they carry compensatory loss-of-function mutations in *SGS1*. Current studies suggest that Top3 plays a key role in genome stability, a topic discussed below.

Not surprisingly, the complement of topoisomerases in mammalian cells is more complex than that found in yeast. The known topoisomerases of mammalian cells are summarized in Table 1. Mammalian cells have one type IB, two type IA topoisomerases and two type IIA topoisomerases that are mainly localized to the nucleus. Additionally, a type IB topoisomerase that localizes specifically to mitochondria has been found. Experiments assessing the role of particular enzymes have been mainly confined to mouse gene deletion experiments. Top1, Top2 α , Top2 β and Top3 α are all required to generate viable offspring. Of these, Top2 β is of particular interest since embryos deleted for both copies of the Top2 β gene die close to birth. Therefore cells lacking this enzyme are viable and mouse embryonic fibroblasts lacking the enzyme grow and divide normally. Embryos lacking this enzyme have defective neuronal migration, resulting in abnormal neuromuscular junctions and a failure to enervate the diaphragm. The role that Top2 β plays in these processes remains to be determined. Top3 β deficient mice give rise to viable offspring, although over successive generations, the fertility of these mice decreases, accompanied by increased aneuploidy in sperm. These results suggest that Top3 β may play a role in genome stability, as has been suggested for type IA topoisomerases in other systems. Deletions of the mitochondrial topoisomerase have not yet been reported. Since deletion of many of the mammalian topoisomerases results in a complete loss of cellular viability, other experimental systems such as RNAi will be needed to dissect the role that specific topoisomerases play in various cellular processes.

Biological Functions: Replication

The main topological problem associated with DNA replication and accurate chromosome segregation arises from the winding of the two DNA strands in a double helix. At the time of chromosome segregation, the two parental intertwined strands must become completely separated. In addition to this fundamental requirement for separating replicated strands, the process of DNA replication involves the action of a replicative helicase that unwinds double stranded DNA

DNA Topoisomerases. Table 1 The topoisomerases of eukaryotic cells

Enzyme	Type	Localization	Distribution	Essential for viability in mammalian cells
Top1	IB	Nuclear	All eukaryotes	Yes
Top2 α	IIA	Nuclear	All eukaryotes	Yes
Top2 β	IIA	Nuclear, mitochondrial	Mammalian cells, absent in yeast, <i>Drosophila</i>	Yes ¹
Top3 α	IA	Nuclear, mitochondrial	All eukaryotes	Yes
Top3 β	IA	Nuclear	Mammalian cells, <i>Drosophila</i> , absent in yeast	No
mTop1	IB	Mitochondrial	Mammalian cells, absent in yeast ²	Not known
mSpo11	IIB ³	Nuclear	Most eukaryotic cells	No ⁴

¹ Top2 β $-/-$ mice do not survive to term, but embryonic cells lacking the enzyme are viable

² Mitochondrial topoisomerase I has been identified in mammalian cells, and is clearly absent in sequenced yeast genomes. Other systems have not been examined in sufficient detail to determine whether a separate mitochondrial isoform is present

³ Type IIB topoisomerases have also been found in plants, where they play additional roles besides meiotic recombination

⁴ Mice lacking mSpo11 are viable, but are defective in meiotic recombination and are sterile

ahead of the replication fork. In the absence of a topoisomerase activity, the DNA ahead of the replication fork will become highly positively supercoiled, eventually blocking further unwinding by the helicase, thereby leading to arrest of the replication fork. A topoisomerase acting ahead of the fork is able to relax the accumulating positive supercoils, and allow replication to continue. In principle, either type I or type II topoisomerases can fulfill this requirement (Fig. 4). A separate problem arises near the completion of DNA replication. Varshavsky suggested that steric hindrance might prevent a topoisomerase from acting when two replication forks converge at the completion of DNA replication. The resulting replication product for a circular DNA would be catenated. The separation of intercatenated dimers of unnicked circular double stranded DNA is a reaction that is uniquely accomplished by type II topoisomerases, suggesting that these enzymes would have to be required at the completion of replication.

The detailed steps of DNA replication in eukaryotic cells remain to be fully elucidated. Studies of replication in eukaryotes have relied on *in vitro* systems using viral origins of replication and *in vivo* studies using yeast as a model system. Replication *in vitro* requires a DNA topoisomerase, and in the absence of topoisomerase II, the products are catenated dimers. *In vivo*, yeast cells can complete replication with either topoisomerase I or topoisomerase II. Without topoisomerase II activity *in vivo*, yeast plasmids are fully replicated, but accumulate as catenated dimers. Similarly, yeast chromosome segregation fails in the

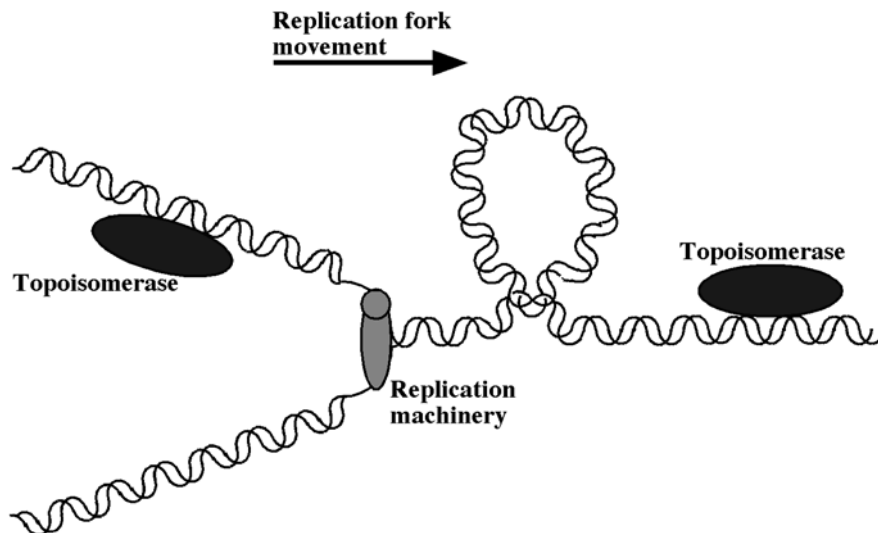
absence of topoisomerase II activity. In the absence of both topoisomerase I and topoisomerase II activities, some replication can still take place in yeast cells although replication that does occur in S phase is very limited.

Topoisomerases in Transcription

The requirement for topoisomerases during replication is clear since the two (parental) DNA strands of a replicated double stranded DNA molecule must be completely separated by the time the molecules are segregated into daughter cells. Although transcription also involves a certain degree of strand separation, the amount of strand separation during transcription is relatively limited, suggesting that DNA topology may not be an important limitation in the movement of RNA polymerase. However, DNA topology may in some instances exert effects on the process of transcription that are independent of providing a "swivel" for the movement of RNA polymerase.

Chromosome Structure and Condensation

The eukaryotic chromosomes at mitosis is a highly compacted structure, while at other points in the cell-cycle, the degree of compaction is relatively small. Therefore, during the process of cell division chromosomes undergo a condensation/decondensation cycle. A key multi-protein complex required for this process is condensin. Biochemical and genetic studies strongly suggest that both topoisomerase II and condensin are required for chromosome condensation, although the precise role of topoisomerases in this process remains



DNA Topoisomerases. Figure 4 Topoisomerases as swivels for DNA replication. Positive supercoils accumulate in front of the replication machinery (either replicative helicases or DNA polymerase holoenzymes) and must be removed for fork progression. Either type I or type II topoisomerases can act as a swivel ahead of replication forks. Topoisomerases can also effect strand separation by acting behind replication forks. Unlinking of replicated molecules at the completion of replication requires a type II topoisomerase.

to be determined. A requirement for topoisomerase II in modulating chromosome condensation suggests that this protein may be an important determinant of chromosome structure. A structure termed the “chromosome scaffold” was operationally defined as a set of proteins that remained following removal of DNA and histones from mitotic chromosomes. Topoisomerase II is one of the major components of chromosome scaffolds, although the biological roles of these structures remains to be worked out (9).

Biological Functions: Recombination and Genome Stability

DNA topoisomerases play clear roles in modulating DNA structure and also play essential roles in DNA replication. A failure to regulate chromosome structure or properly complete DNA replication might be expected to generate damaged structures that might require DNA repair processes for their rectification. It is now widely appreciated that a major function of homologous recombination is to allow restarting of DNA replication when a fork is blocked. In yeast models, a lack of any of the three topoisomerases leads to elevated recombination. Yeast cells lacking *TOP1* or carrying a temperature sensitive *TOP2* allele exhibit elevated recombination, although the hyper-recombination phenotype occurs mainly in rDNA, the locus encoding the RNA components of ribosomes. The hyper-recombination phenotype may arise from alterations in transcription, in replication or in the coordination of these two processes. Yeast cells lacking *TOP3*

also exhibit a hyper-recombination phenotype and *top3* mutants have a hyper-recombination phenotype that is not confined to rDNA. Current models suggest that topoisomerases function to suppress recombination, especially in repeated genetic elements. This may explain the elevated levels of aneuploidy observed in mammalian *top3β* deficient mutants.

Unanswered Questions and Future Directions

Work over the last few years has led to a broad understanding of the enzyme mechanisms of DNA topoisomerases. Although not described in detail in this article, protein structures determined by X-ray crystallography have made it possible to describe in substantial detail how DNA topoisomerases carry out their elaborate strand passage reactions. Studies especially in yeast and mammalian cells, have uncovered new enzymes and made assigning specific roles for the newly discovered enzymes more challenging. Topoisomerases still appear in what seem to be unlikely contexts, such as the phenotypes conferred by mouse cells lacking *top2β*.

Clinical Relevance

Topoisomerases are clearly vital for many aspects of nucleic acid metabolism. Topoisomerases also play an important role as the targets of clinically important anti-bacterial and anti-tumor agents. Fluoroquinolones are broad-spectrum potent anti-bacterial agents. The targets of fluoroquinolones are bacterial type II topoisomerases. Both DNA gyrase and top IV are targeted by quinolones. Agents acting against eukaryotic topoisomerases have

also been identified and are active, clinically used anti-cancer agents. Topotecan and irinotecan are camptothecin derivatives that specifically target topoisomerase I, while agents such as doxorubicin and etoposide target eukaryotic topoisomerase II. Fluoroquinolones, camptothecins and the topoisomerase II targeting drugs have a unique mechanism of action. The most important effect of these drugs is to block their target enzyme at a step in the catalytic cycle where the enzyme is covalently bound to DNA. Hence all of these topoisomerase-targeting agents convert topoisomerases into a form of DNA damage.

► [Replication Fork](#)

Acknowledgements

We thank Ms. Miranda Grey for the preparation of the figures used in this article. Work in the authors' laboratory is supported by grants from the National Cancer Institute and the American Lebanese Syrian Associated Charities (ALSAC).

References

1. Wang JC (2002) Cellular roles of DNA topoisomerases: a molecular perspective. *Nat Rev Mol Cell Biol* 3:430–440
2. Osheroff N (1998) DNA topoisomerases. *Biochim Biophys Acta* 1400:1–2
3. Stone MD, Bryant Z, Crisone NJ et al (2003) Chirality sensing by *Escherichia coli* topoisomerase IV and the mechanism of type II topoisomerases. *Proc Natl Acad Sci USA* 100:8654–8659
4. Li Z, Mondragon A, Di Gate RJ (2001) The mechanism of type IA topoisomerase-mediated DNA topological transformations. *Mol Cell* 7:301–307
5. Wang JC (1998) Moving one DNA double helix through another by a type II DNA topoisomerase: the story of a simple molecular machine. *Q Rev Biophys* 31:107–144
6. Stewart L, Redinbo MR, Qiu X et al (1998) A model for the mechanism of human topoisomerase I. *Science* 279:1534–1541
7. Corbett KD, Berger JM (2004) Structure, molecular mechanisms, and evolutionary relationships in DNA topoisomerases. *Annu Rev Biophys Biomol Struct* 33:95–118
8. Berger JM, Fass D, Wang JC et al (1998) Structural similarities between topoisomerases that cleave one or both DNA strands. *Proc Natl Acad Sci USA* 95:7876–7881
9. Swedlow JR, Hirano T (2003) The making of the mitotic chromosome: modern insights into classical questions. *Mol Cell* 11:557–569

DNA-based Vaccination

REINHOLD SCHIRMBECK, MARTIN SCHLEEF,
JÖRG REIMANN

Institute for Medical Microbiology and Immunology,
University of Ulm, Ulm, Germany
reinhold.schirmbeck@medizin.uni-ulm.de

Definition

Genetic (nucleic acid or DNA) vaccination is an exciting new immunization technique with potential applications in clinical medicine. In ► [genetic immunization](#), plasmid DNA containing antigen-encoding sequences cloned under ► [heterologous promoter control](#) is delivered by techniques that lead to antigen expression *in vivo* and its immunogenic presentation. DNA vaccination has been shown to efficiently prime both humoral (antibody) and cellular (T cell) immune responses against viral, bacterial, parasite and tumor antigens in different animal species, e.g. mice, pigs, cattle, dogs, monkeys and man. A key feature of DNA-based vaccination is its potency to specifically prime ► [Th1-biased T cell responses](#), including MHC-I-restricted, cytotoxic ► [CD8⁺ T lymphocyte](#) (► [CTL](#)) responses. This makes genetic vaccination an attractive candidate for ► [prophylactic](#) or ► [therapeutic immunization](#) against intracellular pathogens and cancer. Protein antigens encoded by DNA ► [vaccines](#) are expressed by transfected cells *in vivo*, thus mimicking viral infection.

Three areas that critically influence the success of DNA vaccination are:

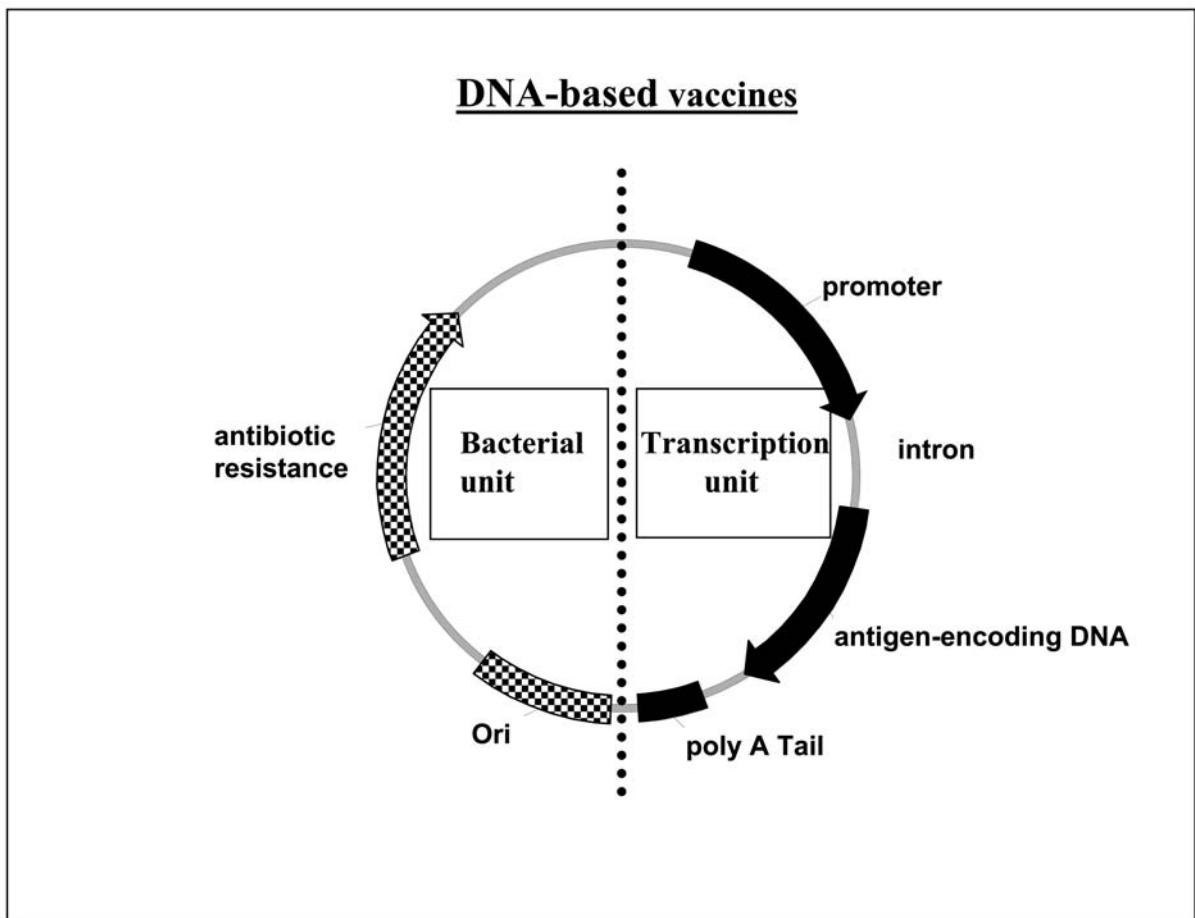
- the design of expression vectors that efficiently produce antigen(s) *in vivo*;
- the large scale production and purification of DNA vaccines;
- techniques of delivery of DNA vaccines that support a level of antigen expression that ensures immunogenicity.

Characteristics

Vector Design

The 'vaccine' delivered in nucleic acid immunization is usually plasmid DNA, though successful vaccination with antigen-encoding mRNA has been reported. The DNA vaccine contains two units, a bacterial unit (necessary for plasmid production) and a transcription unit (encoding the relevant antigen) (Fig. 1). The bacterial unit contains a bacterial drug resistance gene for selecting plasmid-containing bacteria, and an origin of replication to allow production of high copy numbers of the plasmid in appropriate bacterial strains. The transcription unit contains the antigen-encoding sequence cloned downstream from strong ► [promoter/enhancer](#) sequences that support efficient expression of the immunogenic protein in eukaryotic cells. Variables considered in the design of DNA vaccines include:

- a choice of promoter/enhancer sequences to control the level and the tissue specificity of antigen expression;
- the incorporation of innate ► [adjuvants](#) or immune modulators into the coding sequences of the DNA vaccine;



DNA-based Vaccination. Figure 1 Plasmid used for DNA-based vaccines. The plasmid contains a bacterial unit (containing sequences of the origin of replication (ORI) and the antibiotic selection marker) and the eukaryotic transcription unit (containing promoter/enhancer, intron, antigen and poly A signal sequences).

- strategies to control coexpression of different genes (either different antigens or antigen plus costimulator/▶cytokine molecules);
- strategies to target antigen to different subcellular compartments within antigen processing/ presenting cells or to facilitate antigen release from the expressing cell.

Promoter/Enhancer Sequences

Promoter/enhancer sequences used in DNA vaccines to drive antigen expression are often derived from viruses (e.g. cytomegalovirus, simian virus 40, retrovirus) or bacteria (1, 2). Because interferon- γ (IFN γ) release *in situ* during priming of immune responses by DNA vaccine injection can suppress the activity of viral promoters, mammalian cell-derived promoter/enhancer elements have been successfully used to construct DNA vaccines. These include e.g. sequences driving expression of elongation factor-1 α , desmin or metallothionein (1, 2). In addition, polyadenylation and/or

intron sequences (positioned either within the transcription unit or flanking it) can be of critical importance for the immunogenicity of DNA vaccines by stabilizing mRNA transcripts.

Coexpression of Immune Modulators by DNA Vaccines

DNA vaccines can encode cytokines, costimulator molecules, small ▶innate adjuvants or adjuvant-binding domains (2, 3). Stress proteins are potent ‘innate adjuvants’ that facilitate priming of cellular and ▶humoral immune responses. Proteins of the ▶hsp70 and hsp90 families are immunodominant antigens of many bacterial and parasite infections. Construction of DNA vaccines encoding stress proteins fused to antigen has been reported to strikingly enhance the immunogenicity of its T cell ▶epitopes.

A stress protein (hsp73)-facilitated expression system was developed in order to produce large chimeric antigens with enhanced immunogenicity for the specific humoral and cellular immune system (4). The

system is based on the observation that N-terminal sequences of the SV40 large ▶T-Ag show strong, non-covalent binding to hsp73/hsc70 because the N-terminal 77 residues of T-Ag with ▶DnaJ homology have functional docking sites for hsp73. Hence, the N-terminal T-Ag domain can dock to and regulate substrate binding of the constitutively expressed, cytosolic stress protein hsp73. Chimeric antigens constructed from such an N-terminal, hsp73-binding T-Ag domain fused to large C-terminal antigenic domains of >900 residues were efficiently expressed in an immunogenic form. The advantages of the system are:

- high expression of large amounts of chimeric proteins with a long half-life as protein/hsp complexes;
- expression of large protein antigens that bind increasing amounts of hsp with increasing length;
- codelivery of protein antigens that non-covalently bind hsp strikingly enhances and modulates its immunogenic properties.

Using DNA-based vaccination, the relative efficacy of CTL priming to a large panel of ▶MHC class I-binding epitopes processed from antigenic domains expressed either as native antigens or as chimeric antigens fused to the N-terminal hsp73-binding DnaJ domain of the T-Ag were compared. Hsp73-association greatly facilitated induction of humoral and cellular ▶immune responses to antigens or their antigen-domains targeted (4). The stable, non-covalent association of antigen with hsp73 *in situ* during its early expression thus facilitates priming of CTL. Furthermore, this system supported efficient cross priming of ▶TAP-independent and helper T cell-independent CTL responses in murine tumor models due to the codelivery of intrinsic (natural) adjuvants. In addition, priming of antibody responses to endogenous antigen was observed following DNA vaccination. This demonstrates that recombinant DNA technology can be successfully used to design DNA vaccines that codeliver 'natural' adjuvant activity with an extended range of antigenic information thereby facilitating specific stimulation of T cell responses.

Strategies for the Coexpression of Different Genes by DNA Vaccines

A key issue of DNA-based immunization is the design of polyvalent vaccines to increase the amount of immunogenic information that can be delivered by a DNA vaccine. A simple way is to mix two expression constructs before delivery. In more elaborate systems, different genes are coexpressed in complex expression constructs at stoichiometrically defined ratios from one plasmid. This is e.g. achieved by incorporating internal ribosomal entry sequences (▶IRES) between genes to

allow their coexpression under control of a single promoter. Complex expression vectors that contain multiple polycistronic expression units have been successfully designed. This approach offers the prospect of coexpressing multiple antigens, cytokines and costimulator molecules at stoichiometrically defined ratios by a single DNA vaccine. This approach may eventually allow the incorporation of the complete immunogenic information of a virus into a single DNA vaccine and its codelivery with a molecularly defined adjuvant.

Targeting Antigen Expressed by DNA Vaccines to Different Subcellular Compartments of Antigen Processing/presenting Cells or Facilitating Antigen Secretion from the Expressing Cell.

DNA-based vaccines ensure the *in situ* expression of protein antigens with correct posttranslational modifications, three-dimensional conformations or oligomerizations. The integrity of conformational epitopes that stimulate neutralizing antibody (B cell) responses is thereby maintained. DNA vaccines induce antibodies against intracellular proteins (2, 5) by a mechanism not understood. It is generally accepted that exogenous protein antigens efficiently induce antibody responses. DNA vaccines can easily be engineered to optimize the immunogenicity of encoded protein antigens for antibody induction by introducing leader and/or transmembrane anchor sequences. This programs protein antigens to be either secreted or incorporated into the surface membrane. Both modifications have been shown to enhance humoral immune responses induced by DNA vaccines (2, 5).

Production of DNA Vaccines

For manufacturing plasmid DNA of a quality acceptable for clinical use, appropriate production processes are required that have to fulfill the requirements of internationally valid guidelines and laws. Good manufacturing practice (GMP) is applied in production and quality control for pharmaceuticals. These standards are only obtained in specific manufacturing and quality control facilities. The isolation of plasmid DNA for research applications requires high quality materials and technologies to guarantee a reproducible and appropriate quality (6). The main requirements in plasmid purification are:

- the separation of plasmid DNA from contaminating substances derived from either the purification process or the bacterial host cell (e.g. cell wall residues, lipopolysaccharides, bacterial genomic DNA, RNA, proteins, lipids);
- the protection of the desired supercoiled form of the plasmid throughout the manufacturing process;
- the separation of the supercoiled form of the plasmid DNA from other DNA forms to obtain a homogenous product.

The standard process in plasmid production is the stable transformation of the plasmid vector into a defined *E. coli* host strain, its cultivation in a fermenter, the alkaline lysis of the bacterial biomass followed by filtration and chromatography steps to obtain the requirements listed above. In bacterial cultivation for plasmid productions, different plasmid forms appear. The supercoiled plasmid structure shows the most compact structure and is known to be the most active topology.

The product stability in production, storage, formulation, transport and application mainly reflects plasmid topology. The characterization of the DNA topology was for the first time possible using the technology of capillary gel electrophoresis (CGE). The CGE assay is nowadays performed as in-process control in plasmid manufacturing on a routine basis to identify and monitor sensitive process steps in which shear forces can damage plasmid DNA (e.g. sterile filtration). Recently, CGE technology was applied in process to monitor the separation of undesired topologic plasmid variants from the supercoiled fraction.

Storage and transport of DNA vaccines require defined media. Typically, a plasmid DNA solution will be frozen for long-term storage. Optimization studies ongoing in many facilities will include CGE. The use of ►gene gun or jet-injection will require further evaluation of plasmid stability and topology. Monitoring the changes in the supercoiled topology during the application, as well as optimizing the applied pressure with respect to tissue, plasmid and buffer are expected to be essential for the success of the technique.

Delivery of DNA Vaccines

Different techniques for the delivery of low doses of DNA vaccines in immunogenic form are under development. These include the delivery of:

- ►‘naked’ plasmid DNA for injection into muscle or skin;
- ►particle-coated plasmid DNA for delivery to skin or mucous membranes by the gene gun;

The injection of ‘naked’ plasmid DNA into the muscle or skin is an efficient technique to induce humoral and cellular immune responses in mice (1, 2, 5). DNA is administered in a variety of diluents, but the most common is phosphate-buffered saline (PBS). Injection of ‘naked’ plasmid DNA into mice is usually sufficient to establish immune responses. This delivery technique is efficient in mice, but inefficient in other animal species and in man where several injections of mg amounts of DNA often elicit only marginal specific immune responses. New techniques are currently being tested to allow the delivery of low doses of DNA vaccines to a range of animals without affecting the immunogenicity of the vaccine (2).

Low doses of gold particle-coated plasmid DNA can be inoculated intradermally with the gene gun (2, 5). A single inoculation of the DNA vaccine primes readily detectable and long-lasting serum antibody responses and T cell responses. Gene gun vaccination has many advantages compared to i.m. DNA vaccination. It is an easy and reproducible way for DNA injections especially in larger animals and man.

Concluding Remarks

Since its first description in 1992, DNA vaccination has moved with amazing speed from preclinical animal models to clinical trials. Although DNA vaccination involves risks (including insertional mutagenesis, induction of low zone tolerance or autoimmune responses and immune-mediated destruction of tissue transfected *in vivo*), its potential for vaccinology is expanding. DNA vaccines have increasing appeal in the exploration of the increasing universe of molecularly cloned antigens as vaccine candidates.

References

1. Donnelly JJ, Ulmer JB, Shiver JW et al (1997) DNA vaccines. *Annu Rev Immunol* 15:617–648
2. Gurunathan S, Klinman DM, Seder RA (2000) DNA vaccines: immunology, application and optimization. *Annu Rev Immunol* 18:927–974
3. Scheerlinck JY (2001) Genetic adjuvants for DNA vaccines. *Vaccine* 19:2647–2656
4. Schirmbeck R, Kwissa M, Fissolo N et al (2002) Priming polyvalent immunity by DNA vaccines expressing chimeric antigens with a stress protein-capturing, viral J-domain. *FASEB J* 16:1108–1110
5. Schirmbeck R, Reimann J (2001) Revealing the potential of DNA-based vaccination: lessons learned from the hepatitis B surface antigen. *Biol Chem* 382:543–552
6. Schleef M, Voss C, Schmidt T (2002) DNA drugs – production and quality assurance. *Eng Life Sci* 2:157–160

DNA-Dependent RNA Polymerase I

►RNA Polymerase I Transcription

DnaJ

Definition

DnaJ is a bacterial regulatory protein that binds stress proteins.

►DNA-based Vaccination

dNTPs

Definition

The term dNTPs defines 2'-Deoxynucleoside 5'-triphosphates, the “building blocks” of DNA.

- [Nucleotide Biosynthesis](#)
- [SNP Detection and Mass Spectrometry](#)

DOCK-180

Definition

DOCK-180 stands for dedicator of cytokinesis-180, and functions as an activator of the Rho GTPase Rac1 in the Rac GTPase signaling pathway.

- [Signal Transduction: Integrin-Mediated Pathways](#)

Docking

Definition

In the context of protein interaction analysis or structure-based drug design, docking is the process by which the three-dimensional structures of a protein and its ligand (or drug) fit together. The energetically preferred geometry of interaction of a ligand binding to a biological macromolecule is predicted by a computational approach.

Docking (or fitting) in the context of cryo-EM refers to the positioning of atomic models of macromolecules (either as rigid bodies or in a flexible manner) into lower resolution electron density maps.

- [Cryo-Electron Microscopy: Single-Particle Reconstruction](#)
- [Molecular Docking](#)
- [Molecular Dynamics Simulation in Drug Design](#)

Dolichol

Definition

Dolichol is a polyisoprenoid lipid carrier, of the general formula $C_{80}H_{132}O$ to $C_{110}H_{180}O$, which is utilized during the assembly of N-glycans and GPI anchors.

- [Glycosylation of Proteins](#)

Domain

Definition

Domain denotes a part of a polypeptide chain that can fold into a three dimensional structure independently. A domain may be linked to other domains by short flexible stretches of polypeptides. The combination of domains in a single protein determines its overall function.

- [Molecular Motors](#)
- [Protein Databases](#)
- [Two-Hybrid System](#)

Domain Fold

Definition

Large protein molecules are usually composed of compact globular domains that have a limited number of basic architectures. These domain folds may be annotated using programs like SCOP or CATH.

- [Structural Genomics: Structure-to-Function Approaches](#)

Dominant

Definition

Dominant describes the property of an allele to be expressed (phenotypically in heterozygous state; one locus on the homologous chromosomes).

Dominant Allele

Definition

The term dominant allele is applied to the phenotypic effect of a particular allele, in reference to another allele (usually the standard wild-type allele), with respect to a given trait. An allele “A” is said to be dominant with respect to the allele “a” if the A/A homozygote and the A/a heterozygote are phenotypically identical and different from the a/a homozygote.

- [Large-Scale ENU Mutagenesis in Mice](#)

Dominant Inhibitory Mutant/ Dominant Negative Mutant

Definition

Dominant inhibitory (or dominant negative) mutant is used to describe a mutation in a protein, which when expressed, is non-functional, or can interfere with the function of the normal (wild-type) protein. In classical genetic nomenclature, this type of mutation is analogous to an antimorphic allele.

► [Mouse Genomics](#)

► [Xenopus as a Model Organism for Functional Genomics](#)

Dominant Mutation

Definition

Dominant mutation characterizes a mutation that usually results in a gain of gene function and produces the mutant phenotype, even in the presence of the wild-type gene.

► [Protein Interaction Analysis: Suppressor Hunting](#)

Dominant Negative Mutation

Definition

Dominant negative mutation refers to a mutation that dominantly affects the phenotype, by means of a defective protein interfering with and inhibiting the function of the normal (wildtype) gene product in the same cell.

► [Heritable Skin Disorders](#)

► [Tumor Suppressor Genes](#)

Donor

Definition

In Förster resonance transfer (FRET), the donor is defined as the molecule passing energy to another molecule (acceptor) causing its excitation.

► [FRET](#)

Dopamine

Definition

Dopamine (DA) is a catecholaminergic neurotransmitter of dopaminergic neurons, as well as the precursor for the synthesis of noradrenalin. Dopaminergic neurons that have their cell bodies in the ventral tegmental area (VTA) are the key players in the ► [reward system](#). They send their axons to the nucleus accumbens, the frontal cortex, and parts of the striatum. Actions that increase DA release tend to be repeated. There is evidence that the release of the neurotransmitter DA increases after repeated drug use, and is reduced during withdrawal from the drug. Loss of dopaminergic neurons is seen in Parkinson's disease, and therapies are based on ► [dopamine replacement](#) in that disease.

► [Addiction, Molecular Biology](#)

► [Parkinson's Disease: Insights from Genetic Cause](#)

Dopamine Replacement Therapy

Definition

Dopamine replacement therapy means medication that increases the stimulation of dopamine receptors in the brain. This is achieved by either directly increasing the dopamine, via administering dopamine precursors (with medications such as carbidopa/levodopa), or dopamine agonists, or by medications that inhibit the metabolism of dopamine. These treatments are currently the mainstay of therapy in Parkinson's disease.

► [Parkinson's Disease: Insights from Genetic Cause](#)

Dopamine Responsive Parkinsonism

► [Parkinson's Disease: Insights from Genetic Causes](#)

Dormancy

Definition

The term is taken from bacteriology, indicating survival of a pathogen or a population of cancer cells with minimal activity and lack of growth because of unfavourable circumstances.

► [Cancer Invasion and Metastasis: Cellular, Molecular and Clinical Aspects](#)

Dosage Compensation

Definition

Dosage compensation describes a mechanism by which organisms adjust the dose of X-chromosome-linked transcripts, in order to equalize levels of sex-chromosome products in males and females of one species. In mammals, dosage compensation operates by maintaining only a single active X chromosome in each cell, while in *Drosophila* it operates by hyperactivating the single male X chromosome.

► [Chromatin Acetylation](#)

► [X-Chromosome Inactivation](#)

Doss Porphyria

Definition

Doss porphyria is an autosomal recessively inherited disease, caused by deficiency of the δ -Aminolevulinic acid dehydratase, which leads to an acute hepatic porphyria in the compound heterozygous state.

► [Acute Intermittent Porphyria](#)

Double Mutant

Definition

Double mutant describes an individual that is carrying homozygous mutations in two different genes. In the context of mouse genetics, a double mutant is usually produced by crossing individual heterozygous for mutations in two different genes.

► [Gene Duplications](#)

Double-Strand Break Repair

ALBERT PASTINK¹, MALGORZATA Z. ZDZIENICKA²

¹Department of Toxicogenetics, LUMC, Leiden, The Netherlands

²Department of Molecular Cell Genetics, The L. Rydygier Collegium Medicum, Nicolaus Copernicus University, Bydgoszcz, Poland
a.pastink@lumc.nl

Definition

Faithful maintenance of the integrity of the genetic information in response to DNA damage is critical for proper functioning of all living organisms. Double-strand breaks (DSBs) in DNA, i.e. breaks occurring simultaneously in both strands of the DNA double helix, are among the most severe types of damage which, if left unrepaired, can lead to broken chromosomes and cell death. DSBs occur as intermediates in several cellular processes, including meiotic recombination in germ cells and during the assembly of ► [immunoglobulin](#) and T-cell receptor genes to form functional genes in lymphocytes by ► [V\(D\)J recombination](#). In proliferating cells, DNA replication is a major source of DSBs in the genome. DSBs can also be induced by exposure to DNA damaging agents such as ionizing radiation and radiomimetic chemicals like bleomycin, both of which are frequently used in cancer therapy.

The presence of a DSB in the genome of a eukaryotic cell triggers activation of complex and highly coordinated DNA damage response pathways, which function to restore the integrity of the genome through ► [cell-cycle checkpoints](#) and ► [DNA repair](#). Defects in these mechanisms can lead to cell death or the accumulation of mutations and ► [chromosomal aberrations](#). There is abundant evidence from cytogenetic and molecular studies that tumor development in humans is frequently associated with the formation of chromosomal aberrations such as translocations. Presumably, these aberrations arise as a consequence of DSBs and their inappropriate repair. Two main pathways, namely, homologous recombination (HR) and non-homologous endjoining (NHEJ) have evolved in eukaryotes for the repair of DSBs. While largely distinct from one another, they function as complementary mechanisms to repair DSBs. Repair through HR requires the presence of an undamaged homologous (identical) DNA sequence. Such sequence information is present not only on the homologous chromosome but also in replicating cells on the sister chromatid. The homologous sequences are used for accurate repair of the nucleotide sequence at the site of the DSB. The process of NHEJ is based on the simple rejoining of the two DNA ends of a DSB, but this process is less precise than HR.

Characteristics

Homologous Recombination

Repair of DSBs through homologous recombination (HR) was first observed in the yeast *Saccharomyces cerevisiae*. Survival studies showed that diploid yeast cells are more resistant to ionizing radiation than haploid cells. The observations that repair of DSBs induced by ionizing radiation requires the presence of two DNA copies suggested that DSBs could be repaired through HR (also called recombinational

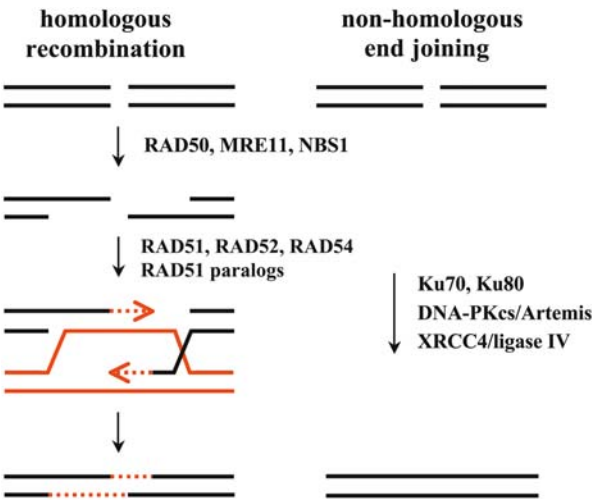
repair). This assumption was confirmed by the isolation of ionizing radiation sensitive (RAD) mutants. These mutants belong to the RAD52 ►epistasis group and are characterized by increased sensitivity to ionizing radiation and chemical compounds capable of inducing DSBs. In addition, most of these mutants have severe defects in mitotic and meiotic recombination as manifested by the formation of nonviable spores. Homologs of these genes have been identified in higher eukaryotes including man. At present the RAD52 group in mammals encompasses 12 genes (Table 1).

The classical model for HR is based on studies in yeast and other fungi. After the induction of a DSB, the first step is the formation of 3'-single-strand ends by the action of specific nucleases. Subsequently, one of the 3'-ends invades an undamaged homologous DNA molecule leading to the displacement of one of the DNA strands (D-loop). The invading 3'-end is then used as a start for DNA synthesis. Later on, after enlargement of the D-loop, the other 3'-end can also be used as a start for DNA synthesis. Finally, the intermediate structures in recombination are resolved and repair is completed by ligation of the ends (Fig. 1).

Double-Strand Break Repair. Table 1 Proteins involved in double-strand break repair in mammals

homologous recombination		non-homologous endjoining
BRCA1	XRCC2:	Ku70
BRCA2	XRCC3	Ku80
RAD50	RAD51B	DNA-PKcs
MRE11	RAD51C	Artemis
NBS1	RAD51C	Ligase IV
RAD51		XRCC4
RAD52		
RAD54		
RAD54B		

With the exception of BRCA1 and BRCA2, proteins required for homologous recombination belong to the RAD52 group and were identified on basis of homology with yeast proteins. XRCC2, XRCC3, RAD51B, RAD51C and RAD51D are referred to as RAD51 paralogs because these proteins display 20-30% identity to RAD51. A sixth RAD51 paralog, DMC1, functions only in meiosis. In addition to the proteins known to be directly involved in HR, there are several other factors required for repair through this pathway, e.g. replication protein A, resolvases of intermediate structures in recombination, DNA polymerase and DNA ligase.



Double-Strand Break Repair. Figure 1 Schematic representation of homologous recombination and non-homologous endjoining, the two main pathways involved in repair of double-strand breaks in DNA in eukaryotes. (a) The first step in HR after the induction of double-strand breaks, is the specific formation of 3'-single-strand DNA ends at the site of the break. One of the 3'-ends invades a homologous undamaged DNA molecule leading to the formation of a D-loop. Both 3'-ends are then used to prime DNA synthesis. The synthesis of new DNA leads to restoration of DNA sequences that were lost as a consequence of the induction of a double-strand break. Finally, the intermediate structures in recombination (also called joint molecules) are resolved and repair is completed by ligation of the ends. Repair through HR results in correct repair of the break. (b) The first step in NHEJ is binding of DNA-PK (consisting of Ku70, Ku80 and DNA-PKcs) and Artemis to the site of the break. After aligning, the ends are ligated by the DNA Ligase IV/XRCC4 complex. Repair by NHEJ frequently results in mutations at the site of the break due to the removal or addition of a few nucleotides.

Proteins encoded by the RAD52 group genes have been purified and analyzed in great detail. The RAD50 protein acts in a complex containing also MRE11 and NBS1. This trimeric complex has been implicated in the formation of 3'-single-strand ends during the initial steps in HR. The RAD50 complex is also involved in cell-cycle regulation after the induction of DSBs in DNA. RAD51 is the eukaryotic homolog of the RecA protein in *E. coli* and plays a central role in recombination. Like RecA, the RAD51 protein binds to the 3'-single-strand ends at the site of the break, and is crucial in the search of a homologous template DNA and the subsequent invasion step leading to the formation of joint molecules between broken ends and the undamaged template DNA molecule. Other

proteins belonging to the RAD52 group such as the RAD52 and RAD54 proteins and the RAD51 ▶ **paralogues** act as accessory factors in this process by stimulating joint molecule formation. Much less is known about the resolution of the intermediate structures in the final stages of HR. With a few exceptions, inactivation of RAD52 group genes in experimental animals is not tolerated, indicating that HR is essential in proliferating cells. In addition to the RAD52 group genes, other factors such as the breast cancer associated proteins BRCA1 and BRCA2 have been implicated in HR in mammals (see below).

During meiosis when the homologous chromosome is used as a template, HR leads to crossing over (i.e. exchange of segments between paternal and maternal chromosomes) and plays a fundamental role in the generation of genetic diversity. In somatic cells the sister chromatid is the preferred template for HR. Therefore, HR is the predominant pathway for DSB repair in late S and G2 phase cells and is less important in G1 phase cells. Since sister chromatids are identical in sequence, HR results in precise restoration of the original sequence of the damaged DNA molecule, i.e. the HR process is error-free. However, HR involving homologous chromosomes can lead to homozygosity for certain mutant alleles present in one chromosome and not in its homolog. This phenomenon termed “▶ **loss of heterozygosity**” or LOH has been observed for various ▶ **tumor-suppressor genes** in cancers.

Non-Homologous Endjoining

DSB-repair *via* non-homologous endjoining (NHEJ) is based on ligation of the two ends of a DSB without the requirement for extensive sequence homology between the DNA ends. This mechanism was first identified in mammals using rodent cell mutants characterized by an increased sensitivity to ionizing radiation. NHEJ is important in mammals to repair DSBs generated by exogenous agents, in particular during the G1 phase of the cell cycle. In addition, NHEJ plays an essential role in V(D)J recombination in differentiating lymphocytes and defects result in severe forms of immunodeficiency in rodents as well as in humans. Biochemical and genetic studies identified six proteins, which are required for NHEJ. The DNA-dependent protein kinase complex (DNA-PK) consisting of Ku70, Ku80 and the catalytic subunit DNA-PKcs has a central role in NHEJ. Ku70 and Ku80 are abundant nuclear proteins with strong affinity for free DNA ends. Binding of Ku to DNA-ends facilitates recruitment of DNA-PKcs and the Artemis protein to the site of the break to form an activated DNA-PK complex. Activated DNA-PK phosphorylates a large number of proteins *in vitro*, but the precise role of DNA-PKcs *in vivo* remains unclear. Binding of Ku to DNA ends also facilitates binding of the ligase IV-XRCC4 complex to the site

of the break and stimulates intermolecular ligation (Fig. 1). The endonucleolytic activity of the Artemis protein is important for processing of DNA ends and hairpin structures. Opening of hairpins is a crucial step in V(D)J recombination in differentiating B- and T-lymphocytes.

Mice deficient for ligase IV or XRCC4 die during early embryogenesis as a consequence of massive neuronal apoptosis, indicating that these gene products are essential for development. However, mice lacking Ku70, Ku80, DNA-PKcs or Artemis could be generated. These animals display the SCID (severe combined immunodeficiency) phenotype and increased sensitivity to ionizing radiation. Mice deficient for Ku70 or Ku80 have reduced body weight and are also characterized by premature ageing implying additional functions for the Ku proteins. Inactivation of NHEJ has been found to be associated with increased rates of lymphomas in mice. Detailed analysis of tumor material indicated that misrepair of DSBs generated during V(D)J recombination may be one of the initiating steps in the formation of these tumors. At the cellular level, impairment of NHEJ leads to enhanced levels of fragmented chromosomes and chromosome aberrations, indicating an essential role for NHEJ in maintaining genome stability.

Clinical Relevance

A direct link between defects in DSB repair and increased tumorigenesis has come from studies of the *BRCA1* and *BRCA2* (breast cancer susceptibility) genes. Women carrying germline mutations in one of these genes have a risk for developing breast cancer of up to 85% by the age of 70. Both *BRCA1* and *BRCA2* act as tumor-suppressor genes since in tumor material the wild type allele has also been mutated. Both the *BRCA1* and *BRCA2* proteins have been found associated with the RAD51 protein, which plays a central role in the repair of DSBs through HR. In addition, *BRCA1* and *BRCA2* function in several other cellular processes. Loss of normal *BRCA1* or *BRCA2* function in mammalian cells leads to the accumulation of spontaneous chromosome breaks, accompanied by cell-cycle checkpoint-mediated growth arrest. However, a total inactivation of *BRCA1* or *BRCA2* genes is not tolerated in experimental animals. Recently, it has been found that mutant alleles of *BRCA2* are responsible for the defect in one of the complementation groups of Fanconi anemia (FA), a rare disorder with chromosomal instability and proneness to cancer. Thus, apparently, residual *BRCA1* and *BRCA2* activities are essential for cell proliferation. Both *BRCA1* and *BRCA2* are specific for higher eukaryotes and have not been identified in yeast.

Cancer-prone disorders involving genes belonging to the RAD52 group are ataxia-telangiectasia-like disease

(ATLD) and Nijmegen breakage syndrome (NBS), which are caused by ►[hypomorphic mutations](#) in *MRE11* and *NBS1*, respectively. Both human syndromes are characterized by a high incidence of cancer and at the cellular level by radiosensitivity and defects in activation of proper cell-cycle checkpoints. A similar phenotype is also observed in ataxia-telangiectasia (AT) patients, a classical human disorder with a defect in response to DSBs. The ATM kinase, defective in these patients, initiates a signal-transduction cascade, by phosphorylating several proteins, including NBS1, p53 and histone H2AX in response to DSBs. The cancer predisposition seen in AT, NBS and ATLD patients underscores the importance of ATM and the RAD50/NBS1/MRE11 protein complex in avoiding increased genome instability caused by DSBs.

Defects in NHEJ have been found in a very limited number of SCID patients. In these patients mutations in the *Artemis* gene have been identified. Cells derived from these patients display an increased sensitivity to ionizing radiation and defective V(D)J recombination. In contrast, patients defective in the *Lig4* gene show radiosensitivity but do not show severe immunodeficiency since there is only a partial defect in the gene. This suggests that the residual activity of Ligase IV is sufficient to repair the smaller number of breaks that arise during V(D)J recombination but not the larger number induced by ionizing radiation. Since the number of analyzed patients defective in NHEJ is still very limited, it cannot be concluded yet whether mutations in these genes lead to an increased cancer incidence in humans. Studies using mouse mutants, however, clearly demonstrated that defects in DSB-repair through either NHEJ or HR result in enhanced cancer incidence and underscore the importance of faithful repair of DSBs in maintaining genome stability.

►[DNA Repair Mechanisms](#)

References

1. Pastink A, Eeken JC, Lohman PH (2001) Genomic integrity and the repair of double-strand DNA breaks. *Mutat Res* 480-481:37–50
2. Mills KD, Ferguson DO, Alt FW (2003) The role of DNA breaks in genomic instability and tumorigenesis. *Immunol Rev* 194:77–95
3. Thacker J, Zdzienicka MZ (2003) The mammalian *XRCC* genes: their roles in DNA repair and genetic stability. *DNA Repair* 2:655–672
4. D'Andrea AD, Grompe M (2003) Fanconi anaemia/BRCA pathway. *Nat Rev Cancer* 3:23–34
5. Bartek J, Lukas C, Lukas J (2004) Checking on DNA damage in S phase. *Nat Rev Mol Cell Biol* 5:792–804
6. Krogh BO, Symington LS (2004) Recombination proteins in yeast. *Annu Rev Genet* 38:233–271

7. Lombard DB, Chua KF, Mostoslavsky R et al (2005) DNA repair, genome stability, and aging. *Cell* 120:497–512

Double-Stranded RNA-Dependent Protein Kinase

►[PKR](#)

Double-Stranded RNA-Mediated Gene Silencing

►[RNA Interference in Mammalian Cells](#)

Down Syndrome

Definition

Down syndrome is caused by an extra copy of chromosome 21; also called trisomy 21.

►[Chromosome 21, Disorders](#)

►[Molecular Aging Research](#)

Downstream

Definition

The term is used to describe the relative positions of sequences on a nucleic acid or protein molecule. "Downstream" defines the direction in which a nucleic acid or protein molecule is synthesized, i.e. on the 3' side of any given site in DNA or RNA, and on the C-terminal side of any site within a polypeptide, i.e. a protein is synthesized stepwise from its amino-terminal end to its carboxyl terminal end. With reference to the direction of genetic transcription, downstream is in the 3' direction on the non-template strand, but in the 5' direction on the template strand.

DPE

Definition

The DPE (down stream core promoter element) is a sequence motif that is commonly found in core promoters. The DPE cannot act independently, but functions along with the initiator (Inr) element.

► [Core Promoters](#)

Drosophila as a Model Organism for Functional Genomics

MARC HILD¹, RENATO PARO²

¹Novartis Institutes for BioMedical Research, Cambridge, MA, USA

²ZMBH, University of Heidelberg, Heidelberg, Germany
marc.hild@pharma.novartis,
paro@zmbh.uni-heidelberg.de

Definition

The fruit fly *Drosophila melanogaster* has an established and successful record as a valuable genetic model system for studying complex biological phenomena. For example, several basic signaling pathways have been discovered in this organism, and a lot of progress on understanding the interconnections between the components of the signaling cascades is still made using this system. Since the completion of the sequence of the *Drosophila* genome we now know that, compared to other known invertebrate genomes, *Drosophila* has the most genes in common with humans (1). Consequently, this has increased the interest of the pharmaceutical industry in using this organism as a powerful and economical model system for target screens and drug testing. Functional genomics, the new research paradigm employing a multitude of high-throughput methodologies (2) promises to add function to the raw, pure digital blueprint of life that was assembled in recent years, by finishing the genome sequences of the human and major model organisms. By integrating results from the analyses of the transcriptome (expression profiling/microarray analysis, ► [ChIP on chip](#)), the proteome (2D-gel electrophoresis and mass-spectrometry analysis) and the interactome (two-hybrid screens, affinity tag and mass-spectrometry analysis, ► [protein chips](#)) of model organisms, it will be possible to establish a more solid understanding of the intricate developmental and

cellular networks that are at the basis of multicellular organisms. Another advantage to this approach will not only be added by the availability of classical ► [gain- and ► loss-of-function screens](#)/mutants, but also by the newly discovered phenomenon of ► [RNA interference](#) (RNAi), which for the first time allows for an easy, systematic genome-wide ► [knock-down](#) of every gene in a model organism. Bioinformatics analyses of such datasets should finally unravel, and give an overview of the underlying mesh of interwoven networks of regulatory interactions in such a comprehensive manner as was not possible before (3).

Characteristics

The main advantages of *Drosophila melanogaster* as a model organism for functional genomics can be summarized as follows:

Bioinformatic Tools

FlyBase, a database of genetic and molecular data that was introduced in 1993 is vital for today's success in *Drosophila* research. While the primary information represented is from *Drosophila melanogaster*, it also presents an increasing amount of data from the family of other *Drosophilidae*. The database includes information on genes and mutant alleles, information about the expression and properties of transcripts and proteins, information on the functions of gene products, information about natural and engineered transposons and other molecular constructs, descriptions of *Drosophila* stocks held in stock centers and research laboratories, images that illustrate *Drosophila* anatomy and development terms, a bibliography of *Drosophila* references and the complete, annotated and curated genome sequence.

Resources

Stock centers, which provide thousands of different mutant lines to labs worldwide, represent a valuable resource for *Drosophila* research. In many cases it is not necessary to produce new mutant alleles, since for most genes mutant lines are already available, or will be in the near future, as there are several large scale gene disruption projects already under way. In addition to mutant alleles, the stock centers offer ► [P-element](#) insertion lines that allow for the controlled over-expression of genes as well as enhancer- or protein-trap lines, specific for a variety of tissues or cellular structures. Likewise, there is a huge collection of cDNA libraries, a gene collection containing sequenced cDNAs for 70% of all *Drosophila* genes, genomic clones such as BACs, cosmids, P1 or YACs, as well as a vast assortment of antibodies and an increasing number of cell lines.

Genetic Tools

Drosophila melanogaster offers a multitude of exceptional genetic tools that allow a thorough and

systematic manipulation of genetic networks. Examples include the availability of a series of differently marked ▶[balancer chromosomes](#), permitting an unproblematic, stable inheritance of mutant alleles in a heterozygous state. In addition, the absence or presence of the marker can be used to trace the fate of the balancer chromosome(s) during crossings. Transgenic lines are generated using the system of P-element mediated transformation, and the mobilization of carefully designed P-elements may be used for extensive, genome-wide gain- or loss-of-function screens. Mutants from such screens are often used as a starting point for additional ▶[enhancer/suppressor screens](#). This can, for example, further deepen our understanding of interconnected signaling pathways, by identifying (all) possible genes involved in modifying a specific signal. Using the ▶[binary Gal4-expression system](#), *Drosophila* not only offers temporal and spatial control on the expression of transgenes coding for proteins but also for the expression of RNAi constructs, and thus the controlled knockdown of endogenous genes. Lethal mutations may be studied by inducing homozygous mutant cell clones in a heterozygous background by means of the ▶[FLP/FRT system](#).

Genomic Tools

In addition to providing the rich source of information on genetic networks acquired by traditional genetic approaches, *Drosophila* research has also now successfully moved into the genomic age. Based on the success of genome-wide protein interaction mapping in yeast, *Drosophila* was the first multicellular eukaryote for which a comprehensive protein interaction map was established. Using the yeast two-hybrid system, Giot et al. (4) screened most of the *Drosophila* proteome (more than 10,000 of the ~13,600 proteins) to produce a draft map of 7,048 proteins and 20,405 interactions, which was refined to a high-confidence map of 4,780 unique interactions involving 4,679 proteins. Following this pioneering work, researchers established more detailed interaction maps by focusing on genes orthologous to human cancer-related proteins, signaling proteins or cell-cycle regulators. While the information acquired so far is still limited, the enormous potential is already obvious, and these types of studies will eventually offer the complete bird's eye view on the complicated mesh spanned by all possible protein-protein contacts. Validation may come from expression profiling data and/or genome-wide RNAi experiments performed in *Drosophila* cell lines. The gene sequences on microarrays needed for such profiling experiments are already available. At the moment, scientists can choose between commercially available oligonucleotide based systems (e.g. Affymetrix, MWG, Operon, Agilent, NimbleGen), a novel

long-oligo set developed by the INDAC (International *Drosophila* Array Consortium), chips based on the BDGP cDNA collection (*Drosophila* Gene Collection, DGC) covering about 70% of all annotated ORFs, and arrays using PCR-fragments representative of every predicted gene in the genome (Heidelberg FlyArray, Eurogentec, White lab, Incyte set). For all this methodology there are a growing number of successful applications, paralleled by an ever increasing amount of expression profiling data deposited in public database, representing a perfect starting point for further data mining projects. Examples include studies on gene expression changes during the life cycle of *Drosophila*, in dorsal-ventral patterning, in mesoderm development, during ageing, in alcohol abuse, in insecticide resistance, in nutritional control, during circadian gene control and during immune responses. Studies in yeast demonstrated the value of combining gene expression analyses with a genome-wide localization of transcription factor binding using ChIP on chip methodologies. In addition to commercially developed, oligo based systems (Affymetrix, NimbleGen), the *Drosophila* research community is actively working on the generation of an array system that will contain the complete genome pursuing two approaches: (1) generation of a large oligonucleotide set for the generation of overlapping PCR fragments spanning the complete genome (K. White, S. Russell, and G. Micklem), and (2) use of the clone collection developed for the generation of the genome sequence (E. Furlong, S. Celniker, J. Hoheisel).

In contrast to mammalian cells, the simple addition of dsRNA to *Drosophila* cells gives fully penetrant phenotypes, opening up the way for systematic ▶[reverse genetic](#) approaches. After an initial, limited study (based on DGC Release 1, ~40% coverage) demonstrated the feasibility of such experiments, another study identified several hundred genes involved in cell growth/viability in different *Drosophila* cell lines (5), scanning the whole transcriptome by preparing dsRNAs to all predicted genes using the Heidelberg FlyArray PCR fragment set. Based on this pioneering work, novel RNAi facility were established at Harvard Medical School (Boston, MA, USA) and the German Cancer Centre (DKFZ, Heidelberg, Germany), not only providing unrestricted access to datasets of previous genome-wide RNAi screens and tools for the design of new, optimized dsRNAs, but also inviting researchers to propose and perform additional screens which will greatly benefit from this standardized set of reagents, and the established excellence in RNAi studies at Harvard and Heidelberg. Consequently, additional RNAi sets were developed by academic laboratories (P. O'Farrell, R. Vale, G. Davis, E. Foley, B. Eaton, N. Stuurman, S. Rogers, distributed by OpenBiosystems) and commercial sources (Ambion).

***Drosophila* as a Model Organism for Functional Genomics. Table 1** Weblinks for Functional Genomics in *Drosophila*

Name	Description	URL
BDGP <i>in situ</i> Database	Patterns of gene expression in <i>Drosophila</i> embryogenesis.	► http://www.fruitfly.org/cgi-bin/ex/insitu.pl
Bloomington Stock Center	The Bloomington Stock Center collects, maintains and distributes <i>Drosophila melanogaster</i> strains for research.	► http://flystocks.bio.indiana.edu/
DrosDel Project	Construction of a new <i>Drosophila</i> deletion collection and a European <i>Drosophila</i> network.	► http://www.drosdel.org.uk/
<i>Drosophila</i> ArmView2	Browse genomic sequence maps by chromosome arm with interactive gifs.	► http://www.fruitfly.org/cgi-bin/annot/arm_view.pl
<i>Drosophila</i> RNAi Screening Center at Harvard medical school	The goal of <i>Drosophila</i> RNAi Screening Center (DRSC) is to make the technology of genome-wide RNAi screens available to the community. The site offers complete datasets, protocols and contact information to perform your own screens.	► http://flyrna.org/
e-RNAi	Web-based tool to evaluate and design optimized (efficiency and specificity) dsRNA constructs. Moreover, the application also gives access to published predesigned dsRNAs.	► http://www.dkfz-heidelberg.de/signaling/e-mai/
Fly BLAST	Sequence similarity searching of <i>Drosophila</i> -specific datasets.	► http://www.flybase.net/blast
Fly Enhancer	Fly Enhancer is a search engine designed to find clusters of binding sites (or any sequences of nucleotides) in the <i>Drosophila melanogaster</i> genome.	► http://flyenhancer.org/Main
FlyBase	A Database of the <i>Drosophila</i> Genome.	► http://www.flybase.org/
FlyBase Genome Browser	Query and browse genomic sequence maps with interactive gifs.	► http://www.fruitfly.org/cgi-bin/annot/gbrowse
Flybrain	An Online Atlas and Database of the <i>Drosophila</i> nervous system	► http://flybrain.uni-freiburg.de/
FlyGrid	General repository for <i>Drosophila</i> interaction datasets. Currently, Fly GRID contains more than 20000 interactions all of which are available both online, and through Osprey.	► http://biodata.mshri.on.ca/fly_grid/servlet/SearchPage
FlyMove	FlyMove is resource to study the development of the fruit fly <i>Drosophila melanogaster</i> . It will provide you with a variety of images, movies and interactive Shockwaves, helpful in understanding complex developmental processes	► http://flymove.uni-muenster.de/
Flytrap	A GFP Protein Trap Database	► http://flytrap.med.yale.edu/
Flyview	FlyView is the beginning of an image database on <i>Drosophila</i> development and genetics, especially on expression patterns of genes (enhancer trap lines, cloned genes).	► http://flyview.uni-muenster.de/
Gene Disruption Project	P-screen database containing information on the ongoing P-element Screen/Gene Disruption Project of the Bellen/Rubin/Spradling labs	► http://flypush.imgen.bcm.tmc.edu/pscreen/

***Drosophila* as a Model Organism for Functional Genomics. Table 1** Weblinks for Functional Genomics in *Drosophila* (Continued)

Name	Description	URL
GenomeRNAi	GenomeRNAi is a database of phenotypes obtained in systematic RNA interference (RNAi) experiments in cultured <i>Drosophila</i> cells. The phenotype database can be searched by keywords, RNAi identifiers or gene sequences. Searches with homologous sequences from human or <i>C. elegans</i> are also possible.	► http://www.dkfz-heidelberg.de/signaling/ernai/ernai.html
Homophila	Human disease to <i>Drosophila</i> gene database.	► http://superfly.ucsd.edu/homophila
Protein Interaction Maps - PIMS	Information on the mapping of the protein-protein interactions encoded by the <i>Drosophila melanogaster</i> genome.	► http://proteome.wayne.edu/PIMproject1.html
Proteome analysis@EBI	Integr8 is a browser for information relating to completed genomes and proteomes, based on data contained in Genome Reviews and the UniProt proteome sets. It provides access to species descriptions, literature, statistical analysis and summary information about each complete proteome; and integrates data from a variety of sources, including InterPro, CluSTR and GO.	► http://www.ebi.ac.uk/integr8/OrganismHomePage.do?orgProteomeId=17
Szeged <i>Drosophila</i> Stock Centre	The Szeged Stock Centre collects, maintains and distributes <i>Drosophila melanogaster</i> P insertion strains for research.	► http://expbio.bio.u-szeged.hu/fly/index.php
The Heidelberg FlyArray	Query and browse genomic sequence maps with interactive gifs. The site includes, among others, a comparative view on different annotations of the <i>Drosophila</i> genome as well as on conserved regions between <i>D. melanogaster</i> and <i>D. pseudoobscura</i> , an EST clustering and details on the Heidelberg PCR fragment set which is the basis for genome-wide RNAi screens performed at Harvard medical school and a whole-transcriptome microarray.	► http://hdflyarray.zmbh.uni-heidelberg.de/
Inparanoid	Inparanoid provides information on orthologs while differentiating between inparalogs and outparalogs.	► http://inparanoid.cgb.ki.se/index.html
The Interactive Fly	A cyberspace guide to <i>Drosophila</i> genes and their roles in development.	► http://www.flybase.org/allied-data/lk/interactive-fly/aimain/1aahome.htm
Whole Genome comparison of <i>Drosophilidae</i> genomes	This site provides a single source for sequences, assemblies, annotations and analyses of the genomes of members of the fruitfly genus <i>Drosophila</i> . It is meant as resource for Drosophilists and other researchers interested in comparative analysis of these species and their genomes.	► http://rana.lbl.gov/drosophila/

With all these tools at hand (or at least close to), *Drosophila* research is well prepared to tackle the next challenging steps in functional genomics.

Clinical Relevance

An increasing number of researchers have tried to exploit the power that the *Drosophila* genetic system offers, to dissect pathways of disease pathogenesis and

identify targets for therapeutic intervention (6). The knowledge that 75% of all known human disease genes are conserved in the *Drosophila* genome (7) and the experimental advantages of the *Drosophila* system (8) formed the basis for several investigations, trying to decipher the normal and pathogenic functions of genes implicated in, for example, neurodegenerative diseases. Consequently, a multitude of transgenic models were

produced to analyze disease pathways including Parkinson's disease, Alzheimer's disease, Huntington's disease, and several spinocerebellar ataxias in an *in vivo* model. Information gained from these models implicate protein folding and degradation pathways in Parkinson's disease and the polyglutamine repeat disorders, as well as kinases and apoptotic pathways that may modulate neurodegeneration in tauopathies. Functional genomics in *Drosophila* will therefore contribute substantially to the elucidation of complex disease pathways/networks, and speed up the process of target identification as well as target validation. Finally, there is an increasing interest in the *Drosophila* system as a basis for drug screens, phenotype screens, which circumvent the classical target identification process and directly aim for lead substances that qualitatively improve the disease phenotype in the fly model. Table 1.

References

1. Rubin GM, Yandell MD, Wortman JR et al (2000) Comparative Genomics of the Eukaryotes. *Science* 287:2204–2215
2. Bader GD, Heilbut A, Andrews B et al (2003) Functional genomics and proteomics: charting a multidimensional map of the yeast cell. *Trends Cell Biol* 13:344–356
3. Alm E, Arkin AP (2003) Biological networks. *Curr Opin Struct Biol* 13:193–202
4. Giot L, Bader JS, Brouwer C et al (2003) A protein interaction map of *Drosophila melanogaster*. *Science* 302:1727–1736
5. Boutros M, Kiger AA, Armknecht S et al (2004) Genome-wide RNAi analysis of growth and viability in *Drosophila* cells. *Science* 303:832–835
6. O'Kane CJ (2003) Modelling human diseases in *Drosophila* and *Caenorhabditis*. *Semin Cell Dev Biol* 14:3–10
7. Reiter LT, Potocki L, Chien S et al (2001) A systematic analysis of human disease-associated gene sequences in *Drosophila melanogaster*. *Genome Res* 11:1114–1125
8. Bier E (2005) *Drosophila*, the golden bug, emerges as a tool for human genetics. *Nat Rev Genet* 6:9–23

Drosophila as a Model to Study Cardiac Disease Genes

ROLF BODMER, ETHAN BIER
UCSD and The Burnham Institute, La Jolla, CA, USA
rolf@burnham.org

Definition

The fly *Drosophila melanogaster* is an invertebrate model system that is exquisitely suited to genetic

studies (<http://flybase.net>). For more than 100 years, genetic methods in *Drosophila* been developed and refined. These powerful tools have provided detailed mechanistic insights into a variety of widely applicable biological problems, including creation of the first genetic maps, discovery of sex-linkage, elucidation of genetic recombination, demonstration of induced mutation by radiation and more recently, a nearly complete understanding of the initial embryonic patterns and the molecular logic of setting up the [primary body axes](#). The insights gained from these studies provided the fundamental concepts for unraveling the genetic basis of vertebrate development and [organogenesis](#).

Characteristics

Human, mouse, *Drosophila* and many other genome sequences have now been completed. As a consequence, a comprehensive cross-genomic analysis was conducted between *Drosophila* and all human disease genes known to have at least one mutant allele listed in the Online Mendelian Inheritance in Man (OMIM; <http://www.ncbi.nlm.nih.gov/omim>) using an interactive database called Homophila: <http://homophila.sdsc.edu> (5). Of the current 2400 human disease genes, 74% have homologs in *Drosophila* and nearly a third of them are as highly conserved as genes known to be functionally equivalent. These *Drosophila* genes include homologs of genes causing a broad spectrum of human diseases, including neurological disorders, cancer and cardiovascular disease. Here we concentrate on genes involved in [congenital heart disease](#) (CHD). Many genes have been identified which can lead to CHD or susceptibility to cardiac dysfunction when mutated. Among them are genes encoding factors involved in cardiac development (e.g., [transcription factors](#)), cardiac muscle contraction (e.g., cytoskeletal proteins), and conduction of electrical signals (e.g., [ion channels](#)). Although many genes causing CHD have been identified, it is estimated that most of the suspected genetic factors have yet to be identified. An important challenge in elucidating the genetic basis of CHD is to identify new genetic loci and to understand how allelic variants interact combinatorially to cause a genetic risk for cardiac arrest and failure.

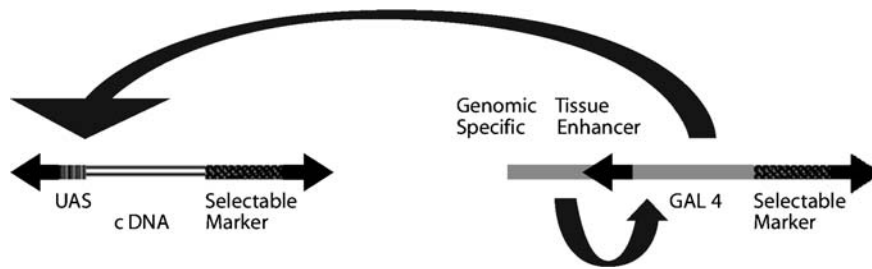
Of the CHDs, [cardiomyopathies](#) typically result from defects in genes encoding contractile proteins, whereas many conduction defects result from malfunctioning mutations in ion channels and the genes encoding them (Table 1). In addition, genes controlling early stages of heart development have been highly conserved between vertebrates and invertebrates (1), indicating that the common ancestor of flies and humans had already invented the basic contractile fluid pumping organ from which all hearts in the animal kingdom eventually evolved. Thus, there is good reason to believe that

***Drosophila* as a Model to Study Cardiac Disease Genes. Table 1** Select list of cardiac disease genes with *Drosophila* counterparts

Disease Category	Disease	OMIM #	e-value	Fly Genes	Protein Function
Cardiomyopathy	Cardiomyopathy, fam. Hypertr., 1	160760	e-300	Mhc, zip	Cardiac myosin heavy chain- β
	Cardiomyopathy, hypertr., midventr. chamber type (3)	160790	2e-34	Mlc-c	Essential light chain of myosin
	Cardiomyopathy, hypertr., mid-left ventr. Ch. type (3)	160781	3e-42	sqh	Myosin ATPase
	Cardiomyopathy, fam. hypertrophic, 9	188840	e-300	bt	Titin: myosin light chain kinase
	Cardiomyopathy, fam. hypertr., 4	600958	e-70	CG18242	Cardiac myosin binding protein C
	Cardiomyopathy, fam. hypertr., 3	191010	2e-63	Tm1	Tropomyosin
	Cardiomyopathy, fam. hypertr., 2	191045	5e-12	up	Troponin complex
	Cardiomyopathy, fam. hypertr., 3	191044	e-10	wupA	Troponin I, cardiac form
	Cardiomyopathy, idiopathic dilated	102540	e-300	Act5C, 42A 57B, 79B, 87E, 88F	Smooth muscle actin
	Cardiomyopathy, dilated, X-linked, Becker & Duchenne muscular dystrophy	310200	e-300	Dys	Dystrophin
	Cardiomyopathy (1), Myopathy, desminopathic	125660	3e-37	Lam	Desmin
	Barth Syndrome	302060	3e-61	tafazzin	P-lipid & glycerol acyl transferase?
Conduction	Jervell & Lange-Nielsen syndrome	192500	e-64	CG12215	Potassium channel
	Long QT syndrome-2	152427	e-300	sei	Potassium channel
	Long QT syndrome-3	600163	e-300	para	Sodium channel
	Atrial septal defect with atrioventr. cond. defects	600584	2e-26	vnd	Transcription factor
	Stress-induced polymorphic ventricular tachycardia	604772	e-300	Rya-r44F	Ryanodine Receptor
	Arrhythmogenic right ventricular dysplasia	600996	e-300	Rya-r44F	Ryanodine Receptor

many of the *Drosophila* homologs of human cardiac disease genes will function similarly to their human counterparts, and that new genes that are identified in *Drosophila* that interact with these disease gene homologs will themselves be candidate cardiac disease

genes. Mutations already exist in many *Drosophila* homologs of human disease genes, and transgenic constructs can be readily introduced into flies to mis-express genes of interest in various spatially and temporally specific patterns using the [▶UAS-Gal4](#)



Drosophila as a Model to Study Cardiac Disease Genes. Figure 1 The UAS/Gal4 Expression System: (1) Cross UAS-cDNA stock to Gal4 driver stock. (2) Gal4 transcription factor activates UAS-cDNA in progeny in the pattern dictated by the tissue-specific enhancer driving Gal4.

system (2). More sophisticated GAL4-dependent mis-expression schemes have also been devised, which allow for conditional mis-expression of genes in certain tissues and for rapid isolation of dominant alleles of a gene of interest (3, 4). *Drosophila* is therefore ideally suited for analysis of gene function through loss-of-function genetics and mis-expression studies. In contrast to this efficient manipulation of genes in flies, knocking out human disease-related genes in mice is costly, time consuming and frequently results in very modest mutant phenotypes. It also is non-trivial to mis-express transgenes in mice in regulated temporal and tissue specific patterns. Given the limitations of model systems and the conservation of principle control functions that extends to insects, the genetic versatility of *Drosophila* is a viable alternative to mammalian experimental assay systems such as mice for studying the function of human disease genes. Furthermore, *Drosophila* is currently the only simple model system for examining heart function, since other invertebrate genetic models such as *C. elegans* lack a heart.

An important type of genetic scheme in *Drosophila* is the so-called second-site enhancer-suppressor screen. The idea of this type of screen is to start off with a partial loss-of-function mutation in one component of a pathway and then search for ▶heterozygous mutations in other genes that greatly increase or decrease the severity of the phenotype resulting from this first mutation. The ability to pick such ▶synergistic interactions in rapid F1-generation screens allows the researcher to survey the entire genome for mutations in genes that interact in a common process. Classically, comprehensive genetic screens in *Drosophila* make use of relatively random chemical or radiation mutagenesis methods. More recently, large-scale mutagenesis by transposon insertion has offered several advantages over point mutagens. An insertional mutant greatly facilitates the molecular characterization of the nearby gene and, in addition, enhancer-trapping transposon vectors have been designed that report on the expression of genes next to the site of transposon integration. A growing number of transposon insertions

have been recovered in major genomic loci in *Drosophila* (▶<http://flybase.net>), which greatly facilitates mutant analysis. The UAS-Gal4 system has also been applied to phenocopying loss-of-function mutations by mis-expressing constructs that lead to transgenic RNA interference (▶RNAi).

▶Evolutionary Conservation of Heart Formation

The discovery of the ▶homeobox transcription factor Tinman in *Drosophila* and its conservation in vertebrates provided the first evidence that heart development was controlled by homologous pathways in the animal kingdom. The elucidation of conserved regulatory genetic networks involved in cardiac specification led to molecular models of cardiac diseases in humans (6). Remarkably, basic aspects of cardiac morphogenesis have also been conserved during evolution. Despite the reversal of the dorsal-ventral axis between vertebrates and invertebrates and obvious differences in terminal morphology, the cardiac progenitors in all heart-forming creatures originate from bilateral symmetrical mesodermal tissues located most distally from the axis of gastrulation, which then fuse into a primitive heart tube at the midline (1). Over the past decade has work in *Drosophila* provided a detailed molecular understanding of the complex interplay between positional cues provided by inductive factors secreted from the ectoderm (Wnts and TGF-βs) and transcription factors expressed in the mesoderm (homeodomain protein Tinman, GATA factor Pannier and T-box proteins), which specify and position the heart field within the embryo. The key determinants of ▶cardiogenesis in the fly also play fundamental roles in early cardiac specification in vertebrates and they do so in similar ways. For example, in flies and vertebrates Dpp/BMP signals cross germ layers to maintain Tinman/NKX2-5 expression. A similar relationship exists between Tinman/NKX2-5 and Pannier/GATA-4. Moreover, mutations have been identified in humans in many of these conserved heart-specifying genes, such as NKX2-5, GATA-4 and TBX-5, which cause various forms of



***Drosophila* as a Model to Study Cardiac Disease Genes.** Figure 2 Performance assay using an electrical pacing paradigm. Flies are arranged touching foil electrodes (left) and the heart failure rate (right) is measured following a 30-second pacing protocol at 6 Hz in one week old flies (left bar: 1) and five week old flies (right bar: 5). Note the three- to four-fold increase in failure rate with age (20-80% failure rate) (see also ref. 7).

CHD. This deep conservation of genetic control mechanisms makes *Drosophila* a versatile genetic model system to identify interacting combinations of candidate genes responsible for CHD.

Drosophila as a Model for Cardiac Function

The fly heart offers a number of advantages as a screening device for genes involved in CHD. In addition to the general efficiency and versatility of performing genetic experiments in flies, the fly heart is a simple and highly ordered linear tube in which morphological changes can easily be detected (1). It consists of two major cell types: the inner contractile myocardial cells, which form a central cavity, the lumen of the heart and the outer, non-myogenic pericardial cells. Importantly, insects have a tracheal system; thus, the heart is not needed for distributing oxygen within the body, which bypasses the problem of early lethality in vertebrates due to cardiac malfunctions. In contrast to our understanding of *Drosophila* heart development, not as much is known yet about how cardiac physiology is controlled. Nevertheless, mutations in ion channels, gap junctions and other membrane proteins involved in ionic homeostasis influence the *Drosophila* heart rate and/or rhythm as they do in humans. Moreover, a large-scale screen revealed dozens of mutants that affect the fly's heart function, and a majority of the corresponding genes have human homologs, which may very well turn out to be candidate CHD genes (R.J. Wessells and R.B., unpubl.).

A number of methods have recently been developed to examine heart function in *Drosophila* larvae and adults, which include measuring the rate of the heartbeat using an edge tracing system, the rhythmicity of the heart beat using an autocorrelation analysis, and

the rate of heart failure (cardiac arrest or fibrillation) in response to stress provided by external [▶electrical pacing](#) to increase the heartbeat frequency (Fig. 2) (7). Using these assays, it is now possible to target UAS-transgenes of interest or UAS-RNAi constructs exclusively to the heart, in order to determine the heart-autonomous effect of mis-regulated or compromised function of one or more genes on cardiac performance. This approach is particularly useful for heart-specific screens for genes involved in cardiac performance and senescence.

Drosophila is an Essential Gene Identification Tool in the Post-Genomic Era

Nearly complete genome sequences are available for humans and a number of model organisms. An important goal in the genomics field is to identify genes that interact to cause complex phenotypes such as human heart disease. It is likely that *Drosophila* and other model systems will play an increasingly important role in the functional analysis of candidate genes involved in [▶polygenic disorders](#). The advantage of these model organisms is that it is possible to survey large numbers of interacting genotypes, which may potentially be involved in CHD. To study the nature of polygenic traits these model systems can be used as *in vivo* test systems to identify small sets of interacting genes for further analysis in a vertebrate model.

Using Model Systems to Identify Gene Functions Involved in Polygenic Traits

One of the challenges of complex traits is to determine which genes among several plausible candidates at multiple loci contribute most to a phenotype in question. Interactions among human disease gene

homologs involved in cell autonomous functions have been successfully identified in yeast. For example, genes required for mismatch repair of DNA cause cancer. Some of these proteins are indispensable for mismatch repair while others function redundantly to assemble or execute the function of complexes binding to mismatched bases. Another example is provided by congenital disorders of glycosylation, which affect consecutive steps in the synthesis of carbohydrate chains. In this case, based on mutant analysis in yeast and rescue of the defect by the human homolog, a screen was devised to identify other genes acting in the same glycosylation pathway in yeast. Subsequently, it was found that mutations in human homologs of these newly identified yeast genes caused disorders of glycosylation. Thus, model genetic systems can be viewed as first pass *in vivo* filters or as “genetic bioassays” to search for relevant genetic interactions among a potentially large set of genes. Interactions among this more manageable set of genes can then be rigorously tested in a vertebrate system such as the mouse or zebrafish. Yeast is ideal for determining genetic interactions between conserved gene networks acting within single cells, but *Drosophila* is the simplest genetic model for studying polygenic traits involved in heart function.

► **RNA profiling** experiments provide an important post-genomic source of information regarding function of organs such as the heart. Initial datasets of this kind identify several hundred genes whose levels change in a statistically significant fashion. If one considers a subset of these genes whose levels change by more than 10-fold, the number often drops to 1–2 dozen genes. This is a manageable number of genes to examine in all pair-wise combinations for genetic interactions in *Drosophila*. If a small subset of interacting combinations is identified, one can then test for synergistic interactions involving more genes. Experiments of this scale are impractical in any vertebrate model system. Once a key subset of interacting genes have been identified in the *Drosophila* genetic model system, synergistic interactions can then be tested in a suitable vertebrate model system.

How do multiple genes contribute to a complex trait such as CHD? When the activities of two or more genes that function in a single genetic pathway are reduced, mutant phenotypes may be observed. This is a classic condition that experimental geneticists often create when screening for second site mutations that enhance the effect of a reference mutation. In simple situations of this kind, complete loss of function of two different genes in the pathway leads to the same phenotype, which may be similar to the combined reduced function of both genes. Another common basis for multigenic effects is in cases where duplicated genes provide overlapping activities. As the vertebrate genome has

expanded by rounds of large-scale duplication, such ► **genetic redundancy** is often encountered as is well exemplified by the overlapping functions of four nearly complete copies of the HOX gene cluster, which is present in only a single copy in *Drosophila*.

Synergistic interactions can also occur between genes acting in different pathways or performing diverse cellular functions, which need to be coordinated to accomplish a biological process. For example, normal cell migration involves genes functioning in signal/pathway recognition, regulated cytoskeletal reorganization, target identification and terminal differentiation. Evidence for the existence of such cooperation between heterologous genetic pathways is provided by contiguous gene syndromes, in which the gene dose of several structurally unrelated but neighboring genes are altered by either deletion or duplications of small chromosomal regions. In Down syndrome (DS), for example, small duplications of regions of chromosome 21 have been associated with CHD and deletions of small chromosomal intervals have been implicated in cardiac phenotypes caused by DiGeorge and Williams syndromes. A possible explanation for the coordinated function of adjacent structurally unrelated genes is that they are co-regulated by a shared cis-regulatory element. Indeed, several genes associated with CHD in DS, or genes contained within the deleted interval of DiGeorge syndrome are co-expressed in the heart. Regarding DiGeorge and other syndromes, there has been some doubt as to whether these diseases result from misregulation of multiple contiguous genes or from one gene alone, even though human *TBX1* and *ELASTIN* have been identified as causing CHD phenotypes similar to those observed in DiGeorge or Williams syndromes, respectively. There is also evidence, however, that phenotypes resulting from monosomy of small deletions encompassing these genes are considerably more severe than those due to the single mutations. Furthermore, mutations in another gene (*CRKL*) within the DiGeorge critical deletion also cause CHD in mice, suggesting that both *TBX1* and *CRKL* may participate in manifestation of the full syndrome.

Clinical Relevance

Because homologous genes control early developmental events, as well as functional components of the *Drosophila* and vertebrate hearts, the fly is the simplest model system for assaying genes involved in human CHD. The wide variety of genetic tools available to *Drosophila* researchers offers many technical advantages for rapidly screening through large numbers of candidate genes. Therefore, an important future direction is likely to be the use of *Drosophila* as a vehicle for analyzing polygenic traits. Indeed, this may be the most important long-term use of *Drosophila* as an aid in human genetics. One can anticipate a time in the not too

distant future when mutant lines exist for every gene in vertebrate systems such as mice and zebrafish. However, one of the enduring problems not easily addressed by such resources will be the tracking of complex traits defined by polygenic variants. For this level of genetic analysis, simple model genetic systems including yeast, *C. elegans* and *Drosophila* will undoubtedly play a crucial ongoing role. Among them, *Drosophila* will be critical for examining gene networks and interactions involved in organogenesis and clearly the system of choice with regard to cardiac development, function and aging since it is the only one with a fluid pumping heart.

References

1. Bodmer R (1995) Heart development in *Drosophila* and its relationship to vertebrate systems. *Trends Cardiovasc Med* 5:21–28
2. Brand AH, Perrimon N (1993) Targeted gene expression as a means of altering cell fates and generating dominant phenotypes. *Development* 118:401–415
3. Duffy, JB (2002) GAL4 system in *Drosophila*: a fly geneticist's Swiss army knife. *Genesis* 34:1–15
4. Guichard A, Srinivasan S, Zimm G et al (2002) A screen for dominant mutations applied to components in the *Drosophila* EGF-R pathway. *Proc Natl Acad Sci USA* 99:3752–3757
5. Reiter LT, Potocki L, Chien S et al (2001) A systematic analysis of human disease-associated gene sequences in *Drosophila melanogaster*. *Genome Res* 11:1114–1125
6. Seidman CE, Seidman JG (2002) The coming of age of cardiovascular science. *Cold Spring Harb Symp Quant Biol* 67:543–550
7. Wessells RJ, Fitzgerald E, Cypser JR et al (2004) Insulin regulation of heart function in aging *Drosophila*. *Nat Genet* 36:1275–1281

Drosophila Melanogaster

Definition

Drosophila melanogaster, commonly called fruit fly, is frequently used in genetic developmental studies.

- [Drosophila as a Model Organism for Functional Genomics](#)
- [Drosophila as a Model to Study Cardiac Disease Genes](#)
- [Molecular Aging Research](#)

DRPLA

Definition

DRPLA (Dentatorubral pallydoluysian atrophy) refers to a rare neurodegenerative disease, primarily reported

to occur in Japan, with a spectrum of multiple system degenerations resembling Huntington's disease and spinocerebellar atrophy.

- [Polyglutamine Disease, the Emerging Role of Transcription Interference](#)
- [Huntington's Disease](#)

Drug Abuse

- [Addiction, Molecular Biology](#)

Drug Substitution and Anti-Craving Medication

Definition

To prevent the withdrawal and craving that often provoke relapse, drug addicts are treated with either substitution or ►[anti-craving compounds](#). Both substitution and anti-craving drugs prevent relapses, and through this effect will help to regain control over behavior. Treatment options for heroin addiction include medications (methadone, LAAM, synthetic opiates), as well as behavioral therapies integrated with other supportive services. Several other medications for use in heroin treatment programs are under study. (See also ►[acamprosate](#) and naltrexone for relapse prevention in alcoholism; nicotine patch and gum for tobacco addiction).

- [Addiction, Molecular Biology](#)

Drug Target

Definition

Drug target is typically a protein but sometimes a RNA, which is produced by an essential gene. Such a target can be used in conjunction with high throughput screening, to discover a new drug that inhibits the target, and thus cures a certain disease.

- [Protein Interaction Analysis: Variations of the Yeast Two-Hybrid System](#)

Drug Tolerance

- [Tolerance \(Drug Tolerance\)](#)

Drug-Seeking Behavior

Definition

The increase of extracellular ►[dopamine](#) in the nucleus accumbens, striatum and the frontal cortex is of pivotal importance for the initiation and the maintenance of drug-seeking behavior. Most of the drug-taking behavior is not cognitively perceived and controlled.

►[Addiction, Molecular Biology](#)

Drusen

Definition

Drusen are extracellular deposits that accumulate below the retinal pigment epithelium and are risk factors for developing age-related macular degeneration.

►[Proteomics in Ageing](#)

DS

►[Entropy Change \(DS\)](#)

DSC

►[Differential Scanning Calorimetry](#)

DSL-Family

Definition

DSL stands for Delta, Serrate, Lag-2. The DSL family of signal proteins comprises mainly of ligands of the Notch receptors. They are characterized by the presence of the DSL domain in the extracellular domain.

►[Notch Pathway](#)

DSN

►[Dejerine-Sottas Neuropathy](#)

Duchenne Muscular Dystrophy

ERYNN S. GORDON, ERIC P. HOFFMAN

Children's National Medical Center, Washington, DC, USA

ehoffman@cnmcresearch.org

Definition

Duchenne muscular dystrophy (DMD) is an X-linked recessive neuromuscular disorder that shows progressive weakness of skeletal muscle. DMD primarily affects males, with symptoms generally presenting by four years of age. Muscle biopsies from individuals with Duchenne muscular dystrophy show a complete absence of the protein ►[dystrophin](#). A milder form of DMD is ►[Becker muscular dystrophy](#) (BMD), in which there is a decrease in, but not loss of, dystrophin. Patients with BMD typically have a later onset and a slower disease progression than patients with DMD.

Duchenne muscular dystrophy (DMD) is the most common form of muscular dystrophy, making it one of the most common inherited lethal disorders (►[incidence](#) 1/3,500 males). Although DMD affects all populations equally, the incidence has decreased in developed countries in recent years due to genetic counseling and prenatal testing. The high incidence of DMD is in large part due to the high ►[mutation rate](#) in the dystrophin gene with a new mutation occurring in 1/10,000 eggs and sperm. Patients with Duchenne muscular dystrophy (as well as patients with other types of dystrophinopathies) show marked elevation of serum ►[creatinine kinase](#) levels (CK), a blood marker for muscle damage.

Characteristics

Boys with Duchenne muscular dystrophy generally start to show noticeable symptoms between four and six years old. In families where an individual has already been diagnosed with Duchenne, parents may notice symptoms earlier. The main symptoms reported by parents include delayed walking, and difficulties in walking, getting up from the floor and walking up stairs. Additional symptoms noticed by parents include learning and speech problems. The first muscles affected by DMD are those closest to the body (proximal muscles). Weakness in the pelvic area can cause boys to appear to waddle in their walk. In order to rise from the floor, boys use the classic Gower's maneuver, which entails using hands on the floor and then the knees to push oneself into the standing position.

Over time, strength is lost and the ability to walk is lost. In general, boys with DMD become wheelchair bound between seven and fifteen years of age, with the

average being eleven years of age. However, the use of steroids in the treatment of DMD has been proven to slow the disease progression and may in time significantly alter this time frame. Along with weakness of the muscles, patients with DMD may develop ►**cardiomyopathy** (weakness of the heart muscle). Finally, patients often die of respiratory failure in their early twenties. The respiratory problems in DMD are caused by weakness of the muscles between the ribs that aid in breathing. Throughout the disease progression the muscles of the face as well as muscles controlling speech are not affected. Although mental retardation is not associated with DMD, in general, boys with Duchenne muscular dystrophy have a lower IQ than average; the mean IQ for boys with DMD is about 85, while the general population mean is 105. In addition, specific learning disabilities, such as listening comprehension difficulties are common in boys with DMD.

Manifesting Carriers

Women, unlike men, have two copies of the X chromosome. Therefore, women carrying a disease causing gene change in the dystrophin gene on one of their X chromosomes, are not generally affected by DMD because the second X chromosome compensates for the non-working copy of the gene (instructions). Females who are symptomatic are known to have ►**skewed X inactivation**. Since women have two copies of the X chromosome, and men only have one copy, women turn off one copy of the X chromosome in each of their cells. In general, we expect each female cell to decide randomly which X they will turn off, resulting in random X-inactivation. In some cases, cells of a woman always turn the same X chromosome off (skewed X-inactivation), if the X chromosome with a non-working copy of dystrophin is the one left active, a woman will show symptoms of Duchenne. Although symptoms are not common in female carriers of DMD, some women may have muscle cramping and others may show abnormalities on electrocardiogram. It is therefore important to identify carriers of DMD in a family with an affected individual both for the purposes of genetic counseling and potential medical complications.

Cellular and Molecular Regulation

The dystrophin gene is the largest gene identified to date. Genes are made of four different base pairs repeated in a meaningful code. At about 2.5 million base pairs, the dystrophin gene is nearly ten times bigger than the next largest gene. Genes have different sections referred to as ►**introns** and ►**exons**. When a gene is being converted to a protein the introns are cut out, while the exons are used as the instructions. The majority of the dystrophin gene sequence is in non-coding introns (about 200 base pairs of intron for every

base of exon, over a total of 79 exons). The complete protein coding sequence is about 11,000 base pairs and encodes a large protein that weighs 427 kD. The protein product, named dystrophin, is completely absent in patients with DMD and is present but abnormal or reduced in patients with BMD. Although it is the lack of dystrophin in muscles that causes the disease symptoms, different forms of dystrophin have also been detected in the eyes, liver, and other neurons.

All Duchenne muscular dystrophy patients show a complete loss of function of the dystrophin protein. The most common type of gene change in both disorders is a deletion of one or more exons of the gene. Because of the large size of the introns, relative to exons, nearly all deletion mutations start and end in an intron. There are two deletion hot spots, one in the region of exons 44–50, and a second near the beginning of the gene (exons 2–13). On a molecular level, the main difference between DMD and BMD is the ►**reading frame** of the gene. Genes are read in groups or “words” of 3 base pairs. When a mutation in a gene alters the 3 base pair “word” to 2 base pairs, then the reading frame is put out of phase, and the protein reading frame becomes incorrect (frame-shift). In general, patients with DMD have deletions that shift the reading frame, while patients with BMD have deletions that keep the reading frame intact. There are also other key regions of the gene that can affect how severe the disease will be. Approximately 45% of DMD patients and 30% of BMD patients do not have deletions but instead have duplications or point mutations. Duplications, accounting for 5% of cases of DMD or BMD, follow the same reading frame rules as deletions.

Clinical Relevance

There are many reasons for narrowing a diagnosis from the general, muscular dystrophy, to the specific, Duchenne muscular dystrophy. This information is needed for providing families with the appropriate prognosis, treatment, and genetic counseling. Currently genetic testing is available for Duchenne muscular dystrophy, however due to the enormous size of the gene it remains difficult to detect the specific gene mutation in all families. The gold standard for testing is still considered to be the lack of dystrophin protein on muscle biopsy.

Genetic counseling and prenatal diagnosis has reduced the number of families with multiple relatives with Duchenne muscular dystrophy in the USA. However, the high new mutation rate of the Duchenne muscular dystrophy gene in all eggs and sperm means that the disease will always occur at a relatively high rate. If parents have a single child with Duchenne muscular dystrophy, and no other affected family member, there is a 66% chance that the mother is a carrier of the disease, and a 33% chance that she is not a carrier (the

mutation happens in just a single egg in the latter case). For women who are defined as definite carriers for Duchenne muscular dystrophy 50% of male pregnancies will be affected with the disease, and 50% of female pregnancies will be carriers. For those individuals concerned about being carriers or passing the disease on to future generations, DNA testing, prenatal diagnosis and ►[assisted reproduction](#) are options currently used by some families.

References

1. Hoffman EP (1997) The muscular dystrophies. In: Rosenberg RN, Prusiner SB, Di Mauro S et al (eds) *The Molecular and Genetic Basis of Neurological Disease*, 2nd edn. Butterworth-Heinemann, pp 877–912
2. Hoffman EP (2001) Dystrophinopathies. In: Karpati G, Hilton-Jones D, Griggs RC (eds) *Disorders of Voluntary Muscle*. Cambridge University Press, pp. 385–432
3. Hoffman EP (1998) Muscular Dystrophies. In: Jameson JL (ed) *Principles of Molecular Medicine*. Humana Press, pp 859–864

Duplication

Definition

Duplication means the transposition of a chromosomal region copy next to the original copy.

►[Chromosome 21, Disorders](#)

Dural Ectasia

Definition

Dural ectasia describes a widening or ballooning of the connective tissue sheath (dura) surrounding the spinal cord.

►[Marfan Syndrome](#)

Duty Cycle

Definition

Duty cycle is defined as the time in which a mass spectrometer is busy with a certain measurement before it can proceed to the next experiment.

►[Mass Spectrometry: Quantitation](#)

Dvl

►[Dishevelled](#)

Dynamic Instability

Definition

Dynamic instability describes a pattern of growth of microtubules or of F-actin. At one end of these polarized structures, abrupt stochastic transitions among phases of growing, shortening and pause occur.

►[Cytoskeleton](#)

Dynamic Mutation

Definition

The term was introduced to distinguish the unique properties of expanding, unstable DNA repeat sequences from other forms of mutations. Dynamic mutations are the molecular basis for a number of human genetic diseases and for all 'rare' chromosomal fragile sites. Common properties of simple tandem repeat DNA sequences in different diseases and fragile sites may allow insights into this unique form of DNA instability.

►[Repeat Expansion Diseases](#)

Dynamic Range

Definition

Dynamic range defines the range of signal intensities that can accurately be quantified; more technically this refers to the range of signal amplitudes over which an electronic communication channel can operate within acceptable limits of distortion. The range is determined by system noise at the lower end, and by the onset of overload at the upper end. In protein chemistry, this refers to the variation of the copy number of proteins present in a cell, reaching from a few copies to several hundred thousands of copies. The dynamic range also describes the concentration range between lower and

upper limits of detection, which can be used for quantitative analysis of pre-defined metabolites.

- Metabolomics
- Microarrays in Pancreatic Cancer
- Protein Microarrays as Tool for Protein Profiling and Identification of Novel Protein-Protein Interactions

Dysautonomia

Definition

Dysautonomia describes a loss of one or more of the functions of the autonomic nervous system and may include orthostatic hypotension, constipation, urinary hesitancy or frequency among others.

- Parkinson's Disease: Insights from Genetic Cause

Dyshormogenesis

- Thyroid Disorders, Genetic Basis

Dyslipidemia

Definition

Dyslipidemia (sometimes also referred to as “hyperlipidemia”) is a disorder of the lipoprotein metabolism, including lipoprotein overproduction or deficiency. Dyslipidemias may be manifested by elevation of the total cholesterol, the “bad” low-density lipoprotein (LDL) and the triglyceride concentrations, and a decrease in the “good” high-density lipoprotein (HDL) concentration in the blood. Dyslipidemia plays a role in the pathogenesis of diseases like diabetes and atherosclerosis.

- Familial Hypercholesterinemia

Dysmyelination

Definition

Dysmyelination describes a developmental defect of myelination. Oligodendrocytes or Schwann cells fail to myelinate axons completely or in part.

- Glial Cells and Myelination

Dysostoses

Definition

Dysostoses refer to disorders affecting single sites/bones of the embryo due to abnormal early development and patterning.

- Bone Disease and Skeletal Disorders, Genetics

Dystroglycan

Definition

Dystroglycan is a receptor that binds to multiple basement membrane proteins. Dystroglycan is composed of an extracellular, highly glycosylated α -chain, and a β -chain spanning the cell membrane. These two chains are encoded by a single gene. Protein sequence comparisons suggest that dystroglycan is a single unique molecule in that it is not member of a family. The α -chain and β -chain core elements are associated with several other proteins, forming a large dystroglycan complex. This receptor system found in muscle, epithelium and some other tissues, links the cytoskeleton to the extracellular matrix. Binding of extracellular ligands to dystroglycan activates signalling pathways, gene transcription and modifies cell behavior. Mutations in the gene encoding dystroglycan lead to severe developmental anomalies.

- Extracellular Matrix
- Gut Epithelium

Dystrophia Myotonica

- Myotonic Dystrophy Type 1

Dystrophin

Definition

Dystrophin is a large, rod-like cytoskeletal protein found at the inner surface of muscle fibers. It provides structural support to the cellular membrane in muscle

cells. Dystrophin is also found in cells of cardiac tissue, brain tissue, liver and other tissue types. Defects of the dystrophin gene cause progressive muscular dystrophy. Dystrophin is missing in Duchenne Muscular Dystrophy (DMD) patients or reduced in amount or changed in shape, respectively, in Becker Muscular Dystrophy (BMD) patients.

► [Duchenne Muscular Dystrophy](#)

Dystrophin Glycoprotein Complex

Definition

The dystrophin glycoprotein complex is a large oligomeric complex composed of ► [dystrophin](#) and other sarcolemmal (glyco)proteins. The complex has emerged as an important structural unit of muscle. It establishes a critical link between the sarcolemma (inside) and outside (extracellular matrix) of muscle fibers.

► [Limb Girdle Muscular Dystrophies](#)

E2F

Definition

E2Fs are the key transcription factors for cell cycle progression that promote expression of many cell growth-related genes, including DNA replication factors. E2F function is suppressed in non-cycling cells through binding to the Rb tumor suppressor protein. G1 cyclin/Cdks phosphorylate Rb protein, which leads to release of E2F and subsequent release of cell cycling.

► [DNA Replication Initiation](#)

E3–Ubiquitin Ligase

Definition

E3–ubiquitin ligases are enzymes that act together with an ubiquitin conjugating enzyme (E2) to couple the small protein ubiquitin to lysine residues on a target protein, marking that protein for destruction by the

- [proteasome.](#)
- [Adherens Junctions](#)
- [Limb Girdle Muscular Dystrophies](#)
- [Ubiquitination](#)

Early Pressure

Definition

Early pressure refers to the pressure on eggs activated with UV-treated sperm a few minutes after fertilisation. It destroys the meiotic spindle and leads to homozygous diploid embryos.

► [Mutagenesis Approaches in the Zebrafish](#)

EBD

► [Electron Beam Deposition](#)

EBI

Definition

The European Bioinformatics Institute (<http://www.ebi.ac.uk>) is part of EMBL.

► [Protein Databases](#)

E-Box

Definition

E-box is an enhancer element with the sequence CACGTG, which is found in the promoters or introns of many genes important to development, or in clock-regulated genes.

► [Circadian Clocks](#)

EC

Definition

EC (short form for **E**nzyme **C**ommission) describes a classification of enzymes according to their mechanism of action by the Enzyme Commission of the International Union of Biochemistry and Molecular Biology (IUBMB).

ECD

Definition

Electron capture dissociation (ECD) is a method of dissociation, specific to the FT-ICR mass spectrometry.

It is compatible with analysis of labile post-translational modifications (e.g. phosphorylation, glycosylation), and allows their characterization.

► [Proteomics in Cancer](#)

Ecdysone

Definition

Ecdysone is a steroid hormone in insects that induces molting (Gr.ek: out of + dyo: to clothe).

► [Nuclear Hormone Receptors: Regulators of Gene Transcription and Cellular Signaling](#)

ECG

► [Electrocardiogram \(ECG\)/Electrocardiography](#)

ECM

► [Extracellular Matrix](#)

► [Microarrays in Pancreatic Cancer](#)

Ectodomain Shedding

Definition

Ectodomain shedding indicates the release into the medium of extracellular parts of transmembrane molecules as a result of proteolysis.

► [Cancer Invasion and Metastasis: Cellular, Molecular and Clinical Aspects](#)

Ectopia Lentis

Definition

Ectopia lentis is a dislocation of the eye lens.

► [Marfan Syndrome](#)

EEG

Definition

An electroencephalogram (EEG) is a recording to detect abnormalities in the electrical activity of the brain.

► [Prader Willi and Angelman Syndromes](#)

EF Hand

Definition

This term refers to a configuration of E and F helices found in voltage gated calcium channels.

► [Autosomal Dominant, Polycystic Kidney Disease](#)

Effector

Definition

Effector is a ligand that modulates the function of a protein receptor by inducing structural changes in the protein.

► [Structure-Based Drug Design](#)

EGF

Definition

EGF stands for Epidermal Growth Factor, a growth factor that stimulates proliferation of a large number of cell types.

► [Epidermal Growth Factor Receptor](#)

► [Growth Factors](#)

EGFP

► [Enhanced Green Fluorescent Protein](#)

EGFR

► [Epidermal Growth Factor Receptor](#)

Egg

Definition

Eggs are ovulated female germ cells that are capable of being fertilized and developing into new individuals that exhibit characteristics of the species.

► [Mammalian Fertilization](#)

eIF-4F

Definition

eIF-4F (eukaryotic initiation factor 4F) refers to a trimer complex containing the cap-binding protein eIF-4E (eukaryotic initiation factor 4E), the RNA helicase eIF-4A, and the adaptor protein eIF-4G. It is responsible for translation initiation, and it probably also directs mRNA circularization by binding to poly(A)-binding protein (PABP).

► [RNA Stability](#)

► [Translational Control](#)

EJC

► [Exon Junction Complex](#)

Electrical Pacing

Definition

Electrical pacing is the stimulation of the heart rate to a faster pace using an electrical stimulator.

► [Drosophila Model of Cardiac Disease](#)

Electrocardiogram (ECG)/ Electrocardiography

Definition

Electrocardiography describes the graphic recording of electrical currents traversing the heart muscle

(electrocardiogram; ECG) and the study and interpretation of ECGs.

► [Heart](#)

► [Hyper- and Hypoparathyroidism](#)

Electromechanical Coupling

Definition

Electromechanical coupling refers to the molecular events that lead from electrical stimulation to contraction of a muscle cell.

► [Heart](#)

► [Muscle Contraction](#)

Electron Beam Deposition

Definition

EBD is an effect that is observed in scanning electron microscopy. An electron beam focused for a longer time at one position on a surface induces deposition of material originating from the gas phase of the not-perfect vacuum.

► [Atomic Force Microscopy](#)

Electron Crystallography

Definition

Electron crystallography is the primary method to determine the structure of 2D crystals. Images, as well as diffraction patterns, are recorded at low electron-doses and keep the sample at low temperature. Combining the information from many images and diffraction patterns taken at different projection angles allows the 3D structure of a crystal to be determined at atomic resolution.

► [Two-dimensional Crystallization of Membrane Proteins](#)

Electron Density

Definition

The spatial distribution of electronic charge in a molecule or in a quantum mechanical description,

the spatial distribution of the probability of transient electron positions.

► [X-Ray Crystallography—Basic Principles](#)

Electron Microscopy

► [EM](#)

Electron Multiplying CCD Camera

Definition

Charged coupled devices (► [CCD](#)) read noise increases proportional to the readout speed. This disadvantage is overcome in the electron multiplying charge coupled devices (EMCCDs). They use a special extended register, the so-called multiplication register, to provide a high gain before image readout. Electrons are accelerated from pixel to pixel in the multiplication register by applying higher-than-typical CCD clock voltages, thus generating secondary electrons via an impact-ionization process. The probability of electron generation is small (around 0.01). Although this number is small, the overall gain may be very high (above a factor of 1000) due to a large number of pixels in the multiplication register. EMCCDs amplify low light signals above the read noise floor of the CCD, which otherwise defines the detection limit for the system. This enables performance with effectively zero read noise at extremely high frame rates. Their low noise, high speed, high quantum efficiency and robustness to over-exposure make them ideally suited to single particle tracking applications.

► [Fluorescence Microscopy: Single Particle Tracking](#)

Electron Tomography

MANFRED AUER

Laboratory of Sensory Neuroscience, The Rockefeller University, New York, NY, USA
 auerm@mail.rockefeller.edu

Definition

► [Electron tomography](#) is a transmission electron microscope-based imaging technique, ideally suited

for the three-dimensional (3-D) structural determination of cellular volumes at molecular resolution. The 3-D structure of molecular machines in their natural surrounding of a cell can be revealed, even if the cellular complex is neither abundant, robust, homogeneous nor arranged in a regular array. A series of 2-D projection images of the tilted specimen is recorded, aligned and reconstructed into a 3-D volume, also known as a tomogram.

Although still an expert technique, setups for data collection and 3-D reconstruction are becoming widely available. Visualization of tomographic data is less straightforward due to its 3-D nature, low signal-to-noise ratio, ► [macromolecular crowding](#) and the complexity of cellular volumes. Novel 3-D data visualization and exploration, noise reduction, automated ► [segmentation](#) and ► [skeletonization](#) approaches have been developed, which in combination with interactive model building assist in interpretation of cellular tomograms.

Further developments including virtual reality approaches, real-time annotation and automated model structure fitting, together with biochemical ► [tagging](#) and ► [labeling](#) approaches are needed to achieve high-throughput of this powerful method in cellular proteomics. Automated high throughput electron tomography will enable the comparison between healthy and diseased tissue on the level of proteins and will lead to the understanding of molecular pathogenesis at a structural level.

Electron tomography in combination with other structural techniques will allow the 3-D visualization of the architecture of cells and tissues and their cellular machines at a molecular level.

Description

2-D Electron Microscopy and Sample Preparation for EM

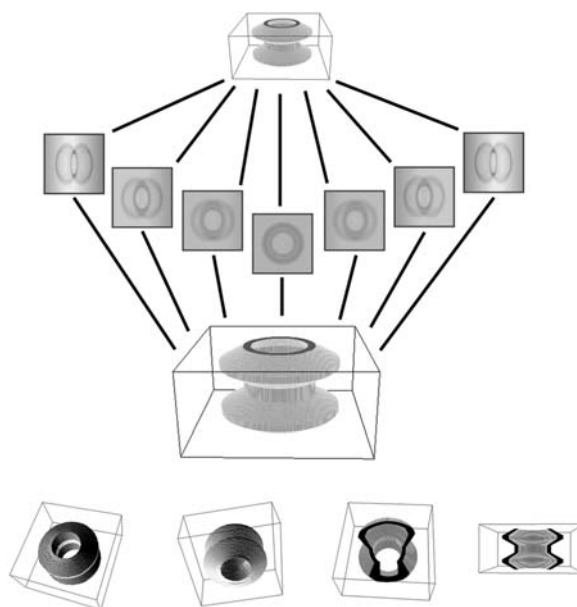
Microscopic imaging has been critical for the understanding of healthy and diseased cells and tissues. Fluorescence confocal light microscopy dominates biological imaging, due to its ability for real-time, non-invasive imaging of cellular events, using genetically encoded fluorescence reporter molecules. Electron microscopy (EM) reaches beyond the diffraction limit of light microscopy and has led to a detailed understanding of the ultrastructure of cells and tissues. EM requires thin sectioning of the specimen and imaging under high vacuum conditions, due to the strong interaction of accelerated electrons with matter. For this reason, tissue samples are embedded into a plastic resin, which is then polymerized to ensure successful sectioning using an ultramicrotome. The hydrophobic nature of most plastic resins requires an exchange of cellular water with an organic solvent prior to embedding, and hence the

cellular components are usually chemically fixed in order to avoid protein denaturation and extraction. In addition, heavy metal salts are added throughout the tissue processing procedure to obtain a better contrast upon EM imaging. Not only is this protocol tedious and lengthy, it is also vulnerable to sample preparation artifacts, such as denaturation, aggregation, component rearrangement and extraction. Considerable efforts have been made to minimize such sample preparation artifacts, including the introduction of instantaneous cryo-immobilization methods, such as high pressure freezing or slam-freezing, as well as low-temperature solvent substitution methods, such as freeze-substitution and progressive lowering of temperature dehydration. Moreover, a number of embedding resins with specific hydrophilicity profiles are commercially available and can be polymerized by heat, chemical cross-linkers or UV light. A truly artifact-free sample preparation is the fast immobilization (freezing) of the cellular components, sometimes referred to as cryo-fixation, followed by observation of the unstained, frozen-hydrated specimen using **►cryo-EM** imaging. While this method has been very successful for the tomographic imaging of very thin specimens, cells and tissues have to be sectioned in their frozen-hydrated state. Cryo-sectioning, once it becomes a routine technique, could revolutionize the art of sample preparation and avoid artifacts altogether.

Electron Tomography – Taking EM to the Third Dimension

Although sections for EM imaging are very thin, they still exceed the dimensions of individual molecules by at least one magnitude and since electron microscopes project a 3-D volume onto the 2-D plane of a recording device such as film or **►CCD**, the information about the third dimension is lost; all features along the electron path are superimposed. Among the cryo-EM 3-D reconstruction techniques used for the study of macromolecular complexes (1), electron tomography is the most general method, since it doesn't require the complex to be abundant, robust, homogeneous or arranged in a regular array. 3-D information can be obtained by recording projection images of the specimen from different angles; the images then need to be aligned and combined into a 3-D volume, e.g. by **►weighted back-projection** or algebraic reconstruction techniques (for a thorough introduction to all aspects of electron tomography see ref. 2). The principle of electron tomography is similar to that of **►CAT** scan medical imaging, however the microscope stays stationary and the EM specimen is rotated to discrete angles (Fig. 1).

Years of pioneering work both on the hardware as well as the software side have resulted in commercially available setups for semi-automated data collection.



Electron Tomography. Figure 1 Principle of electron tomographic imaging. Projection images of a 3-D object are recorded, aligned and then back-projected into a 3-D volume, which can then be inspected by suitable computer graphics. The 3D volume can be rendered, rotated and analyzed to reveal the architecture.

The stability and accuracy of electron microscope stages, as well as the sensitivity of detection devices such as CCD have been significantly improved, along with software for low-dose imaging and automated data acquisition, minimizing the electron dose that is spent on tracking and focusing of the image (3). Due to the radiation sensitivity of biological material, automation has played a crucial role because the dose has to be kept to a minimum. The tilt range is usually restricted to $\pm 75^\circ$ due to the thickness of cell slices at high tilt angles and the limitations introduced by the specimen support (EM grid), resulting in an incomplete sampling of the projection space (missing wedge). Whole cells and tissues could, in principle, be reconstructed by combining tomograms of neighboring areas or serial tissue sections, but the size of the resulting 3-D data set would be gigantic.

Visualization of 3-D Data

While data collection and 3-D reconstruction are becoming routine tools, the further development of visualization tools is crucial for the study of increasingly larger and more complex 3-D volumes. Tomographic 3-D volume can be inspected slice by slice, especially when supported by non-orthogonal slicing and slice averaging methods. However, the 3-D nature of the tomographic data is better appreciated by

►**rendering** methods, such as volume or surface rendering.

Volume rendering retains the volume data and sums it from back to front with appropriate weights (ray casting). The information of each volume element (voxel) is incorporated, resulting in a transparent, ghost-like impression of the electron density.

Surface rendering creates the illusion of depth by the removal of hidden surfaces, as well as the use of perspective, lighting options, shading, color and texturing effects. Such 3-D scenes are more familiar to our brain and therefore often preferred for display of the 3-D data. However, surface rendering shows the object only at a certain chosen density threshold and is often not very helpful for highly complicated objects with internal cavities. For a photorealistic impression, ray tracing is the method of choice, but it is computationally very expensive, since it follows the paths of each individual light ray as it interacts with the surface of the object. Hence, ray tracing cannot be achieved in real-time and is therefore not suited for data inspection, but for data presentation and publication.

Stereo display techniques as well as a looped video animation of the rendered object, particularly when rotated around an axis, can substantially ease the understanding of structural detail.

Physical models, although out of fashion, are a powerful alternative to computer graphics display and even complicated models can – in principle – be produced by stereolithography.

Noise Reduction

The poor signal-to-noise ratio usually observed in tomograms complicates automated feature extraction, as well as the visualization of the volume. Since averaging techniques cannot be applied to unique cellular 3-D volumes, other methods for signal enhancement such as median ►**filtering**, wavelet filtering, bilateral filtering and nonlinear anisotropic diffusion filtering are often used.

Data Exploration

The study of cellular volumes resembles the exploration of uncharted territory and, while the individual protein components may be familiar, there is often no preconception of the architecture of cellular machines. Real-time rendering of whole tomograms is therefore key for interactive data exploration and has recently been implemented using multi-resolution interactive geometry. Further developments towards a cellular “flight simulator” are needed, which would allow 3-D navigation through the densities, zooming up and down in scale from individual proteins to the whole cell,

as well as interactive, semi-automated fitting of candidate protein structures to the density envelopes and validation of the fit. Such a virtual-reality-based program would be of enormous value in the analysis of tomographic data.

Data Simplification: Interpretation of Tomograms Through Segmentation, Skeletonization and Model Building

Tomograms of cellular volumes contain thousands of individual proteins, all densely packed and mostly organized into cellular machines. This macromolecular crowding complicates the analysis and therefore the complexity of the scenery needs to be reduced to allow biological interpretation. Segmentation methods can electronically extract multi-protein complexes from their cellular surrounding, allowing an unobstructed view into the machinery’s architectural organization.

While manual segmentation is still widely used, automated segmentation methods are desirable due to the labor-intensive and somewhat subjective nature of manual segmentation. Robust automated segmentation algorithms based on density thresholds (watershed immersion (4)) and density gradients (boundary segmentation, see ref. 5) have been proposed and allow fast and more objective segmentation of cellular volumes.

Skeletonization approaches (5) have been developed in order to simplify a complex 3-D density pattern into a stick model, as a cartoon is a simplified version of the real world. Skeletonization also allows quantitative analysis of the object including distances and angles.

Model building has long been the key in interpreting biological structures and has led to a level of insight that was not available from the obtained density distribution alone. If the exact molecular makeup of a complex is known and the resolution is sufficient, protein structures can be fitted into the density maps, either manually using interactive 3-D graphics or semi-automatically. Another approach, ►**template matching** has been proposed for data exploration and analysis. It seems particularly well suited for segmented volumes, whereas it may be problematic in cases where the features are not sufficiently separated from one another or not distinctly shaped.

Molecular Component Identification: the Quest for a ►GFP for Electron Microscopy

Due to the complexity of a cell and its unknown exact molecular composition, the localization of an individual protein molecule within the tomographic volume is not a straightforward task. In other words, cellular volumes do not come annotated. In some cases, the cellular location and association with easily identifiable features, such as actin filaments, as well as the size and shape of the densities may give clues to the identity of a

protein, but even a good fit of a likely candidate protein to the density is no proof.

What is needed is a widely applicable reporter system that recognizes the protein of interest, ideally through a genetically encoded tag. This reporter system must provide sufficient contrast to be unambiguously detected in the electron microscope, be small enough to ensure access to the protein, be stable throughout the sample preparation, show a high affinity, yet be specific for its target to avoid false-negative and false-positive signals, respectively.

The recent development of stable and small, but recognizable immuno-gold-cluster labels is certainly a step in the right direction. However, for each epitope, a suitable antibody must be found, which can be a rather tedious endeavor and may sometimes fail altogether.

In contrast, an ideal method would be based on a genetically encoded tag that is recognized by an electron-dense label. The labeling protocol would only have to be worked out once and would then be applicable to all systems, provided the tag-modified protein is expressed and functional. In the ►[FlAsH/ReAsH](#) system a small genetically encoded tag is recognized by a non-toxic membrane permeable dye, resulting in a steep increase in fluorescence. The use of the fluorescence photooxidation technique allows ►[correlative microscopy](#), meaning that a system can be first studied by confocal fluorescence microscopy and subsequently processed for EM (6). The only but important downside of this approach is that the stain is not localized to one point but is rather diffuse. For the exact 3-D localization of a candidate protein, an electron-dense point label would be desirable. The commercially available Ni-NTA-Nanogold may become such an ideal labeling system. It is designed to recognize hexa-histidine tags genetically engineered onto the protein of interest. Once in place, such a system based on tags and labels would allow the exact 3-D-localization of proteins within a cell and therefore become a central part of cellular proteomics approaches.

References

1. Auer M (2000) Three-dimensional electron cryo-microscopy as a powerful structural tool in molecular medicine. *J Mol Med* 78:191–202
2. Frank J (1992) *Electron tomography*, Plenum Press, New York
3. Baumeister W, Grimm R, Walz J (1999) Electron tomography of molecules and cells. *Trends Cell Biol* 9:81–85
4. Volkman N (2002) A novel three-dimensional variant of the watershed transform for segmentation of electron density maps. *J Struct Biol* 138:123–129
5. Bajaj C, Yu Z, Auer M (2003) Volumetric feature extraction and visualization of tomographic molecular imaging. *J Struct Biol* 144:132–143

6. Gaietta G, Deerinck TJ, Adams SR et al (2002) Multicolor and electron microscopic imaging of connexin trafficking. *Science* 296:503–507

Electronic Transition Energy

Definition

Electronic transition energy is defined as the energy difference between electronic wavefunction energy levels. For MAD applications, the relevant transitions are between electrons in the K- or L-shells and the free electron state.

►[3D Structure Determination by X-Ray](#)

Electrophoresis

Definition

Electrophoresis denotes the movement of molecules in the presence of an electric field. The velocity of the molecules depends on the mass and charge of the molecules.

►[2D-PAGE Database](#)

►[Proteomics in Microfluidic Systems](#)

►[Two-Dimensional Gel Electrophoresis](#)

Electroporation

Definition

Electroporation is a high efficiency transformation method, where DNA or other large molecules are introduced into a cell by means of a brief high voltage pulse (voltage differential across the cell), which renders the plasma membrane of cells temporarily permeable, and allows genetic material (e.g. DNA) to enter.

►[Large-Scale Homologous Recombination Approaches in Mice](#)

►[Morpholino Oligonucleotides, Functional Genomics by Gene ‘Knock-down’](#)

►[Mouse Genomics](#)

►[Neural Crest Cells and their Derivatives](#)

►[Recombinant Protein Expression in Yeast](#)

►[YAC and PAC Maps](#)

Electrospray Ionization Mass Spectrometry

Definition

Electrospray ionization (ESI) is the most commonly used method for “soft” ionization of peptides and proteins in quadrupole, ion-trap, or time-of-flight mass spectrometers. The ionization is performed by application of a high voltage to a stream of liquid emitted from a capillary. The highly charged droplets are shrunk, and the resulting peptide or protein ions are sampled and separated by the mass spectrometer.

- Mass Spectrometry: ESI
- Mass Spectrometry: MS/MS
- Proteomics in Microfluidic Systems
- SNP Detection and Mass Spectrometry

ELISA

Definition

ELISA defines a method that uses an antibody to detect an antigenic molecule in a sample. In a typical ELISA, plastic wells with sticky walls are used to immobilise molecules of a biological sample. A specific antibody is covalently linked to an enzyme and is added to the wells, where it is retained on the walls if immobilised antigenic molecules are present. This is detected by a color-change reaction catalysed by the linked enzyme (Linked Immuno-Sorbent Assay), converting a colourless substrate into a coloured reaction product.

- Affinity Chromatography and *In Vitro* Binding (Beads)
- Monoclonal Antibodies

EM

Definition

- Cryo-Electron Microscopy: Single Particle Reconstruction
- Electron Tomography

EMBL

Definition

The European Molecular Biology Laboratory (<http://www.embl-heidelberg.de>), located in Heidelberg, Germany.

- Protein Databases

Embryo Transfer

Definition

Embryo transfer is the insertion of an embryo into the uterus of a surrogate or biological mother, usually after *in vitro* fertilization.

- SRY – Sex Reversal

Embryoid Body

Definition

Embryoid body refers to a three-dimensional spheroid structure that is formed by the aggregation of undifferentiated embryonic stem cells *in vitro*, which subsequently differentiate into a wide variety of specialized cell types.

- ES Cell Differentiation as a Model System for Functional Genomics

Embryonic Lethality

Definition

Embryonic lethality is defined when the embryo dies during its development in the womb and is not born alive.

- Mouse Genomics

Embryonic Stem Cell

Definition

Embryonic stem (ES) cells are derived from the inner cell mass of mammalian embryos, usually mouse embryos. An ES cell self-renews and has the potential to develop into all or nearly all of the tissues in a mammal, maintaining its ability to contribute to the germline of an organism. ES-cells can be indefinitely maintained in culture. When introduced into another embryo, ES-cells can contribute to the host tissues. ES-cells that efficiently contribute to the germline of the host have so far only been established in mice.

- ▶ [Cell Division](#)
- ▶ [Cre/Lox P Strategies](#)
- ▶ [ES Cell Differentiation as a Model System for Functional Genomics](#)
- ▶ [Mouse Genomics](#)
- ▶ [Stem Cells - Overview](#)
- ▶ [Transgenic and Knockout Animals](#)

EMCCD

- ▶ [Electron Multiplying CCD Camera](#)

Emission Spectrum

Definition

The emission spectrum is a plot of the emitted light wavelength versus the intensity of emitted light.

- ▶ [Fluorescence Microscopy: Spectral Imaging vs. Filter-Based Imaging](#)

EMT

- ▶ [Epithelial-to-Mesenchymal Transition](#)

5' End/3' End

Definition

A single stranded molecule of DNA will have an orientation determined by its sugar-phosphate backbone. The links between the nucleotides in the molecule are through the 5' and 3' carbons in the ribose moiety. In a double stranded DNA molecule, the strands are antiparallel, so that at a break in the molecule, the 5' end of one strand will be the 3' end of the complementary strand. Nucleases (including restriction endonucleases) cleave DNA leaving a 5' phosphate group and a 3' hydroxyl group.

- ▶ [Shotgun Libraries](#)

5' End Processing

- ▶ [mRNA Capping](#)

Endocarditis

Definition

Endocarditis is an inflammation of the lining membrane of the heart.

- ▶ [Marfan Syndrome](#)

Endochondral

Definition

Endochondral refers to growth and formation of new bone via a temporary cartilaginous bone formation template in the growth plate.

- ▶ [Bone Disease and Skeletal Disorders, Genetics](#)

Endocrine

Definition

Endocrine describes the secretion of a factor (hormone, growth factor) from a cell into the blood stream, which affects cells in distant organs or tissues.

- ▶ [Growth Factors](#)

Endocrine Cells

Definition

Endocrine cells are cells in glandular organs that secrete factors into the blood system. Examples are cells of the pituitary or thyroid gland. In contrast, ► **exocrine cells** secrete their factors into ducts ultimately leading to the exterior of the body.

► **Microarrays in Pancreatic Cancer**

Endocytic Pathway

► **Endocytosis/Endocytic Pathway**

Endocytosis

Definition

Endocytic pathway/endocytosis comprises of a series of trafficking steps through which materials are internalized into eukaryotic cells, without passing through the plasma membrane, and are then transported to endosomes, ► **endosomal compartment** and/or ► **lysosomes**. The plasma membrane folds around material outside the cell, resulting in the formation of a sac-like vesicle into which the material is incorporated, often in a highly concentrated form. The best characterized mechanism of such internalization involves the coat protein clathrin in conjunction with the adaptor complex AP-2.

► **Cell Polarity**

► **Exocytosis**

► **Functional Assays**

► **Microvilli**

► **Morpholino Oligonucleotides, Functional Genomics by Gene ‘Knock-down’**

► **Rho, Rac, Cdc42**

► **Vesicular Traffic**

Endoderm

Definition

Endoderm is the innermost of the three primary germ layers of an embryo, developing into the gastrointestinal tract, the lungs, and associated structures.

► **Lung**

Endogenous Proviruses

Definition

Endogenous proviruses are proviruses that result from integration of retrovirus DNA into the genome of a germ line cell, and subsequently can be inherited by subsequent generations as though it were a normal component of the host genome. The retrovirus gene env encodes the envelope glycoprotein consisting of ► **SU** and ► **TM** that mediates interaction of the virus with and entry into the host cell.

► **Retroviruses**

Endonuclease

Definition

Endonuclease is an enzyme that cleaves its nucleic acid substrate at internal positions in the nucleotide chain. Ribonucleases are specific for RNA, deoxyribonucleases for DNA. Bacterial restriction endonucleases are crucial in recombinant DNA technology, for their ability to cleave double-stranded DNA at highly specific sites.

► **mRNA Stability**

► **Transposons**

Endoplasmic Reticulum

Definition

Endoplasmic reticulum (ER) comprises an extensive network of interconnected membrane tubules within the cytoplasm of eukaryotic cells, which is a major site for intracellular protein synthesis. Proteins synthesized on membrane-bound ribosomes (“rough ER”) are inserted into the ER lumen, where they undergo assisted folding and may be subject to posttranslational modifications (e.g. proteolytic processing and core glycosylation) and oligomerization. An ER quality control mechanism ensures that only properly folded proteins leave the ER to the Golgi complex in COP II vesicles.

► **Diabetes Insipidus, a Water Homeostasis Disease**

► **Glycosylation of Proteins**

► **Vesicular Traffic**

Endosomal Compartment

Definition

The endosomal compartment is a complex set of vesicles and tubules, of various sizes, extending from the cell periphery (early endosomes) to the peri-nuclear region (late endosomes). In the endocytic process, the endosomal compartment is the place where the ligand is dissociated from the receptor, and decisions are made about whether to destine the receptor for degradation or to recycle it to the cell surface; ► [endocytosis](#)

► [Vesicular Traffic](#)

Endosome

Definition

Endosome is a subcellular vesicle that is involved in the uptake into the cell of membrane and extracellular material.

► [Endosomal Compartment](#)

► [Morpholino Oligonucleotides, Functional Genomics by Gene ‘Knock-down’](#)

Endosteum

Definition

Endosteum is the lining of the cortical bone surface facing the bone marrow. The endosteum also lines the trabeculae. It consists of osteoprogenitor cells.

► [Bone and Cartilage](#)

Endothelial Cells

Definition

Endothelial cells are the major cell type forming the single-layered inside lining of blood vessels, lymphatic vessels, and the heart, the endothelium. During wound healing endothelial cells are stimulated by various cytokines to migrate and proliferate and to build new capillary sprouts.

► [Angiogenesis](#)

► [Kidney](#)

► [Morbus Wegener](#)

► [Wound Healing](#)

Endothelium

► [Endothelial Cells/Endothelium](#)

Enhanced Green Fluorescent Protein

Definition

Enhanced green fluorescent protein (EGFP) refers to a recombinant fluorescent protein that is widely used for transfection studies. The protein can be localized in transfected cells without the need for staining or addition of co-factors. It is often used to determine the efficiency of gene transfer.

► [GFP \(Green Fluorescent Protein\)](#)

► [RNA Interference in Mammalian Cells](#)

Enhanceosome

Definition

Complex of an enhancer subsegment of approx. 60 bp with several DNA binding transcription factors. The enhanceosome is a functional entity of transcriptional activation, due to cooperative factor binding, assisted by “architectural” proteins that bend DNA.

► [Enhancer](#)

Enhancer

WALTER SCHAFFNER

Institute of Molecular Biology, Zurich University,
Zurich, Switzerland

walter.schaffner@molbio.unizh.ch

Synonyms

Transcription enhancer, transcriptional enhancer

Definition

An enhancer (transcription enhancer, transcriptional enhancer) is a regulatory DNA segment of about 200 base pairs that is typically found in multicellular eukaryotes. It can strongly stimulate (“enhance”) the transcription of a linked ► [transcription unit](#), i.e. it acts *in cis*. An enhancer can activate transcription over very long distances of many thousand base pairs, and from a position upstream or downstream of the site of

transcription initiation. Enhancers have a modular structure by being composed of multiple binding sites for transcriptional activator proteins. Many enhancers control gene expression in a cell type-specific fashion. Expression of one and the same gene can be controlled by several remote enhancers. Conversely, a given enhancer may also serve to activate more than one gene.

Characteristics

Viral Enhancers as Harbingers of Cellular Enhancers

The first enhancer was described in 1981 in the genome of a small DNA virus, simian virus 40 (SV40) as a DNA segment that was not only essential for viral early gene expression but in experimental conditions was able to stimulate the expression of a heterologous gene linked to it more than one hundred fold (1, 2). Naturally this enhancer is located between 100 bp and 300 bp upstream of the transcription initiation site for the viral “early” messenger RNA. In experimental situations it was able to stimulate the expression of the reporter gene over long distances of several thousand base pairs, both from positions upstream or downstream of the site of transcription initiation. Other enhancers were discovered soon thereafter in both DNA viruses and retroviruses.

Already at this early stage it was proposed that enhancers were not a viral peculiarity but that similar enhancers would also be associated with cellular genes (1). The discovery of an enhancer associated with immunoglobulin (Ig) heavy chain genes marked another breakthrough in the understanding of eukaryotic gene regulation, mostly for two reasons. First of all, it was active only in B lymphocyte-type cells and thus represented the first genetic element of eukaryotes to control cell type-specific gene activity. Furthermore, this enhancer is located in the second intron, i.e. within the transcription unit, and thus demonstrated that enhancers can indeed be located downstream of the transcription initiation site in a natural setting (3, 4) (Fig. 1 a, b, c, d). Over the following years, many other cellular enhancers were discovered, some close to the ►promoter/transcription initiation site, others tens, if not hundreds of kilobases away from the gene to be activated (strictly speaking, the term “gene” should be replaced here by “transcription unit” because even a very remote enhancer is, by definition, an integral part of the gene). Enhancers are widespread in nature; in all multicellular eukaryotes from worms to mammals, control of gene activity *via* enhancers seems to be the rule rather than the exception. Most enhancer studies so far have involved genes transcribed by RNA polymerase II, but enhancers have also been described for ribosomal RNA genes, which are transcribed by RNA polymerase I.

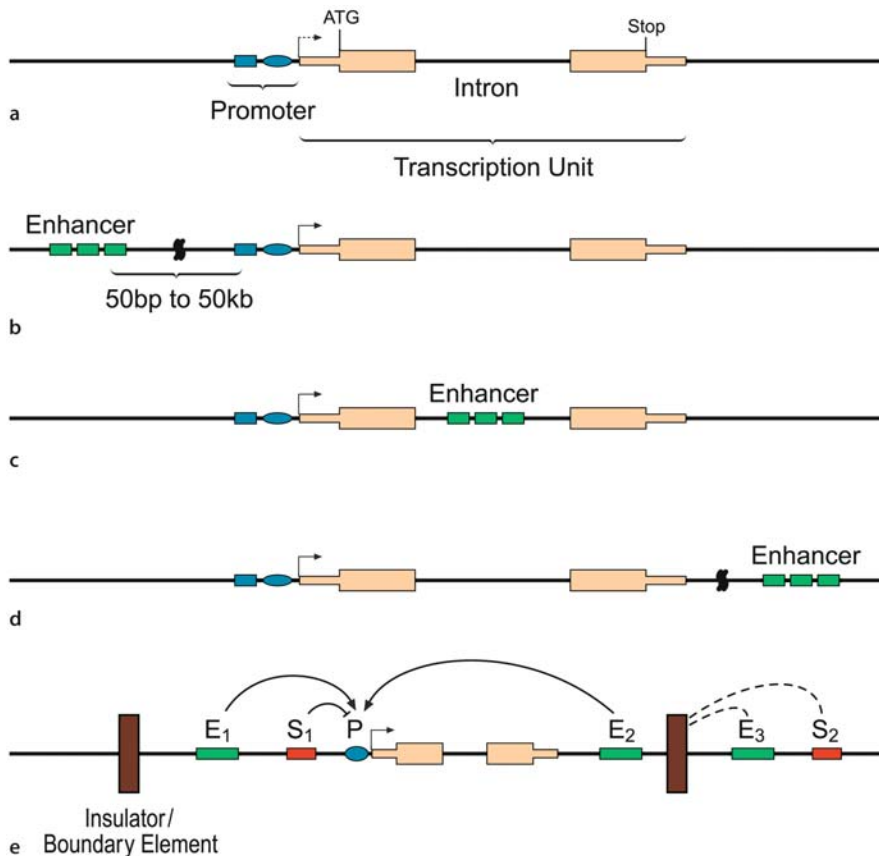
For stage- and cell type-specific gene expression, one and the same gene may rely on multiple enhancers. The immunoglobulin heavy chain genes are no exception; besides the originally discovered enhancer in the second intron, this locus contains another B-lymphocyte-specific enhancer far downstream of all IgH transcription units, as well as additional cytokine-inducible enhancers which play an important role in the so-called heavy chain “class switch” during an immune response. The most extensive studies on this topic have been carried out in the fruit fly *Drosophila*. There, genes for specific developmental regulators are known to be controlled by multiple enhancers in most cases. For example, the seemingly simple stripe pattern of the fly embryos is not controlled by one enhancer, but multiple enhancers are needed to form but one of the stripes. One of the key regulators of *Drosophila* development are ►transcription factors of the Pax family. The most extensive enhancer study so far was done with the Pox neuro gene that controls multiple steps in development (5) (Fig. 2).

In baker's yeast (*Saccharomyces cerevisiae*), a unicellular eukaryote of simpler organization than most other eukaryotes, regulatory sequences that are functionally equivalent to enhancers of higher eukaryotes have been discovered. These are called “upstream activating sequences” (UASs) and also serve as binding sites for transcriptional activators. The UASs also activate transcription of a linked transcription unit independent of their orientation and over distances of up to approx. 1000 bp (6). However, UAS motifs are usually in close proximity to the initiation site of transcription and do not activate from a position downstream of the transcription unit. Even in prokaryotes, long-range effects of regulatory DNA elements were found (7), but unlike the situation in higher eukaryotes, enhancer effects appear to be the exception rather than the rule, in prokaryotes.

Enhancers Have a Modular Structure

A typical enhancer is about 100 bp to 300 bp long (or longer) and consists essentially of a string of binding sites for DNA-binding activator proteins or transcription factors. Each cell of a higher organism harbors hundreds of different such transcription factors. A transcription factor contains a DNA-binding domain to recognize a short DNA-sequence motif of approx. 10 bp and an “activation domain”. The latter establishes contacts with other components of the transcription apparatus and consequently raises the rate of transcription of the corresponding gene. At least some activation domains also facilitate, most probably indirectly, the process of transcription elongation by ►RNA polymerase II.

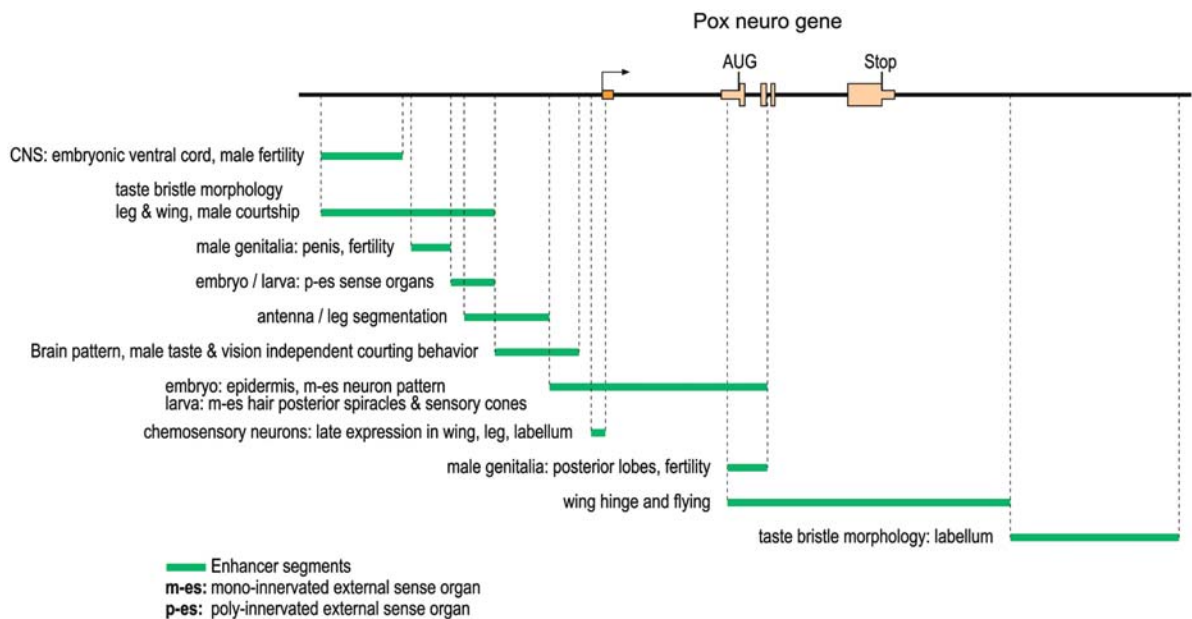
Unexpectedly, and despite the eukaryote's larger genomes compared to the ones of prokaryotes, the



Enhancer. Figure 1 Possible positions of enhancers in eukaryotic genes. (a) A gene without enhancer that would merely contain a promoter consisting of one proximal binding site (green box) for a transcription activator protein, a TATA box and an Inr (initiation region) would be poorly, if at all, transcribed (dotted arrow). (b–d) An enhancer (c) green boxes schematically indicate the multiple binding sites for activating transcription factors) can be located at different positions: near the promoter region, or overlapping with the latter, as is the case in different viral genes, as well as metallothionein genes (b); far upstream of the transcription unit, as in the albumin gene; d), within a gene, usually as part of an intron, such as in immunoglobulin (Ig) genes (c); or even far downstream of the transcription unit, as, e.g., in T-cell receptor and Ig genes (d). (e) Enhancers, promoters, silencers and boundary elements control gene expression. Similarly to enhancers (simplified as a single green box), silencers (red box) can influence gene activity over long distances. They also consist of an array of binding sites, but for transcription repressor proteins. A gene flanked by so-called boundary elements (= insulators) is protected from undesired effects of neighboring enhancers or silencers. In this scheme, one might envision a situation where two enhancers (E1 and E2) activate transcription from the promoter (P, in blue) in two different cell types. In a third cell type, the gene is completely repressed by the silencer S1. Neither the nearby enhancer (E3) nor the silencer (S2) can exert any effect because the gene is bracketed by a pair of insulator/boundary elements (brown).

selectivity of DNA recognition by eukaryotic transcription factors is often less pronounced than that of bacterial regulators. This deficiency in specificity, however, is compensated by a greater number of binding sites and factor interactions; while bacterial, and even some yeast genes are regulated by a single activator, most eukaryotic enhancers have to interact with multiple factors to become active. By serving as a platform for the binding of specific transcription factors, enhancers thus ensure correct development in all multicellular organisms (5).

Although a typical enhancer contains binding sites for several different transcription factors (8), synthetic enhancers can, in accordance with the concept of a modular structure, readily be generated by linking together multiple copies of one and the same DNA sequence motif. From this point of view, every binding site for a transcription protein (or a protein dimer) can be considered a minimal enhancer unit, even though on its own it would have low, or even no activity (9). Multiple binding sites however show a strong synergy in transcriptional activation. At least in some cases, the



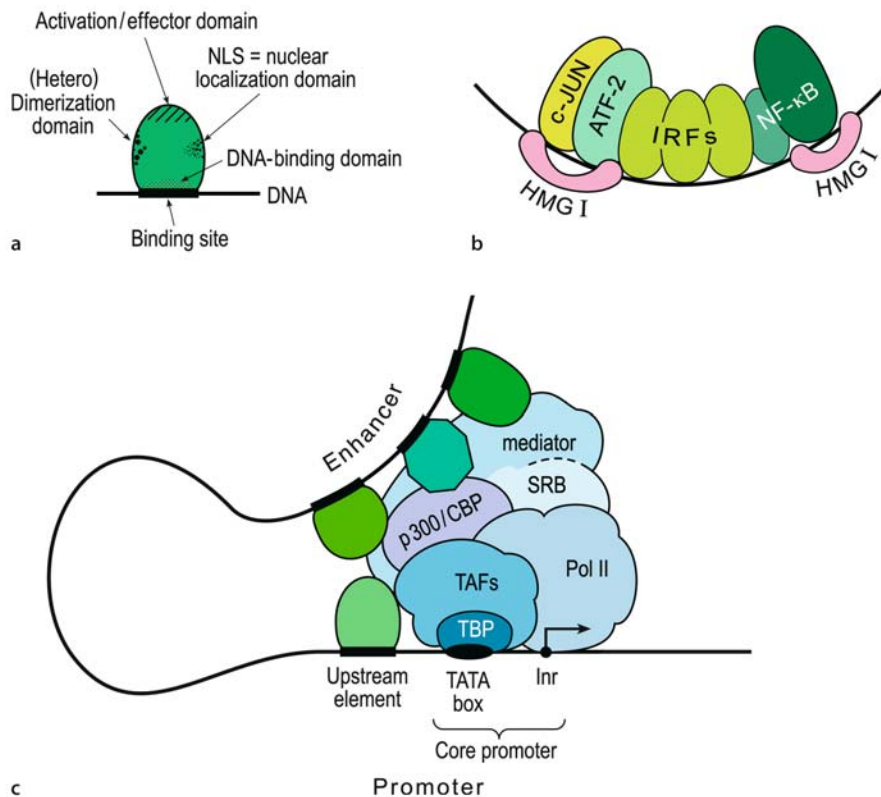
Enhancer. Figure 2 Many enhancers control cell fate in embryonic development and cell differentiation. Genes in control of embryonic development tend to be regulated by particularly complex patterns of enhancers, silencers and boundary elements. Shown here is the multitude of enhancers regulating the expression of a paired-box transcription factor-gene, *pox neuro* in the fruit fly *Drosophila*. Various enhancers are required for ensuring gene activity at a multitude of specific stages, in different tissues/cell types (5).

minimal functional unit consists of a larger DNA segment of some 60 bp, on which multiple transcription factors stereospecifically interact with each other and the DNA, with the assistance of structural (“architectural”) proteins, which suitably bend the DNA. Such a superstructure has been termed an **enhanceosome** (10) (Fig. 3).

In bacteria, RNA polymerase directly recognizes the promoter, and this interaction may be facilitated by the nearby binding of an activator protein. In contrast, in eukaryotes not even the DNA-bound transcription factors are able to directly bind and thereby recruit RNA polymerase II for transcription initiation. Rather, they only bind indirectly *via* mediator proteins (adaptor proteins) that are part of large multiprotein complexes. Because of the multiple layers of interactions and because of their sheer size, such protein complexes are relatively insensitive to the exact position of their DNA-binding partners. This explains why a whole enhancer, or parts of it, and in many cases even single binding sites for transcription factors, can be inverted relative to other DNA-sequences in an enhancer or a promoter without loss of activity. This remarkable orientation independence is a general attribute of eukaryotic regulatory DNA sequences and was first documented for the enhancer/promoter region of a histone gene in the sea urchin, a DNA-segment of –100 to –300 from the transcription start dubbed the “modulator” (11).

Mechanism of Action: Certainly Recruitment – but How?

Unlike bacterial RNA polymerases, eukaryotic RNA polymerase II is unable to recognize the naked DNA of a promoter; rather, the promoter has to be “tagged” by DNA binding proteins. However, there are additional layers of complexity in that not even such DNA-binding proteins are able to recruit the polymerase to the promoter; they in turn bind to a multitude of intermediary proteins, referred to as co-activators, mediator complex, adaptors or SRB proteins, which are organized in large multiprotein complexes, and on their part recruit the RNA polymerase II to the initiation site (12) (Fig. 3). For example, many genes depend on the closely related, large adaptor proteins CBP and p300. These proteins were originally identified as binding partners of the DNA binding transcription factor **CREB**, but this turned out to be but one of many interactions with transcription factors. DNA-binding transcriptional activators may also interact with the general transcription factors (GTFs) that include the promoter-proximal TFIIB, the TATA-box-binding protein (TBP) and/or so-called TAF-proteins, which are complexed to the latter. The question as to how transcription from a given promoter can be dramatically induced by a remote enhancer has been a source of speculation ever since the discovery of the enhancer effect. Several models have been proposed to explain



Enhancer. Figure 3 Enhancer-binding transcription factors: recruitment and DNA looping. (a) A typical DNA-binding transcription factor (= transcriptional activator = transactivator) contains a DNA binding site to dock onto the enhancer or upstream promoter site, a so-called **NLS** (nuclear localization sequence) for transport to the nucleus, an activation/effector domain which interacts with the co-activator/mediator proteins and often an additional domain for homo- or hetero-dimerisation. (b) Superstructures composed of a stereospecific array of DNA sites and transcription factors, aided by “architectural protein” HMG I(Y) which bends the DNA to facilitate assembly of this complex, are found in the enhancer/upstream region of interferon beta genes. Such superstructures act as functional entities and have been referred to as enhanceosomes (10). (c) The transcription factors bound to both enhancer and promoter, including the generally used TATA-box binding protein (TBP) and TFIIB, recruit further general transcription factors, as well as adaptor and mediator proteins via multiple protein-protein-contacts. The mediator complex helps to recruit RNA polymerase II to the initiation site, resulting in the formation of the extremely large initiation complex. Besides RNA polymerase, several other enzymatic activities are associated with this initiation complex, including protein kinases and acetyl transferases. The latter primarily acetylate histones, resulting in a loose chromatin structure permissive for transcription.

the mode of activation, three of which were most often considered, namely, (i) today’s most popular “looping” model, (ii) the “linking” (or “oozing”) model and (iii) the “polymerase scanning” model. Regarding the looping model, several experiments support the concept of DNA-looping between enhancer and promoter (13, 14). Most probably, the enhancer-bound transcription factors bind indirectly, *via* a multitude of cofactors, to the ones bound in the promoter region and thereby form a large initiation complex, with the intervening DNA being looped out (Fig. 3). Under experimental conditions in a cell-free system, a promoter on its own can establish such an initiation complex, but *in vivo*, on a native chromosome, it depends on the assistance of

the enhancer. The linking model (also known as “oozing”) has recently gained momentum. The linking model proposes that the binding of certain activating proteins to the enhancer regions causes a modified chromatin structure, which thereafter spreads along the DNA in either direction until ultimately a transcriptionally competent promoter is encountered and included in the “open” configuration (15). A combination of the two models can also be envisaged, whereby DNA looping between enhancer and promoter is not permanent but rather a transient on-off phenomenon. Even though in some cases RNA polymerase II is also found associated with an upstream enhancer, where it may or may not initiate a low level of transcription,

there is hardly any evidence for a sliding, or scanning by the polymerase from there until it reaches the “real” promoter.

DNA-bound factors, the mediator proteins and the transcription apparatus synergize in the recruitment of chromatin “remodeling” complexes such as SWI-SNF. The concept of local or long-range modification of chromatin structure, starting at an enhancer and extending until the promoter is reached, is also supported by the findings that several transcription-associated proteins contain, among other enzymatic activities, ►**histone acetyltransferase** (HAT) activity. This serves to acetylate specific residues on histones and transcription factors. Acetylation of histones generally loosens chromatin structure and makes the DNA more accessible to the transcription apparatus. An obvious indication of this is the increased sensitivity of such chromatin on treatment with DNase I. Transcribed “active” DNA generally is DNase I sensitive, so this condition correlates with DNA being packaged into acetylated histones. On top of this general sensitivity, enhancer and promoter regions are, as a consequence of the binding of transcription factors, hypersensitive to DNase.

Enhancer Shuffling, Enhancer Rearrangements

In higher eukaryotes, genes which control embryonic development and other traits are usually controlled by multiple enhancers. It has been argued that DNA rearrangements leading to the displacement of enhancers within the genome (“enhancer shuffling”) are a major driving force in evolution to generate new expression patterns of genes. Besides such “shuffling”, enhancers themselves can undergo alterations, as is most dramatically shown in the evolution of viruses. Selection for a new cell type specificity was shown to be accompanied by extensive alterations in the enhancer DNA sequence and relatively small changes in protein-coding sequences. Gross deletions/duplications of enhancer subsegments are most commonly observed (16 and references therein). Duplications of enhancer regions, which indicate relatively recent changes in gene expression, have also been found in some cellular enhancers and may contribute, for example, to variation in disease resistance. While an enhancer translocated to a new position in the genome can result in a new gene expression pattern, not every arbitrarily chosen enhancer-promoter combination is functional, since promoter and enhancer can co-evolve for optimal function. There are well-documented cases of exclusive enhancer/promoter-specificity (17). On the other hand, novel enhancer-promoter combinations that interact productively have been found from yeast to mammals and may be the rule, rather than the exception. In a similar way, displacement of an enhancer in somatic cells by chromosome translocation

can result in over-expression of cell growth-promoting genes and thus contribute to malignant cell transformation. The same dire consequences can result from the insertion of a retrovirus with a strong transcription enhancer near a regulatory gene. Due to the long-range activity of enhancers, the insertion site might well be located dozens, if not hundreds of kilobases away from the affected cellular gene.

Enhancers in the Context of Eukaryotic Gene Regulation

Thanks to a combination of genetics, specific mutagenesis and biochemistry, many aspects of eukaryotic transcription control have been elucidated – but a number of questions remain. One of these concerns the mode of activation; does an enhancer affect a gene in an all-or-none fashion whereas the fine-tuning of transcription is left to the promoter, or do enhancers of different strength confer gradual levels of transcription to an individual gene? The latter possibility seems more plausible at first sight, but at least in particular cases, such as in hemoglobin genes, there is evidence for an all-or-none “switch” of individual genes whenever the concentration activating transcription factors reaches a critical level. In the latter case, gene expression in the beta-globin locus is controlled by a cluster of no less than five enhancer-type DNA segments that are collectively referred to as the LCR (►**locus control region**) and that are 20–50 kb away from the beta-globin family members. The LCR ensures the complete stage- and cell-type-specific expression of the different beta-chains (18). Facing the possibility of an activation over such distances of dozens of kilobases, the question arises whether activation can only occur on a contiguous DNA molecule (“*in cis*”), or whether an activation *in trans*, from one DNA to another, is also possible. Although enhancers typically exert their activity *in cis*, cases of an activation *in trans* have been reported. This phenomenon, called ►**transvection**, is of particular importance in *Drosophila*, where the homologous chromosomes remain paired in somatic cells, allowing an enhancer to influence the corresponding allele on the homologous chromosome.

Because of its property of causing a loose, permissive chromatin structure, an enhancer can be exploited for additional activities. A particular stage of development can be associated with the emergence of a new origin of DNA replication. During immune cell maturation, an immunoglobulin gene enhancer can cause cytosine demethylation in CG-dinucleotides, sequence-specific DNA-rearrangement (VDJ joining), and “class switch recombination”. Furthermore, the amazing process of local immunoglobulin gene hypermutation in B-cells is dependent on an active transcription enhancer.

Besides transcription enhancers, the immunoglobulin gene locus harbors so-called “matrix attachment regions” (MARs). These are [A+T] rich DNA-segments, which do

not necessarily enhance transcription on their own, but in an experimental situation can facilitate enhancer activity over long distances in transgenic animals, apparently by counteracting the formation of transcriptionally inactive chromatin.

The transcription enhancer's counterpart is referred to as a ►[silencer](#). This term defines DNA segments that are able to inactivate genes *in cis* over long distances (19) (Fig. 1e). Instead of activators, a silencer interacts with repressor proteins, which unlike bacterial repressors do not necessarily have to bind to the promoter region, but inactivate corresponding genes through different mechanisms, recruitment of histone deacetylases, specific histone methylases and co-repressors that in turn can block the activity of the basal transcription apparatus. Enhancers and silencers were originally described as completely different entities. But especially in the control of cell-cycle and embryonic development, situations have been described where in the absence of an activation signal, one and the same DNA-binding transcription factor can bind a co-repressor instead of a co-activator, such that the distinction between enhancer and silencer becomes ambiguous. Therefore, an enhancer in one stage or cell-type would act as a silencer in other circumstances. In other cases of overlapping enhancer/silencer functions, binding sites for activator and repressor proteins are not the same but are rather interdigitated in a given segment of DNA. However, such overlaps do not seem to be the rule, and in most cases, enhancers and silencers have been described as separate entities.

Wherever activation or repression over long distances is possible, there also must be mechanisms to limit undesirable effects on noninvolved genes. The so called boundary elements or insulators serve this purpose, namely, to protect a gene from the effect of a neighboring enhancer or silencer or from negative effects of heterochromatin (20). Insulators are again DNA segments with an array of binding sites for specific regulatory proteins. For optimal insulation from neighboring elements, a gene has to be bracketed by a pair of ►[insulator/boundary elements](#) (Fig. 1e). Experimental deletion of an insulator can leave an undefined transition between inactive heterochromatin and active euchromatin; the border is then settled more or less randomly in every individual cell. Once such a border is established in early embryonic development, it is retained throughout following cell divisions, which leads to clonal patches of cells that may or may not express a particular gene in the border region. Such patchy gene expression patterns are referred to as PEV (position effect variegation) and have been most extensively studied in the fruit fly, e.g., as “mottled” eye color.

Finally, to avoid confusion of terms, it should be mentioned that during the process of RNA-splicing, certain RNA sequences, especially in exons, bind specific proteins that facilitate splicing at a nearby site. Such ►[“splice enhancer”](#) sequences thus allow for the usage of an otherwise suboptimal splice site. Splice enhancers play an important role in the regulation of alternative splicing. Fundamental differences from transcription enhancers (and, accordingly, from silencers) are that splice enhancers consist of (single stranded) RNA, rather than DNA, and thus are not active in reverse orientation and that they act over only a short distance (dozens of base pairs) to facilitate splicing.

►[Core Promoter](#)

►[Gene Duplications](#)

►[NFκB Pathway](#)

►[RNA Polymerase I](#)

►[Transcription Factors and Regulation of Gene Expression](#)

References

1. Banerji J, Rusconi S, Schaffner W (1981) Expression of a beta-globin gene is enhanced by remote SV40 DNA sequences. *Cell* 27:299–308
2. Moreau P, Hen R, Wasylyk B et al (1981) The SV40 72 base repair repeat has a striking effect on gene expression both in SV40 and other chimeric recombinants. *Nucleic Acids Res* 9:6047–6068
3. Banerji J, Olson L, Schaffner W (1983) A lymphocyte-specific cellular enhancer is located downstream of the joining region in immunoglobulin heavy chain genes. *Cell* 33:729–740
4. Gillies SD, Morrison SL, Oi VT et al (1983) A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell* 33:717–728
5. Boll W, Noll M (2002) *Drosophila* Pox neuro gene: control of male courtship behavior and fertility as revealed by a complete dissection of all enhancers. *Development* 129:5667–5671
6. Struhl K (1999) Fundamentally different logic of gene regulation in eukaryotes and prokaryotes. *Cell* 98:1–4
7. Reitzer LJ, Magasanik B (1987) Transcription of *glnA* in *E. coli* is stimulated by activator bound to sites far from the promoter. *Cell* 45:785–792
8. Serfling E, Jasin M, Schaffner W (1985) Enhancers and eukaryotic gene transcription. *Trends Genet* 1:224–230
9. Barolo S, Posakony JW (2002) Three habits of highly effective signaling pathways: principles of transcriptional control by developmental cell signaling. *Genes Dev* 16:1167–1181
10. Agalioti T, Lomvardas S, Parekh B et al (2000) Ordered recruitment of chromatin modifying and general transcription factors to the IFN-beta promoter. *Cell* 103:667–678

11. Grosschedl R, Birnstiel ML (1980) Identification of regulatory sequences in the prelude sequences of an H2A histone gene by the study of specific deletion mutants *in vivo*. *Proc Nat Acad Sci USA* 77:1432–1435
12. Myers LC, Kornberg RD (2000) Mediator of transcriptional regulation. *Annu Rev Biochem* 69:729–749
13. Mueller-Storm HP, Sogo JM, Schaffner W (1989) An enhancer stimulates transcription *in trans* when attached to the promoter via a protein bridge. *Cell* 58:767–777
14. Dillon N, Grosveld F (1993) Transcriptional regulation of multigene loci, multilevel control *Trends Genet* 9:134–137
15. Bulger M, Groudine M (1999) Looping versus linking: toward a model for long-distance gene activation. *Genes Dev* 13:2465–2477
16. Schirm S, Jiricny J, Schaffner W (1987) The SV40 enhancer can be dissected into multiple segments, each with a different cell type specificity. *Genes Dev* 1:65–74
17. Li XX, Noll M (1994) Compatibility between enhancers and promoters determines the transcriptional specificity of gooseberry and gooseberry neuro in the *Drosophila* embryo. *EMBO J* 13:400–406
18. Wijgerde M, Grosveld F, Fraser P (1995) Transcription complex stability and chromatin dynamics *in vivo*. *Nature* 377:209–213
19. Brand AH, Breeden L, Abraham J et al (1985) Characterization of a ‘silencer’ in yeast: a DNA sequence with properties opposite to those of a transcriptional enhancer. *Cell* 41:41–48
20. Kellum R, Schedl P (1991) A position-effect assay for boundaries of higher order chromosomal domains. *Cell* 64:941–950

thus allow recognition of additional components of regulatory networks as a whole.

► [Drosophila as a Model Organism for Functional Genomics](#)

Enophthalmos

Definition

Enophthalmos is the recession of the eyeball into the orbital cavity.

► [Marfan Syndrome](#)

Ensemble Average

Definition

An ensemble is a statistically representative set of configurations. Averages of molecular properties calculated over all members of the ensemble are used to evaluate macroscopic (thermodynamic) quantities according to principles of statistical thermodynamics.

► [Molecular Dynamics Simulation in Drug Design](#)

Enhancer Trap

Definition

Enhancer trap describes a strategy by which a targeting vector traps or senses a nearby enhancer through activation of a reporter gene. The inserted vector sequence acts as a tag that facilitates rapid cloning of the nearby gene.

► [Medaka as a Model Organism for Functional Genomics](#)

Enhancer/Suppressor Screens

Definition

Such modifier screens are aimed at identifying all genes that may enhance or suppress a specific phenotype, and

Enteric Nervous System

Definition

Enteric nervous system refers to a branch of the autonomic nervous system which consists of neurons and glia, organized into two main plexuses (submucosal and myenteric) localized throughout the gut and its derivatives.

► [Neural Crest Cells and their Derivatives](#)

Enterobacteriaceae

Definition

Enterobacteriaceae are a family of Gram-negative, asporogenous, facultative anaerobic bacteria. They are typically rod-shaped and can be motile or non-motile.

The metabolism is chemoorganotrophic and may be respiratory or fermentative, depending on growth conditions. They occur frequently as parasites, pathogens or commensals in humans and animals. They are typically found in the intestine, particularly the small intestine. Many are important opportunist causes of nosocomial infections. Important genera: *Enterobacter*, *Escherichia*, *Salmonella*, *Shigella*, *Yersinia*.

► [Protein Interaction-Phage Display](#)

Enthalpy/Enthalpy Change (DH)

Definition

The term enthalpy defines the heat (or energy) content of a system. Enthalpy change (DH) defines a thermodynamic state function, which represents the difference in internal energy between a given state and the reference states absent pressure/volume or other work.

► [Protein-Protein Interaction](#)
 ► [Thermodynamic Properties of DNA](#)
 ► [Two Hybrid System](#)

Entrainment

Definition

Entrainment describes the process by which the period and phase of a biological oscillator, such as the circadian clock in the suprachiasmatic nucleus of the brain, synchronizes to an environmental rhythm such as the daily light-dark cycle.

► [Circadian Clocks](#)

Entrez

Definition

Entrez is a WWW-based database retrieval program created by the National Center for Biotechnology Information (NCBI), a division of NIH.

► [Protein Databases](#)

Entropy Change (DS)

Definition

Entropy change (DS) defines a thermodynamic state function, which represents a measure of the change of order in a given state relative to the reference state.

► [Thermodynamic Properties of DNA](#)

ENU

Definition

ENU stands for N-ethyl-N-nitrosourea. It is an alkylating chemical mutagen that can transfer its ethyl group to oxygen or nitrogen radicals in DNA, without any metabolic processing required for its activation. This causes point mutations in a single DNA strand.

► [Large-Scale ENU Mutagenesis in Mice](#)
 ► [Mutagenesis Approaches in Medaka](#)
 ► [Mutagenesis Approaches in the Zebrafish](#)

env

Definition

The retrovirus gene env encodes the surface trimer consisting of SU (surface) and TM (transmembrane) and mediates interaction with and entry into the host cell.

► [Retroviruses](#)

Envelope

Definition

Envelope refers to the virion membrane surrounding the capsid of some viruses, including ► [retroviruses](#).

The envelope is formed from a part of the host cell membrane by budding.

Enzyme Catalyzed Post-Translational Hydroxylation of Proteins

CHRISTOPHER SCHOFIELD,

THOMAS MURRAY-RUST

The Oxford Centre for Molecular Sciences and
The Chemistry Research Laboratory, Oxford, UK
christopher.schofield@chem.ox.ac.uk
thomas.murray-rust@chem.ox.ac.uk

Definitions

The hydroxylation of proteins as catalyzed by dioxygen dependent enzymes is now established as an important ►post-translational modification. The importance of the oxidative hydroxylation of amino acid residues in collagen biosynthesis has been long recognized, but the role of post-translational hydroxylation in signalling and degradation pathways has recently become clear. Whilst the full extent of post-translational hydroxylation relative to other modifications, including phosphorylation, glycosylation, acetylation and methylation, has yet to be established, the results arising work on the hypoxic sensing pathway in animals suggests that hydroxylation may be significantly more extensive than previously perceived. Manipulation of post-translational hydroxylation offers prospects of new therapies for the treatment of cancer, cardiovascular diseases and diseases associated with connective tissue.

Characteristics

Types of Protein Hydroxylation

Initially it was thought that the oxygen of post-translationally introduced alcohols was derived from water, however pioneering work by Hayaishi and contemporaries demonstrated that it originated from dioxygen. Oxidative hydroxylation, i.e. the conversion of a C-H bond to a C-OH (alcohol) group, must be mediated via a ►reactive oxidizing species (ROS). In principle protein hydroxylation may be mediated either by 'free' ROS, e.g. superoxide, peroxide radicals or macromolecular bound ROS, i.e. enzymes. Protein hydroxylation via 'free' ROS has been proposed to occur in oxidative responses involved in stress and aging and it is also known to be involved in the inactivation of oxidizing enzymes. However, whilst 'free' ROS may well mediate specific post-translational

modifications such as disulfide formation and the oxidation of tyrosine residues, evidence for their involvement in the selective hydroxylation of unactivated C-H bonds is lacking. Given the activation energy barrier for hydroxylation of an unactivated C-H bond it would seem unlikely. The oxidation, including hydroxylation, of nucleic acids by free ROS has been more thoroughly studied and has been plausibly linked with tumour development.

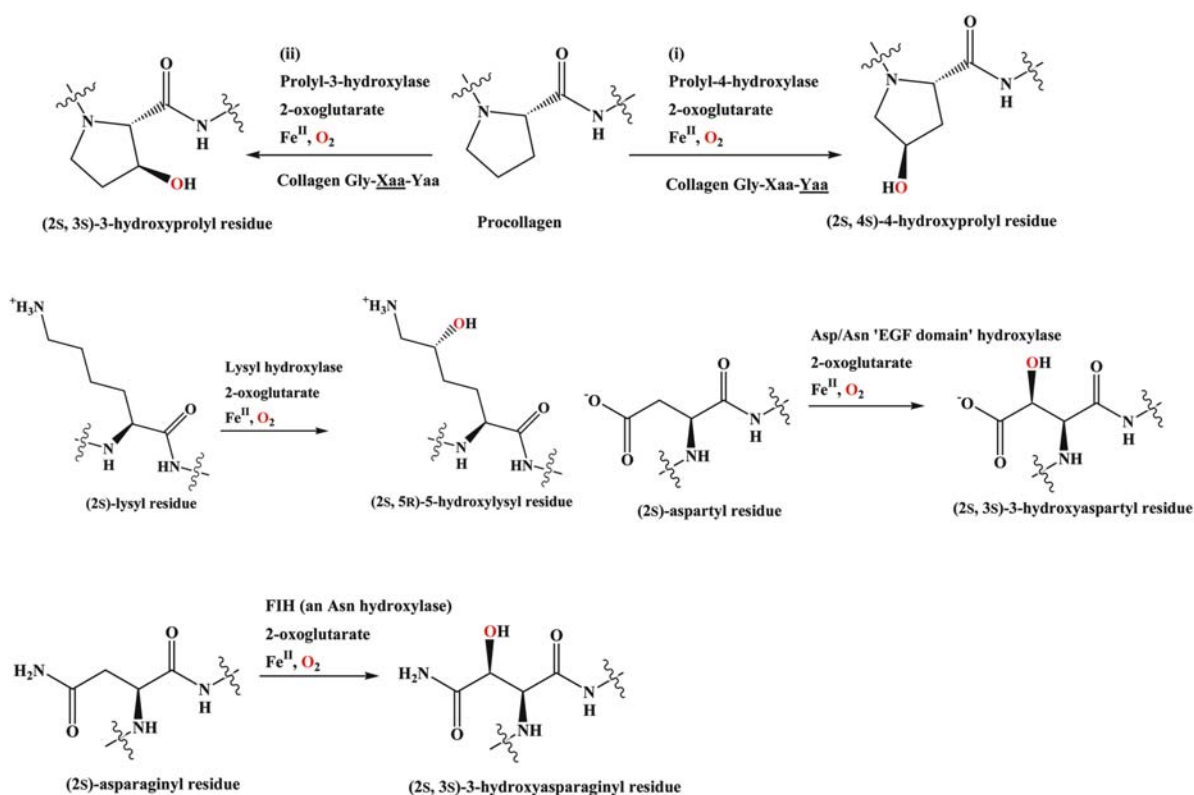
The involvement of protein ►hydroxylases in the biosynthesis of 4-hydroxyprolyl residues was first demonstrated in the 1960s. The following post-translational modifications involving oxygenase-mediated hydroxylation of an unactivated C-H bond have been identified: (Fig. 1)

1. Prolyl residues to 4*R*-(*trans*)-hydroxyprolyl residues (in collagen biosynthesis and transcription factors involved in hypoxic sensing),
2. Prolyl residues to 3*S*-(*trans*)-hydroxyprolyl residues (in collagen biosynthesis),
3. Lysyl residues to 5*R*-hydroxylysyl residues (in collagen biosynthesis),
4. Aspartyl residues to 3*S*-hydroxyaspartyl residues (in epidermal growth factor-like [EGF-like] domains),
5. Asparaginyl residues to 3*S*-hydroxyasparaginyl residues (in EGF-like domains and transcription factors involved in hypoxic sensing),
6. Formation of α -hydroxylated derivatives of C-terminal glycyl residues in protein amidation. (Fig. 12)

Characteristics of Protein Hydroxylases

To date, the enzymes identified as catalyzing oxidative post-translational hydroxylations of C-H bonds in side-chains of protein residues belong to a single superfamily termed the 2-oxoglutarate (2-OG) dependent oxygenases. Almost all members of this family employ ferrous iron [Fe(II)] as a cofactor and 2-oxoglutarate and dioxygen as cosubstrates (Fig. 2). [Hydroxylation of C-terminal peptidyl glycine residues is catalysed by different enzymes and is discussed below]. These enzymes are widespread and, in addition to post-translational modifications, catalyze many reactions in metabolic and regulatory processes including the hydroxylation of free amino acids, steps in the biosynthesis of antibiotics in microorganisms and plant signalling molecules, fatty acid metabolism and DNA repair. Although the 2-oxoglutarate oxygenases catalyze a range of oxidative reactions including desaturation and oxidative cyclizations, their most common reaction is hydroxylation, either of small molecules (e.g. gamma-butyrobetaine) or macromolecules including proteins and DNA.

Crystallographic and sequence analysis studies have revealed that many 2-oxoglutarate dependent oxygenases, including most of those involved in post-translational

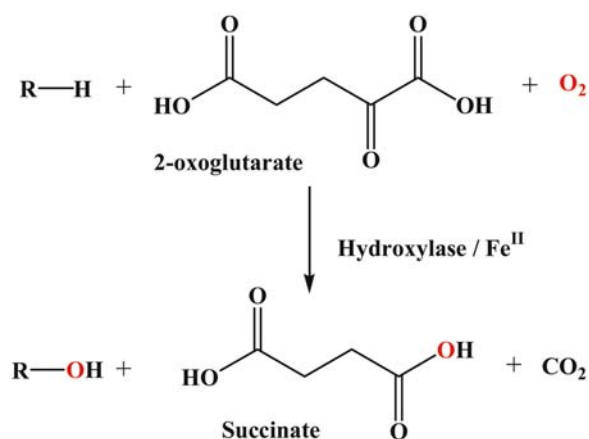


Enzyme Catalyzed Post-Translational Hydroxylation of Proteins. Figure 1 Known post-translational hydroxylations catalyzed by Fe(II) oxygenases: (i) prolyl (Gly-Pro-Pro) to (2S, 4R)-4-hydroxyprolyl; (ii) prolyl (Gly-Pro-Pro) to (2S, 3S)-3-hydroxyprolyl; (iii) lysyl to (2S, 5R)-5-hydroxylysyl; (iv) aspartyl to (2S, 3S)-3-hydroxyaspartyl; (v) asparaginyl to (2S, 3S)-3-hydroxyasparaginyl.

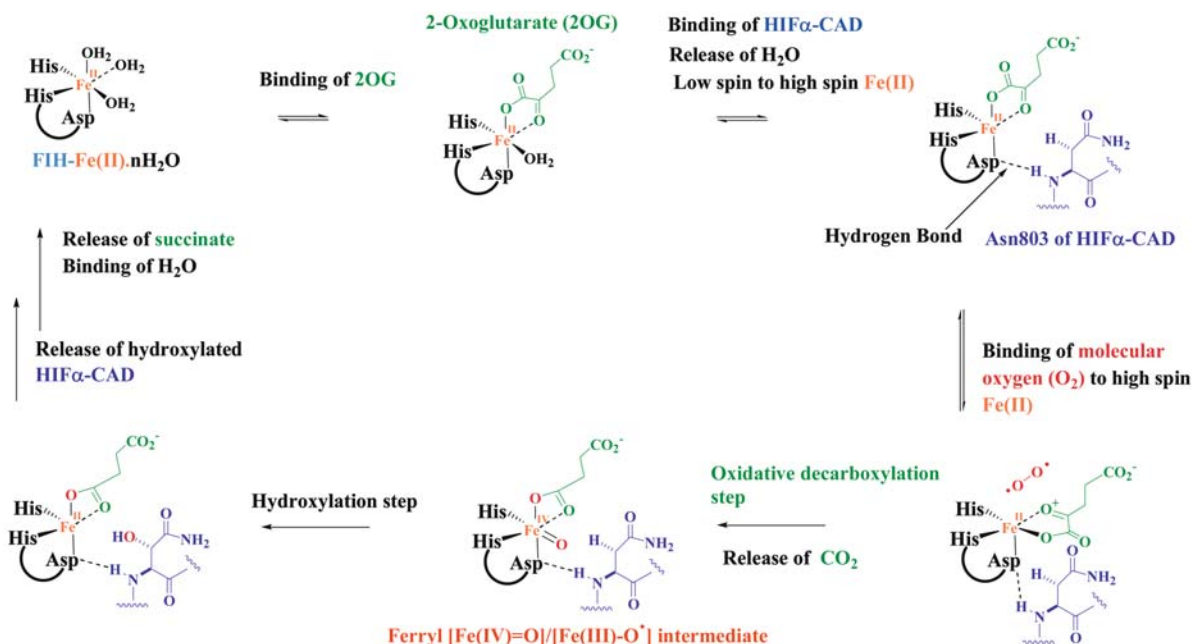
hydroxylation, contain a core of eight beta-strands arranged in a double-stranded beta helix (or jelly roll) motif. In the light of the crystallographic data, spectroscopic and kinetic analyses suggest that many, but not necessarily all of the family share closely related mechanisms in which the enzyme-Fe(II) complex first binds 2-OG followed by substrate. Binding of the latter triggers the enzyme for reaction with dioxygen. Subsequently, a ferryl [$\text{Fe}(\text{IV})=\text{O}$] \rightleftharpoons $\text{Fe}(\text{III})-\text{O}^\bullet$ intermediate, which acts as the ROS, is produced by oxidative decarboxylation of 2-oxoglutarate (Fig. 3).

For some 2-OG oxygenases, e.g. procollagen prolyl-4R-hydroxylase, catalytic activity is highly dependent upon the presence of ascorbate. In other cases, the role of ascorbate can be substituted by other reducing agents and/or the addition of ascorbate makes little difference to the catalytic rate. One role of ascorbate is to complete 'uncoupled' reaction cycles in which a ferryl [$\text{Fe}(\text{IV})=\text{O}$] intermediate is generated by reaction of 2-oxoglutarate and oxygen either in the absence of (protein) substrate or in the presence of a substrate that is incorrectly bound to enable hydroxylation (Fig. 4).

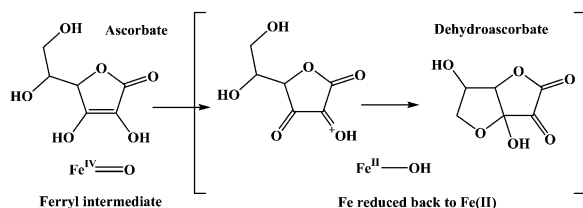
In this process the ascorbate is oxidized (either to dehydroascorbate or semi-dehydroascorbate) and the reactive intermediate is reduced to Fe(II) enabling the



Enzyme Catalyzed Post-Translational Hydroxylation of Proteins. Figure 2 The stoichiometry of a 2-oxoglutarate catalyzed hydroxylation reaction.



Enzyme Catalyzed Post-Translational Hydroxylation of Proteins. Figure 3 Outline mechanism for a 2-oxoglutarate catalyzed hydroxylation reaction, illustrated for factor inhibiting HIF (FIH).



Enzyme Catalyzed Post-Translational Hydroxylation of Proteins. Figure 4 A role for ascorbate (vitamin C) in completing 'uncoupled' reaction cycles by 2-oxoglutarate oxygenases.

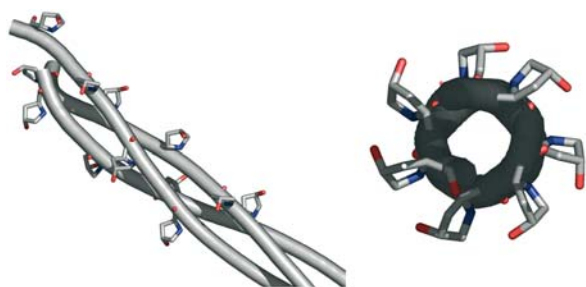
initiation of another catalytic cycle. In the absence of ascorbate, or other appropriate reducing agents, the enzyme-bound ferryl [Fe(IV)=O] intermediate or an oxidizing species derived from it can cause inactivation *via* oxidative processes including self hydroxylation or proteolytic cleavage. However, the *in vitro* dependence of 2-oxoglutarate oxygenases on ascorbate is highly variable and there is some evidence that in the case of enzymes that have a strict requirement, ascorbate may be involved in forming a productive enzyme-substrate complex.

Post-Translational Hydroxylation and Collagen Biosynthesis Overview

Collagen is one of the most abundant and important structural proteins in animals. Genomic analyses also

suggest that collagen related structures are present in most major organisms including plants and viruses. The limited occurrence of collagen type proteins in lower eukaryotes and their absence in archaeobacteria suggests that horizontal gene transfer may account for their presence in lower organisms. In plants, the collagen-like macromolecules – including extensins, proline-rich proteins, lectins and arabinogalactan proteins – function structurally within cell walls. The hydroxyprolyl residues of these plant macromolecules are often glycoacceptors and sugar attachment is critical to their ability to provide tensile strength to the cell. Three post-translational hydroxylations have been identified in animal collagen, 4*R*-hydroxy-*L*-proline, 3*S*-hydroxy-*L*-proline and 5*R*-hydroxy-*L*-lysine, while in plants the main modification appears to be 4*R*-hydroxy-*L*-proline. In animals, 4*R*-hydroxy-*L*-proline is the most common modification followed by 5*R*-hydroxy-*L*-lysine and the rare 3*S*-hydroxy-*L*-proline.

Collagen is structurally distinctive due to the presence of domains possessing an unusual triple stranded tertiary structure. This comprises three parallel left-handed strands wound around a central axis to form a triple helix with a right-handed pitch. This is present in the *ca.* 20 different types of collagen of which the most abundant is Type I collagen which contains *ca.* 300 repeating Gly-Xaa-Yaa triplet motifs, which are critical to the tertiary structure and thermostability of collagen. The packing of the collagen triple helix requires every



Enzyme Catalyzed Post-Translational Hydroxylation of Proteins. Figure 5 Views of collagen. Left, the triple helix is shown in cartoon form with the 4*R*-hydroxy-*L*-prolyl residues displayed. Right, a view along the axis of the triple helix, with the 4*R*-hydroxy-*L*-prolyl residues shown (oxygen in red, nitrogen in blue).

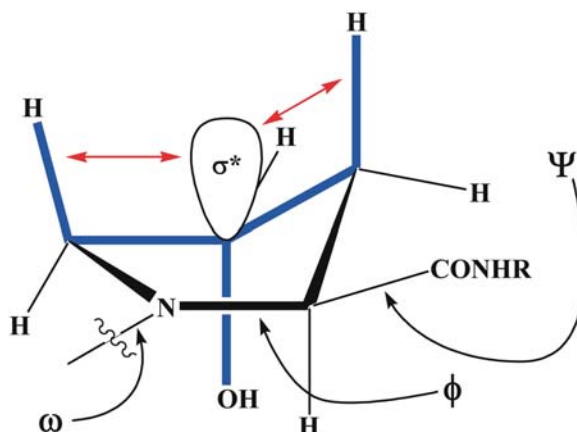
third residue to be glycine. The Xaa residue is often *L*-proline and the Yaa residue 4*R*-hydroxy-*L*-proline. The tertiary structure of collagen enables it to form strong and flexible supramolecular fibrils that can be cross-linked.

Types and Roles of Collagen Hydroxylations

The hydroxyl group of the 4*R*-hydroxy-*L*-proline residues of the triplet motifs was once thought to confer stability on the structure by its involvement in intra-helix hydrogen bonding and/or by stabilising a network of water bridges. However, structural work and studies employing the systematic substitution of the triplet residues are consistent with stabilisation by 4*R*-hydroxy-*L*-proline in the Yaa position occurring at least in part *via* a stereoelectronic effect termed the gauche effect. This arises from the propensity of electronegative groups on adjacent carbons to adopt a gauche relationship (Fig. 6). As a consequence the presence of an electron-withdrawing group at the 4*R*-position of proline stabilizes a particular conformation (or pucker) of the proline pyrrolidine ring (*Cy-exo*) such that the main chain torsion angles (omega, phi, psi) are preorganised to favour the collagen structure. Central to the development of this proposal was the observation that substitution of 4*R*-fluoro-*L*-proline at the Yaa position [of the (Gly-Xaa-Yaa) collagen triplet] resulted in collagen that was even more stable than that with 4*R*-hydroxy-*L*-proline at this position.

Fewer studies have been carried out on the role of the rare 3*S*-hydroxy-*L*-proline residues on collagen stability. In contrast to 4*R*-hydroxy-*L*-proline, it appears to effect a slight destabilization in its natural Xaa position, an effect that is increased when it is in the unnatural Yaa position.

Although details remain to be resolved, the available evidence implies that the pattern and extent of lysyl hydroxylation in the non-triple helical domains of



Enzyme Catalyzed Post-Translational Hydroxylation of Proteins. Figure 6 The conformation of (2*S*, 4*S*)-4-hydroxyprolyl residues in collagen. The favoured conformation is proposed to maximise overlap between the C-O (σ^*) antibonding orbital and the two adjacent C-H (σ) bonding orbitals. This preorganises the fixation of the bond angles ω , ϕ and ψ at the ideal values for triple helix formation.

collagen are important in determining collagen cross-linking pathways that are tissue specific. Hydroxylysine is essential for the formation of the covalent pyridinium cross-links pyridinoline and deoxypyridinoline (among mature collagen molecules). Pyridinoline derives from three hydroxylysyl residues, whereas deoxypyridinoline derives from one lysyl and two hydroxylysyl residues. These and other cross-links tune the properties of collagens to specific roles, e.g. hydroxylysyl-pyridinoline is a typical cross-link of skeletal tissue and it is also thought to play a major part in the hardening of sclerotic tissue. The patterns of lysyl hydroxylation may be regulated by expression of different lysyl hydroxylase genes.

The Enzymes of Collagen Hydroxylation

The expression of both the prolyl-3- and 4-hydroxylases has been shown to be dependent on the oxygen tension and, in the case of the prolyl-4-hydroxylases, shown to be mediated by HIF1 α . (See below)

Procollagen Prolyl-4-Hydroxylase (EC 1.14.11.2)

The substrates for both procollagen prolyl-4-hydroxylase and prolyl-3-hydroxylase are repeated amino acid triplets in procollagen, with (-Gly-Pro-Pro)_n being the natural substrate for the prolyl-4-hydroxylases. The (-Gly-Pro-4Hyp)_n generated by this reaction is in turn the prime substrate for the prolyl-3-hydroxylases. It has been shown that the prolyl-4-hydroxylases only act on procollagen that has not formed a triple helical structure, and also that helical and nonhelical procollagen are secreted from the cell at different rates.

Studies on the quaternary structure of collagen prolyl-4-hydroxylases have shown that they can adopt different oligomerisation states but that the active enzyme exists in vertebrates as an $\alpha_2\beta_2$ heterotetramer. The α subunit exists in various isoforms [$\alpha(I)$, $\alpha(II)$ and $\alpha(III)$] and contains the residues directly involved in prolyl hydroxylase activity.

The β subunits have been identified as protein disulfide isomerase (PDI) and are essential for both catalytic activity and productive folding of the α subunits. The PDI/ β subunit is typically expressed in 10–100 fold excess over the α subunits in most cells and tissues and also functions in different pathways independently of them. It has been proposed that there is no mixing of the α subunits within tetramers, i.e. there are [$\alpha(I)_2\beta_2$] and [$\alpha(II)_2\beta_2$] complexes which form the Type I and Type II prolyl hydroxylases respectively.

The prolyl-4-hydroxylases are widely expressed, and of the two types ($\alpha(I)$ and $\alpha(II)$) it is the Type I enzymes which are believed to contribute most towards the total *in vivo* prolyl hydroxylase activity in humans – estimated at 60–90%. Prolyl-4-hydroxylases have been identified in many vertebrates as well as bacteria, plants and viruses and play a critical role in correct development of structural proteins across species. In humans, prolyl-4-hydroxylases are expressed in a wide variety of tissue and cell types and Type I prolyl-4-hydroxylase is the main form in most cell types, including skeletal myocytes and smooth muscle cells. Type II prolyl-4-hydroxylase is the main form in chondrocytes, capillary endothelial cells and cultured umbilical cells, as measured using immunofluorescence and Western blotting. The subcellular location of prolyl-4-hydroxylase α subunits is predominantly in the endoplasmic reticulum, although PDI is more widely distributed.

Procollagen Prolyl-3-Hydroxylase (EC 1.14.11.7)

Prolyl-3-hydroxylases are less common than the prolyl-4-hydroxylases and have been less extensively studied. They have only been definitively identified in vertebrates and evidence for their existence in plants and viruses is lacking. The substrate of prolyl-3-hydroxylase is the prolyl residue in the repeated amino acid sequence (-Gly-Pro-4Hyp-) of procollagen. Little structural information on the prolyl-3-hydroxylases has been reported but gel filtration has indicated its molecular weight to be roughly 160 kD. The tertiary and quaternary structures are not known at this time and so far it has not been cloned and expressed *in vitro*.

Procollagen Lysyl-5-Hydroxylases (EC 1.14.11.4)

The substrate of procollagen lysyl-5-hydroxylase is procollagen and contains the sequence (-Gly-Xaa-Lys-). Lysyl hydroxylation is involved in the formation of all the different types of collagen, as it is critical for both

the crosslinking and attachment of sugars to the collagen chain. In addition to hydroxylase activity, one human lysyl hydroxylase has recently been shown to possess both glucosyl- and galactosyl-transferase activity, and hence has the ability to generate glucosylgalactosylhydroxyllysyl residues. There are three isoforms of lysyl hydroxylase in humans (LH1-3), with subunit weights of *ca.* 82 kD, one of which has two different splice variants (LH2a and LH2b). LH1 has recently been cloned and expressed in a eukaryotic cell line (CHO-K1).

The three human lysyl hydroxylase isoforms are found in a wide variety of tissue types including brain, heart, spleen, artery and liver. LH1 is constitutively expressed in the widest range of tissues while LH2b is expressed at different levels in different tissue types. The subcellular localization of LH1 is known to be in the lumen of the endoplasmic reticulum but nothing is known about the other isoforms.

Protein Hydroxylation and the Hypoxic Response

When the tissues of multicellular organisms are subjected to conditions of low oxygen tension, e.g. by travel to high altitude or impaired blood supply, there is a physiological reaction known as the hypoxic response. Critical to this response is the transcription factor HIF (hypoxia-inducible factor) which is a heterodimeric protein comprising HIF1 α and HIF1 β . HIF is conserved in a wide range of organisms ranging from *C. elegans* to humans. Whilst HIF1 β is a constitutive nuclear protein (identical in sequence to the aryl hydrocarbon receptor nuclear translocator) and is unaffected by oxygen tension, levels of HIF α are highly sensitive to dioxygen levels. Under normoxic conditions (sufficient oxygen), HIF α is not normally detectable, but under hypoxic conditions levels of HIF α rise and enable transcription of the array (>100 in humans) of genes involved in the hypoxic response.

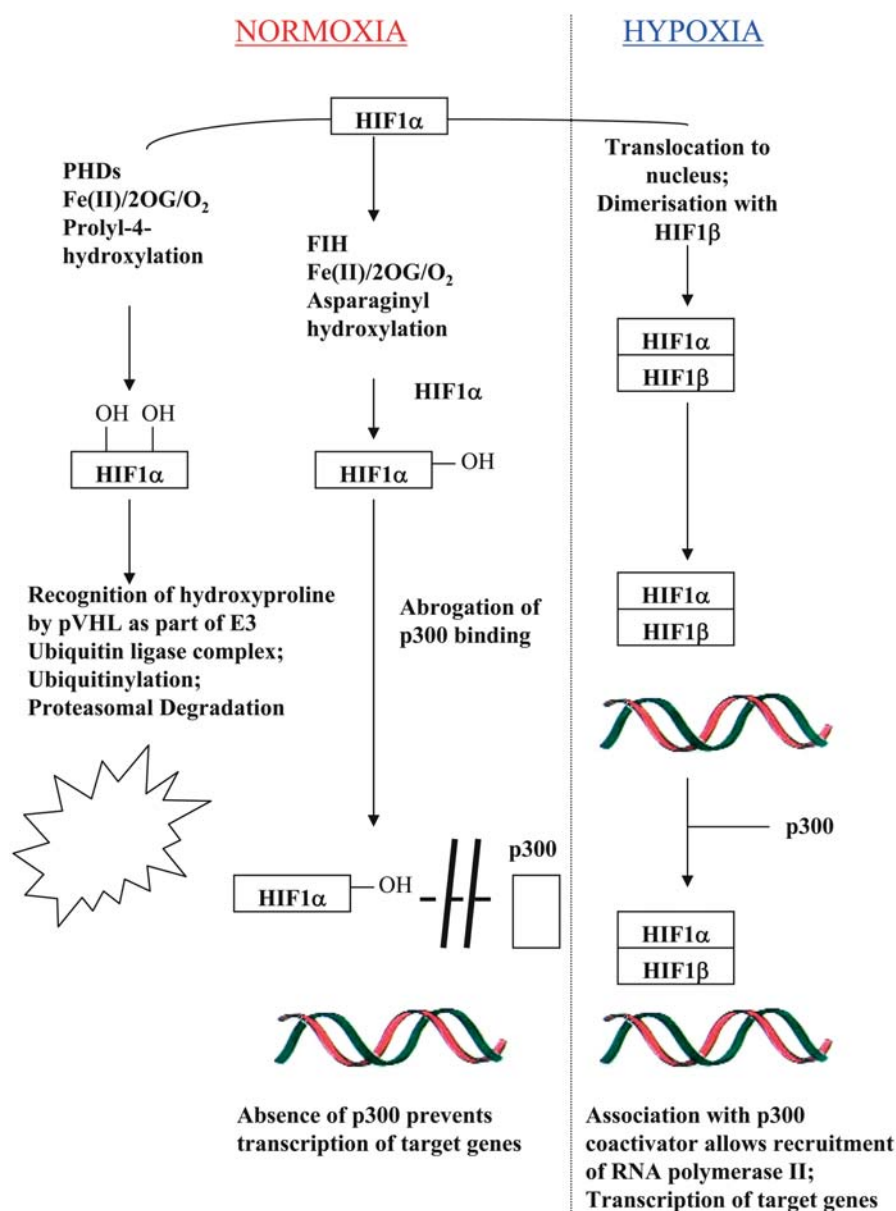
Three human variants of HIF α encoded by different genetic loci have been identified, two of which, HIF1 α and HIF2 α , are closely related. Both HIF1 α and HIF2 α possess a central oxygen dependent degradation domain that is absent in HIF1 β . The oxygen dependent degradation domain is composed of two sub-domains each of which can independently enable degradation of this HIF subunit.

Under hypoxic conditions, HIF1 α translocates to the nucleus where it heterodimerises with HIF1 β to form functionally active HIF1 $\alpha\beta$. The HIF1 $\alpha\beta$ dimer then recruits the transcriptional coactivator p300, that together with other proteins recruits RNA polymerase II to initiate transcription. Target genes that are up-regulated by HIF include those involved in glycolysis, [angiogenesis](#), vasculogenesis and erythropoiesis. Modulation of the hypoxic response in order to enable selective expression of hypoxically regulated genes

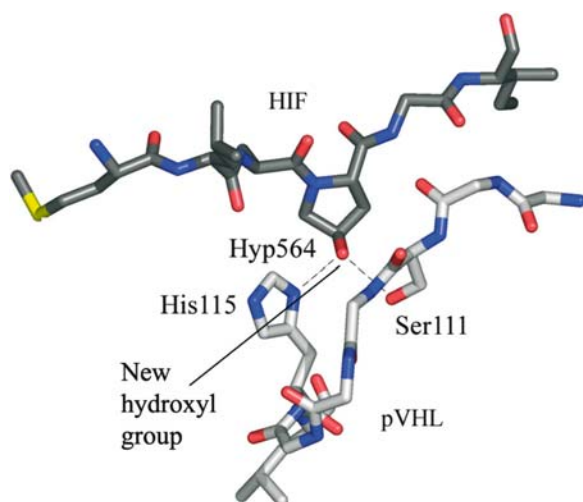
thus offers the possibility of new therapies for cancer and cardiovascular disease.

Recent work has demonstrated that post-translational modification, catalysed by a set of 2-oxoglutarate dependent oxygenases, is responsible for two separate mechanisms enabling the oxygen sensitive regulation of HIF α activity. Under normoxic conditions, HIF1 α can be hydroxylated by three closely related prolyl hydroxylases (PHD1-3, prolyl-hydroxylase domain-containing proteins) and an asparaginyl hydroxylase (factor inhibiting HIF, FIH). As shown in Fig. 7, both the stability and the transcriptional activity of HIF1 α are regulated by its hydroxylation.

The PHDs catalyse hydroxylation of two critical prolyl residues, Pro402 and Pro564 (using the human numbering system) in the oxygen dependent degradation domain of HIF1 α . Hydroxylation of HIF1 α in this manner enables it to form a complex with the von Hippel-Lindau tumour suppressor protein stabilised *via* two buried hydrogen bonds involving the new hydroxyl group. Since pVHL is the recognition subunit of an E3 ubiquitin ligase complex, this enables degradation of ubiquitin-tagged HIF1 α *via* the 26S proteasome. FIH enables its control of HIF α activity through the hydroxylation of an asparaginyl residue (Asn803 in human HIF1 α) in the C-terminal transactivation



Enzyme Catalyzed Post-Translational Hydroxylation of Proteins. Figure 7 Schematic of the hypoxic response.



Enzyme Catalyzed Post-Translational Hydroxylation of Proteins. Figure 8

The structure of a peptide corresponding to the C-terminal oxygen-dependent degradation domain (C-ODD) of HIF1 α after prolyl hydroxylation, complexed to pVHL. The hydroxyl group of Hyp564 (red, arrowed) forms hydrogen bonds with the side chains of His115 and Ser111 of pVHL, thus allowing pVHL to specifically recognise hydroxylated HIF.

domain (CAD) required for its transcriptional activity; this transactivation domain binds to the CH1 (cysteine/histidine rich) domain of p300/CREB binding protein. Hydroxylation at the β -position of Asn803 blocks the interaction between CAD and the p300/CBP coactivator.

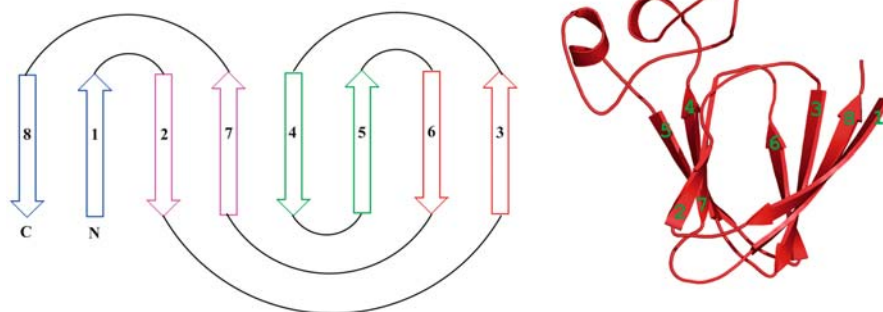
The relative importance to the hypoxic response of hydroxylation of the two prolines (HIF1 α P402, P564) in the ODDD compared to the Asn803 of the CAD is not yet fully resolved and it may be that PHDs and FIH are tissue and/or substrate dependent allowing selective regulation.

Crystallographic analysis of FIH has revealed that, like other 2-OG oxygenases, it contains a single Fe(II) coordinated in an octahedral manner. His199, Asp201 and His279 provide side-chains that bind the iron in a facial triad (Fig. 11). The 2-OG cosubstrate binds to the iron in a bidentate manner *via* its 1-carboxylate and 2-oxo groups. In the absence of HIF substrate, a water molecule is also ligated to the metal.

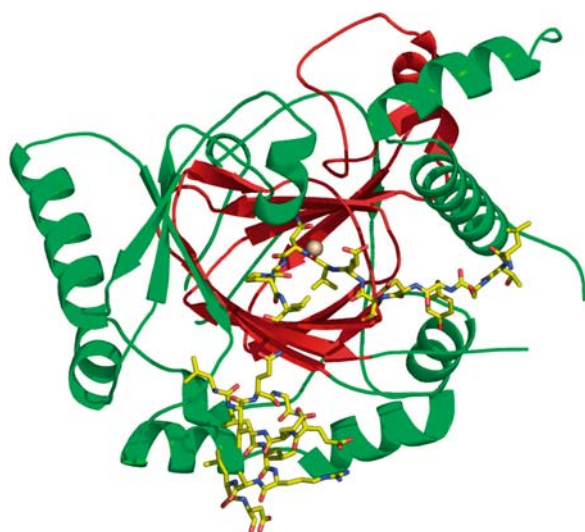
In contrast to many 2-OG oxygenases in which a Ser/Thr residue and an Arg residue (on the eighth strand of the DSBH), bind the 5-carboxylate of 2-OG, in FIH the residues that form electrostatic interactions with the 2-OG side chain are the side chains of Lys214 (on the fourth strand of the DSBH), Thr196 and Tyr145. This and other structural features identify FIH as a member of a distinct subfamily of human 2-OG dioxygenases. Furthermore, since FIH contains a JmjC homology region, the sequence of FIH defines it as one of the jumonji transcription factors. It is of current interest to define the functions of the JmjC proteins that are related to FIH. Members of the JmjC transcription factors family have been implicated in cell growth and heart development and possess a conserved motif present in the 2-OG oxygenases, suggesting that, like FIH and the PHD isozymes, some of them might be iron oxygenases involved in transcriptional regulation or another type of signalling.

Other Protein Hydroxylases

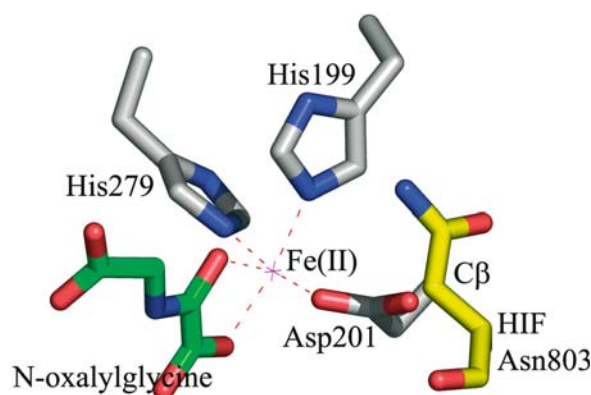
Vitamin K dependent proteins, including some coagulation factors, are modified by post-translational modification at the β -position of specific aspartyl and asparaginyl residues in their epidermal growth factor domains. Although the modified residues may be involved in calcium ion binding, the available evidence suggests that their hydroxylation does not seem to affect their affinity for the metal. The biological significance of these hydroxylations has not been elucidated, and they may be related to signalling since



Enzyme Catalyzed Post-Translational Hydroxylation of Proteins. Figure 9 (i) Schematic of the topology of a double stranded beta helix (DSBH) fold; (ii) An example of the formation of the DSBH motif fold found in most 2-OG dioxygenases. The figure shows the DSBH of FIH.



Enzyme Catalyzed Post-Translational Hydroxylation of Proteins. Figure 10 View from the crystal structure of FIH complexed with the C-terminal transactivation domain of its substrate HIF (in yellow). The double stranded beta helix motif in the centre of the enzyme is coloured in red and the iron is in grey-brown. Note that Asn803 of HIF1 α is adjacent to the iron.



Enzyme Catalyzed Post-Translational Hydroxylation of Proteins. Figure 11 Active site residues of FIH showing the side chains of the Fe-binding residues, Asn803 of HIF which is hydroxylated at its β -carbon and N-oxalylglycine (a 2-OG analogue used for crystallization) binding Fe in a bidentate manner.

EGF-like (epidermal growth factor-like) domain substrates occur in blood clotting factors and other proteins such as Notch. Recent results have showed that EGF Asp/Asn hydroxylase null mice crossed with APCmin, an intestinal tumour model, develop more and larger intestinal polyps than hydroxylase WT/APCmin mice. Although previous work has shown

increased levels of the hydroxylase in some tumour tissues, this finding suggests that loss of hydroxylase function may promote tumour formation.

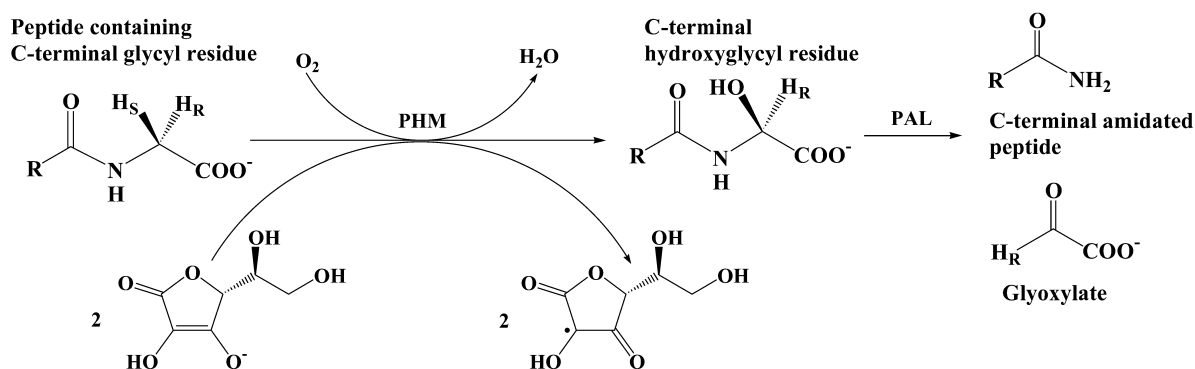
The EGF hydroxylase has been isolated and shown to be a 2-oxoglutarate oxygenase. Sequence alignments suggest that it probably does not contain the typical $\text{HXD}\dots\text{H}$ motif for iron binding and is likely to form a new structural sub-group. The substrate selectivity of the human epidermal growth factor hydroxylase contrasts with FIH since it accepts both Asp and Asn residues within sequence motifs as substrates, whereas FIH is limited to the latter.

Hydroxylation of C-Terminal Peptidyl Glycine Residues During Amidation

Protein hydroxylation is also involved in signalling pathways mediated by peptide hormones and neurotransmitters due to the prevalence of C-terminal amides in these molecules. Amidation is effected in a two-stage process originally thought to involve a one step mechanism but now known to require two catalytic domains within a bifunctional enzyme. The enzyme is known as peptidyl glycine α -amidating monooxygenase (PAM) and consists of two separate catalytic units: peptidyl glycine α -hydroxylating monooxygenase (PHM; EC 1.14.17.3), which converts the C-terminal glycyl residue to a hydroxyglycyl intermediate, and peptidyl hydroxyglycine N-C lyase (PHL), which cleaves it to produce a C-terminal amide and glyoxylate. The C-terminal part of the protein contains transmembrane and hydrophilic domains. The entire protein is encoded by one gene and the translated precursor protein formed can be processed to form the two subunits by proteolytic cleavage.

Crystal structures of a hydroxylase have revealed two copper ions complexed by separate jelly roll motifs. Like the 2-OG oxygenases, these enzymes use molecular oxygen as a cosubstrate, but do not use 2-OG. Instead they employ two molecules of ascorbate to effect 2-electron reduction of the dioxygen producing a reactive oxidizing species and two molecules of semidehydroascorbate. It is unclear whether the disproportion of semidehydroascorbate to ascorbate and dehydroascorbate is enzymatically controlled (a semidehydroascorbate reductase is known) or non-enzymatic.

There has been interest in PAM due to the stimulation of neuronal survival and olfactory neurogenesis by pituitary adenylyl cyclase-activating polypeptide (PACAP38), an amidated neuropeptide. It was observed that Mottled (Brindled) mice lacking a functional ATP7A copper transporter (hence providing a partial loss-of-function PAM knockout) had reduced levels of amidated PACAP and a concomitantly decreased number of olfactory neurons. Other neurotransmitters that rely on C-terminal amidation for their biological activity include adrenomedullin, galactin and allatostatin.



Enzyme Catalyzed Post-Translational Hydroxylation of Proteins. Figure 12 C-Terminal amidation of peptide hormones and neurotransmitters. The first step is abstraction of the *pro-S* hydrogen from α of the glycyl residue, followed by hydroxylation. The second step of the process comprises cleavage of the hydroxylated peptide to yield a C-terminal amide and glyoxylate.

Future Prospects

Until recently protein hydroxylation was thought to be a relatively specialized structural post-translational modification with the only widespread example occurring in collagen biosynthesis. The recent discovery of the role of protein hydroxylation in the HIF signalling pathway and its role in C-terminal amidation, combined with genomic analyses has raised the possibility that post-translational hydroxylation may be much more widespread than once thought. Whilst it is probably unlikely that the extent of post-translational hydroxylation will rival those of phosphorylation or glycosylation, it seems probable that, as with acetylation and methylation, protein hydroxylation will play an extensive role in signalling pathways. To date the comparative lack of studies on hydroxylation and other oxidative modifications including desaturation may reflect the lack of appropriate methodology for their detection, a problem that is being addressed by modern proteomic techniques.

Clinical Relevance

Clinical Relevance of Collagen Hydroxylation

Scurvy is a famous connective tissue disease with symptoms including haemorrhaging due to capillary degradation, breakdown of dentin and gum tissue (leading to loosening and loss of teeth) and joint pain. Unlike specific diseases associated with lysyl-5-hydroxylase (see below), the underlying cause of scurvy is deficiency of ascorbic acid (vitamin C) which causes procollagen prolyl-4R-hydroxylase to malfunction leading to a lack of mature collagen. Scurvy is treated directly by ascorbate or by application of a diet rich in ascorbate. The latter approach was famously undertaken by the British Navy upon the advice of

Sir Gilbert Blane in the 18th Century, following observations by James Lind that a daily ration of lemon juice was effective in combating scurvy.

Several subtypes of Ehlers-Danlos syndrome involve defects in collagen, and they occur at a frequency of *ca.* 1 in 20,000 people. Ehlers-Danlos syndrome Type VI is an autosomal recessive connective tissue disease with symptoms including muscular hypotonia and ocular problems. It is characterized by defects in collagen and is caused by mutations to procollagen lysyl-5-hydroxylase leading to deficient hydroxylation of lysine residues in tissue collagens, particularly those of the skin. The ratio of collagen cross-linking amino acids, hydroxylysyl pyridinoline and lysyl pyridinoline is biased towards the latter in Ehlers-Danlos syndrome Type VI patients and an abnormal ratio is used to diagnose the condition. Vitamin C has been tried as a therapy for the treatment of Ehlers-Danlos VI but, while in some cases it may be effective, it is not a standard treatment. Vitamin C stimulated production of pyridinoline and deoxypyridinoline in fibroblasts from patients with EDS VI has been observed and may reflect the stimulation of the activity of an alternative lysyl hydroxylase activity or that some mutations only partially inactivate the known procollagen lysyl-5-hydroxylase.

Bruck syndrome, also associated with lysyl hydroxylase, is a very rare genetic disease, the symptoms of which include bone fragility due to aberrant bone collagen formation. Bruck syndrome is associated with a deficiency in a telopeptide bone-specific lysyl hydroxylase (TLH), which has recently been identified as PLOD2, one of the lysyl hydroxylase isoforms, by identification of two missense mutations in exon 17 of the gene encoding PLOD2. There is also evidence that increased mRNA levels of PLOD2 correlate with

the increased crosslinking of collagen observed in fibrotic diseases and hence the link between hydroxylation and disease states is beginning to be understood. Chronic fibroses of major organs, commonly occurring after tissue damage, are significant clinical problems ranging from skin disfigurement to liver cirrhosis and can lead to progressive disability and death. For many of these diseases there are no effective therapies. These diseases characteristically involve overproduction of extracellular matrix material (such as collagen) and hence the specific inhibition of collagen hydroxylases may be a viable therapy. The inhibition of enzymes that are responsible for collagen biosynthesis and its regulation are thus of pharmaceutical relevance. Significant efforts have been made to produce inhibitors of collagen hydroxylases, in particular procollagen prolyl-4R-hydroxylase. As yet no compounds are in clinical use and the side effects of inhibiting the biosynthesis of mature collagen may be significant.

Clinical Relevance of HIF Hydroxylases

von Hippel-Lindau disease is a heritable cancer syndrome associated with a germline mutation of the VHL tumour suppressor gene on the short arm of human chromosome 3. It occurs in about 1 in 36,000 live births and its inheritance leads to a high individual risk of disease involving various benign and malignant tumours of the central nervous system, kidneys, adrenal glands, pancreas and reproductive organs. Although other factors may be involved, the lack of pVHL in patients is thought to disrupt the degradation of HIF α via the PHD/pVHL pathway leading to an uncontrolled hypoxic response. In VHL disease sufferers, the regulatory action of FIH may be disproportionally important and boosting FIH activity, e.g. *via* gene or protein therapy, may be a possible avenue for therapeutic evaluation. Levels of HIF α are increased in ischemic tissues and other diseased tissues including the pre-eclamptic placenta, the hypoxic retina and inflamed joints. Stimulation of the natural HIF mediated response may be useful in the treatment of important diseases including limb ischemia, heart attack and stroke. Regions of tissue that have died or are compromised as a result of oxygen starvation (e.g. after ischemic insult such as a heart attack) could potentially be revascularised *via* stimulation of HIF mediated transcriptional activity. Additionally, up-regulation of HIF activity may be used to induce expression of hypoxically regulated genes of medicinal importance such as **erythropoietin**, a recombinant form of which is currently used on a widespread scale for the treatment of anaemia. Perhaps the most attractive target for stimulation of the HIF activity is small molecule mediated inhibition of the PHDs and FIH, but

alternative targets for small molecules include blocking the interaction between HIF1 α and pVHL.

Modulation of the HIF system is also a potential target for cancer treatment. Growing tumours characteristically contain an oxygen tension gradient due to poor vascularisation and resulting regions of hypoxia within the tumour. As a consequence HIF α levels can be up-regulated to enable vascularisation of the tumour. Inhibition of the HIF transcriptional activity may thus be useful in the treatment of tumours. This might be achieved by small molecule mediated blocking of the interaction between HIF and p300 or gene therapy approaches that modulate the activity of the HIF system including the use of the HIF hydroxylase genes.

►Hypoxia Inducible Factors

References

1. Jenkins CL, Raines RT (2002) Insights on the conformational stability of collagen. *Nat Prod Rep* 19:49–59
2. Prockop DJ, Kivirikko KI (1995) Collagens: molecular biology, diseases, and potentials for therapy. *Annu Rev Biochem* 64:403–434
3. Myllyharju J, Kivirikko KI (2001) Collagens and collagen-related diseases. *Ann Med* 33:7–21
4. Schofield CJ, Zhang Z (1999) Structural and mechanistic studies on 2-oxoglutarate-dependent oxygenases and related enzymes. *Curr Opin Struct Biol* 9:722–731
5. Hegg EL, Que L Jr (1997) The 2-His-1-carboxylate facial triad – an emerging structural motif in mononuclear non-heme iron(II) enzymes. *Eur J Biochem* 250:625–629
6. Hewitson KS, McNeill LA, Elkins JM et al (2003) The role of iron and 2-oxoglutarate oxygenases in signalling. *Biochem Soc Trans* 31:510–515
7. Pugh CW, Ratcliffe PJ (2003) Regulation of angiogenesis by hypoxia: role of the HIF system. *Nat Med* 9:677–684
8. Semenza GL (2002) HIF-1 and tumor progression: pathophysiology and therapeutics. *Trends Mol Med* 8:S62–67
9. Klinman JP (1996) Mechanisms Whereby Mononuclear Copper Proteins Functionalize Organic Substrates. *Chem Rev* 96:2541–2562
10. Prigge ST, Mains RE, Eipper BA et al (2000) New insights into copper monooxygenases and peptide amidation: structure, mechanism and function. *Cell Mol Life Sci* 57:1236–1259

Enzyme Linked Immuno-Sorbent Assay

►ELISA

Enzymes

Definition

Enzymes are complex proteins produced by living cells, which promote a specific biochemical reaction by acting as catalysts. Enzymes accelerate, enable or control chemical reactions in living systems without being consumed in the reactions. Enzymes are typically composed of a protein part (apoenzyme) which confers specificity, and a nonprotein part (coenzyme) which is necessary for activity.

► [Enzyme Catalyzed Post-translational Hydroxylation of Proteins](#)

► [Immunochemical Methods, Localization](#)

► [Proteases and Inhibitors](#)

► [Protein Databases](#)

► [Reverse Transcriptase](#)

Epidermal Growth Factor Receptor

Definition

Epidermal growth factor receptor (EGFR) refers to a transmembrane glycoprotein receptor of 170 kD (also called HER1/ErbB1) which belongs to the EGFR family of receptors, containing an intracellular tyrosine kinase domain. It can be activated by various ligands including epidermal growth factor (► [EGF](#)) and transforming growth factor- α (TGF α).

► [Breast Cancer](#)

► [Growth Factors](#)

Epidermis

Definition

The epidermis is the outermost, keratinizing layer of the skin and provides the first barrier of protection from the invasion of foreign substances into the body. The principal cell of the epidermis is the ► [keratinocyte](#). The epidermis is subdivided into the basal layer, the spinous layer, the granular layer and the cornified layer. Only keratinocytes in the basal layer are able to proliferate, subsequently they gradually migrate from the bottom to the surface, thereby losing their proliferation capacity. The dead cells are finally sloughed off. This process is called terminal differentiation.

► [Heritable Skin Disorders](#)

► [Wound Healing](#)

Epidermolysis Bullosa (Simplex)

Definition

Epidermolysis bullosa (EB) is a group of genetic skin disorders, in which the tissue is rendered fragile by the presence of mutations in one of a range of structural proteins that normally contribute to the resilience of the skin. Clinically, they all show splits or breakdown of the epidermis (epidermolysis) with fluid-filled blister formation (bullosa). They are distinguished microscopically by the plane of cleavage of the skin. Epidermolysis Bullosa simplex (EBS) shows an intra-epidermal split, and is caused by mutations in basal cell keratins. Junctional EB splits within the basal lamina due to mutations in laminin 5 genes or basal cell integrin genes. Dystrophic EB (the most serious type) shows a sub-epidermal split below the basement membrane, and is caused by mutations in type VII collagen.

► [Heritable Skin Disorders](#)

► [Intermediate Filaments](#)

Epidermolytic Hyperkeratosis

Definition

Epidermolytic hyperkeratosis is an autosomal dominant congenital ► [ichthyosis](#), characterized by skin blistering and hyperkeratosis (thickening of the epidermis).

► [Heritable Skin Disorders](#)

Epidermolytic Palmoplantar Keratoderma

Definition

Epidermolytic palmoplantar keratoderma is an autosomal dominant disorder with thickening of palms and soles with histological signs of ► [cytolysis](#).

► [Heritable Skin Disorders](#)

Epigenetic (Epigenetics)

Definition

Epigenetic describes any modification of gene activity during development of the organism from the fertilized

egg to the adult without directly affecting its DNA sequence. An epigenetic modification may indirectly influence the expression of the genome. Epigenetic changes can, for instance, be caused by DNA methylation, chromatin remodeling, the inactivation of one X chromosome in females, or by gene silencing processes (epigenetic silencing). Epigenetic changes that occur in somatic cells are difficult to distinguish from rare mutations at the genetic level.

“Epigenetics” aims to describe heritable changes in gene function that occur without a change in the sequence of the DNA.

- ▶ [Chromatin Acetylation](#)
- ▶ [Hereditary Diseases Genetic Basis](#)
- ▶ [Nucleosomes](#)
- ▶ [PNA Chips](#)
- ▶ [Prader Willi and Angelman Syndromes](#)
- ▶ [Transgene Silencing](#)

Epigenetic Determinants

Definition

Epigenetic determinants are factors that are inherited, but not coded for, in the DNA.

- ▶ [Centromeres](#)

Epigenetic Modification

Definition

Epigenetic modification denotes a covalent modification of the protein, or the DNA that composes the chromatin, but is independent of the DNA sequence. The known modifications include acetylation, methylation, phosphorylation, poly(ADP-ribosylation), and ubiquitination. The usual targets are histones, but DNA can also be modified by methylation. Epigenetic modifications influence the stability of the chromatin. These marks differentiate the active and inactive genes in the genome, and play a role in the transmission of the activity of the genes during cell division. Epigenetic marks are sometimes heritable even through meiosis.

- ▶ [Genomic Imprinting](#)

Epiglottis

Definition

Epiglottis describes a thin cartilaginous flap that covers the entrance to the larynx during swallowing, and prevents food from entering the trachea.

- ▶ [Bone and Cartilage](#)

Epiphysis

Definition

Epiphysis designates the ends of the long bones.

- ▶ [Bone and Cartilage](#)

Epistasis

Definition

Epistasis refers to the interaction of genes, where particular alleles of one gene interfere considerably with the phenotypic expression of other genes. A gene is said to be epistatic when its presence suppresses the effect of a gene at another locus.

- ▶ [C. Elegans as a Model Organism for Functional Genomics](#)
- ▶ [Epistasis in Cystic Fibrosis](#)
- ▶ [Genetic Epidemiology](#)
- ▶ [SLE Pathogenesis Genetic Dissection](#)

Epistasis Group

Definition

Epistasis group comprises of a group of genes that operate in the same pathway. Classification of genes in epistasis groups is based on the analysis of single and double mutant strains. If the phenotype (e.g. increased sensitivity to ionizing radiation) of the double mutant is the same as that of the single mutants, the two genes are assigned to the same epistasis group. When the double mutant shows additive or synergistic effects relative to both single mutants, the genes are placed in different epistasis groups.

- ▶ [Double-Strand Break Repair](#)

Epistasis in Cystic Fibrosis

FRANCESCO SALVATORE¹, GIUSEPPE CASTALDO^{1,2}

¹CEINGE – Biotecnologie Avanzate, scrl and Dipartimento di Biochimica e Biotecnologie Mediche, University of Napoli “Federico II”, Napoli, Italy

²Facoltà di Scienze Matematiche, Fisiche e Naturali, University of Molise, Isernia, Italy
salvator@unina.it

Synonyms

CF; Cystic fibrosis; cystic fibrosis of the pancreas; mucoviscidosis; Epistatic genes; interacting genes; modifier genes; modulator genes

Definition

Cystic Fibrosis

Cystic fibrosis (CF) is the most frequent lethal inherited disorder among Caucasians, having a ►prevalence of 1:2,500 newborn; it is inherited as an autosomal recessive trait. The prevalence of ►carriers in the population is about 1:25. The disease is due to alterations in the cystic fibrosis transmembrane regulator (►CFTR) protein (1). The protein is encoded by the *CFTR* gene that maps on chromosome 7q31.2, contains 27 ►exons and encompasses 230 ►kb. The heterogeneous clinical expression of the disease seems to be modulated by genes that could exert epistatic effects and that are inherited independently of the *CFTR* gene (1).

Epistasis

The term “epistasis” was coined by Bateson, about a century ago (2). When expressed as a specific allelic variant, an epistatic locus/gene “masks” the alleles of another locus/gene. Epistasis has recently been reviewed particularly as regards the statistical procedures used to evaluate its effect in humans (2). The clinical expression of various genetic diseases seems to be influenced by the epistatic interaction of genes; the most notorious example is inhibition of “sickling” in homozygotes for sickle cell anemia, exerted by the alpha thalassemia concomitant genotype (3). Epistasis is also involved in such other diseases as diabetes and immune deficiencies.

Characteristics

Modifier Genes of the CF Phenotype

More than 1,000 *CFTR* gene mutations have been detected in CF patients. About 85% of CF patients have ►pancreatic insufficiency; the pancreatic status of CF patients depends on the *CFTR* mutation and a set of

mutations specifically associated to pancreatic insufficiency has been identified (1). In contrast, there is no general consensus about correlations between the *CFTR* genotype and the pulmonary, liver and gastrointestinal expression (►meconium ileus) of the disease. A different clinical expression of the disease has been observed in CF patients bearing the same *CFTR* genotype and in patients who are homozygotes for nonsense mutations (1). Furthermore, a discordant ►phenotype has been observed in CF siblings (1, 4). These data suggest that environmental factors do not play a major role in CF and that epistatic genes other than *CFTR* modulate the CF phenotype (1). A locus that modulates gastrointestinal expression was identified in mice and subsequently in humans. We suggested that this locus had a dominant effect in 9 CF patients discordant for meconium ileus (1). Among the putative epistatic genes of CF pulmonary expression, genes encoding proteins of the surfactant layer of the respiratory apparatus may play a relevant role. Antimicrobial proteins, e.g., human beta defensins 1-2 (hBD1 and 2), lactoferrin, lysozyme, histatin and cathelicidin are salt-sensitive antibacterial peptides of the airway layer that are said to belong to the innate immune system. Thus, the severity of lung expression in CF and the susceptibility of these patients to opportunistic infections could depend on the levels of sodium and chloride in the airway layer. However, findings about sodium and chloride levels at the airway surface of CF patients are discordant and other studies, possibly with more standardized techniques, are necessary to resolve this issue. The beta-defensins (hBD) 1-4 have been proposed as modulators of airway infections in CF patients. These small salt-sensitive peptides are expressed at normal levels in sputum from CF patients, but hBD2 was found to be enhanced in the bronchoalveolar fluid of CF patients. This suggests that hBD2 expression can be induced by inflammation, whereas hBD1 appears to be involved in the innate defense system and is expressed independently of bacterial stimulation. Beta-defensin antibacterial activity depends on the levels of salt in the medium. Thus, it is possible that these proteins are normally secreted but their activity is reduced in CF patients. In a collaborative study led by Dr. J.J. Cassiman in Leuven, we recently demonstrated (unpublished data) a higher rate of ►polymorphisms in *hBD1*, 2 and 4 genes in both CF patients and normal subjects as compared with most other human genes. We are now examining these polymorphisms in relation to the biological activity of beta-defensins and to the pulmonary phenotype of CF patients. More recently, a study of 34 CF sibling pairs (4) with a discordant or concordant clinical expression revealed a novel locus at the 3' end of the *CFTR* gene that could modulate the CF phenotype (stature, food intake and energy homeostasis). Genes involved in local

innate and adaptive immunological defense and inflammation could also act as epistatic genes in CF patients. A severe pulmonary expression was observed in CF patients bearing anti *P. aeruginosa* IgG3 polymorphisms associated with high protein expression and in CF patients with a polymorphic allele in the promoter region associated with enhanced production of the pro-inflammatory cytokine TNF-alpha. The latter allele is also associated with an increased risk of chronic bronchitis in non-CF patients. Pulmonary oxidative stress plays a relevant role in the pathogenesis of lung damage in CF patients. *Glutathione transferase* is involved in detoxifying hydroperoxides resulting from oxidative stress and gene polymorphisms associated with low levels of the enzyme thereby give rise to a more severe pulmonary phenotype in CF patients. A more severe pulmonary expression was reported in CF patients who expressed a polymorphic variant with low levels of mannose binding protein, a lectin involved in bacterial [opsonization](#) and [phagocytosis](#).

A body of evidence suggests that epistatic genes also influence the risk of liver expression in CF patients. Mutations of the *mannose binding lectin* gene are associated with a higher risk of severe liver disease in CF patients. In a multicenter study led by Dr. K. Friedman (1), we found that Z or S mutations of the *alpha-1 antitrypsin (A1AT)* gene confer a 3- to 7-fold higher risk of severe liver disease in CF patients.

Alpha-1 antitrypsin has been implicated in the pulmonary expression in CF patients. A better pulmonary function was reported in CF patients bearing the S and Z alleles of the *alpha-1 antitrypsin* gene, but the study was biased because it did not include CF patients who had died in early childhood from severe pulmonary disease (1). The same group of researchers retrospectively analyzed DNA from blood spots of CF patients who had died in early childhood from severe pulmonary disease and confirmed that A1AT deficiency is not associated with a more severe pulmonary disease (1). Finally, a less severe pulmonary phenotype was described in CF patients bearing polymorphisms of the A1AT enhancer. These data contrasted with the higher risk of opportunistic colonization reported in CF patients carrying mutations of the *A1AT* gene, although the pulmonary expression of the disease was not significantly different. Thus, multicenter studies of large numbers of CF patients are required to establish the role of *A1AT* as a modulator of the lung phenotype.

Cellular and Molecular Regulation

CFTR Gene and Mutations

Protein CFTR is an anion channel protein belonging to the [ATP-binding cassette \(ABC\) transporter](#) family, and is widely expressed at the apical side of cells bearing polarized [epithelia](#) involved in secretory/absorption

activities. The anion channel acts as a regulator of chloride levels across the cell membrane and may be involved in the regulation of other membrane conductance pathways. CFTR is activated by beta-2 adrenoreceptor stimulation followed by cAMP activation. *CFTR* mutations are grouped into four classes based on the molecular alteration at the protein level. Class 1 mutations (i.e., nonsense mutations, frame-shift mutations due to deletions or insertions and mutations that cause altered splicing) typically alter protein synthesis. The *CFTR* W1282X, R553X and G542X class 1 mutations each have a frequency of >2.0% among CF chromosomes within most populations studied. Class 2 mutations (including DF508, the most frequent *CFTR* mutation) alter *CFTR* protein processing. The protein is mostly degraded in the endoplasmic reticulum; it does not reach the epithelial membrane to express its normal function. Class 3 mutations include G551D, which is frequent in Northern Europe. These mutations affect the regulatory domains of the *CFTR* protein. Class 4 mutations (usually missense mutations) alter the ionic conductivity of the chloride channel and so reduce the time of channel opening and ionic flux. In summary, mutations of different classes produce different effects on the *CFTR* protein: class 1 and 2 mutations have a more dramatic effect because they are related to the absence or to an altered localization of the *CFTR* protein, whereas class 3 and 4 mutations are associated with residual *CFTR* activity. However, as previously indicated, the genotype-phenotype correlation in CF patients is not clear and epistatic genes are probably involved in modulation of the clinical expression of CF. In any event, the altered protein usually causes dense mucous epithelial secretions and thus a clinical expression that typically involves the gastrointestinal tract with the pancreas and liver, the pulmonary apparatus and the reproductive tract.

CF Polymorphisms

Several *CFTR* gene polymorphisms have been related to reduced synthesis of the channel protein. A case in point is IVS8 poly T, in which the 5T allele gives rise to alternative splicing of the mRNA. The 5T allele is more frequently observed in patients bearing atypical CF forms, where the prevalent clinical expression concerns the reproductive system and in patients with only chronic pancreatic insufficiency or symptoms not related to the main pulmonary or gastrointestinal system. In addition, third-base polymorphisms (not producing changes in the amino acid) within the coding regions of the *CFTR* gene could derange the splicing and thus reduce the levels of the *CFTR* protein (5).

Clinical Relevance

Clinical cases of CF range from death in early infancy from meconium ileus or respiratory complications to a mild form of the disease where patients live into their

40s. Pulmonary, liver and gastrointestinal expression (meconium ileus) can be observed in CF patients. Respiratory expression is responsible for about 95% of CF mortality, but the severity of pulmonary expression varies among patients. Pulmonary expression is typically due to mucous obstruction and recurrent infections resulting from the altered mucous composition and the consequent reduced mucociliary clearance. Several opportunistic bacteria gradually damage the lungs, essentially by stimulating the release of proteinases. Pulmonary function tests, i.e. forced expiratory volume (FEV) and forced vital capacity (FVC), are used to monitor pulmonary function in CF patients and to classify the severity of CF expression. The prevalence of liver involvement in CF patients ranges from 4.0% to 25–30%. The pathogenesis of liver expression is not clear, but intrahepatic biliary ducts are probably obstructed by inspissated bile, gradually leading to hepatic fibrosis and focal biliary cirrhosis. Multilobular cirrhosis with portal hypertension occurs in only 2–7% of CF patients. Biochemical markers have little predictive value for the development of severe liver disease in CF patients and specific CFTR mutations have not been linked to this outcome. Consequently, it is not yet possible to predict which CF patients will develop severe liver disease. Meconium ileus is present in 15%–20% of CF patients. It could be due to the dehydration of fetal intestine. The distal intestine obstruction syndrome (DIOS), present in about 5% of CF patients, is considered to be an adult equivalent of meconium ileus. Attempts have been made to relate the occurrence of meconium ileus or DIOS to the CFTR genotype but the question is still open. The data reported herein suggest that a number of epistatic genes contribute to produce the highly varied CF phenotype. The study of these genes could have an impact on the prognosis and care of CF patients, i.e. patients at high risk for pulmonary or liver expression could be monitored and treated with specific therapies before irreversible damage is produced. Multicenter studies on large numbers of cases with homologous genotypes and well-defined phenotypes will give more reproducible and definitive data.

References

1. Salvatore F, Scudiero O, Castaldo G (2002) Genotype-phenotype correlation in cystic fibrosis: the role of modifier genes. *Am J Med Genet* 111:88–95
2. Cordell HJ (2002) Epistasis: what it means, what it doesn't mean, and statistical methods to detect it in humans. *Hum Mol Genet* 11:2463–2468
3. Higgs DR, Aldridge BE, Lamb J et al (1982) The interaction of alpha-thalassemia and homozygous sickle-cell disease. *N Engl J Med* 306:1441–1446

4. Mekus F, laabs U, Veeze H et al (2003) Genes in the vicinity of CFTR modulate the cystic fibrosis phenotype in highly concordant or discordant F508del homozygous sib pair. *Hum Genet* 112:1–11
5. Pagani F, Stuani C, Tzetis M et al (2003) New type of disease causing mutations: the example of the composite exonic regulatory elements of splicing in CFTR exon 12. *Hum Mol Genet* 15:1111–1120

Epithelia

Definition

Epithelia are the sheet tissues of the body surrounding organs and lining body cavities. They are single layered (simple epithelia) or multilayered (stratified) and have no intraepithelial blood vessels. Epithelia are clearly polarised, bounded on one side by an acellular space, and on the other side by a specialised carpet of ►[extracellular matrix](#) that defines their orientation. By definition, they occur at interfaces between two dissimilar environments, and the cells in epithelia are usually steadily turned over. Cells are held together by ►[adherens junctions](#) and ►[desmosomes](#), which are connected across each cell's cytoplasm by actin and keratin filaments.

►[Adherens Junctions](#)

►[Cell Polarity](#)

►[Desmosomes](#)

►[Epithelial Cells](#)

►[Gut Epithelium](#)

►[Intermediate Filaments](#)

►[Lung](#)

►[Microvilli](#)

Epithelial Cells

APARNA LAKKARAJU,

ENRIQUE RODRIGUEZ-BOULAN

Margaret M. Dyson Vision Research Institute and Departments of Cell Biology and Ophthalmology, Weill Medical College of Cornell University, New York, NY, USA

boulan@med.cornell.edu

Definition

An epithelium is a layer of cells that lines body cavities and tubular structures and forms a physical barrier between the organism and the external world. Epithelial cells are tightly packed into two-dimensional sheets, which rest on basement membranes. A major subdivision is into stratified (e.g. skin) and simple epithelia. A distinguishing feature of simple epithelial cells is the polarized phenotype, i.e. the apical or luminal surface is structurally and functionally distinct from the basolateral or serosal side. By virtue of their specialized intercellular junctions, epithelial cells form selective permeability barriers between tissue compartments and maintain cellular homeostasis. Epithelia play a crucial role in tissue and organ morphogenesis during embryonic development. Based on their predominant functions, simple epithelial cells can be classified into absorptive (kidney tubules and intestine), exchange (lung and capillary endothelium), secretory (hepatocytes and cells of the thyroid gland and pancreas) or protective (skin). The rest of this article will deal with simple transporting epithelia.

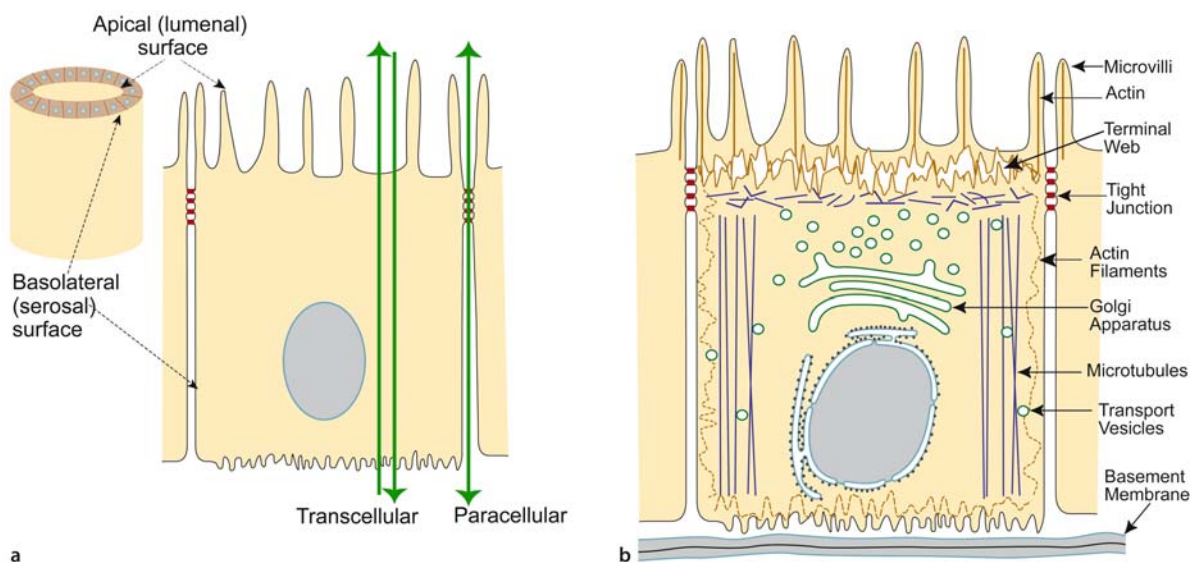
Characteristics

Polarized epithelial cells are characterized by a) specialized intercellular junctions, b) cohesive interactions between the cells and the basement membrane,

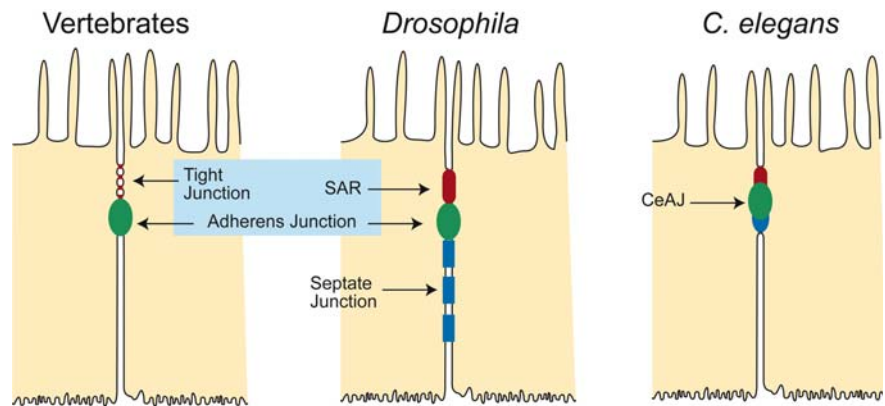
c) asymmetric distribution of cytoskeletal elements and d) differential sorting of proteins to the apical and basolateral plasma membranes. Figure 1 depicts a simple transporting epithelium and the organization of a polarized epithelial cell.

Intercellular Junctions

Several types of junctions mediate cell-cell interactions and contribute to the barrier function characteristic of epithelia (Fig. 2) (1). All simple epithelial cells have an adhesive belt called the zonula adherens that encircles the cell just below the apical surface and forms the ▶adherens junction (AJ). In vertebrates, tight junctions (TJ) or zonula occludens (ZO) form apical to the AJ and mark the morphological and functional boundary between the apical and basolateral membrane domains. Adherens junctions and desmosomes, which lie below the AJ, are thought to provide mechanical strength to the cells. ▶Gap junctions are formed by ▶connexins, which form hemichannels that link the cytoplasm of adjacent cells and allow for intercellular communication. In *Drosophila*, epithelial cells have septate junctions basal to the AJ, which form membrane contacts that extend over a large area of the lateral plasma membrane. Although TJs are absent in *Drosophila*, the sub-apical region, which is apical to the AJ, has protein complexes that colocalize with vertebrate TJs



Epithelial Cells. Figure 1 (a). A simple tubular epithelium showing the apical and basolateral surfaces demarcated by tight junctions (TJs) which create a physical barrier to the movement of macromolecules. Asymmetric protein distribution in the apical and basolateral surfaces enables intracellular vectorial solute transport. While transcellular transport is vectorial, energy-dependent and relies on cell-specific transporters and channels, paracellular transport occurs by passive diffusion down a concentration gradient and does not exhibit directionality. The relative contribution of each route to total transepithelial flux depends on the physical characteristics of the solute because TJs exhibit charge and size selectivity. (b) The organization of an epithelial cell depicting the main structural features that contribute to the polarized phenotype. See Section II for details.



Epithelial Cells. Figure 2 Junctional complexes in epithelial cells. The vertebrate epithelial cell contains the tight junctions (TJ) and the subjacent adherens junction (AJ) that form the apical junctional complex (blue box). Although *Drosophila* epithelial cells do not have TJs, many proteins in the sub-apical region (SAR) colocalize with TJs. The septate junction is present basal to the AJ in *Drosophila*. The epithelial cells of *C. elegans* have a single junction, the *C. elegans* apical junction (CeAJ). See Table 1 for the molecular composition of these junctions.

(see below). In the nematode *Caenorhabditis elegans*, the *C. elegans* apical junction (CeAJ) resembles the AJ of *Drosophila* and vertebrates (Fig. 2). In vertebrates, TJs form a selective permeability barrier that restricts the paracellular movement of molecules and water (gate function). Tight junction proteins include the transmembrane occludin and claudins and peripheral scaffolding proteins like ZO-1 (Table 1). As seen in electron micrographs, these proteins appear as parallel strands that form an anastomosing network

around the apical pole between adjacent cells and form the physical basis for the diffusion barrier. A qualitative estimate of a junction’s gate function is provided by the average number of strands along the apical-basal axis; a precise quantitative estimate of this gate function is provided by the [trans-epithelial electrical resistance](#) (TER). In addition to the gate function, TJs also form an intramembranous fence that prevents the intermixing of apical and basolateral plasma membrane proteins and lipids, which helps to maintain cell surface polarity

Epithelial Cells. Table 1 Components of junctional complexes in vertebrates and invertebrates

Vertebrate		<i>Drosophila</i>		<i>C. elegans</i>	
Trans-membrane	Scaffolding	Trans-membrane	Scaffolding	Trans-membrane	Scaffolding
TJ CRB1 occludin claudins 1-24 JAM	ZO 1-3 Par3 (Bazooka) Par6 aPKC Cdc42 ERMs Stardust (Pals1) Discs lost (PATJ)	SAR Crumbs	Par3 (Bazooka) Par6 aPKC Cdc42 Stardust Discs lost DMoesin	CeAJ CRB1 E-cadherin	α -catenin β -catenin Par3 Par6 PKC-3
AJ E-cadherin	α -catenin β -catenin	AJ E-cadherin (shotgun)	α -catenin β -catenin (armadillo)		
		SJ Neurexin IV Fasciclin III	Scribble Discs large Lgl		

(fence function). As the TJ barrier is physiologically regulated, its disruption may contribute to disease pathogenesis (see below).

Cell Adhesion

Epithelial cells interact with one another *via* ►cadherin-mediated cell-cell adhesion and with the ►extracellular matrix (ECM) *via* ►integrin-mediated cell-matrix adhesion on the basal side of the cells. Accumulation of an endogenously secreted ECM generates an initial level of cellular asymmetry and orients the apicobasal axis of polarity that is necessary for vectorial transport of solutes and ions. Contacts between cells or between a cell and the ECM trigger the compartmentalization of proteins into contacting (basolateral) and free (apical) surface domains (2).

All adhesion complexes share a common architecture, a set of transmembrane proteins that interact with extracellular components (ECM or neighboring cells) on the outside and are anchored to other proteins or the cytoskeleton on the inside. Extracellular domains of the epithelial cell-specific E-cadherin mediate Ca^{2+} -dependent homophilic interactions between cells, mainly at the AJs. Cadherins also have two cytoplasmic domains, one of which binds β -catenin and the other p120 catenin, a regulator of cadherin clustering. β -catenin anchors cadherin to the actin cytoskeleton *via* α -catenin. Another actin-binding protein, vinculin, also helps in anchoring cadherins to the actin cytoskeleton and organizes cadherin into circumferential AJs. Cell-cell adhesion mediated by E-cadherin plays an important role in specifying membrane asymmetry and is required to restrict the localization of apical and basolateral proteins to their respective domains.

Interactions between epithelial cells and the ECM are mediated by the integrin superfamily of adhesion receptors. These receptors, which consist of an α subunit non-covalently linked to a β subunit, bind all major constituents of the ECM including collagens, laminins and fibronectin. Integrins serve as transmembrane linkers, mediating interactions between the ECM and the ►actin cytoskeleton with the help of proteins such as α -actinin, talin, tensin and filamin. These linkages lead to a clustering of integrins and the formation of focal adhesions. Integrin binding to the ECM activates intracellular signaling pathways, which inform the cell about the character of the underlying matrix, and influences intracellular gene expression. Integrin-mediated signaling pathways mainly involve focal adhesion kinase (FAK) and integrin-linked kinase (ILK). Activation of FAK triggers intracellular signaling cascades such as the extracellular signal-regulated kinase (ERK) pathway, which regulates cell

proliferation and survival. Activation of ILK activates transcription factors LEF/TCF and AP-1.

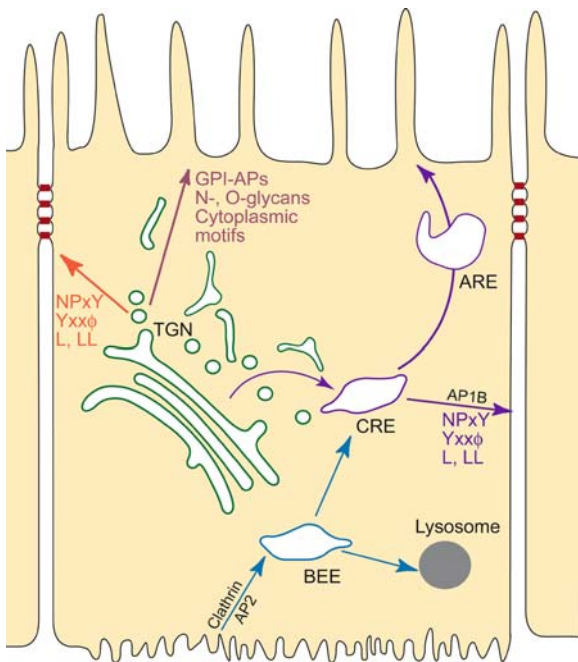
►Cytoskeleton

As stated above, the actin cytoskeleton associates with both cadherin and integrin adhesion receptors. These interactions lead to localized cytoskeletal assemblies that strengthen cell-cell and cell-ECM contacts and serve as scaffolds for the recruitment of signaling molecules (see below). From the standpoint of polarity, these localized cytoskeletal assemblies result in the immobilization of different proteins at the basal and lateral domains, generating subcompartments within the basolateral membrane without a definite barrier in between them.

Distinct cytoskeletal elements function in the structural and functional organization of the apical and basolateral membrane domains in epithelial cells. The apical plasma membrane has numerous extensions, or microvilli ($\sim 1 \mu\text{m}$ long and $0.1 \mu\text{m}$ in diameter), that protrude into the extracellular milieu and create an increased surface area for secretion and absorption. At the core of each microvillus lies a bundle of parallel actin filaments cross-linked by the actin-bundling proteins villin and fimbrin. These actin filaments extend to the base of the cell and are embedded in a terminal web of fodrin and actin, which in turn is linked laterally to the zonula adherens by proteins such as α -actinin and vinculin. At the basolateral membrane, the principal cytoskeletal protein fodrin, actin and other associated proteins form a protein lattice on the cytoplasmic face of the membrane. Fodrin also binds ankyrin which binds with high-affinity to integral membrane proteins such as Na^+ , K^+ -ATPase and $\text{Cl}^-/\text{HCO}_3^-$ exchanger. This fodrin-based lattice may participate in the establishment of the basolateral membrane by directing the retention and accumulation of proteins that have an affinity for the fodrin lattice. Epithelial monolayers participate in organ morphogenesis through junctional complexes and the contractile cytoskeleton that is formed by actin filaments attached to AJs. Contraction of these filaments draws the apical surfaces of the connected cells together and causes an invagination of the epithelium at that point. Non-contractile cytoskeletal elements such as the intermediate filaments form an intracellular scaffold from one side of the cell to the other at desmosomes. This scaffold helps maintain the structural integrity of the epithelium against shear stress. Microtubules in epithelial cells are organized into a subapical web and longitudinal bundles to the base of the cells. These microtubules participate in the spatial organization of cytoplasmic organelles and serve as tracks in the vectorial transport of vesicles to and from the plasma membrane.

Polarized Membrane Traffic

Newly synthesized membrane proteins are synthesized in the endoplasmic reticulum and transported through the Golgi apparatus to the cell surface. Some plasma membrane proteins, particularly on the basolateral surface, are quickly internalized into endosomes and recycled back to the original cell surface (Fig. 3). Other proteins undergo transcytosis: after internalization into endosomes, the protein is not recycled back to its original surface, but is delivered to the opposite domain. The biosynthetic, recycling and transcytotic routes of plasma membrane proteins are controlled by



Epithelial Cells. Figure 3 Sorting pathways in epithelial cells. From the trans-Golgi network (TGN), proteins can be directly transported to either the apical (pink) or basolateral (orange) domains based on their sorting signals i.e., “selective delivery”. Alternatively, cargo can be first sent to the common recycling endosome (CRE, purple) from where the proteins are differentially sorted. At the basolateral plasma membrane, proteins are either “selectively retained” or re-endocytosed and transported to basolateral early endosomes (BEE, blue). These proteins can be recycled back to the basolateral surface, sent to lysosomes for degradation or transcytosed to the opposite plasma membrane. Transcytosing proteins and recycling apical membrane proteins are ferried from the CRE to the apical recycling endosome (ARE). Clathrin adaptor proteins such as AP1B and AP2 decode sorting signals and function in the polarized trafficking of proteins. For a detailed explanation of these pathways and other intermediate organelles that might be involved, please see Ref.5.

specific sorting mechanisms. These involve the recognition of specific “zip codes” (sorting signals) in the transported proteins by specialized “sorters” that mediate the incorporation of the protein into carrier vesicles (see below). All epithelial cells use biosynthetic sorting, recycling *via* ►endosomes and selective delivery/transcytosis to sort proteins, but the relative contribution of each pathway varies among different epithelia. For example, in hepatocytes, most proteins are delivered to the basolateral membrane, after which a subset is selectively retained there, while others undergo transcytosis to the apical membrane. Some proteins may be additionally stabilized at specific membrane domains by interactions with scaffolding proteins, which often contain specialized interacting domains such as ►PDZ domains.

Basolateral proteins are sorted into post-Golgi transport vesicles based on specific motifs in their cytoplasmic domains that are thought to interact with adaptor coat proteins (Fig. 3). In contrast, apical sorting signals are less well understood and seem to be present in either the luminal domain or the membrane anchor. ►Glycosylphosphatidylinositol (GPI) anchors, N- or O-glycosylation or clustering of proteins into ►lipid rafts have been reported to confer apical targeting on some proteins, although exceptions exist in all cases. Transport vesicles are delivered from the trans-Golgi network (TGN) to the relevant plasma membrane with high fidelity. Microtubule motors, the actin cytoskeleton and specification of vesicle docking and fusion with the correct membrane all contribute to the accuracy of polarized protein sorting. The molecular nature of the sorting signals and vesicle targeting and fusion machinery are discussed below.

Polarized membrane traffic also underlies the role of epithelia as active barriers. Epithelial cells modify their environment by secretion, ion and nutrient transport and transcytosis of macromolecules across the cell. Secretion can occur at either the apical or basolateral surface and can be constitutive (production of ECM components) or regulated. Ion transport, which is essential for maintaining cell volume and osmotic balance, occurs with the help of transporters and channels. Transcytosis of macromolecules such as immunoglobulins from one biological compartment to another aids in host defense against infectious agents (see below).

Specialized Epithelia

Tissue-specific modifications of polarized epithelial cells produce a wide variety of specialized cell types. An example of this is the retinal pigment epithelium (RPE), which lies beneath the photoreceptors and forms the outermost layer of the retina, i.e. the blood-retinal barrier. Unlike most other epithelial cells, the microvilli on the apical surface of the RPE are not free, but

are interdigitated with photoreceptor outer segments. The RPE supports photoreceptors by processing retinoids, supplying them with nutrients and transporting metabolites out of the retina. Cells of the RPE participate in the circadian phagocytosis of shed outer segment tips of photoreceptors, thus aiding photoreceptor regeneration. Indeed, these cells are the most phagocytically active cells in the body. Dysfunction of the RPE plays a principal role in the pathogenesis of age-related macular degeneration.

Another example of specialized epithelia is the choroid plexus epithelium (CPE), part of the choroid plexus which lies within the ventricles of the [▶brain](#). These cells have tightly packed microvilli at the apical surface and TJs that prevent the paracellular transport of solutes between the blood and cerebrospinal fluid (CSF) i.e., the blood-CSF barrier. In addition to supplying the brain with hormones and nutrients, the CPE is partly responsible for secreting the CSF, which bathes the central nervous system and provides protective mechanical cushioning for the brain through its buoyant action. Polarized secretion of CSF occurs as a result of unidirectional flux of Cl^- and Na^+ , which in turn creates an osmotic gradient that drives the secretion of water. Aquaporin 1, which mediates water transport across cell membranes, is present on the apical surface of the CPE. Impaired choroidal transport may contribute to the pathogenesis of Alzheimer's disease. In contrast to most epithelia where Na^+ , K^+ -ATPase is on the basolateral membrane, both the RPE and CPE express this pump exclusively on their apical membranes. This "reverse" polarity is necessary for maintaining a high Na^+ environment required for photoreceptor function and for Na^+ secretion into and K^+ absorption from the CSF.

Hepatocytes are another type of specialized epithelial cell with a different polarity axis; the canalicular membrane domain, which corresponds to the apical surface in epithelial cells, is enclosed by adjacent cells while the sinusoidal and lateral domains correspond to the basolateral surface of epithelial cells. The apical membranes of hepatocytes line the bile canalicular tract. The sinusoidal membranes face the blood circulation. The sinusoidal and canalicular membrane domains are separated by tight junctions and exhibit distinct lipid and protein polarity as seen in other epithelia. This polarization is necessary for the diverse functions performed by hepatocytes, such as canalicular bile secretion and sinusoidal secretion of serum proteins into the blood. Hepatocellular polarity is disturbed in diseases like cholestasis.

Molecular Interactions

The generation of epithelial polarity in a multicellular organism requires highly coordinated molecular

interactions (3, 4). These interactions function at two focal points: one, the apical junctional complex that initiates intercellular contacts and regulates the identity of membrane domains and two, the polarized sorting and targeting of proteins.

The Apical Junctional Complex

The AJC, which is conserved between *C. elegans*, *Drosophila* and vertebrates, refers to the multi-protein assemblies that make up the adherens and tight junctions. The AJC is a multifunctional apparatus that provides strong adhesion between adjacent cells and forms the diffusion barrier characteristic of epithelia (Fig. 2 and Table 1). In addition, signaling networks that control epithelial cell proliferation, polarity and differentiation converge at the AJC. The constituents of the AJC share a basic structural motif; an integral membrane protein binds intracellular scaffold proteins that in turn bind the actin cytoskeleton and interact with other complexes. Genetic analyses in *Drosophila* and *C. elegans* have identified proteins in AJC sub-complexes that are crucial for epithelial cell polarity: cadherins/catenins, Crumbs (Crb)-Stardust (Sdt)-Patj, Bazooka (Baz/Par 3)-Par 6-atypical protein kinase C (aPKC) and Scribble (Scrib)-Lethal giant larvae (Lgl)-Discs large (Dlg). These protein sub-complexes are functionally integrated into a hierarchical model that regulates the specification of apical and basolateral membrane domains. The Baz-Par 6-aPKC complex initiates the formation of the apical membrane, whose lateral spread is controlled by the Scrib-Lgl-Dlg complex. The Crb-Sdt-Dlg complex is apical to the Baz complex and helps maintain the apical membrane identity. The Scrib-Lgl-Dlg complex establishes the identity of the lateral membrane and is antagonized by the Crb complex. *Drosophila* embryos that lack Crb or Sdt fail to establish a continuous AJ and exhibit lethal defects in cell polarity and tissue organization. Loss of Bazooka or aPKC results in an inability to establish polarity and a failure to form AJs. Defects in the Scrib-Lgl-Dlg complex cause an expansion of the apical membrane into more lateral positions and lead to a multilayered epithelium. In vertebrate TJs, the transmembrane proteins occludin, claudins and the junctional adhesion molecule (JAM) interact with the cytoplasmic PDZ domain-containing proteins ZO-1, ZO-2, ZO-3 and Par3, which serve as scaffolds for other cytoskeletal proteins and signaling molecules.

The [▶Rho-family small GTPases](#) Cdc42 and Rac1 link actin to components of the AJC and regulate actin polymerization with the help of the Arp2/3 complex and other proteins like N-WASP. Cdc42 contributes to microtubule polarization by recruiting IQGAP1, which binds the microtubule tip protein CLIP170 and

promotes microtubule capture. IQGAP1 interferes with homophilic cadherin interactions and reduces epithelial cell adhesion. Cdc42 and Rac1 counter this effect of IQGAP1 by inhibiting the latter's interactions with E-cadherin and β -catenin.

Ezrin, radixin and moesin (ERM) are a family of membrane-associated proteins that link the actin cytoskeleton to proteins on the cell surface. Ezrin is predominantly found in the apical microvilli of polarized epithelial cells while radixin is found in the microvilli of bile canaliculi. By virtue of their interactions with F-actin, ERM proteins participate in maintaining epithelial polarity, membrane transport and signal transduction. In *Drosophila*, moesin localizes with Crumbs and moesin mutants exhibit loss of polarity and disassembled junctions. Recent studies in mice show that ezrin is essential for maintaining the apical terminal web, which anchors the microvilli and stabilizes intercellular junctions. In the absence of ezrin, microvilli are irregular and the apical junctions are elongated.

Polarized Protein Sorting and Transport

As stated previously, short sequences in the cytoplasmic or luminal domains of proteins are thought to dictate their sorting into vesicles destined for transport to the basolateral or apical membrane from the TGN (Fig. 3). Basolateral sorting signals usually contain a crucial tyrosine residue within the sequence NPxY or Yxx ϕ (where x is any amino acid and ϕ is an amino acid with a bulky hydrophobic side chain). Non-tyrosine-based (mono- and dileucine) and dihydrophobic sorting signals have also been identified for some basolaterally targeted proteins. The Yxx ϕ motif and sorting signals with a pair of leucine residues preceded by an acidic cluster bind to tetrameric clathrin adaptor proteins such as AP-1 at the TGN or endosomes. A second type of dileucine signal binds another family of clathrin adaptors called the GGAs (Golgi-localized, γ -ear-containing, ARF-binding proteins).

Once the apical and basolateral membrane proteins are sorted into distinct subdomains, they are packaged into carrier vesicles that bud from tubular extensions of the TGN. Polarized budding of vesicles from the TGN is thought to be regulated by Cdc42. Fidelity of vesicle docking and fusion with the correct membrane domain is ensured by the [▶SNARE complex](#) where cognate vesicle v-SNAREs bind target membrane t-SNAREs. Endogenous t-SNAREs have polarized distributions in epithelial cells; syntaxin 3 is apical while syntaxin 4 is predominantly basolateral. Recent studies in [▶Madin-Darby canine kidney \(MDCK\) cells](#) suggest that a multiprotein complex consisting of mammalian homologs of the yeast exocyst (sec6/8) complex may function in targeting vesicles to the basal-lateral

membrane. During polarization, following cadherin-mediated adhesion, the exocyst is recruited to the apex of the lateral membrane, i.e., the region of basolateral vesicle docking and fusion. The actin cytoskeleton is important for organizing vesicle delivery to the basolateral membrane, while microtubules participate in long-range vesicle transport ([▶Vesicular Traffic](#)) and organize vesicle delivery to the apical membrane (5, 6).

Regulatory Mechanisms

Epithelial Polarity and Neoplasia

Approximately 95% of lethal human carcinomas are epithelial in origin and are characterized by loss of cell polarity. Cells at the leading edge of these tumors undergo an [▶epithelial-to-mesenchymal transition \(EMT\)](#), which makes them less adhesive and more migratory. Structural alterations in tight junctions have been demonstrated in ovarian, liver, colon and breast cancers. Tumor-promoting factors such as tissue plasminogen activator cause leakiness in the TJs and allow the passage of many solutes. Thus, regulation of epithelial cell polarity is tightly coupled to proliferation and invasion. In *Drosophila*, *discs large*, *scribble* and *lethal giant larvae* have been identified as tumor-suppressor genes; *dlg*, *scrib* or *lgl* homozygous mutants exhibit uncontrolled proliferation. Interestingly, human Dlg and Scrib are targeted for proteasomal degradation by oncogenic viruses such as the human papilloma virus, which causes cervical cancer.

Pathogen Entry and Host Defense

Epithelial cells lining the respiratory, gastrointestinal and urogenital tracts are continually challenged by bacteria, viruses and toxins. While tight junctions and the glycocalyx that covers the apical surface provide a barrier to invasion, bacteria and other pathogens have evolved numerous mechanisms to circumvent the epithelial barrier and colonize the host tissue (7). Bacteria such as *Vibrio cholerae* secrete enzymes that break down one or more transmembrane proteins that form the AJC, resulting in loss of barrier function. Others like pathogenic strains of *E. coli* synthesize toxins that activate Rho GTPases Cdc42 and Rac1, leading to cytoskeletal rearrangement and internalization of TJ proteins JAM, ZO-1 and occludin. *Helicobacter pylori* translocates an effector protein CagA into the host cell that forms a complex with JAM and ZO-1 and disrupts the AJC. The diffusion barrier is compromised, resulting in leakage of solutes from the interstitium to the apical surface where the bacteria colonize. For pathogens like *Shigella flexneri* and *Pseudomonas aeruginosa*, loss of epithelial cell polarity is essential for infection.

Epithelial Cells. Table 2 Pathological manifestations of altered trafficking and polarity of membrane proteins

Disease	Defect	Consequences
Cystic Fibrosis	Loss of C-terminal CFTR PDZ-binding domains (apical sorting determinant)	Missorting of CFTR protein; Impaired chloride secretion and lethal alterations in lung epithelial function.
Wilson disease (Copper toxicity)	Mutations prevent sorting of P-type ATPase ATP7B into apically targeted vesicles.	Mislocalization of ATP7B which mediates copper clearance in renal epithelia.
Familial hypercholesterolemia	Mutations affect proper targeting and internalization of the LDL receptor.	LDLr is mistargeted to the canalicular (apical) membrane of hepatocytes instead of the sinusoidal membrane. FH predisposes patients to heart failure and early death.
Nephrogenic diabetes insipidus	Mutations in aquaporin-2 that redirect it to lysosomes for degradation.	Polydipsia, polyuria, renal insufficiency.
Liddle's syndrome	Prolonged stabilization of epithelial Na ⁺ channel on apical membrane caused by a loss of the motif that tags the protein for rapid endocytosis and ubiquitination.	Severe salt-sensitive hypertension; excessive Na ⁺ reabsorption by kidney epithelia.
Polycystic kidney disease (autosomal dominant)	Defective basolateral targeting due to depletion of the exocyst (Sec6/8) complex; aberrant signaling caused by mutant polycystin-1.	Dedifferentiation of kidney epithelia; formation of large fluid-filled cysts in the kidney; loss of renal function.

Epithelia have non-specific mechanisms for host defense including mucus secretions, gastric acid and lysozyme that can neutralize pathogens. Specific defense mechanisms involve immunoglobulins IgA, IgG and IgM, which are secreted by local B-lymphocytes underlying the epithelium and are actively transported across the cell (transcytosis) by specific receptors. They are then secreted into the lumen where they act specifically to prevent infection. IgA and IgM bind the polymeric immunoglobulin receptor on the basolateral membrane of the epithelial cells and the ligand-receptor complex is then endocytosed and transported through a variety of intracellular compartments to the apical surface. The secreted immunoglobulin prevents the interactions of pathogens with the epithelial cell surface by interfering with their motility or by competing with the pathogens for binding sites on the cells (8).

Sorting Defects in Disease

Regulatory pathways that maintain the epithelial phenotype are crucial for the normal functioning of organs such as the liver, intestine and kidneys. Defects in these pathways can lead to mistargeting of proteins

and a loss of polarity and consequently, manifestation of disease (Table 2).

►Gut Epithelium

References

1. Knust E, Bossinger O (2002) Composition and formation of intercellular junctions in epithelial cells. *Science* 298:1955–1959
2. Schock F, Perrimon N (2002) Molecular mechanisms of epithelial morphogenesis. *Annu Rev Cell Dev Biol* 18:463–493
3. Yeaman C, Grindstaff KK, Nelson WJ (1999) New perspectives on mechanisms involved in generating epithelial cell polarity. *Physiol Rev* 79:73–98
4. Nelson WJ (2003) Adaptation of core mechanisms to generate cell polarity. *Nature* 422:766–774
5. Rodriguez-Boulant E, Kreitzer G, Musch A (2005) Organization of vesicular trafficking in epithelia. *Nat Rev Mol Cell Biol* 6:233–247
6. Mostov K, Su T, ter Beest M (2003) Polarized epithelial membrane traffic: conservation and plasticity. *Nat Cell Biol* 5:287–293
7. Vogelmann R, Amieva MR, Falkow S et al (2004) Breaking into the epithelial apical-junctional complex – news from pathogen hackers. *Curr Opin Cell Biol* 16:86–93
8. Rojas R, Apodaca G (2002) Immunoglobulin transport across polarized epithelial cells. *Nat Rev Mol Cell Biol* 3:944–956

Epithelial-to-Mesenchymal Transition

Definition

Epithelial-to-mesenchymal transition (EMT) (synonym: epithelial-to-mesenchymal conversion) describes a common process during morphogenesis and differentiation. It includes a series of coordinated events that result in cell-cell and cell-matrix interactions and cytoskeletal organization. EMT is a common event in early embryonic development, whereby tightly bound columnar (epithelial) cells lose their intercellular attachments and convert into motile, multipolar (mesenchymal) cells. EMT also accompanies cancer progression from the non-invasive to the invasive state. The process not only comprises of the transition from the epithelioid (e) to fibroblastic (f) morphology, indicating epithelial cell plasticity, but also a change in molecular markers with loss of E-cadherin and gain of vimentin.

- ▶ Cadherins
- ▶ Cancer Invasion and Metastasis: Cellular, Molecular and Clinical Aspects
- ▶ Epithelial Cells
- ▶ Neural Crest Cells and their Derivatives

capable of being recognized by an antibody, and therefore detected using immunochemical techniques.

- ▶ Immunochemical Methods, Localization
- ▶ Functional Assays

EPR

Definition

EPR stands for Electron Spin Resonance. This magnetic resonance method provides structural, kinetic and dynamic information on functionally important states of a protein.

- ▶ Photoreceptors

Epitope

Definition

Epitope refers to a short amino acid sequence of an antigen that is recognized by the variable region of immunoglobulins, or (along with an MHC class I molecules) by the T cell receptor.

- ▶ Autoimmune Diseases
- ▶ Biochemical Engineering of Glycoproteins
- ▶ DNA-based Vaccination
- ▶ Immunological Methods, the Use of Antibodies to Discover the Function of Novel Gene Products
- ▶ Recombinant Protein Production in Mammalian Cell Culture

Definition

EPSP (excitatory postsynaptic potential) and IPSP (inhibitory postsynaptic potential) denotes the depolarising or hyperpolarising electrical signal, respectively, generated by the activation of postsynaptic receptors. The summation of EPSPs/IPSPs represents the input integration and information processing in a given neuron.

- ▶ Neurons

Epstein-Barr Virus

Definition

Epstein-Barr-virus (EBV) is the member of the γ subfamily of herpes viruses that infects and persists in B-lymphocytes, which are stimulated to growth by the expression of viral latency-associated gene products. The virus is the cause of infectious mononucleosis, a self-limiting lymphoproliferative disease. In addition, EBV is highly associated with several malignant diseases (endemic form of Burkitt's lymphoma, Hodgkin's lymphoma, T-cell lymphoma and nasopharyngeal carcinoma).

- ▶ Viral Oncogenesis

Epitope Tag

Definition

Epitope tag refers to a short, specific amino acid sequence fused to one end of a protein of interest that is

Equilibrium

Definition

Equilibrium describes a condition in which the state functions have constant values throughout the system.

► [Thermodynamic Properties of DNA](#)

ER

- [Endoplasmic Reticulum](#)
- [Estrogen Receptor](#)

ERK

Definition

ERK stands for “Extracellular-signal-regulated kinase”, a protein kinase that is activated in response to growth factors.

- [RNA Polymerase III](#)
- [Signal Transduction: Integrin-Mediated Pathways](#)

ERM Protein (Ezrin, Radixin, Moesin)

Definition

ERM proteins (ezrin, radixin, moesin) act as plasma membrane-actin cytoskeleton linkers and are involved in cell morphogenesis, migration and signalling. The common amino and carboxy-terminal domains of ERM proteins mediate homo and heterotypic protein interactions.

- [Focal Adhesions](#)
- [Microvilli](#)

Erythropoietic Porphyrrias

Definition

A group of diseases that include erythropoietic protoporphyria and congenital erythropoietic porphyria.

- [Acute Intermittent Porphyria](#)

Erythropoietic Protoporphyria

Definition

Erythropoietic protoporphyria is a ferrochelatase deficiency, which in autosomal dominant inheritance leads to a chronic porphyria with skin symptoms.

- [Acute Intermittent Porphyria](#)

Erythropoietin

Definition

Erythropoietin (Epo) is a glycoprotein growth factor secreted by the capillary epithelium of the kidney in response to hypoxia. Circulating EPO binds to EPO receptors on the surface of erythroid progenitors in the bone marrow, to stimulate proliferation and maturation to functional erythrocytes. Recombinant human erythropoietin is used to treat anemias, e.g. anemias caused by renal failure or anemia in ► [AIDS](#) and cancer patients.

- [Enzyme Catalyzed Post-translational Hydroxylation of Proteins](#)
- [Growth Factors](#)

ES Cell Differentiation as a Model System for Functional Genomics

KENNETH R. BOHELER

Kenneth R. Boheler, Laboratory of Cardiovascular Science, NIA/NIH, Baltimore, MD, USA
bohelerk@grc.nia.nih.gov

Definition

The pace of sequencing whole genomes has far exceeded our knowledge of gene function, resulting in a set of biotechnological tools consisting of clones, sequence data across genomes and a full catalog of genes from multiple species. Unlike genomics, which is the study of genes and gene function, ► [functional genomics](#) can more aptly be defined as a set of biotechnological tools employed either in the context of an entire genome (e.g., gene trapping) or based on genome-wide information that can be employed to

understand basic mechanisms of gene regulation, structure and function. The development of functional genomics in the context of the latter has specifically led to high throughput assays that permit rapid analysis of a cell's ►**transcriptome** (i.e., entire RNA population in a sample) to elucidate expression patterns of genes (known and unknown) as well as the regulation of gene function. Recently, functional genomics in the context of high throughput assays has been employed to examine ►**embryonic stem cell** lines. These cell lines, derived from the early embryo, retain properties of embryo founder lines, have the capacity to undergo unlimited self-renewal and are capable of multilineage differentiation (►**pluripotentiality**). The specific transcripts/proteins that distinguish ►**stem cells** from non-pluripotent cells remain incompletely known and therefore some mechanisms responsible for self-renewal and pluripotentiality are unknown. Since the identification of signals that regulate the transition from undifferentiated to differentiated cells is fundamental to the understanding of cellular diversity and development, functional genomic approaches have proved critical for defining the molecular basis of multilineage differentiation and to impart the rationale for future hypothesis-driven studies.

Characteristics

Embryonic Stem Cells

Pluripotent stem cell lines derived from early mouse embryos can be maintained as relatively homogeneous stem cell cultures and expanded indefinitely. Three types of pluripotent, self-renewing stem cell lines have been established from mammalian embryos, embryonic carcinoma (EC), embryonic stem (ES) and embryonic germ (EG) cells. EC cell lines are derived from undifferentiated stem cell components of spontaneously arising or experimentally induced germ cell tumors, EG cell lines are derived from primordial germ cells, isolated primarily from the genital ridge and ES cell lines originate from pre-implantation embryos. ES cell lines, in particular, maintain a stable karyotype in culture and retain the character of embryo founder cells. When injected into a blastocyst, ES cells reintegrate fully into embryogenesis, in such a way that chimeric mice contain ES-derived cell progeny among all tissue cell types, including gametes. Undifferentiated ES cells are amenable to genetic engineering *via* random insertion events or homologous recombination, which allows the introduction of precise genetic modifications (knock-ins or knock-outs) in the mouse germ-line.

Embryo-derived stem cell lines (EC, EG and especially ES cell lines) readily differentiate *in vitro* into virtually any cell of endodermal, mesodermal or ectodermal origin. This differentiation process recapitulates many aspects of development, including early but poorly

accessible stages of mammalian embryogenesis. Differentiation of embryo-derived stem cell lines *in vitro* normally (except neurogenesis) requires an initial aggregation step to form three-dimensional spheroid structures, termed ►**embryoid bodies** (EBs) that can differentiate into a wide variety of specialized cell types, including cardiomyocytes, smooth and skeletal muscle cells, hematopoietic cells, cartilage, melanocytes and neurons. Importantly, differentiation of ES cells and formation of EBs can be regulated by the addition of growth factors and additives, the number of starting cells within a developing EB and the time and duration of plating.

Functional Genomics

Functional genomic assays are explicitly designed to generate gene expression profiles and identify gene transcripts that show altered abundance. The assays are not intended to provide mechanistic information, but are intended to provide new information to help explain the physiological role of a given gene-protein in a living organism, a task that has often proven difficult in early mammalian development. The conventional reductionist approach of studying one gene at a time is thus complemented by a more global or integrative approach that considers many genes at once. This ultimately facilitates gene, gene function and pathway discovery. An ideal transcriptome analysis should assay every transcript in a pure cell population prepared or cultivated under standardized conditions. Although the mouse and human genomes have been sequenced, some genes and gene transcripts remain to be identified, precluding an ideal analysis. Embryo-derived stem cell lines that are stable in culture represent an excellent source of cells, but because cell cultivation techniques are not always consistent between laboratories (differences in culturing conditions or reagents, isolation techniques, etc), some variation is expected. Differentiation of embryo-derived stem cells *in vitro* also leads to a heterogeneous population, making interpretation of transcriptome analyses challenging in the absence of complimentary data. Finally, the assumption in performing functional genomic analyses of ES cells during differentiation is that self-renewal and pluripotentiality involve a degree of regulation (e.g., transcriptional, post-transcription) that alters mRNA abundance.

Functional genomic techniques capable of analyzing mRNA transcript abundance include sequencing of expressed sequence tags (ESTs), subtractive hybridization, differential display, competitive PCR, cDNA or oligonucleotide ►**arrays** (►**microarrays**), and ►**serial analysis of gene expression** (SAGE). Only the latter two methods emerge as high throughput technologies suitable for gene expression analysis. Both techniques

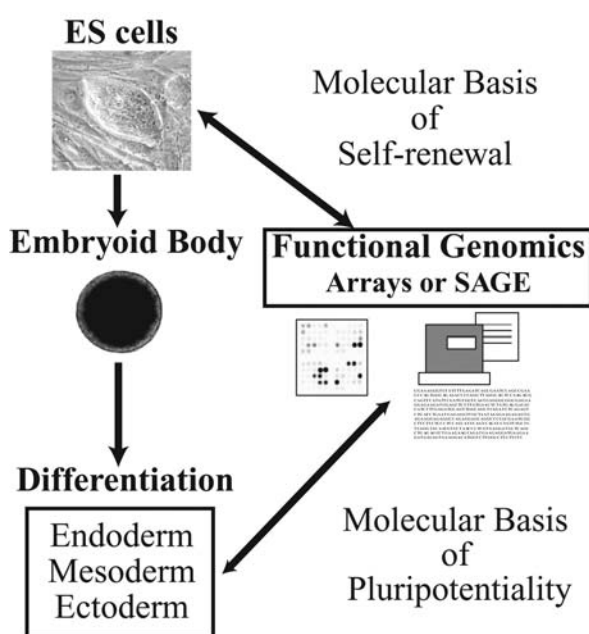
are useful for comparisons, but the quantitative nature and uniformity of SAGE greatly facilitates comparative and comprehensive analyses among laboratories, particularly when the datasets are available in the public domain (<http://www.ncbi.nlm.nih.gov/sage/>). SAGE, a sequence-based technique, has the potential to detect and quantify the expression of all cell transcripts (known and unknown), but it is limited by the fidelity and cost of sequencing (1). It is a quantitative sampling technique that is ideal for examining poorly described cell transcriptomes. Two major principles underlie SAGE. First, short DNA sequences are sufficient to identify individual transcripts, and second, concatenation (linking together) of the short DNA sequences (i.e., SAGE tags) increases the efficiency of identifying unique transcripts in a sequence-based assay. The technique generates large numbers of short (e.g., 10 bp) tags, originating from the last (most 3') unique location of an enzyme recognition site in a single transcript. Sequences that satisfy a strict size criterion are analyzed, so that multiple tags corresponding to different transcripts are recognized in a single sequencing run. Currently more than 2,538,124 (476,880 unique) and 534,283 (229,136 unique) SAGE tags of human and mouse origins respectively, have been deposited in the SAGE database (<http://www.ncbi.nlm.nih.gov/SAGE/index.cgi?cmd=printstats>). Within these datasets, many SAGE tags do not correspond to sequences located within currently predicted genes. While some of the mismatches may be due to sequencing errors inherent in the SAGE method, studies utilizing SAGE catalogs have altered the previous estimates of biological complexity and led to the identification of novel genes in mammalian genomes (see below).

► **DNA microarrays**, a hybridization-based technique, permit a systematic and quantitative transcriptome analysis (2). Microarrays can only examine expressed sequences that have already been identified either from clones or previous sequencing assays. This assay is thus limited by the probes present on an array and by signal intensity; poorly expressed transcripts are often not reproducibly quantified by this technique. Specifically, arrays are tools that when employed in hybridization assays, either with radioactive or non-radioactive probes, can assay small amounts of DNA or RNA. Microarray experiments, when compared against other hybridizations performed in a similar manner, permit the elucidation of which transcripts show altered abundance (increased or decreased expression).

Functional Genomics and Application to Embryonic Stem Cells

The identification of signals that regulate pluripotentiality and self-renewal is fundamental to the

understanding of stem cell biology (Fig. 1). The earliest attempt to quantify the functionally active genome of ES cells employed SAGE. Murine R1 ES cells were cultivated under conditions that resulted in a relatively pure cell population while maintaining critical characteristics of undifferentiated ES cells (3). The SAGE catalog contained 44,569 unique SAGE tags, of which 35.2% did not match any sequence previously deposited to any EST dataset. The number of unmatched sequences has now declined to ~10%, but the vast majority (>60%) of transcripts still encodes proteins with only a putative or a completely unknown function. Although SAGE had initially been employed to define the transcriptome of the R1 ES cell line, only a limited number of SAGE libraries were, until recently, available for comparative purposes. This precluded



ES Cell Differentiation as a Model System

for Functional Genomics. Figure 1 Embryo-derived stem cells (ES, EC and EG) are pluripotent stem cell lines that self-renew without apparent limitation *in vitro*. Undifferentiated ES cells on primary cultures of embryonic fibroblasts (feeder layers), when allowed to form cell aggregates known as embryoid bodies, spontaneously differentiate into cells typical of all three primary germ layers (endoderm, ectoderm and mesoderm). Functional genomic high throughput assays (Microarrays or SAGE) generate gene expression profiles and identify transcripts with altered abundance. The results from these assays provide useful information that can be employed to define the molecular basis of pluripotentiality and self-renewal and to provide a rationale for experiments to test newly developed hypotheses.

defining the molecular basis for the embryonic stem cell phenotype.

The analysis did however highlight areas of genomic controversy. Pilot surveys of early mouse embryonic tissues and pre-implantation embryos had suggested that a large number of developmental regulated genes might be expressed only in embryonic tissues. The findings that a large percentage of unknown transcripts were present in R1 ES cells and subsequently in P19 EC cells supported the idea that many genes, known and unknown, are uniquely expressed in pre-implantation embryos. Other groups independently identified SAGE tags that did not correspond to known genes. This led to the use of novel SAGE tags to discover previously unidentified genes in the genome. A systematic large-scale analysis of the genome by SAGE coupled to complementary approaches for gene identification is now envisioned to overcome current computational inadequacies of gene prediction.

SAGE data can also be employed to estimate the total number of transcripts present in ES cells (4). In the R1 ES cell SAGE catalog described above, the number of transcripts increases rapidly (as an inverse-square power law) as the detected transcript abundance is reduced. Although it is difficult, for statistical reasons, to estimate accurately the total number of unique transcripts, a simple correction for sampling and sequencing error indicates that more than 54,000 unique transcripts must be present in this cell population. Since a wide variation in the number of unique transcripts is possible, it is also likely that many very low abundance transcripts have been missed. Monte-Carlo simulations based on the inverse-square model take this latter possibility into account and indicate that as many as 130,000 unique transcripts are compatible with the R1 ES cell sampling profile. Because ~10% of the tags in this SAGE library do not map with any previously described EST dataset, the number of unique transcripts (splice variants or novel gene transcripts) that have not yet been identified remains quite high (~6,000–13,000). This highlights the limited degree of knowledge available to explain the molecular basis of embryonic stem cells.

To define the molecular signature of stem cells, two groups have recently employed microarrays to compare ES cells with adult stem cells of hematopoietic and neuronal origin (5, 6). These analyses identified 216 and 283 transcripts, respectively that were enriched in all three stem cell libraries. When these two datasets were compared, only 5.4% of the factors overlapped, indicating important variations due to culturing conditions, array and hybridization protocols, data analysis and potentially contaminating cells. When the “stemness” transcripts were grouped, however, a more common theme emerged. Stem cells, relative to other

cell types, expressed a large number of transcripts that could be described as signaling factors, transcription/translation factors and proteins associated with DNA repair, protein degradation and protein folding. When compared with differentiated cell types, stem cells expressed a higher number of gene transcripts with unknown function. Some of the “stemness” factors clustered to chromosome 17, suggesting that characterization of the genomic regions that regulate stem cell-associated factors will further our understanding of the regulatory networks required to maintain undifferentiated stem cell populations (5).

Finally, embryo-derived stem cells (EC, ES and EG) *in vitro* readily differentiate to all cell lineages. This is particularly important for the cardiovascular field, where remarkably little is known about the genetic program that directs the differentiation of cardiomyocytes from precursor cells and where the lack of established cardiac cell lines precludes most genetic studies to address this problem. Functional genomic approaches coupled with P19 EC differentiation to cardiomyocytes have therefore been applied to the program of cardiac development. Two independent groups took advantage of the ability of P19 EC cells treated with dimethylsulfoxide (1%) to produce a high percentage of cardiomyocytes *in vitro*, to perform a transcriptome analysis to uncover relationships among known genes and identify novel transcripts that mediate cardiomyocyte differentiation. One group employed cDNA microarrays (7) and the other SAGE (8). The data were remarkably consistent, highly complementary, and revealed that only a small percentage of genes are regulated during the induction of cardiomyocyte differentiation. Although the complexity of differentiation involved a limited number of gene transcripts, the multifaceted results (gene activation, inhibition, bi- and multi-phasic regulation) indicate that the process is more complicated than originally envisioned. Because neither study employed a pure cell population for the transcriptome analysis, it was impossible to identify any factor(s) that specifically controlled cardiac differentiation *in vitro*. These large-scale genomic analyses of gene expression were unable to provide mechanistic information about induction of cardiomyocyte progenitor cells, but the results did provide the rationale and basis for future hypothesis-driven studies. Future studies will require lineage restricted cell surface markers or genetic manipulations (introduction of antibiotic resistance genes or marker genes like the [▶green fluorescence protein](#)) of ES cells coupled with fluorescence activated cell sorting ([▶FACS](#)) and/or antibiotic selection to enhance the isolation of specific subpopulations of ES-derived cell progeny (e.g., subtypes of cardiomyocytes). ES cell differentiation can

thus be effectively employed as a model system for functional genomics to define the molecular basis for multilineage differentiation.

Clinical Relevance

In vivo, stem cells or committed progenitor cells allow organ or tissue self-repair; however, tissue damage or disease (► [Alzheimer's disease](#), diabetes, chronic heart failure, stroke, end-stage kidney disease and ► [cancer](#)) that disrupt cell function require therapy in the form of transplantation or cell replacement. Today's most urgent problem in transplantation is a lack of suitable donor organs and tissues. The recent descriptions of human ES cells suggest that cell replacement therapy is theoretically achievable by the transfer of isolated and defined ES cell progeny to a target organ in sufficient numbers and of good enough quality for them to survive and restore function. To be realized, however, stem cells must exhibit full proliferative potential and *in vitro* differentiation capacity, two of the areas recently evaluated by combining ES cell differentiation with functional genomics. The mechanisms regulating differentiation remain poorly understood and may be influenced by epigenetic or genetic modifications *in vitro*. The identification of signaling molecules that regulate ES cell self-renewal, pluripotentiality and differentiation to specific subtypes should promote our understanding of these processes not only in embryo-derived stem cells, but also in adult stem and progenitor cells. Thus, by continuing basic human and non-human embryonic and adult stem cell research and comparing gene expression patterns and phenotypes, the future application of human stem cells (ES, adult stem and progenitor cells) in clinical settings may become viable. These analyses are critical if we are to demonstrate that ES-derived progeny can be purified, transplanted and lead to normal physiological and functional recovery *in vivo*.

References

1. Velculescu VE, Zhang L, Vogelstein B et al (1995) Serial analysis of gene expression. *Science* 270:484–487
2. Schena M, Shalon D, Davis RW et al (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270:467–470
3. Anisimov SV, Tarasov KV, Tweedie D et al (2002) SAGE identification of gene transcripts with profiles unique to pluripotent mouse R1 embryonic stem cells. *Genomics* 79:169–176
4. Stern MD, Anisimov SV, Boheler KR (2003) Can Transcriptome Size be Estimated from SAGE Catalogs? *Bioinformatics* 19:443–448
5. Ramalho-Santos M, Yoon S, Matsuzaki Y et al (2002) "Stemness": Transcriptional profiling of embryonic and adult stem cells. *Science* 298:597–600
6. Ivanova NB, Dimos JT, Schaniel C et al (2002) A stem cell molecular signature. *Science* 298:601–604
7. Peng C-F, Wei Y, Levsky LV et al (2002) Microarray analysis of global changes in gene expression during cardiac myocyte differentiation. *Physiol Genomics* 9:145–155
8. Anisimov SV, Tarasov KV, Riordon D et al (2002) SAGE identification of differentiation responsive genes in P19 embryonic cells induced to form cardiomyocytes *in vitro*. *Mech Dev* 202:25–74

ES Cells

► Embryonic Stem Cell

ESCRT

Definition

The "endosomal sorting complexes required for transport" (ESCRT) is a cellular protein complex involved in the fusion of vesicles to form multi-vesicular bodies. Many enveloped viruses use the complex to complete the process of budding and release from the host cell.

► Retroviruses

ESE

Definition

ESE is the abbreviation for a short regulatory sequence that is located in an exon.

► Alternative Splicing

ESI

► Electrospray Ionisation Mass Spectrometry

ESI-MS

► Electrospray Ionization Mass Spectrometry

Essential Hypertension

►Hypertension

EST

►Expressed Sequence Tags

Estrogen Receptor

Definition

Estrogen receptor (ER) is a transcription factor of the nuclear receptor superfamily which is activated by its ligand, estrogen, and then binds to certain genes containing estrogen responsive elements (EREs) that activate their transcription. ER thus exerts various effects, notably proliferation of mammary epithelial and carcinoma cells.

►Breast Cancer

Etherphospholipids

Definition

Etherphospholipids comprise of a special class of phospholipids, characterized by the presence of an ether-bond at the *sn*-1 position.

►Peroxisomal Disorders

Ethical Issues in Medical Genetics

HANS MARTIN SASS

Kennedy Institute of Ethics, Georgetown University,
Washington DC, USA

sassh@georgetown.edu

Definition

Ethical issues in medical genetics arise from (a) new and previously unknown information on individual and familial genetic heritage and (b) new knowledge in manipulating genetic traits associated with disorder or disease or for enhancement. Moral subjects to make decisions and assess the ethical parameters of new information and knowledge are individuals, families, communities, legal experts, politicians even human-kind; the assessment of who, for a special issue or in a given situation, is or should be the prime moral agent is a highly debated and controversial ethical, cultural and political issue.

Existing and developing knowledge in genetic information and genetic manipulation is a scientific revolution of Copernican dimension and will have far reaching consequences cultural and social consequences in individual and collective attitudes, on values and virtues, lifestyles, hopes and expectations, also in structuring systems of health promotion and disease management.

Characteristics

Towards Risk Factor Health Assessment and Education

New knowledge in science and technology, such as in molecular medicine and new professional and personal applications of new knowledge, such as those in genomics and proteomics, require integrated assessment of values and technology. New knowledge in human genetics and pharmacogenetics will have a great impact on future predictive and preventive health care services, on individualized and more efficacious drug development and prescription, on lay health education in individualized health risk management and in health protection and enhancement, addressing interfaces of physical exercise, nutrition, medication and prudent lifestyle management. The dimensions of those changes cannot yet be described in terms of ►[bioethics](#) and will depend on individual and collective ethical and cultural responses and on regulation in supporting, suppressing or rejecting new fields of research or of transfer of results into medical research, clinical practice and health care management.

Traditional concepts of health may become obsolete together with health policies. In molecular medicine, health cannot simply be understood anymore as 'a state of complete physical, mental and social wellbeing and not merely the absence of disease or infirmity', as the WHO defines, rather as a process of challenge and response, a process of balancing, which needs health literacy, protection, promotion and management by the individual person. Health is not just a status; rather the balanced result of health-literate and risk-competent care of one's own physical, emotional and social

wellbeing and wellfeeling, achieved in competent understanding, modification and enhancement of individual genetic, social and environmental properties, with the support of health care professionals and through equal access to health care services, including information, predictive and preventive medicine. Therefore, a re-evaluation and a re-prioritizing of traditional principles of care, confidentiality, beneficence, ►informed consent and harm will have to occur. Long neglected patient's ethics and health care ethics of the lay person have to become a prime topic for bioethics research, education and application in the clinical, primary care and public health care settings. Future medical ethics will have to focus on modified bioethical principles such as duty to inform, duty to be told and to know, health education, ►health literacy, health care competence, informed request, ►informed contract and the ethics of ►data availability. These new challenges to health care are challenges in health care education of health care experts and of individual citizens, their families and communities. Risk factor medicine will have to focus more on health care ethics of the citizen and the lay person than exclusively on physician's ethics; information and counseling might play a more prominent role in medical education and medical ethics, changing repair mentality in physicians' attitudes and statutory health care systems.

►Carrier Ethics and Family Health Care

New diagnostic knowledge in human genetics and pharmacogenetics make responsibility sharing with citizens as future patients or actual patients possible and ethically mandatory (1). Based on Western and Eastern traditions of responsibility and self-determination within the individual's cultural environment, the obligation to know about one's genetic heritage and its advantages, disadvantages, risks and uncertainties is a precondition for living a self-determined, risk-competent educated life and for enjoying the fullest possible individual quality of life. These dormant cultural and ethical maxims might lead to a new emphasis to inform, to educate, to counsel and to support for the health care experts and to a right to be told and – in those cases, where others are not impacted negatively – a right to either follow or to refuse health care advice. There does not seem to be an obligation to tell, if diagnostic findings cannot result in advice or prescription; but one could make the argument that even in cases of interventional futility citizens have the right to request genetic information for whatever personal reasons. Medical ethics support a right to know and an obligation to tell, if health risks are present or predictable. However, it might not be ethically advisable to call for a legal obligation to follow health care advice; such an obligation might become more pressing if health care costs are shared by insurance.

As genetic diagnosis provides potentially important information for family members in regard to health risk, health status, potentially helpful preventive measures and information essential for individual self-determination and quality of life decisions, each and every person diagnosed will need to consider as part of her or his carrier ethics her or his responsibilities towards family. In modern Western cultures, emphasis is put on privacy of patients or those diagnosed, while traditional Asian professional and family cultures focus more on responsibilities towards family, filial love, parental love and love and obligation within the wider family (2).

As not only severe genetic disorders, but also higher than average risk of hypertension, of cancer and of metabolic disorders may run in families, information about those risks will be influenced by yet to be re-evaluated or newly developed family health care ethics. As a result, family relations will be influenced by new sources of guilt-feelings, shame, accusations, self-denials, maybe divorce, suicide and the breakup or a new bonding-together of families. The golden rule would suggest, not hiding behind traditional attitudes towards secrecy and privacy, but to openly and aggressively inform, educate, teach and support dialogue and discourse in families and in society.

In complex issues of family ethics, privacy, disclosure, right not to know and duty to know, diagnosed carriers would be the prime moral agents to make educated and responsible choices (a) to disclose, (b) to refuse disclosure of all or of some information, (c) to postpone hard choices in informing family members. There will be hard cases, where information might be lifesaving to family members who might be carriers. Proposed WHO guidelines (3) on ethical issues in medical genetics recommend violating the principle of confidentiality in favor of informing and consulting family members; but guidelines of national and professional organizations differ from such a stern position. Additionally, future principles and virtues in responsible parenthood might include decisions whether or not (a) to have children at all, (b) to have prenatal testing and eventually elected abortion following positive testing, or (c) to do nothing and trust in future breakthroughs in medical treatment of as yet untreatable disorders (4).

As molecular medicine will not only allow for better prediction, prevention and management of disorders or diseases, ethical issues of health enhancement such as anti-aging and the improvement of certain physical, intellectual or mental capacities will become an issue of individual choice and of controversy over principles of justice, in particular when not the protection or promotion but the enhancement of health is requested to be paid for by public or insurance funds; definitions of protection, promotion and enhancement of health

will cause public debate and controversies among citizens and ethicists.

Pharmacogenomics and Individualized Drug Treatment

Since the discovery of blood types, reliable diagnosis of blood types, typing and screening has become an essential part of emergency medicine and surgery and has saved millions of lives. For over a century, the ethical and medical benefit of the principle of data availability has been very well documented. Now, molecular medicine demonstrates that we do not differ only in types of blood but in many other individual properties, such as in cytochrome P450 isoforms, controlling drug metabolism, causing non-efficacy, side-effects with nutrition and other medication, even death, in some types of metabolizers (► [Pharmaco-genomics](#)).

When individual pharmacogenetic profiles for medication-typing can be established (5) in the same way that we can easily establish individual profiles in blood-typing, personalized drug delivery is possible and ethically required. The fears that genotyping for drug metabolism will lead to discrimination are not convincing, they are theoretical, ethically unfounded. Blood typing did not lead to discrimination, even though some individuals have blood types, which are rarer, at least in certain populations, and therefore might have less access to blood replacement. Blood profiles and medication profiles do not describe disorders, i.e. an individual aberration from a generic image; rather they constitute different types or variations, none of which is the 'normal' one. We have a model of variation, not one of order and disorder. It would have been a crime against humanity and an unexcusable wrong towards fellow humans if lives had not been saved by technically possible but legally or culturally unaccepted blood transfusion based on proper blood-typing.

Genotyping for drug compatibility causes no significant other ethical concerns than those associated with blood typing, clinical reliability of typing procedures, equal access to typing services, no (medication prescribing) intervention prior to typing. The new scenario of metabolism typing has consequences for the traditional and accepted bioethical setup of vaccine development, clinical trials, prescription procedures and nutritional advice. Given actual pharmacogenetic knowledge, it seems to be unethical not to introduce molecular medicine into drug development and to establish efficacy, dosage and side effects for major types of metabolizers.

Individuals have a civil right to information about their individual proteonomic and enzymatic properties for metabolizing drugs and nutrition. This informational right would best be served by providing individual drug and nutrition cards and access to information and education for professionals and the lay person.

Informational Property Rights and Data Profiling

Personal data, including genomic and proteonomic data on health and health care are the informational property of the individual. In other areas of life we share selected informational properties with others for our own benefit or convenience, such as with credit card providers, supermarkets, libraries, online-merchants and insurers of various kind. But, we rightly worry about protection of private data; we have laws and regulations protecting privacy, which work most, but not all of the time; we accept these risks as we balance risk with benefit. There is no good argument, to treat individual rights on health data differently from other ► [informational property rights](#).

Citizens are informational property owners of data concerning individual health status and health care. It would most probably be in their best interest to have health care cards and to share information with professionals in a protected framework as data availability becomes an important integrated part of data protection. Health care professionals cannot provide quality service if denied access to information necessary for providing safe and efficacious service. Technically, it would be extremely difficult to provide the best possible predictive, preventive and therapeutic service to those who do not share personal health status data with professionals.

Individual health care cards might be an efficacious and just instrument to promote as yet undeveloped cultures and attitudes of individualized health literacy and care. As we as individuals differ in more than enzymes and protein metabolism, it would only be reasonable to allow citizens to have individual health care cards containing information on individual genetic or acquired properties, abilities, disabilities and disorders such as risk of hypertension or diabetes. Data availability is the precondition for good diagnosis and prognosis and subsequently for prevention and treatment.

Research Ethics in Molecular Medicine

Traditional models of informed consent and proband-researcher or patient-physician interactions in communication and cooperation have to be restructured for adequacy in molecular medical research. The soft-paternalism principle of informed consent would be best replaced by a principle of informed contract, detailing for researchers and probandi or patients rights and obligations and liberating probandi and patients from their passive role of just consenting. In particular, this should apply to issues of research in drug metabolism, DNA-sampling and -storing. Moreover, disease-specific research cannot be justified without taking into account the probable benefits to the patient or her or his family. Modern medical research will discover quite a lot of information about pedigree and family members; new ethical issues can rarely be addressed in the old format by simple individual consent.

For genotyping in highly defined populations of patients suffering from certain subgroups of cancer or other diseases and receiving specific medication, it has been debated whether traditional models of informed consent would be enough for multipurpose longterm DNA-banking. Informed consent forms rarely address issues of multipurpose screening and longterm storage. It has been suggested that for genotyping, only specific informed consent should be requested and that further use should be covered by new specific re-consent. Generic consent forms – in particular for prenatal and newborn screening – were proposed, but others criticized such an approach as lowering the standards of informed consent.

As the probability of benefits in cross-purpose genotyping and of future yet to be specified testing and re-testing is of great moral importance for the individual patient, patient groups and the progress of clinical research, a contract model might be more appropriate, describing the obligation of the researchers to inform the patient on all or some of their findings and establishing a contract spelling out the obligations towards the patient and her or his family (6). Within such a contract, patients or their legal representatives must be informed on standard data-protection. In order to solve complex issues of privacy and disclosure, the right not to know and the duty to know, a contract must provide that patients can make their own choices (a) for mandating disclosure of individual predictive, preventive or therapeutic knowledge, (b) for refusal of all or some information and (c) for postponing such a decision for later based on then existing individual circumstances or clinical results. The moral issues of informing and protecting family members will similarly have to be addressed within the contract by allowing the patient to choose among a number of procedures by which family members of various degree may or may not be involved, informed or invited. Such an informed contract model would be based on the stakeholder concept which is usually the preferred scheme for complex management of risk and uncertainties elsewhere. An example of how the ethics of diagnostic research is changing towards a more appropriate appreciation of individual and family health care competence and decision making as partners in research is the HUGO statement (7) on choices in data availability within a firm model of data protection and privacy protection. 'The choices and privacy of individuals, families and communities should be respected. (a) Choices may be with regard to donation, storage and use of samples and the information derived therefrom (e.g. specific, related or other uses subject to authorization by an ethics committee, etc). Informed consent may include notification of uses (actual or future) or opting out or, in some cases, blanket consent. (b) Mechanisms should be established to ensure respect

for such choices. (c) Participants should be informed about the degree of identifiability of their data (e.g. coded, anonymized, aggregate, etc) and the security mechanisms in place to ensure confidentiality. (d) Participants should be told that samples or the information derived therefrom may be shared with other researchers including those from other countries, with commercial entities and through publication and availability on the WEB'.

Individual Values and Cultural Diversities

Modern medicine, including molecular medicine, recognizing the principle of autonomy and self-determination as a most basic human and civil right, allows for clinical research and medical treatment only if the proband or patient has given free and informed consent based on individual concepts of risk, benefit, values, fear and hopes. Global recognition of the informed-consent principle correlates with the vision of universal human rights, as expressed by the United Nations Declaration, 1948 and is a fruit of the processes of enlightenment and emancipation since the European Age of Reason. But many cases have been well documented demonstrating that the principle of autonomy does not work all the time and can be used in an abusive or exploitative manner. (8) If people do not clearly understand risks and benefits associated with research and treatment, oral or written consent is void. (7) If people feel an 'obligation' to sign forms, such consent is not given freely. (4) If researchers cannot or do not adequately inform probands or physicians their patients, signed forms are only a smoke-screen to hide the fact that true informed consent is not given even if the legal requirements seem to be satisfied.

We already see standard deviations, exemptions and modifications form the general rule. (8) Consent for minors is given by their ethical and/or legal representatives, mostly the parents. (7) Competent adults may designate another person to give consent on their behalf, either immediately or under certain conditions in the future. (4) The consent required from psychiatric patients is related to their particular disease and situation at a given time; however, there are well-developed treatment contracts signed by patients and their caretakers for possible future situations known to the patient. Other cultures, even though they might be in transition to being more and more influenced by cultures of individualistic ethics, still have a strong sense of family ethics and family decision making for the good of the individual family member, thus traditionally giving consent for the good of an individual family member as family consent (2). Individualistic European models are seen as an intrusion into a different trust-and-responsibility structure if used only legalistically. Family consent is not without risk. It might be the

elder male or female head of the smaller or larger family accepting responsibility for his or her relatives and having been trusted by them to make those decisions, even far-reaching ones such as those concerning marriage, education or job training. Some individuals or branches of the family might not trust the proxy decision maker; elders might violate the trust with which they are endowed. These are the situations that have led to the rise of emancipation and enlightenment in the Age of Reason. But there are still families and communities around for whom the model of an individual person making autonomous decisions by herself and for herself alone, is considered unethical, not supported by culture and values, actually decadent and perverse.

In order to implement some of the visions and goals of the 'informed consent' principle it might be helpful to at least work with a formula of 'informed consent plus X', X being different supportive features depending on the special cultural and traditional attitudes and modes of decision making, supporting but not replacing individual consent.

Researchers required to inform and educate and to gain consent, would be well advised to use the existing trust-and-responsibility structure for information and education and for contracting with the community and/or families and/or individuals. Benefits for the community, for the families and the individuals should be spelled out in detail (6).

Common sense suggests recognition that probands, even more so patients, expect 'something' in return for their participation; this 'something' can be personal recognition, personal attention, individual or general health care advice, better hygiene, fresh water or nursing care for the community. Recent CIOMS guidelines for medical research in 'populations and communities with limited resources' recommend: 'the sponsor and the investigator must make every effort to ensure that: the research is responsive to the health needs and the priorities of the population or community in which it is to be carried out and that any intervention or product developed or knowledge generated will be made reasonably available for the benefit of that population or community' (8). The Nuffield Council on Bioethics (9) suggests 'that sponsors of research should require that the development of local expertise in healthcare is an integral component of research proposals. Consideration should be given to the extent to which any strengthening of local healthcare facilities can be done in such a way that the changes are sustainable in the local context once the research is over'. Given the diversity of individual and collective cultures in decision-making, one size of consent does not fit all. It seems to be clear, that the classical model of informed consent has outlived its useful life as a general standard

for all, for each and every personal, familial, communitarian, cultural or legal situation. Where the basic cultural attitudes and legal preconditions are not in place to make the classical form of informed consent the preferred and most useful tool, it cannot be made a requirement that medical experts first of all change cultures and attitudes and then proceed with their medical work. It is also not acceptable that medical experts turn a blind eye to the missing of essential prerequisites for making informed consent work. Everyone has to work on implementing human rights and free decision making by competent and risk-literate adults; this task cannot be put on the physicians alone. Also, there might be true ethical situations where coherent trust-and-responsibility structures within families or communities are well developed by cultural or religious tradition and in the history of ideas supported and proven to work well in quite a number of cases. In those situations it would be culturally and ethically insensitive and counterproductive to destroy a working network of trust, hope, responsibility and reliability in order to replace it by a model developed under different cultural and historical conditions.

Models of contract rather than one-sided still soft-paternalistic individualistic consent might work better in those situations, both in medical research and in health care. Of course, all culturally sensitive models need 'escape clauses' or 'conscience clauses' allowing each and every individual to decide for herself or himself on the basis of individual self determination about her or his place in a moral community and in regard to community values and decision making. Molecular medicine will need to be aware of cultural traditions and moral trends among citizens, regulators and bioethicists when developing and offering new options in health care, health promotion and disease management.

Culturally sensitive introduction of new knowledge always was associated with moral and cultural risk and uncertainty. Open and fair discourse, communication and cooperation among all stakeholders in health care will be successful only if personal choices of moral agents at their respective levels are respected. The identification and acceptance of moral decision makers on different levels has yet to be made and will not occur without ethical and political controversy.

References

1. Wertz D, Nippert I, Wolff G (2003) Patient and professional responsibilities in genetic counseling. In: Sass HM, Schroeder P (eds) *Patientenaufklärung bei genetischem Risiko*. Muenster Lit, pp 79–92
2. Sass HM (2003) Common moral priorities and cultural diversities. *Formosan Journal of Medical Humanities* 4(1+2):6–21

3. WHO (1998) Proposed International Guidelines on ethical issues in medical genetics and medical services. WHO (doc.ref.who/hgn/gl/eth/98.1), Geneva
4. Kielstein R, Sass HM (2002) Genetics in kidney disease; how much do we want to know? *Am J Kidney Disease* 39:637–652
5. Paul NW, Roses AD (2003) Pharmacogenetics and pharmacogenomics: recent developments, their clinical relevance and some ethical, social and legal implications. *J Molec Med* 81:135–140
6. Sass HM (2001) A Contract Model for Genetic Research and Health Care for Individuals and Families. *Eubios* 11:130–132
7. HUGO (2003) Statement on genetic screening. *Eubios* 13:99
8. CIOMS (2002) International Ethical Guidelines for Biomedical Research involving Human Subjects, Geneva, CIOMS
9. Nuffield Council on Bioethics (2002) The Ethics of Research related to Healthcare in Developing Countries. Nuffield Council, London

Etiology

Definition

Etiology describes the causes or study of direct and predisposing causes of a disease.

- ▶ [Microarrays in Pancreatic Cancer](#)
- ▶ [Proteomics in Cardiovascular Disease](#)

Euchromatin

Definition

Euchromatin defines an extended chromatin structure with open nucleosome organization, accessible to RNA polymerase and permissive to transcription. It contains most of the genes.

- ▶ [Nucleosomes](#)
- ▶ [Transgene Silencing](#)

Evanescent Electric Field

Definition

Evanescent electric field refers to a state when light is totally internally reflected at the boundary of two media

with different refractive indices. The electric field of the incident light beam extends beyond the reflecting surface.

- ▶ [Surface Plasmon Resonance](#)

Evolution Time

Definition

Each NMR experiment consists of a sequence of radio frequency pulses with delay periods in between them. Timing, frequencies, and intensities of these pulses distinguish the different NMR experiments from one another. During some of the delays, the nuclear spins are allowed to freely rotate for a given length of time, known as the evolution time. Indirect dimensions are created by systematic variation of the evolution time.

- ▶ [Multidimensional NMR Spectroscopy](#)

Evolutionary Conservation

Definition

Evolutionary conservation refers to processes or gene/protein functions that are likely to have been present in the common ancestor of two species.

- ▶ [Drosophila Model of Cardiac Disease](#)

Ex Vivo

Definition

The term ex vivo is used for cells or tissues that have been modified outside of the body.

- ▶ [Heritable Skin Disorders](#)

Excitable Membranes

- ▶ [Ion channels/Excitable membranes](#)

Excitation Spectrum

Definition

The Excitation spectrum is the plot of incident light wavelength versus the intensity of light absorbed by the fluorochrome.

► [Fluorescence Microscopy: Spectral Imaging vs. Filter-based Imaging](#)

Exclusion

Definition

Bacterial plasmids, including those on which the standard cloning vectors are based, have replication systems based on a replication origin counting mechanism. The total copy number of a given origin is regulated and thus multiple clones in the same vector (or more generally with the same replication origin) are not maintained stably in the same cell. Newly transformed cells may contain many plasmids, but individual colonies of their progeny reliably contain one each.

► [YAC and PAC Maps](#)

Executive Function

Definition

Executive functions describe higher-order neurobehavioral conditions that enable humans to plan, sequence, initiate, and sustain behavior towards some goal, incorporating feedback and dynamically adapting behaviour to a certain condition.

► [Parkinson's Disease: Insights from Genetic Cause](#)

Exocrine Cells

Definition

Exocrine cells or organs form glands that secrete enzymes or other biological active compounds into a system of ducts that lead ultimately to the exterior of the body. Examples are cells of the salivary, sweat or gastric glands. ► [Endocrine cells](#), in contrast, place

their secretions into the internal environment – the blood.

► [Microarrays in Pancreatic Cancer](#)

Exocyst

Definition

The exocyst is a multimeric complex of eight subunits (Sec3, Sec5, Sec6, Sec8, Sec10, Sec15, Exo70 and Exo84) which was initially identified in yeast, where it is involved in targeting and tethering of secretory vesicles to sites of polarized exocytosis. Mammalian cells express homologues of all eight yeast subunits. The mammalian exocyst associates with the *trans*-Golgi network and the plasma membrane. In polarized cells it appears to be involved in basolateral traffic.

► [Protein and Membrane Transport in Eukaryotic Cells](#)

Exocytic Pathway

ROMAN S. POLISHCHUK¹,
ALEXANDER A. MIRONOV¹, ALBERTO LUINI²

Department of Cell Biology and Oncology, Laboratory of Molecular Neurobiology, S. Maria Imbaro (Chieti), Italy

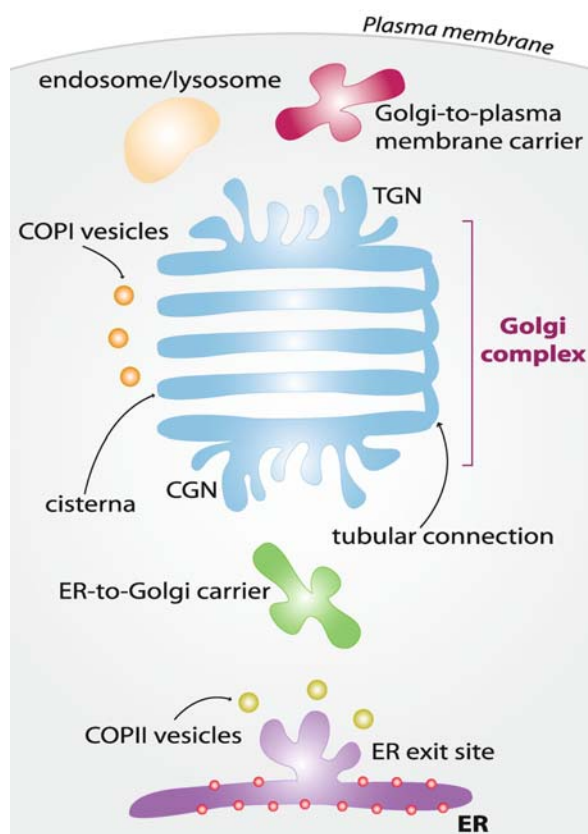
luini@negrisud.it

Definition

All cells need to secrete a variety of macromolecules in order to modify their environment and to protect themselves. At the same time, cells synthesize new proteins and lipids to replace those that are consumed within the various intracellular compartments. During their synthesis on ribosomes, proteins are either transferred into the lumen of the endoplasmic reticulum (ER) or incorporated into the ER membrane itself and are transported through the several membranous compartments that form the exocytic (also known as the secretory or biosynthetic) pathway. This includes the main secretory/sorting station, the Golgi complex (Fig. 1). After crossing the Golgi complex, proteins are sorted and shipped to their final cellular destinations, such as the apical or basolateral cell surfaces, the secretory granules or the endocytic compartments (1). During this journey through the cell, proteins and lipids are modified by a variety of enzymes, which are located

¹The first two authors contributed equally to this work

²Corresponding principal investigator



Exocytic Pathway. Figure 1 Scheme illustrating the organization and the components of the exocytic pathway.

mostly in the Golgi complex. Glycosylation is the most common and important of these post-translational modifications. With this in mind, it can be seen that the exocytic pathway has two main functions: (i) to transport proteins and lipids to their appropriate cellular compartments as needed and (ii) to complete the protein expression process by adding certain specific chemical elements, which are generally crucial for protein structure and function, to the amino acid backbones of proteins.

Characteristics

The exocytic process needs to be characterized at both the molecular and the morphofunctional levels. While the molecular machineries underlying the main elementary traffic steps (e.g. membrane bending and fusion) are now relatively well understood (1, 2, 3), the overall morphofunctional organization of traffic, in terms of the structures and dynamics of the secretory organelles themselves, remains less clear. However, following the development of new and powerful morphological approaches, this has begun to be unraveled in the last few years. It should be mentioned

here that the traffic model that dominated the field until recently has now become inadequate to explain a number of new observations. This vesicular traffic model, according to which secretory proteins move across successive compartments of the secretory pathway within small vesicles that bud from one compartment and fuse with the next, is now being replaced by different traffic principles that are being tested in several laboratories. The morphofunctional organization of traffic will be the focus of this review.

Exit from the ER

Secretory proteins must travel from their site of synthesis, the ER, to distant cellular sites, one of which is the plasma membrane. The question as to how exit from the ER is organized remains an issue of debate. It is known that exit is mediated by numerous proteins (2, 3), among which the set of **coat proteins** known as COP-II has a prominent role; however, the precise way in which COP-II is involved in the creation of transport carriers remains unclear. According to the original vesicular model, COP-II drives the formation of coated vesicles (of about 70 nm in diameter) at specialized membrane domains that are known as the ER exit sites (2, 3). These vesicles then fuse with each other to form larger (1–5 μm across) pleiomorphic carriers that move to the Golgi complex. According to an alternative and more recent model, however, the secretory proteins are directly incorporated into large transport carriers that emerge from the ER by COP-II mediated *en bloc* protrusion of the ER surface at the ER exit sites (3, 4). While the vesicular model is supported by a large amount of molecular and *in vitro* evidence, it does not easily account for the export of molecular complexes from the ER that are much larger than a vesicle in size, such as procollagen, lipoprotein particles or lipid droplets. These cargoes are abundant and of physiological importance in mammals. Their export is easily explained by the transport carrier model.

After formation and before moving away from the ER, newly formed carriers (Fig. 1) undergo a process of 'maturation', a key part of which is the replacement of the COP-II coat proteins with a different 'coat', **COP-I**. Once COP-I is bound, the carrier proceeds from the peripheral exit sites towards the Golgi complex, which is positioned near to the centre of the cell (4). This centripetal movement is mediated by microtubules and a molecular motor protein, dynein, and is thought to involve the physical detachment of the carrier from the parental ER membranes (4). It is also conceivable, however, that at least in some cases, cargoes might flow towards the Golgi along pre-existing tubular structures (6). Once the cargo carriers arrive at the centre of the cell, they must discharge their contents into the Golgi complex.

The Golgi Complex

The Golgi complex is the central sorting station on the secretory pathway and has been a main focus of research in cell biology for many decades. The Golgi complex receives newly synthesized proteins and lipids from the ER, transports them across its cisternae and then distributes them to the plasma membrane, the endosomal/lysosomal system or the secretory granules (Fig. 1). The Golgi complex also serves as a chemical 'factory' for post-translational modifications of proteins and lipids (mostly by glycosylation). In the lumen of the cisternae of the Golgi complex glycosyltransferases further modify the oligosaccharide chains of secretory proteins (a process initiated in the ER). As well as glycosyltransferases, the Golgi complex contains other enzymes, including pro-protein convertases (such as furin), which cleave protein precursors into their mature forms, and kinases.

From the structural point of view, the Golgi complex consists of a series of 'stacks' made up of 4 to 8 flat cisternae. These cisternae are laterally interconnected by tubular networks (the 'non-compact zones') and together they form a continuous membranous ribbon that is held in the pericentriolar space (5). Despite this complex architecture, the Golgi complex is highly dynamic and can combine enormous rates of membrane flux with the ability to change its shape rapidly and even to disassemble and reassemble, under a variety of physiological and pathological conditions (4).

The glycosylation enzymes reside within the cisternae of the individual stacks. In general, enzymes acting early in the biosynthetic pathway localize to the *cis* and medial portions of the Golgi complex, while the late-acting enzymes reside in the medial/*trans* cisternae. However, quantitative immuno electron microscopy has clearly demonstrated that these enzymes are not restricted to just one or two cisternae; rather, they exhibit concentration gradients through the cisternae of the stacks (6).

Based on these considerations, each Golgi stack is traditionally considered to be composed of three main compartments, the *cis* (entry side)-, the medial- and the *trans* (exit side)-Golgi (5, 6). At the *cis* side of each stack, there is a membrane network known as the *cis*-Golgi network (CGN), which is composed of branching tubules connected with the *cis*-most cisterna (6). The CGN is followed by a number of flat cisternae (Fig. 1). Finally, the *trans* side of each stack, known as the *trans*-Golgi network (TGN), often appears as a network of branching tubules in electron microscopy sections (6). The cisternae usually appear as very flattened and wide sacks or tubules with a narrow (10–20 nm) lumen. This geometry may enhance the efficiency of glycosylation, considering the small volume of the cisternae and the relatively large surface area where cargo proteins can interact with the

glycosylation enzymes (6). Apart from the cisternae, the Golgi has significant tubular and vesicular structural elements.

Vesicles of smaller (50–60 nm) and larger (up to 80–100 nm) diameters are considered to be important elements of the Golgi complex. They depend on the activities of either COP-I or the [clathrin](#) coat machinery (1, 2, 3), respectively.

Arrival at and Transit Through the Golgi Complex

There are two main potential mechanisms for the entrance of cargo into the Golgi. The first involves the direct fusion of ER-to-Golgi carriers with the *cis*-Golgi membranes. The alternative is that the ER-to-Golgi carriers do not fuse directly with the *cis*-Golgi, but instead fuse with each other, thus forming a new *cis*-Golgi cisterna that can then progress through the Golgi stack (by cisternal maturation, see below). Finally, as indicated above, the possibility cannot be excluded that at least in some cases, the cargo may flow along pre-existing tubular structures connecting the ER with the *cis*-Golgi.

After entering the Golgi, the cargo must cross the stack before exiting from the *trans*-Golgi side. The mechanisms behind this intra-Golgi transport remain poorly understood (5, 6, 7). There are three main models for traffic through the Golgi: (i) anterograde vesicular traffic, in which cargo traverses the Golgi stacks *via* COP-I vesicles that bud from donor and fuse with acceptor cisternae (2), (ii) cisternal maturation, in which the cargo remains within the lumen of the cisternae, which themselves mature and progress through the stack (5, 7) and (iii) flow *via* continuities, in which the cargo flows through the stack *via* tubular connections between cisternae at different levels of the stacks (4, 6, 7).

The anterograde vesicular model was favored by the traffic community for a long time, but it no longer appears to apply. In the majority of cases, COP-I vesicles appear to be depleted of most of the secretory cargoes that have been examined, and some cargoes are simply much too large to fit into these vesicles (6, 7). As a result of this and other lines of evidence, the anterograde vesicular model has been all but abandoned, and in recent years, the cisternal maturation model has gained prominence (5, 6, 7). Some aspects of this maturation model remain unclear, however. For instance, according to classical maturation schemes, cisternae containing cargo progress through the stack, while the Golgi enzymes maintain their steady-state positions in the cisternae through their retrograde flow within COP-I vesicles in synchrony with the progression of the cisternae (7). However, most (albeit not all) studies have shown that both cargo proteins and Golgi enzymes are depleted in COP-I-dependent vesicles.

Thus, at least such a ‘classical’ version of the maturation model is unlikely to be valid. In our view, flow *via* continuities is also likely to have a key role in traffic. Indeed, due to the extensive use of electron microscopy tomography, evidence is now accumulating that membrane continuities between heterogeneous cisternae do exist (in contrast to previous beliefs), and that they should therefore have a role in traffic through the Golgi (6, 7). For instance, heterotypic intercisternal connections within the same stack have been seen in actively secreting cells (6). The question as to whether these intercisternal connections can explain antero-grade transport of cargo proteins and/or retrograde transport of Golgi enzymes is an important emerging issue for future research on the Golgi complex.

Exit from the Trans-Golgi

After crossing the Golgi, secretory proteins are ‘packaged’ for their departure to their final destinations. The *trans*-face of the Golgi has attracted the attention of cell biologists since its identification as the exit pole of the Golgi complex. To understand the mechanisms of sorting and exit from the Golgi complex, it was essential to identify the specific compartment in which these processes take place. This compartment is now defined as the TGN, although its precise organization remains under debate. Biochemical and immuno electron microscopy studies have revealed that the TGN is involved in the terminal glycosylation of proteins as well as in cargo packaging into membrane carriers that are destined for the plasma membrane or the endosomal/lysosomal system (6). The TGN has generally been considered to be the *trans*-most cisterna of the Golgi stack, which is seen to continue into a large anastomosing tubular network (6). The morphology and size of the TGN apparently depend on the predominant type and amount of cargo protein leaving the Golgi complex (6).

An alternative view of the organization at the *trans*-face of the Golgi complex comes from studies using electron microscopy tomography (5). Surprisingly, these studies demonstrated a lack of the classical TGN; instead, the three *trans*-most cisternae exhibited tubules that extended into the *trans*-space of the Golgi stacks and frequently had coats on their tips. Each of these cisternae appeared to produce only one type of coated vesicular structure, with only the *trans*-most cisterna having clathrin-coated buds. Thus, this suggests that the three *trans*-most cisternae of the Golgi stacks serve as the classical TGN, and that each of them is specialized in the packaging and export of its own, specific, cargo protein(s).

Given this structure, how does the exit of cargo from the TGN take place? Plasma membrane-directed proteins leave the Golgi within Golgi-to-plasma membrane carriers (GPCs), which at the electron

microscopy level have a complex, tubular-saccular morphology (4, 6) (Fig. 1). Recent work has shown that these carriers form by the fission of complex tubular-reticular membranes from specialized areas of the TGN (the GPC precursors) that are extruded from the Golgi complex along microtubules by the kinesin motors (6). These GPC precursors do not concentrate cargo proteins, do not exhibit visible coated regions and remain in physical continuity with the TGN membranes. They exclude TGN proteins that are directed to the endosomal/lysosomal system, such as mannose 6-phosphate receptor, lamp and furin (6).

What then is the role of coat proteins in this export from the TGN? While, as noted, clathrin and the related adaptor proteins, such as AP-1 and AP-3, are excluded from plasma membrane-directed carriers, AP-based coats are apparently required for the exit of the cargo that is targeted to the compartments of the endosomal system (3, 4, 6). Indeed, endosomal cargoes undergo marked concentration within small clathrin-coated membrane domains before their exit from the TGN (3, 6). The mechanisms by which proteins are selectively targeted to their specific final destinations, which include the apical or basolateral plasma membranes, the endosomal system and the secretory granules, are collectively known as the ‘sorting mechanisms’. The COP and AP proteins have major roles in this sorting and the reader is referred to previous reviews for an exhaustive treatment of these issues (3, 6).

Molecular Interactions

The main traffic molecular machineries and their functions have been admirably reviewed by others recently (1, 2, 3). Coat proteins are thought to be the molecular centerpiece in the exocytic pathway because they allow the selective transfer of macromolecules from one membrane compartment to another by concentrating proteins into specialized membrane patches and then deforming these patches into coated carriers. Furthermore, coat proteins may also participate in the association of these carriers with the cytoskeleton and with acceptor organelles by recruiting cytoskeleton motors and the docking-fusion molecules named ►SNAREs.

Regulatory Mechanisms

Throughout the life of a cell, the secretory pathway must maintain its homeostasis and adapt its level of activity to changing cellular needs. Our knowledge of the mechanisms involved in this regulation is still in its infancy. It is clear, however, that all stages of intracellular traffic are coordinated amongst themselves and with other cellular functions (6). In this context, an important emerging theme is the role of cargo as a

potent determinant of Golgi structure and function. For example, in yeast, cycloheximide treatment (which blocks proteins synthesis, and therefore reduces the secretory load) leads to the complete disappearance of the Golgi complex (6). Similarly, in the primitive eukaryote *Giardia lamblia*, the typical Golgi complex is absent during the proliferative stage, when secretion is very limited. However, the Golgi stacks reappear when the cells begin to secrete the glycans that are needed for cyst formation (6). It can also be seen that the level of cargo traffic influences the organization of the mammalian Golgi complex. Here, blockade of protein synthesis with cycloheximide leads to the formation of very thin cisternae in onion-like Golgi stacks (6). Moreover, when cargo proteins are arrested within the ER, the Golgi structure is all lost within a few hours. In contrast, when cargo influx increases sharply, the Golgi complex immediately grows in size (6). Thus, the Golgi complex tends to disappear gradually in the absence of cargo input even in mammalian cells, where it becomes larger when transport intensifies. This phenomenon has been clearly illustrated by the stimulation and inhibition of prolactin transport in lactating rats (6).

► **Vesicular Traffic**

References

1. Mellman I, Warren G (2000) The road taken: past and future foundations of membrane traffic. *Cell* 100:99–112
2. Schekman R (2002) Lasker Basic Medical Research Award. SEC mutants and the secretory apparatus. *Nat Med* 8:1055–1058
3. Rothman JE (2002) Lasker Basic Medical Research Award. The machinery and principles of vesicle transport in the cell. *Nat Med* 8:1059–1062
4. Bonifacino JS, Lippincott-Schwartz J (2003) Coat proteins: shaping membrane transport. *Nat Rev Mol Cell Biol* 4:409–414
5. Lippincott-Schwartz J, Roberts TH, Hirschberg K (2000) Secretory protein trafficking and organelle dynamics in living cells. *Annu Rev Cell Dev Biol* 16:557–589
6. Marsh BJ, Howell KE (2002) The mammalian Golgi complex debates. *Nat Rev Mol Cell Biol* 3:789–795
7. Polishchuk RS, Mironov AA (2004) Structural aspects of Golgi function. *Cell Mol Life Sci* 61:146–158
8. Mironov AA, Weidman P, Luini A (1997) Variations on the intracellular transport theme: maturing cisternae and trafficking tubules. *J Cell Biol* 138:481–484

Exocytosis

- **Exocytotic Pathway**
 ► **Vesicular Traffic**

Exon

Definition

An exon is that part of the DNA sequence that is – as part of an eucaryotic gene – transcribed into mature mRNA, which (when positioned 3' to the AUG and 5' to the first ► **stop codon**) can then be translated into protein. This information is interspersed with noncoding sequences which are called introns, i.e. longer or shorter regions of DNA that have no apparent function (non-coding-regions). In order to translate a gene into a protein, the non-coding sequence has to be removed in a process called splicing.

- **Alternative Splicing**
 ► **Duchenne Muscular Dystrophy**
 ► **Epistasis in Cystic Fibrosis**
 ► **Familial Hypercholesterolemia**
 ► **Full Length cDNA Sequencing**
 ► **Huntington's Disease**
 ► **Morpholino Oligonucleotides, Functional Genomics by Gene 'Knock-down'**
 ► **Mutagenesis Approaches in Medaka**
 ► **Mutagenesis Approaches in the Zebrafish**
 ► **Repeat Expansion Diseases**
 ► **Schizophrenia Genetics**
 ► **Splicing**
 ► **SRY – Sex Reversal**
 ► **Protein microarrays as Tool for Protein Profiling and Identification of Novel Protein-Protein Interactions**

Exon Junction Complex

Definition

Exon Junction Complex (EJC) refers to a complex of at least six proteins that is deposited upstream of exon-exon junctions on RNAs undergoing splicing. The proteins in the EJC have roles in mRNA export, localization and surveillance.

- **RNA Export**

Exon / Intron Junction

Definition

Exon/intron junction denotes a site within the pre-mRNA sequence where pre-mRNA splicing occurs to eliminate intron sequence segments from mature mRNA.

- **Morpholino Oligonucleotides, Functional Genomics by Gene 'Knock-down'**

Exonic Sequence Element

Definition

Exonic sequence element (ESE) refers to a short regulatory sequence that is located in an exon.

► [Alternative Splicing](#)

Exonic Splicing Enhancer/Silencer

Definition

Exonic splicing enhancer/silencers are specific short sequences in an exon that enhance or block pre-RNA splicing, and play important roles in both constitutive and alternative splicing. An exonic splicing enhancer is generally present at a nearby splicing junction. It functions as a cis-acting sequence element with a binding site for serine/arginine-rich (SR) proteins.

► [Spinal Muscular Atrophy](#)

► [Splicing](#)

Exonuclease

Definition

Exonuclease designates an enzyme that catalyses the hydrolytic removal of deoxyribonucleotide monophosphates one – by – one from the end of a DNA strand. Exonucleases remove nucleotides, specifically either in 3',5' or in 5',3' direction, depending on the enzyme.

► [DNA Polymerases](#)

► [RNA Stability](#)

Exosome

Definition

Exosome defines a complex of at least 10 proteins that has 3',5' exoribonuclease activity.

► [RNA Stability](#)

Export Adapters

Definition

Export adapters are proteins that bind the RNA cargo and bring it to export receptors.

► [Nuclear Import and Export](#)

► [RNA Export](#)

Export Receptors

Definition

Export receptors are proteins that interact with RNA cargo, either by direct binding or by interaction with ► [export adapters](#). Export receptors transport the cargo through the nuclear pore complex by interacting with nucleoporins.

► [Nuclear Import and Export](#)

► [RNA Export](#)

Exportin

Definition

Exportin is a nuclear transport receptor protein which recognizes a nuclear export signal (NES).

► [Nuclear Import and Export](#)

Expressed Sequence Tag

Definition

Expressed sequence tags (EST) are short fragments of a sequence (100 – 1000 bp), derived from a cDNA clone, representing expressed genes. The EST frequency gives information on the expression level of the related genes. ESTs are organised into libraries, one library being the result of one experiment. ESTs are useful for identifying full-length genes, and serve as a landmark for mapping. ESTs can provide important evidence for the confirmation of predicted ► [open reading frames](#) (ORFs).

► [C. Elegans Genome, Comparative Sequencing](#)

► [DNA Chips](#)

- ▶ EST Mining for Expression Analysis
- ▶ Full Length cDNA Sequencing
- ▶ Immunological Methods, the Use of Antibodies to Discover the Function of Novel Gene Products
- ▶ Microarrays in Pancreatic Cancer

Expression

Definition

Expression refers to proteins that are produced (and present) in the cell at a given cell cycle stage and environmental conditions.

- ▶ Proteomics in Microfluidic Systems

Expression Difference Mapping

Definition

Expression difference mapping denotes a method of comparing the differentially expressed proteins in different sample sets by (Surface Enhanced Laser Desorption/Ionization) SELDI technology.

- ▶ Mass Spectrometry: SELDI

Expression Landscape / Mountains

Definition

An expression map of the entire genome is obtained by simultaneously studying the expression of individual genes under different microarray conditions. Studying multiple (ideally all) genes of a genome simultaneously allows coregulated genes to be identified. Clusters of co-regulated genes from numerous different microarray experiments, which are measured under a variety of experimental settings, are called expression mountains.

- ▶ *Caenorhabditis Elegans* as a Model Organism for Functional Genomics
- ▶ *C. Elegans* Genome; Comparative Sequencing

Expression Profile

Definition

Expression profile comprises of the global changes in expression levels of mRNAs comparing experimental

and reference cell or tissue samples usually measured on DNA microarrays. The method used is referred to as expression profiling.

- ▶ *Caenorhabditis Elegans* as a Model Organism for Functional Genomics
- ▶ Genomic Information and Cancer
- ▶ Microarray Data Analysis
- ▶ Microarrays in Colorectal Cancer
- ▶ Microarrays in Pancreatic Cancer

Expression Screening

- ▶ Automated High Throughput Functional Characterization of Human Proteins

Expression System

Definition

A system to produce recombinant proteins in prokaryotic or eucaryotic cells.

- ▶ Protein Tags

Expression Vector/Expression Construct

Definition

Expression vector/ expression construct is a plasmid coding for a gene product or an autonomous DNA molecule into which foreign DNA can be introduced. The vector contains the necessary regulatory sequences to allow transcription and translation of a cloned gene or genes, and thus transcribe and clone DNA. Expression vectors can be delivered to eukaryotic cells by naked nucleic acid transfer viral or non-viral vector.

- ▶ Clinical Gene Transfer
- ▶ Functional Assays
- ▶ High-Throughput Approaches to the Analysis of Gene Function in Mammalian Cells
- ▶ Protein Microarrays as Tool for Protein Profiling and Identification of Novel Protein-Protein Interactions

Expressivity

Definition

Expressivity refers to variability in the phenotypic manifestation of a genotype.

► [Cell Polarity](#)

Extinction Coefficient (ϵ)

Definition

A parameter that describes the intensity of light absorption by a given substance, at a specified wavelength. The extinction coefficient correlates between the concentration (C , molar) of the substance, the optical path length (b , in cm), and the absorption measured (A), in the following equation. $A = \epsilon b[C]$, where ϵ units are $(\text{cm molar})^{-1}$.

► [Amino Acids: Physicochemical Properties](#)

Extracellular Matrix

JAN F. TALTS¹, PETER EKBLOM²

¹Animal and Veterinary Basic Sciences,
The Royal Veterinary and Agricultural University,
Frederiksberg, Denmark

²Department of Cell and Molecular Biology,
Lund University, Lund, Sweden
jft@kv1.dk,
peter.ekblom@medkem.lu.se

Definition

► [Extracellular matrix](#) (ECM) is the name given to the molecular meshwork located between cells in multicellular organisms. It is comprised of sugars and proteins organized into various supramolecular assemblies. Small soluble molecules permanently bound to ECM, such as growth factors, are important constituents of the extracellular microenvironment, but are not normally counted as part of the ECM.

Characteristics

The amount and composition of ECM varies between different tissues. In epithelium or muscle the ratio of tissue volume occupied by ECM to that occupied by cells is less than 1:10, whereas the ratio is more than

10:1 in tendons, cartilage and bone. The ECM provides structural support for cells and also tensile strength, elasticity and compression resistance. Examples include cushioning in cartilage, strength in tendons, elasticity in blood vessels and formation of tissue boundaries and anchorage of cells in epithelial organs. In synergy with growth factors or by themselves, ECM molecules are essential for several signaling events. Thus ECM composition needs to be radically different in different tissues to support these varying functions.

Constituents of the ECM

The molecules comprising ECM can be subdivided into three main categories, proteoglycans, glycoproteins and ► [collagens](#). All morphologically different types of ECM contain some members of these categories. These categories can, in turn, be subdivided based on expression pattern, sugar side chain composition, interactions and typical features of the protein core sequence (1). A recently proposed classification of proteoglycans and glycoproteins (3) is used here. Glycoproteins are divided into glycoproteins of the interstitial connective tissue or ► [basement membrane](#) glycoproteins, whereas proteoglycans are divided into small leucine-rich-, modular nonhyaluronanbinding- or modular hyaluronan and lectin binding-proteoglycans (hyalactans) (Table 1). A few glycoproteins of the interstitial connective tissue, such as fibronectins or fibulin-1 and in disease conditions tenascins are also plasma proteins. Other proteins, more often thought of as plasma proteins, such as vitronectin, fibrinogen, thrombospondin and von Willebrand factor can in their turn be associated with the ECM.

Collagens are a large diverse family of ECM molecules, which account for about one third of the total protein mass in vertebrates. Twenty seven different collagen types have been identified. The many diseases caused by mutations in collagen genes have amply demonstrated their important roles (5). The basic subunits of the collagens are the α -chains. These α -chains contain one or more segments with multiple glycine-X -Y sequence motifs, also known as the collagen sequence. The amino acids X and Y are often proline, hydroxyproline, and hydroxylysine, unique hydroxylated versions of proline and lysine. Three α -chains form the triple helix typical of collagens. Collagens can be divided into fibril-forming and non-fibrillar collagens, but both types consist of homo and heterotrimers of α -chains. In the non-fibrillar collagens these collagen α -helices are often interrupted by non-collagenous sequences. The fibril-forming collagens are synthesized as precursors and processed into mature collagen, followed by fibrillogenesis in which the molecules assemble head to tail and aggregate in a staggered manner into rope-like structures. The non-fibrillar collagens can be divided into microfibrillar

Extracellular Matrix. Table 1 Division of structural glycoproteins and proteoglycans of the ECM into subcategories

Glycoproteins	Basement membranes	Interstitial connective tissue	
	Fibulins	Elastin	
	Laminins	Fibulins	
	Nidogens/Entactin	Fibrillins	
		Fibronectins	
		Matrilins	
		Tenascins	
		Thrombospondins	
Proteoglycans	Small leucine-rich	Modular non-hyaluronan binding	Hyalectans
	Asporin	Agrin	Aggrecan
	Biglycan	Perlecan	Brevican
	Chondroadherin	Testican	Neurocan
	Decorin		Versican
	Epiphygan		
	Fibromodulin		
	Keratocan		
	Lumican		
	Opticin		
	Osteoadherin		
	Osteoglycin		
	PRELP		

collagen VI, network-forming collagens, fibril associated collagens with interrupted triple helix (FACIT), multiplexins, anchoring fibril collagen VII and transmembrane collagens (Table 2).

Organization of the ECM

Most structural ECM proteins are chimeric and share similar domains with each other. Many of them form homotypic aggregates to which other proteins bind, thus forming intricate networks. These highly organized polymers do by themselves determine the specific histoarchitecture of tissues, but they also provide cells with information by binding to receptors. Tissue and cell type specific functionality is due to expression of different family members or splice variants.

Morphologically, two main forms of ECM can be discerned, basement membranes and the interstitial ECM. The interstitial ECM can be divided into subforms. Basement membranes underlie epithelial and endothelial cells and surround Schwann, muscle and fat cells. They are composed of two

self-assembling networks of heterotrimeric molecules, collagen IV and laminin. Both these molecules exist as protein families. Thus, several organ or cell-type specific basement membranes are produced. During formation of basement membranes, these networks interconnect through other molecules such as nidogens and perlecan to form the final mat-like structure (4). Basement membrane molecules have been found in all metazoan phyla examined, suggesting that this molecular structure is at least 500 million years old. In contrast, many interstitial ECM components emerged later (2). The evolutionary constancy and genetic data on embryonic development in model organisms strongly suggest that the invention of basement membranes was one essential step for the evolution of multicellular life. Two other collagens, VII and XVII, are associated with specialized structures of basement membranes underlying stratified epithelia, the anchoring fibrils and hemidesmosomes respectively. Mutations in the collagens, the skin basement membrane ► [laminins](#), receptors or intracellular intermediate

Extracellular Matrix. Table 2 Division of collagens into subcategories

	Types
Fibril forming collagens	I, II, III, V, XI, XXIV, XXVII
Microfibrillar collagen	VI
Network forming collagens	IV, VIII, X
FACIT	IX, XII, XIV, XVI, XIX, XXI, XXII, XXVI
Multiplexins	XV, XVIII
Transmembrane collagens	XIII, XVII, XXIII, XXV
Anchoring fibril	VII

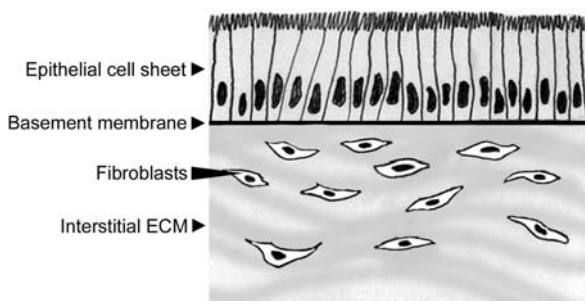
filaments connecting to the hemidesmosomes, lead to various forms of skin blister diseases ([▶epidermolysis bullosa](#)).

The interstitium is the space beneath epithelial sheets, and is composed mainly of ECM (Fig. 1). A similar arrangement is seen under endothelial cells. The word mesenchyme is frequently used for embryonic interstitium and the word stroma for interstitium surrounding tumor cells, but these words (interstitium, stroma, mesenchyme) are often used interchangeably. The main constituent in interstitial ECM is collagen I, produced by fibroblasts located within the interstitium. Other major proteins in this meshwork are fibronectin, fibulins and fibrillins (1, 7). Fibronectin mediates a wide variety of cellular interactions and plays an important role in cell adhesion, migration, growth and differentiation. These functions are especially apparent in matrices under reconstruction, such as during wound healing and in development. Fibulins do not form large homotypic aggregates, but can join other supramolecular structures as diverse as basement membranes, elastic fibers, microfibrils and proteoglycan aggregates. To some degree, these structures are thereby stabilized,

but more importantly the biological functions are modified. Mutational studies in mice have revealed that fibulins are essential for the development of larger or smaller blood vessels (7). Mutations in fibrillin-1 lead to [▶Marfan's syndrome](#) in man, which manifests mainly as skeletal abnormalities, ocular changes and cardiovascular problems. The most serious are the cardiovascular problems, in particular the lethal ruptures of the aorta. Marfan's syndrome is the most prevalent hereditary extracellular matrix disorder, with an estimated incidence of 1 in 7000 and an autosomal dominant inheritance with variable penetrance.

A major stromal ECM subtype is produced by cartilage cells, chondrocytes. Cartilage is specialized for a supportive role in the body. Three types of cartilage can be distinguished, articular (or hyaline) cartilage, elastic cartilage and fibrocartilage. Articular cartilage is found on the articulating surfaces of bones and plays a fundamental role in low-friction movement as well as in absorbing and distributing mechanical load. It forms most of the entire appendicular and axial skeleton in the embryo. The cartilage ECM consists of a fibrillar network based on collagen II fibrils dispersed in an environment of highly negatively charged [▶aggrecan](#) molecules that form large supramolecular aggregates. In this fibrous network, large extremely anionic proteoglycans provide a fixed charge density that maintains a high osmotic pressure, thereby promoting water retention. Surrounding the chondrocytes is a thin layer of pericellular ECM containing molecules that signal cells by interacting with cell surface receptors. Further from the cell, a territorial ECM, consisting of molecules with roles in catalyzing and regulating ECM assembly, is found. Still further out, the interterritorial ECM mainly appears to play a role in contributing mechanical properties.

In elastic cartilage there are high levels of elastic fibers composed mainly of fibrillin and elastin, where elastin is laterally packed in thin ordered filaments in the elastic fiber core and fibrillins together with other



Extracellular Matrix. Figure 1 Scheme of the two main extracellular matrix forms, the thin basement membrane and the interstitial ECM. The basement membrane is thin and underlies epithelial cells, whereas the interstitial ECM surrounds cells and occupies much space.

proteins constitute the associated microfibrils. Elastic fibers are otherwise ubiquitous throughout the body. Fibrocartilage shows many similarities to interstitial connective tissue. It is often interposed between other tissues and articular cartilage, tendons or ligaments. The main constituent in this type of ECM is collagen I. A highly specialized form of stromal ECM is found in bone. Bone is a connective tissue with cells and fibers embedded in a hard and largely unbending substance, well suited for supportive and protective functions. The ECM of bone is composed of collagen I fibers and small quantities of sulfated glycosaminoglycans and several glycoproteins produced by osteoblasts. The collagen fibers are organized in spirals around a central canal. These spirals are also arranged at right angles to each other in every second layer. A unique feature of bone is the inorganic phase, the hydroxyapatite crystals, which are deposited in the osteoid structure as slender needles within the collagen fiber network. Other ECM components play important roles in regulating the deposition and organization of hydroxyapatite.

Molecular Interactions

ECM-ECM Interactions

The major basement membrane molecules, the laminins, the collagen IVs, the nidogens and perlecan all bind each other with various affinities depending on isoform. They can also bind to the cell membrane receptors, initializing signaling cascades and cytoskeletal rearrangements. The emerging picture of basement membrane assembly is that laminin binding to cell surface receptors leads to polymerization of the laminins into a network. Then, collagen IV forms a second network, connecting to laminins *via* nidogens. This connection is greatly stabilized by interactions with perlecan.

Perlecan is a large proteoglycan expressed in both cartilage and basement membranes. It interacts with a number of other proteins including ►integrins, ►dystroglycan, collagens and laminins and also sequesters growth factors, particularly those of the fibroblast growth factor family. The majority of perlecan-null mice die due to subtle basement membrane defects during day 10–12 of embryonic development. Basement membranes are formed, but are leaky. Similar leaky basement membranes in the absence of perlecan can be observed in blood vessels after activation of perlecan-degrading proteinases. Surviving mice show extensive chondrodysplasia with abnormal endochondral ossification. Cartilage formation appears normal, but turnover shows major alterations.

In articular cartilage, hyaluronan is the central molecule to which several aggrecan molecules are connected through link proteins. Aggrecan is a large proteoglycan with many sugar side chains. It is linked to other

aggrecan molecules by interactions with fibulin-2. In tissue, collagen fibers can be demonstrated in the interior of the aggregate. Moreover, it has been demonstrated *in vitro* that aggrecan through its keratan sulfate side chains binds tightly to collagen, thus linking the two major molecular aggregates in cartilage to each other. Many molecules interact with collagen and have the ability to modulate fibril assembly. Two examples of collagen binding proteins are thrombospondin-5 and decorin. Thrombospondin-5 is a pentameric molecule with several binding sites on collagen II. However, the molecule is not large enough to span the distance between the binding sites on the same molecule, suggesting that it would link neighboring collagen molecules to each other. The protein cores of leucine-rich repeat (LRR) proteins are multiple repeats of compact horseshoe-shapes, a conformation suitable for protein-protein interactions. The LRR protein decorin can inhibit collagen fibril formation *in vitro*, and decorin null-mice have fragile skin with reduced tensile strength and larger irregular collagen fibrils, suggesting a role in fibril assembly and maintenance of the fibrillar network.

Interactions with Cell Surface Receptors

The integrin family, containing more than 20 members, was the first identified family of ECM receptors. The first discovered integrin bound to fibronectin, but it was soon found that several ECM components bind to similar types of receptors and that several receptors can bind more than one ECM component. All integrins are transmembrane proteins composed of non-covalently linked heterodimers of an α and a β subunit. All integrins containing an α_v or a β_1 subunit bind ECM proteins. Inside the cell, integrins bind signaling molecules and other proteins connecting to actin, thus linking the ECM to the cytoskeleton. The importance of cell attachment to ECM is shown by the prevention of apoptosis in anchorage-dependent cells by integrin $\alpha_5\beta_1$ adhesion to fibronectin. Further examples of the importance of integrin-ECM interactions are bone resorption where osteoclasts bind osteopontin through their integrin $\alpha_v\beta_3$ receptors and integrin binding to laminin isoforms during basement membrane assembly, as occurs during epithelial, endothelial and muscle development (1, 4).

Another group of cell-surface associated receptors, the ►syndecans, are heparan sulfate proteoglycans and act as co-receptors for growth factor receptors. However, they are also known to bind ECM glycoproteins such as tenascin-C, fibronectins and laminins. Intracellularly they are also connected to the cytoskeleton and signaling pathways involving protein kinase C.

The dystroglycan complex (DGC) has emerged as another type of receptor linking the cytoskeleton to the

ECM. The DGC is an important receptor system for ECM in muscle and epithelia. In muscle, mutations in any of the DGC components or in their extracellular ligand laminin $\alpha 2$ chain will lead to different forms of muscular dystrophies. In epithelial development, the main ligand seems to be laminin-1.

Regulatory Mechanisms

Breakdown of ECM by proteinases is an essential step in embryogenesis, tissue remodeling and metastasis. These events are often associated with elevated production and activity of ►[matrix metalloproteinases](#), of which over 20 have been identified. These proteinases can degrade the main components of ECM and can thus profoundly remodel the ECM (6). The positive regulation of ECM has thus far not been extensively studied. However, tissue homeostasis requires a fine balance of catabolic and anabolic signals. A few examples are given.

Epithelial differentiation correlates with increased mRNA expression of laminin-1 and collagen IV. In embryoid bodies, a model system simulating early embryonic development, fibroblast growth factor induced differentiation is followed by activation of PI3-kinase and protein kinase B (Akt/PKB). Among the downstream targets of Akt/PKB are important transcriptional regulators, such as the FOX multigene family and NF κ B. Akt/PKB activation can thus potentially regulate gene transcription. Synthesis of laminin-1 and collagen IV increased markedly in this and other systems with constitutively active Akt/PKB or the catalytic subunit of PI3K, whereas dominant negative Akt/PKB inhibited transcription of laminin $\beta 1$ and collagen IV. Hence, in many cell types, Akt/PKB activation stimulated by receptor tyrosine kinases might be involved in the positive transcriptional regulation of basement membrane components, including laminin-1.

Upon tissue injury, inflammation occurs with concomitant release of cytokines. Transforming growth factor- β (TGF- β) stimulates fibroblasts and other cells to proliferate and synthesize ECM components. This leads to provisional repair, which under normal conditions results in remodeling and regeneration. A balance in the synthesis of ECM proteins, matrix metalloproteinases and tissue inhibitors of these has to be reached. For collagen I, it is known that its two genes are regulated by TGF- β through the TGF- β -activator and the Smad pathways. Many of the provisional or

stromal ECM molecules such as collagen I, fibulins and tenascins are down-regulated by glucocorticoids. Since glucocorticoids also have effects on TGF- β transcription and secretion it is possible that they affect ECM by down-regulating this growth factor. However, ECM synthesis is not only regulated by different growth factors, but also by the mechanical forces, which directly affect transcription by now partially understood mechanisms.

The mineralized part of bone is the inert skeleton that remains after death, but bone is not a metabolically inert part of the living body. On the contrary, throughout life, bone is actively remodeled in a dynamic process. The catabolic phase includes the triggering of osteoblasts to remove the thin layer of osteoid and to condition the bone ECM to recruit osteoclast precursor cells. This activity includes the deposition of osteopontin, a cell-binding ECM protein. The osteoclast precursor cells develop into resorbing osteoclasts, which dissolve the mineral and digest the organic ECM. The resorption pits resulting from osteoclast activity recruit novel osteoblasts essential for new bone tissue formation. An important feature in skeletal biology is a variable mechanical load that induces cellular responses, leading to remodeling.

►[Bone and Cartilage](#)

►[Bone Diseases and Skeletal Disorders, Genetics](#)

References

1. Aumailley M, Gayraud D (1998) Structure and biological activity of the extracellular matrix. *J Mol Med* 76:253–265
2. Hutter H, Vogel BE, Plenefish JD et al (2000) Conservation and novelty in the evolution of cell adhesion and extracellular matrix genes. *Science* 287:989–994
3. Iozzo RV, Murdoch AD (1996) Proteoglycans of the extracellular environment: clues from the gene and protein side offer novel perspectives in molecular diversity and function. *FASEB J* 10:598–614
4. Kalluri R (2003) Basement membranes: Structure, assembly and role in tumor angiogenesis. *Nat Rev Cancer* 3:422–433
5. Myllyharju J, Kivirikko KI (2004) Collagens, modifying enzymes and their mutations in humans, flies and worms. *Trends Genet* 20:33–43
6. Sternlicht MD, Werb Z (2001) How matrix metalloproteinases regulate cell behavior. *Annu Rev Cell Dev Biol* 17:463–516
7. Timpl R, Sasaki T, Kostka G et al (2003) Fibulins: a versatile family of extracellular matrix proteins. *Nat Rev Mol Cell Biol* 4:479–489

Extrachromosomal Elements

Definition

Extrachromosomal elements are well-defined genetic elements that are not stable constituents of cellular chromosomes, by virtue of their ability to move within chromosomes or to replicate independently of them.

► [Translational Frameshifting, Non-standard Reading of the Genetic Code](#)

Ezrin

Definition

Ezrin is a member of the ezrin/radixin/moesin family of proteins that serve to directly link certain membrane receptors to the cytoskeleton.

► [Adhesion Molecules](#)

► [ERM Protein](#)

► [Focal Adhesions](#)

Fabrication

Definition

Fabrication is the art of building or making from raw materials.

► [Proteomics in Microfluidic Systems](#)

Fabry Disease

Definition

Fabry disease is an X-linked recessive lysosomal storage disease with cardiovascular, renal and lenticular abnormalities, resulting from the deficient activity of the exogalactosidase, alpha-galactosidase A (alpha-Gal A). Phenotypic expression in female heterozygous carriers depends on ► [lyonization](#).

FACS

Fluorescence activated cell sorting, ► [flow cytometry](#).

F-Actin

► [Actin Cytoskeleton](#)
► [Tangier Disease](#)

Factor Inhibiting HIF-1

► [FIH](#)

FAD

► [Familial Alzheimer's Disease](#)
► [Alzheimer's Disease](#)

FADD

► [Fas-Associated Death Domain](#)

FAK

Definition

FAK stands for Focal Adhesion Kinase. It is a non-receptor proteintyrosine kinase implicated in signaling pathways involved in cell motility, proliferation and apoptosis. Activated by tyrosine-phosphorylation in response to either integrin clustering induced by cell adhesion or antibody cross-linking, or via G-protein coupled receptor (GPCR) occupancy by ligands such as bombesin or lysophosphatidic acid, or via LDL receptor occupancy. FAK plays a potential role in oncogenic transformations resulting in increased kinase activity.

► [Signal Transduction: Integrin-Mediated Pathways](#)

False Positives/False Negatives

Definition

False positives (or false negatives, respectively) exist when laboratory tests or biometric scans report, incorrectly, that they have found a signal where none exists in reality (or found no signal although one exists in reality). For example, a positive test for AIDS or cancer, when the person was disease free, would be a false positive. In a ►[two-hybrid screen](#), false positives are physiological protein-protein interactions, which cannot be confirmed by another method.

►[Two-Hybrid System](#)

Familial Adenomatous Polyposis

Definition

Familial adenomatous polyposis (FAP) is an autosomal dominantly transmitted syndrome associated with an increased risk of colorectal cancer. It is characterized by the development of hundreds to thousands of adenomatous polyps of the colorectum during the second and third decade of life. ~80% of the FAP patients harbour truncating germ-line mutations in the ►[APC](#) (Adenomatous Polyposis Coli) tumor suppressor gene.

- [Colorectal Cancer](#)
- [Gut Epithelium](#)
- [Hereditary Nonpolyposis Colorectal Cancer](#)
- [Predictive Genetic Testing](#)

Familial Aggregation

Definition

Familial aggregation refers to the tendency for a trait such as diabetes to appear more frequently than expected in specific families.

Familial Alzheimer Disease

►[FAD \(Familial Alzheimer Disease\)](#)

Familial Amyotrophic Lateral Sclerosis

Definition

Amyotrophic lateral sclerosis (ALS) is a fatal neurological disease that attacks the nerve cells responsible for controlling voluntary muscles. The disease belongs to a group of disorders known as motor neuron diseases, which are characterized by the gradual degeneration and death of motor neurons. Mutations in Cu/Zn superoxide dismutase-1 (SOD) have been identified as a genetic cause of the inherited familial form of ALS.

►[Peptidyl Prolyl Cis/Trans Isomerases](#)

Familial Combined Hyperlipidemia

Definition

Familial combined hyperlipidemia is an inherited disorder of high serum cholesterol and/or high blood triglycerides. Phenotype of IIa, IIb, IV hyperlipoproteinemia (increase in LDL and/or VLDL) is concurrently seen in the same family. VLDL is overproduced and small size LDL are present. This disorder predisposes the person to a greater risk of early coronary artery disease. Frequency of this syndrome is 1:50-1:100 in the general population. Autosomal dominant inheritance is suspected, but the causative gene locus has not yet been determined.

- [High-HDL Syndrome](#)
- [Familial Hypercholesterolemia](#)

Familial DCM

Definition

Familial DCM refers to an (inheritable) ►[Dilated cardiomyopathy](#) (DCM) caused by an inherited gene mutation.

►[Familial Dilated Cardiomyopathy](#)

Familial Dilated Cardiomyopathy

DIANE FATKIN

Sr Bernice Research Program in Inherited Heart Diseases, Molecular Cardiology Unit, Victor Chang Cardiac Research Institute, Darlinghurst, NSW, Australia
d.fatkin@victorchang.unsw.edu.au

Definition

Cardiomyopathies are diseases of ►heart muscle that are associated with cardiac dysfunction. These disorders are classified on the basis of morphology and function into four categories: hypertrophic cardiomyopathy, dilated cardiomyopathy, arrhythmogenic right ventricular dysplasia and restrictive cardiomyopathy. ►Dilated cardiomyopathy (DCM) is characterized by dilation and contractile dysfunction of the left (± right) ventricles. DCM may result from a variety of conditions that impair cardiomyocyte function or cause cardiomyocyte injury or loss, including myocardial ischaemia, infection, inflammation, increased pressure or volume load, and toxic agents. In approximately 50% cases, an underlying cause is unable to be identified and DCM is termed ►idiopathic. Over the past decade, it has been recognized that genetic factors are also an important cause of DCM. When DCM is caused by an inherited gene defect in families, it is known as ►familial DCM. Studies to date have concentrated on ►monogenic forms of familial DCM that are caused by mutations in a single gene. The extent to which genetic factors may alter the susceptibility to acquired causes of DCM or modify disease ►phenotypes has not yet been elucidated.

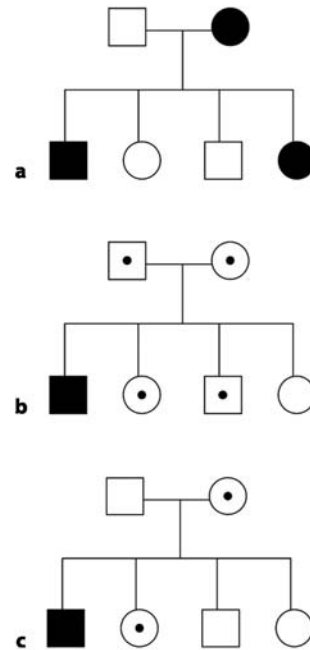
Characteristics

Prevalence

The first studies to evaluate the prevalence of familial DCM were based on obtaining a positive clinical history in the relatives of ►probands and had a low yield (<10%). Subsequent studies have systematically evaluated first-degree relatives using physical examination, ECG and transthoracic echocardiography. It is now estimated that at least 20-35% cases of idiopathic DCM are likely to have a genetic basis (1).

Inheritance

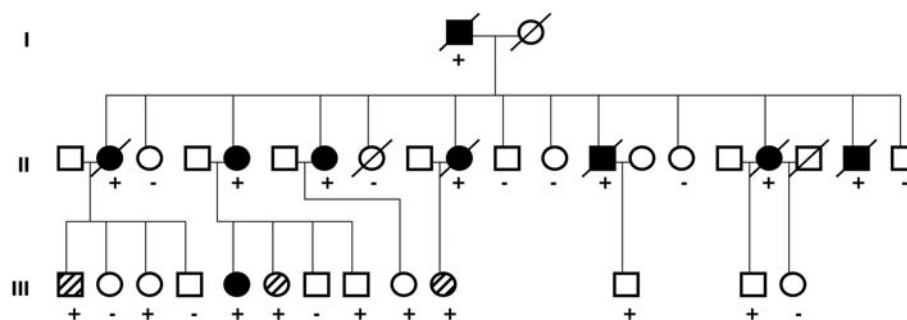
Families with DCM can show autosomal dominant, autosomal recessive, X-linked or maternal modes of inheritance (Fig. 1). Autosomal dominant inheritance is observed most frequently. Several factors may complicate determination of inheritance patterns in familial DCM pedigrees. An apparent absence of disease in younger generations of a family may not necessarily reflect ►genotype status, but rather, the phenomenon of age-related ►penetrance, ie. lack of disease manifestations in genotype-positive individuals until after a certain age (Fig. 2). In autosomal dominant pedigrees, an unaffected individual who has an affected child may be non-penetrant, ie. genotype-positive but phenotype-negative throughout life. Alternatively, DCM in the child may be attributable to a cause other than the familial gene mutation, ie. ►phenocopy. Although



Familial Dilated Cardiomyopathy. Figure 1

Inheritance patterns in families with DCM. Squares denote male family members; circles denote female family members. Clinically affected individuals are shown with solid symbols; clinically unaffected individuals who carry one mutant *allele* are shown by clear symbols marked with a dot. Autosomal dominant inheritance (a) is the most common pattern. In autosomal dominant DCM, affected individuals carry one mutant allele and one normal allele of the disease gene. One mutant allele is sufficient to cause disease. Both males and females can be affected. Each child of an affected parent has a 50% chance of being affected. Autosomal recessive DCM (b) occurs when two individuals who each carry one mutant allele have a child that inherits two mutant alleles. Individuals with one mutant allele are clinically normal and two mutant alleles are required to develop disease. In such a family, each child has a 25% chance of inheriting two normal alleles, a 50% chance of inheriting one mutant allele and a 25% chance of inheriting two mutant alleles. Both males and females can be affected. X-linked DCM (c) results from inheritance of a mutant allele of a gene located on the X-chromosome. If a mother carries a mutant allele, there is a 25% chance of each of four possible outcomes for her children: an affected boy, a normal boy, a carrier girl, or a normal girl. If a father carries a mutant allele, all sons will be normal and all daughters will be carriers.

the traditional view is that female carriers in X-linked disorders remain phenotypically normal, recent observations indicate that these individuals can develop symptoms and signs of DCM from late adolescence to older adult life. Overt expression of disease in female



Familial Dilated Cardiomyopathy. Figure 2 Pedigree of a 3-generation kindred with DCM and conduction-system disease. Symbols with a slash mark denote deceased individuals. Family members with conduction abnormalities without DCM are shown by striped symbols; family members with conduction abnormalities and DCM are shown by solid symbols. The presence (+) or absence (-) of a *LMNA* gene mutation is indicated below each symbol. In generation III, the majority of clinically affected individuals have conduction disturbances with normal left ventricular function. This is typical of the DCM + conduction-system disease phenotype, which is characterized by a prodrome of conduction disturbances from the second to fourth decades with the subsequent development of DCM. In generation III, the absence of clinical manifestations of disease in some of the young genotype-positive individuals can be explained by age-related penetrance. These individuals are at risk of developing overt disease with increasing age.

carriers may mask the presence of X-linked inheritance. Accurate determination of inheritance patterns is important for identification of “at risk” individuals who require cardiac screening, as well as for appropriate genetic counselling, and selection of candidate genes for mutation analysis.

Diagnosis

Familial DCM shows clinical variability. The age of onset and type of symptoms can vary between different families and within members of the same family. Families with DCM can be classified into one of the four phenotype subgroups:

- DCM alone
- DCM with ►conduction-system disease (sinus bradycardia, atrioventricular conduction block, ± supraventricular arrhythmias)
- DCM with skeletal myopathy, ±conduction-system disease
- DCM with other disorders, eg. sensorineural deafness.

Probands and family members should be evaluated by medical history, physical examination, 12-lead electrocardiography and transthoracic echocardiography. Affected individuals in families with the DCM-only phenotype generally present with symptoms and signs of congestive heart failure or cardiac arrhythmias, including dyspnoea, fatigue, palpitations, syncope, pulmonary venous congestion, and peripheral oedema. The diagnosis of DCM is based principally on the findings of increased left ventricular chamber diameters

and reduced fractional shortening on transthoracic echocardiography. In the subgroup of families with the DCM + conduction-system disease phenotype, the majority of affected individuals present with symptomatic or asymptomatic electrocardiographic changes in the second to fourth decades, with the subsequent development of DCM in later life (Fig. 2). Clinical signs of muscle weakness and wasting, and/or elevated serum creatine kinase levels, provide evidence of co-existent skeletal muscle involvement.

Natural History

The natural history of familial DCM is variable, with some individuals having a benign course, while others develop progressive heart failure that may require cardiac transplantation. Sudden death can occur at any age, irrespective of ventricular function. Individuals in families with DCM and conduction-system disease may develop progressive atrioventricular conduction block and require pacemaker implantation. Thromboembolic complications can result from left atrial or left ventricular blood stasis or may be precipitated by cardiac arrhythmias. Genotype-phenotype correlations in familial DCM have not yet been established and will require large populations of genotyped individuals. Although a number of clinical parameters have been proposed as prognostic factors in heart failure populations, it is likely that genotype will be the strongest determinant of outcome in familial DCM.

“Early Disease”

Because of the high prevalence of familial disease, current international guidelines recommend cardiac

screening of all first-degree relatives of probands with “idiopathic” DCM (2). As a result of implementation of screening strategies, a new population of individuals has been identified that have asymptomatic echocardiographic changes that do not fulfil standard criteria for the diagnosis of DCM. Isolated left ventricular dilation is the most common finding, occurring in up to 20% of individuals in families with DCM (3). The significance of these findings is not clear. It has been proposed that left ventricular dilation might represent a pre-symptomatic stage (“early disease”), and that further disease progression would be anticipated. Alternatively, these changes may represent a subclinical form of established disease or be caused by factors unrelated to familial disease. The natural history and optimal management of asymptomatic LV dilation in familial DCM is unknown and requires further evaluation.

Cellular and Molecular Regulation Chromosomal Loci and Disease Genes

Families with DCM have been evaluated genetically using genome-wide linkage analyses in large kindreds and candidate gene screening in small kindreds. Data from these studies have shown that familial DCM is a genetically heterogeneous disorder (4). To date, 32 chromosomal loci and 25 disease genes have been associated with the various forms of inherited DCM (Table 1). The list of gene mutations associated with DCM is already long and is growing rapidly. Surprisingly, very few mutations have been reported in each of these genes. Given the number of individuals in the cohorts screened, it appears likely that these known genes account for only a relatively small proportion of all cases of familial DCM. It is important to note that a number of these genes do not have robust genetic and/or functional evidence supporting disease causality. In some cases, a single mutation has been identified in one family, while in other cases, mutations have been identified in individuals in whom clinical and genetic data from family members are unavailable. Further studies are required to identify additional genes and to establish that putative DCM genes are in fact disease-causing.

Pathogenetic Mechanisms

► **Muscle** contraction is achieved by the sliding movement of thick filaments relative to thin filaments, which is mediated by the cyclical attachment and detachment of myosin cross-bridges. A number of factors regulate actin-myosin interaction, including the calcium-sensitive troponin-tropomyosin complex, intrasarcomeric cytoskeletal proteins and ATP hydrolysis (Fig. 3). Defective generation of contractile force due to mutations in various protein components of the sarcomere has been widely proposed as the

mechanism responsible for familial ► **hypertrophic cardiomyopathy**. Identification of mutations in genes encoding sarcomeric proteins (5), calcium-handling proteins (6) and proteins involved in myocardial energy regulation has indicated that impaired ► **force generation** can also cause familial DCM (Fig. 4). The question of why a number of genes can cause either hypertrophic cardiomyopathy or DCM is intriguing and has yet to be answered, but differences in calcium handling and/or myocardial energetics may be involved.

While some sarcomere protein gene mutations are located in residues involved in actin-myosin interactions, others are located in residues that are predicted to be involved in transmission of generated force to adjacent structures. The cardiomyocyte cytoskeleton provides an intracellular scaffolding that mediates transduction of mechanical force from the sarcomere to the extracellular matrix (Fig. 3). Mutations have been identified in a number of cytoskeletal proteins at various points along this pathway, providing strong support for impaired ► **force transmission** as a mechanism for familial DCM (Fig. 4).

Several studies have highlighted the important additional role of cytoskeletal proteins as determinants of passive myocardial wall properties. For example, titin is an intrasarcomeric protein that contributes to the structural organization of the sarcomere and myocardial elasticity. Mutations in titin are thought to have effects not only on active contraction but also on passive myocardial wall stress and elastic recoil (7). An interacting complex comprised of titin, muscle LIM protein (MLP) and T-cap, that is located at the sarcomeric Z disc, has been proposed to play a critical role in the cardiomyocyte's ability to sense mechanical stretch. DCM-causing mutations in the MLP or T-cap genes are predicted to disrupt this complex and impair stretch responsiveness, resulting in sustained alterations of passive stretch, left ventricular dilation and cell death (8). These remodelling changes reduce the efficiency of contraction and promote left ventricular systolic dysfunction (Fig. 4). Studies in mice suggest that myocardial energetics defects due to mitochondrial dysfunction associated with MLP deficiency could also contribute to contractile dysfunction.

The discovery of mutations in genes encoding the cardiac sodium channel (9), and the SUR2A subunit of the cardiac K_{ATP} channel (10), has recently shown that defective ion homeostasis is yet another disease pathway for familial DCM (Fig. 4). Cardiac K_{ATP} channels are important sensors of metabolic stress within cells. Defective cardiac K_{ATP} channel function compromises stress tolerance and may precipitate intracellular calcium overload and cell death.

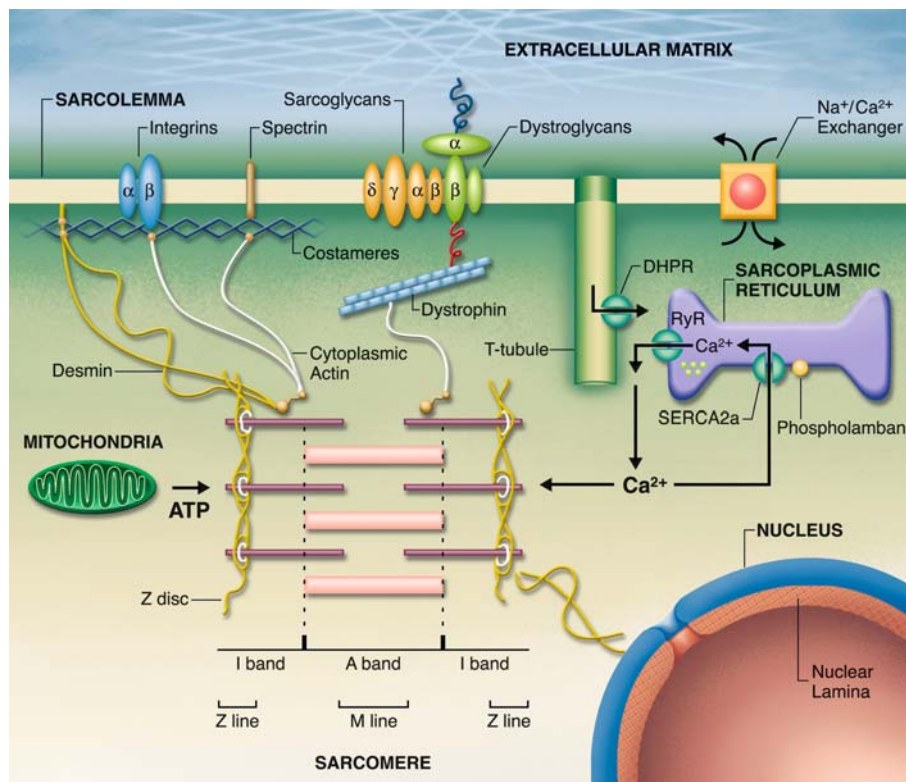
Mutations in the *LMNA* gene, that encodes the nuclear lamina proteins, lamins A and C, have been shown to

Familial Dilated Cardiomyopathy. Table 1 Chromosomal loci and disease genes associated with DCM

Chromosomal Locus	Gene	Protein
A. Autosomal Dominant Inheritance		
i) DCM only		
1q32	<i>TNNT2</i>	Cardiac troponin T
1q42-q43	<i>RYR2</i>	Cardiac ryanodine receptor
1q42-q43	<i>ACTN2</i>	α -actinin 2
2q31	<i>TTN</i>	Titin
2q35	<i>DES</i>	Desmin
3p21-p14	<i>TNNC1</i>	Cardiac troponin C
5q33	<i>SGCD</i>	δ -sarcoglycan
6q12-q16	?	?
6q22	<i>PLN</i>	Phospholamban
9q13-q22	?	?
9q22-q31	?	?
10q21-q23	?	?
10q22-q23	<i>VCL</i>	Metavinculin
10q22-q23	<i>Cypher/ZASP</i>	Cypher/ZASP
10q26	?	?
11p11	<i>MYBPC3</i>	Cardiac myosin binding protein C
11p15	<i>MLP</i>	Muscle LIM protein
12p12	<i>ABCC9</i>	SUR2A subunit, cardiac K _{ATP} channel
12q22	<i>TMPO</i>	Thymopoietin
14q12	<i>MYH7</i>	β -myosin heavy chain
14q12	<i>MYH6</i>	α -myosin heavy chain
15q14	<i>ACTC</i>	Cardiac actin
15q22	<i>TPM1</i>	α -tropomyosin
17q12	<i>TCAP</i>	T-cap
ii) DCM + Conduction-System Disease		
1p1-q21	<i>LMNA</i>	Lamins A and C
2q14-q22	?	?
3p21	<i>SCN5A</i>	Cardiac sodium channel
iii) DCM + Skeletal Myopathy \pm Conduction-System Disease		
1p1-q21	<i>LMNA</i>	Lamins A and C
6q23	?	?
iv) DCM + Sensorineural Deafness		
6q23	<i>EYA4</i>	Eya4

Familial Dilated Cardiomyopathy. Table 1 Chromosomal loci and disease genes associated with DCM (Continued)

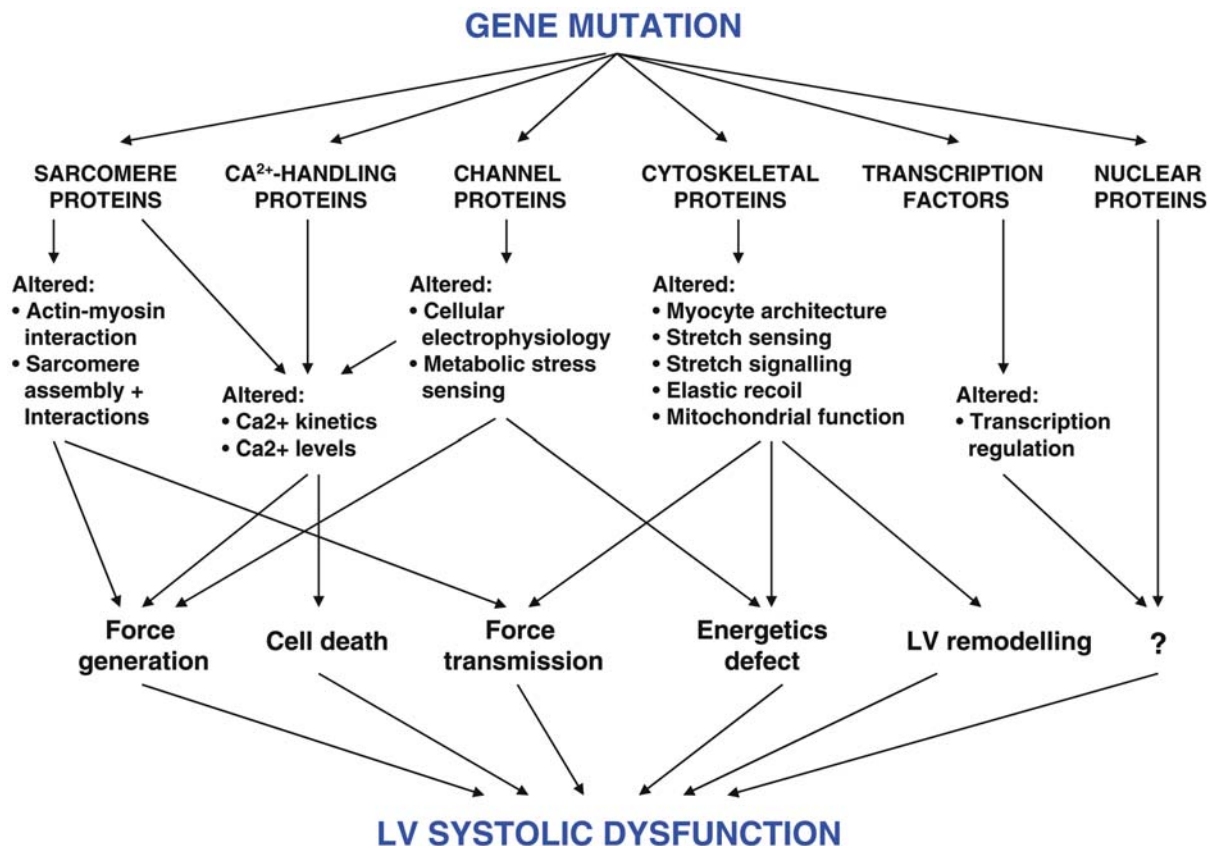
Chromosomal Locus	Gene	Protein
B. Autosomal recessive inheritance		
i) DCM only		
19q13	<i>TNNI3</i>	Cardiac troponin I
C. X-Linked Inheritance		
Xp21 (adult-onset DCM)	<i>DMD</i>	Dystrophin
Xq28 (infantile-onset DCM)	<i>G4.5</i>	Tafazzin



Familial Dilated Cardiomyopathy. Figure 3 Schematic of cardiomyocyte cytoarchitecture. The cardiac sarcomere is comprised of interdigitating thick and thin filaments. Sarcomeres are linked by a scaffolding of cytoskeletal proteins to the sarcolemma, intercalated discs, and nucleus. Muscle contraction results from the sliding movement of thick filaments with respect to thin filaments that is achieved by the cyclical attachment and detachment of myosin crossbridges. This process is energy-requiring and utilises ATP supplied by mitochondria. The intracellular free Ca^{2+} concentration has a critical regulatory effect on both muscle contraction and relaxation. DHPR, dihydropyridine receptor; RyR, ryanodine receptor. With permission reproduced from: Fatkin D, Graham RM. Molecular mechanisms of inherited cardiomyopathies. *Physiol Rev* 2002; 82:945–980.

cause 9 diverse disorders, including autosomal dominant DCM with conduction-system disease (11). The consequences of *LMNA* mutations and the mechanisms underlying the various tissue-specific phenotypes are not fully understood. Lamin A/C deficiency causes

changes in nuclear shape and size, as well as altered distribution and localization of heterochromatin. Based on these findings, three principal hypotheses for laminopathies have been proposed: altered gene expression, altered protein interactions, and increased



Familial Dilated Cardiomyopathy. Figure 4 Putative pathogenetic mechanisms causing familial DCM.

nuclear fragility and susceptibility to mechanical stress. These hypotheses are not necessarily mutually exclusive and it is likely that a number of factors are involved. The recent identification of mutations in the gene encoding the transcriptional coactivator, Eya4 (12), provides further evidence that altered transcription regulation can underlie familial DCM. The downstream genes involved and the pathophysiological basis for contractile dysfunction have yet to be determined. Although attempts have been made to identify a unifying paradigm for the pathogenesis of familial DCM, current data indicate that the molecular defects that trigger the disease process are quite diverse, with multiple interactions between the various putative disease pathways. In addition to these inherited molecular “triggers”, a number of compensatory and decompensatory responses to the presence of left ventricular dilation and contractile dysfunction can promote further disease progression. These responses include activation of neurohumoral factors, re-induction of embryonic genes, altered calcium homeostasis, apoptosis and left ventricular chamber remodelling. While an inherited gene defect is sufficient to cause DCM in families, a complex interaction of other

genetic and environmental factors are likely to modify the clinical manifestations of disease in individual family members.

Clinical Relevance

Heart failure due to DCM is a major health problem in the developed world. The number of individuals with idiopathic DCM has been estimated to be 5-8 cases per 100,000 per year. The incidence of heart failure increases with age, affecting 10 per 100 individuals aged 65 years and older. Despite advances in medical management, hospitalization rates have risen progressively over the past two decades and the 5-year survival after a diagnosis of heart failure is only 50%. Clearly, new approaches to the treatment of patients with heart failure are required. Studies of single gene defects that cause DCM in families may provide insights into pathophysiologic processes applicable to a wide range of more commonly-occurring, acquired forms of DCM. This information will also facilitate the development of new therapeutic strategies directed towards prevention or attenuation of the disease process itself, rather than alleviation of symptoms. Compilation of a comprehensive list of DCM disease

genes is an essential pre-requisite for understanding the molecular pathogenesis of this disorder and for the development of patient screening strategies. Since each family generally has a unique gene mutation, the coding sequence of all known disease genes needs to be analysed in every new proband. With current techniques, this process is time-consuming and expensive. Rapid, automated methods for large-scale genotyping will be required before molecular genetic information can be incorporated into clinical practice. Ultimately, the ability to genotype populations with familial DCM will enable genotype-phenotype correlations to be performed with determination of “high risk” and “low risk” mutations. This prognostic information will assist physicians in selection of appropriate therapies, timing of follow-up visits, counselling on lifestyle modifications, pregnancy counselling, etc. Genotyping would also enable clinical surveillance of the offspring of affected individuals to be rationalized, so that serial diagnostic investigations need only to be performed in genotype-positive individuals. Importantly, the ability to identify young genotype-positive individuals who are phenotype-negative or those who have early disease, provides the opportunity for preventative interventions, with direct benefits to patient outcome and health-care costs.

Acknowledgements

Dr Fatkin is supported by a Senior Medical Research Fellowship from the Sylvia and Charles Viertel Charitable Foundation.

References

1. Grunig E, Tasman JA, Kucherer H et al (1998) Frequency and phenotypes of familial dilated cardiomyopathy. *J Am Coll Cardiol* 31:186–194
2. Mestroni L, Maisch B, McKenna WJ et al (1999) Guidelines for the study of familial dilated cardiomyopathies. *Eur Heart J* 20:93–102
3. Baig MK, Goldman JH, Caforio ALP et al (1998) Familial dilated cardiomyopathy: cardiac abnormalities are common in asymptomatic relatives and may represent early disease. *J Am Coll Cardiol* 31:195–201
4. Fatkin D, Graham RM (2002) Molecular mechanisms of inherited cardiomyopathies. *Physiol Rev* 82:945–980
5. Kamisago M, Sharma SD, De Palma SR et al (2000) Mutations in sarcomere protein genes as a cause of dilated cardiomyopathy. *N Engl J Med* 343:1688–1696
6. Schmitt JP, Kamisago M, Asahi M et al (2003) Dilated cardiomyopathy and heart failure caused by a mutation in phospholamban. *Science* 299:1410–1413
7. Gerull B, Gramlich M, Atherton J et al (2002) Mutations of TTN, encoding the giant muscle filament titin, cause familial dilated cardiomyopathy. *Nat Genet* 30:201–204
8. Knoll R, Hoshijima M, Hoffman HM et al (2002) The cardiac mechanical stretch sensor machinery involves a Z disc complex that is defective in a subset of human dilated cardiomyopathy. *Cell* 111:943–955
9. McNair WP, Ku L, Taylor MRG et al (2004) SCN5A mutation associated with dilated cardiomyopathy, conduction disorder, and arrhythmia. *Circulation* 110:2163–2167
10. Biengraher M, Olson TM, Selivanov VA et al (2004) ABCC9 mutations identified in human dilated cardiomyopathy disrupt catalytic KATP channel gating. *Nat Genet* 36:382–387
11. Fatkin D, Mac Rae C, Sasaki T et al (1999) Missense mutations in the rod domain of the lamin A/C gene as causes of dilated cardiomyopathy and conduction-system disease. *N Engl J Med* 341:1715–1724
12. Schonberger J, Wang L, Shin JT et al (2005) Mutation in the transcriptional coactivator EYA4 causes dilated cardiomyopathy and sensorineural hearing loss. *Nat Genet* 37:418–422

Familial Endocrinopathy

Definition

Familial endocrinopathy describes a disorder in the function of an endocrine gland, which affects more members of the same family than can be accounted for by chance.

►Hyper- and Hypoparathyroidism

Familial Exudative Vitreoretinopathy

Definition

Familial exudative vitreoretinopathy (FEVR) is a hereditary ocular disorder characterized by incomplete vascularization of the peripheral retina. Dragged retina, exudation, vitreous hemorrhage, and retinal detachment may develop. One form of FEVR is caused by defects in Frizzled-4 (Fz4), a presumptive Wnt receptor.

►Wnt/Beta-Catenin Signaling Pathway

Familial Hypercholesterolemia

HENRICK K. JENSEN¹, TROND P. LEREN²

¹Department of Cardiology, Aarhus University Hospital, Skejby Hospital, Aarhus, Denmark

²Medical Genetics Laboratory, Department of Medical Genetics, Rikshospitalet University Hospital, Oslo, Norway

hkjensen@dadlnet.dk, trond.leren@rikshospitalet.no

Synonyms

Autosomal dominant hypercholesterolemia; Hereditary xanthomatosis

Definition

Familial hypercholesterolemia (FH) is an autosomal dominant inherited disorder of ▶**lipoprotein** metabolism characterized by very high plasma concentrations of low-density lipoprotein (LDL) cholesterol, tendon ▶**xanthomas** and increased risk of premature coronary heart disease (CHD). FH is caused by mutations in the LDL ▶**receptor** gene.

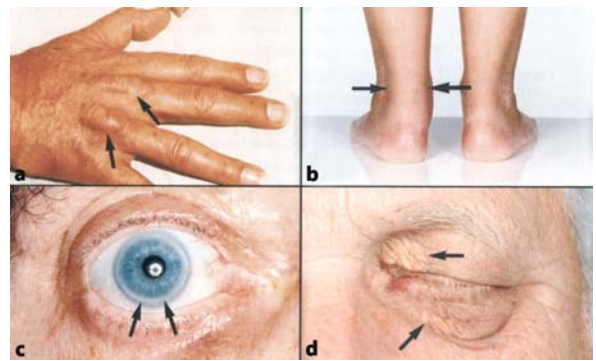
Characteristics

In the late 1930s Müller and Thannhauser recognized an inherited disease with elevated plasma cholesterol, cholesterol accumulation in tendons, and presence of CHD at an early age. Their suggestion that there might be a genetic basis for the hypercholesterolemia was substantiated by several family and metabolic studies during the following years (1). These studies laid the basis for Brown and Goldstein's discovery of the receptor for LDL. This LDL receptor is a cell-surface glycoprotein that plays a pivotal role in the regulation of LDL metabolism (1). For their important and brilliant work Brown and Goldstein were awarded the Nobel Prize for Physiology or Medicine in 1985.

Mutations in the gene encoding the LDL receptor give rise to FH and to date more than 800 pathogenic mutations have been reported (▶<http://www.ucl.ac.uk.fh>; ▶<http://www.umd.necker.fr>). FH is one of the most common autosomal dominantly inherited disorders in humans with an estimated prevalence of heterozygous FH of about 0.2% (1:500) among Caucasians. In contrast, the prevalence of ▶**homozygous** FH is about one in a million. The frequency of heterozygous FH is much higher in certain populations, i.e. one in 67 in the Ashkenazi Jews of Lithuanian descent, one in 100 in the South African Afrikaners, and one in 270 in the French Canadians. These high prevalences are due to ▶**founder gene effects**. Worldwide, an estimated ten million people are afflicted with FH.

In ▶**heterozygotes**, hypercholesterolemia is the earliest manifestation and is present at birth in nearly all subjects. Plasma LDL cholesterol levels are two- to three-fold above normal. Xanthomas especially in the Achilles tendons, in the extensor tendons of the hand, and in the patellar tendons are hallmarks of FH (Fig. 1). They are virtually specific for the disease, and appear with increasing frequency with age. By age 40, approximately 30% of heterozygous FH patients have xanthomas. ▶**Corneal arcus** and ▶**xanthelasmata** are commonly seen but are not specific for FH (Fig. 1).

In untreated FH, premature CHD typically occurs from the third to the fifth decade of life. In a recent



Familial Hypercholesterolemia. Figure 1

Extravascular lipid deposits. (a + b) Tendon xanthomas in the extensor tendons of the hand and in the Achilles tendon, respectively; (c) Corneal arcus; (d) Xanthelasmata.

prospective study, heterozygous FH was associated with a substantial excess mortality from CHD especially in young adults. The data indicate that approximately 50% of FH males and about 20% of females with FH die before age 40. Similar statistics are seen for deaths from all causes because most deaths in FH heterozygotes are due to CHD. Compared to the general population, the coronary standardized mortality ratios are 90 times higher in males under age 40 and 120 times higher in females in the same age group.

The much rarer homozygous condition is particularly severe, with a six- to eightfold increase in plasma cholesterol. These patients develop both cutaneous and tendon xanthomas as well as aortic and coronary atherosclerosis, before the age of 10 years. Typically, they contract CHD in their teens or early twenties.

Genetically determined changes in the ▶**ligands** for the LDL receptor, ▶**apolipoprotein** (apo) B and apoE, may also cause dyslipidemia and premature CHD. Several amino acid substitutions in apoE that prevent normal binding to the LDL receptors result in type III hyperlipoproteinemia which is quite different from FH. In the apoB gene, however, the R3500Q mutation results in a disorder that may mimic FH biochemically and clinically, namely familial defective apolipoprotein B (FDB) (2). The clinical course of FDB, however, is usually milder than that of FH.

Recently, a third locus associated with autosomal dominant hypercholesterolemia, HCHOLA3 at 1p32, has been reported. Two mutations in the gene PCSK9, (proprotein convertase subtilisin/kexin type 9) were identified causing autosomal dominant hypercholesterolemia (3). PCSK9 encodes NARC-1 (neural apoptosis regulated convertase) a newly identified human subtilase that is highly expressed in the liver but the precise implication in cholesterol homeostasis is unknown.

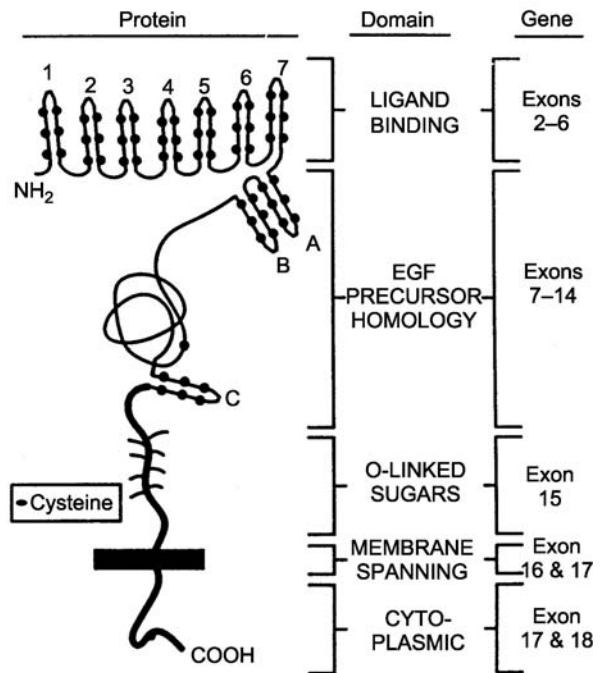
Thus, classical Mendelian inherited elevations of plasma LDL cholesterol may be caused by mutations in at least three different genes, the LDL receptor gene on chromosome 19p13.1-p13.3, the apoB gene on chromosome 2p23-pter and the PCSK9 gene on chromosome 1p32.

A rare **autosomal recessive** form of hypercholesterolemia that clinically resembles FH but is not due to mutations in the LDL receptor gene nor the apoB gene has also been identified (4). The gene locus has been mapped to chromosome 1p35, and a gene encoding a putative LDL receptor adaptor protein has been identified. This adaptor protein contains a phosphotyrosine-binding domain, which in other proteins binds Asn-Pro-X-Tyr motifs in the cytoplasmic tails of cell-surface receptors, including the LDL receptor. Furthermore, the adaptor protein appears to have a tissue-specific role in LDL receptor function, as it is required in hepatocytes but not in fibroblasts.

Cellular and Molecular Regulation

The 5.3 kb human LDL receptor mRNA of which almost half is **untranslated**, encodes a protein of 860 amino acids including an N-terminal signal sequence of 21 amino acid residues which is cleaved from the receptor post-translationally (5). Thus, a mature receptor protein of 839 amino acids is transported to the cell membrane (Fig. 2). The 45 kb LDL receptor gene comprising 18 **exons** and 17 **introns** has been highly conserved through evolution. The regulation of the LDL receptor gene expression is quite complex. Relative to the major transcription start site an 177-base pair fragment of the 5'-flanking DNA has been shown to be sufficient for controlling basal transcription as well as negative feedback regulation by cholesterol and its derivatives. The positive regulatory elements within this region were identified as three GC-rich imperfect 16 bp direct repeats and two TA-rich TATA-like sequences of 7 base pairs each. Repeats 1 and 3 contain **transcription factor** Sp1-binding sites. Interference with Sp1 binding to either repeat severely decreases basal transcription. Regulation of transcriptional activity occurs mainly through a 10-base pair sequence within repeat 2 designated sterol regulatory element-1. When intracellular cholesterol is low, sterol regulatory element-binding proteins bind to the sterol regulatory element-1 and subsequently interact with Sp1 in repeat 3. However, transcription of the LDL receptor gene is regulated not only by cholesterol, but also by nonsterol mediators e.g. hormones, growth factors and cytokines.

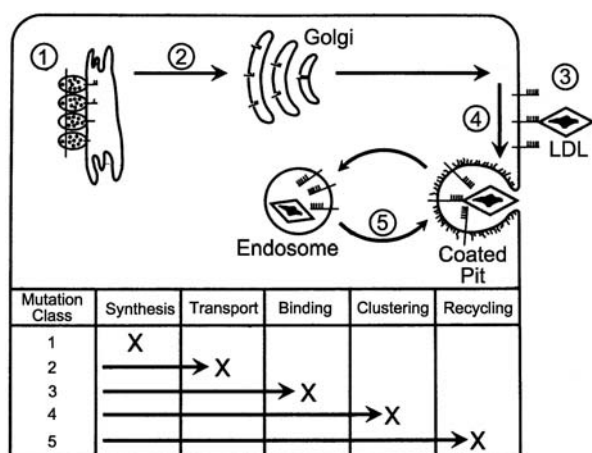
Normal function of the hepatic LDL receptor is required for normal levels of plasma LDL cholesterol. The LDL receptor regulates the concentration of plasma LDL cholesterol by internalizing apo



Familial Hypercholesterolemia. Figure 2 Domain structure of the human LDL receptor protein and its relation to the exon organization of the gene. The domains of the 839-amino acid mature protein are shown at the left and the corresponding exons encoding the protein domains at the right. Exon 1 (not shown) encodes a 21-amino acid signal sequence, which is cleaved from the mature protein during synthesis in the endoplasmic reticulum. (From ref. 5)

B-100- and apo E-containing lipoproteins by receptor-mediated endocytosis.

Mutations in the LDL receptor gene are classified into five functional classes as shown in Fig. 3. Class 1 mutations eliminate the synthesis of the LDL receptor. They include mutations in the **promoter**, nonsense mutations and mutations affecting **mRNA splicing** (null alleles). Class 2 mutations, which represent over 50% of mutations causing FH, result in LDL receptors that fail to be transported from the endoplasmic reticulum to the Golgi complex (class 2A) or are transported very inefficiently (class 2B) (transport-defective alleles). Class 3 mutations result in receptors that reach the cell surface but fail to bind their ligands, apoB or apoE (binding-defective alleles). Class 4 mutations generate LDL receptors that are delivered to the cell surface and bind LDL, but fail to concentrate in coated pits thereby leading to defective receptor-mediated endocytosis (internalization-defective alleles). Class 5 mutations result in LDL receptors that bind and internalize ligands normally, but fail to release LDL intracellularly and do not return to the cell surface (recycling-defective alleles).



Familial Hypercholesterolemia. Figure 3

Classification of LDL receptor mutations based on abnormal function of the mutant protein. These mutations disrupt the receptor's synthesis in the endoplasmic reticulum, transport to the Golgi complex, binding of apoprotein ligands, clustering in coated pits, and recycling in endosomes. (From ref. 5)

Clinical Relevance

FH has severe cardiovascular consequences due to the high levels of LDL cholesterol. However, efficient therapy is available in heterozygous FH that may potentially normalize levels of LDL cholesterol. Because target values for LDL-cholesterol are not achieved by dietary changes advised by the European Atherosclerosis Society or the US National Cholesterol Education Program alone, most patients require lipid-lowering drugs. Several studies have shown that hydroxy methylglutaryl coenzyme A (HMG-CoA) reductase inhibitors (statins), which competitively inhibit HMG-CoA reductase and thereby suppress cholesterol synthesis, have a marked effect on elevated plasma LDL cholesterol levels. If necessary, this effect can be enhanced by combining statins with other lipid-lowering drugs such as bile acid-binding resins, niacin and fibrates. It has been documented that intensive combination therapy leads to significant regression of coronary atherosclerosis. Moreover, aggressive statin monotherapy leads to regression of carotid intima media thickness, whereas lower, conventional statin doses do not. No long-term prospective randomized double-blind placebo-controlled studies of lipid-lowering drug therapy, however, have been performed in FH patients to assess the effect on morbidity/mortality. This is because it has been considered unethical to withhold treatment from patients with a diagnosis of FH. However, the Simon Broome cohort of FH patients in the UK showed a decline in relative risk for coronary

mortality in patients aged 20–59 years—from eightfold before 1992 to 3.7-fold post 1992—when statins were introduced (6). Moreover, data from large clinical cholesterol-lowering trials in the general population, strongly suggest that statin treatment will reduce morbidity and mortality in FH subjects. It is therefore, important to identify FH individuals so that appropriate treatment can be offered to reduce the risk of developing premature atherosclerotic cardiovascular disease.

However, even though efficient lipid lowering therapy is available, too few patients with FH have been diagnosed and too few have been put on lipid lowering therapy. In one study it was shown that only 20% of FH heterozygotes had been diagnosed and only 7% were being adequately treated. The main reason for the failure to put FH patients on lipid lowering drugs is assumed to be difficulties in diagnosing FH using clinical criteria. An alternative to a clinical diagnosis of FH is to apply molecular genetic methods to identify a mutation in the LDL receptor gene. However, these analyses are expensive and are only available in a few countries.

Patients with homozygous FH respond poorly to any drug therapy. In these patients non-pharmacological treatment modalities are necessary, preferable selective removal of apoB-containing lipoproteins by LDL apheresis. Although gene replacement therapy has been attempted to restore the function of the defective LDL receptors in patients with homozygous FH, it has met with little success and is still several years away from being a safe and effective treatment.

References

- Goldstein JL, Hobbs HH, Brown MS (1995) Familial hypercholesterolemia. In: Scriver CR, Beaudet AL, Sly WS (eds) *The Metabolic Basis of Inherited Disease*, 7th edn. McGraw-Hill Book Co, New York, pp 1981–2030
- Innerarity TL, Mahley RW, Weisgraber KH et al (1990) Familial defective apolipoprotein B-100: a mutation of apolipoprotein B that causes hypercholesterolemia. *J Lipid Res* 31:1337–1350
- Abifadel M, Varret M, Rabes JP et al (2003) Mutations in PCSK9 cause autosomal dominant hypercholesterolemia. *Nat Genet* 34:154–156
- Cohen JC, Kimmel M, Polanski A et al (2003) Molecular mechanisms of autosomal recessive hypercholesterolemia. *Curr Opin Lipidol* 14:121–127
- Hobbs HH, Russell DW, Brown MS et al (1990) The LDL receptor locus in familial hypercholesterolemia: mutational analysis of a membrane protein. *Annu Rev Genet* 24:133–170
- Simon Broome Steering Committee (1999) Steering committee on behalf of the Simon Broome Register Group. Mortality in treated heterozygous familial hypercholesterolemia: implications for clinical management. *Atherosclerosis* 142:105–112

Familial Hypertrophic Cardiomyopathy

Definition

Familial hypertrophic cardiomyopathy is an autosomal dominantly transmitted disorder, which is characterized by asymmetric left ventricular wall thickening with histological features of myocyte hypertrophy, myofibril disarray and interstitial fibrosis.

► [Familial Dilated Cardiomyopathy](#)

Familial Spastic Paraplegia

► [Hereditary Spastic Paraplegia](#)

Fanconi Anaemia

Definition

Fanconi anaemia (FA) is an autosomal recessively inherited chromosomal instability disorder, which is characterized by congenital abnormalities, defective haematopoiesis and a high risk of developing acute myeloid leukaemia and certain solid tumours. It can be caused by mutations in at least eight different genes (complementation groups). Affected persons usually develop severe aplastic anaemia by the age of 9 years.

► [DNA-Repair Mechanisms](#)

FAP

► [Familial Adenomatous Polyposis](#)

Farnesyl

Definition

A 15 carbon unit made up of three isoprene (dimethyl allyl) units.

► [Protein Prenylation](#)

Farnesylation

Definition

Farnesylation describes a post-translational modification of proteins which is characterized by an attachment of an isoprenoid (► [farnesyl](#) or geranylgeranyl diphosphate) to the C-terminal cysteine residue.

► [Ras Signalling](#)

FAS

Definition

The synonyms of FAS are ATP1, CD95, APO-1 and FAS1. FAS is a membrane protein with a molecular weight of 40 kD, and belongs to the tumor necrosis factor receptor superfamily member 6 (TNFRSF 6). The gene maps to 10q24 and is mutated in the Autoimmune Lymphoproliferative Syndrome. The protein mediates apoptosis through sequential activation of ICE-like caspases (Casp 3–5), and may be involved in autoimmune diabetes.

► [Apoptosis](#)

► [Apoptosis, Regulation and Clinical Implications](#)

► [Catalytic RNA](#)

► [TNF Receptor / Fas Signaling Pathways](#)

Fas Signaling

► [TNF Receptor/Fas Signaling Pathways](#)

Fas-Associated Death Domain

Definition

The Fas-associated death domain (FADD) is a universal adapter protein in apoptosis, which mediates signaling of all known death domain-containing members of the tumor necrosis factor (TNF) receptor superfamily.

- Apoptosis
- Tangier Disease

Fast Exchange Limit

Definition

Describes an interaction where the life-time of a ligand in a molecular complex is small, compared to the time-scale defined by the chemical shift difference of a nuclear resonance in the free and in the bound state.

- [Protein-Ligand-Interaction by NMR](#)

FASTA

Definition

FASTA is an alignment program for protein sequences that is one of the many heuristic algorithms proposed to speed up sequence comparison. The basic idea is to add a fast prescreen step to locate highly matched segments between two sequences, and then extend these matching segments to local alignments using more rigorous algorithms such as Smith-Waterman.

- [Protein Databases](#)

Fat Metabolism

Definition

Fat metabolism describes the breakdown or synthesis of body fat.

- [Caenorhabditis Elegans as a Model Organism for Functional Genomics](#)

Fatty Acid Acylation of Proteins

MICHAEL VEIT, MICHAEL F.G. SCHMIDT
Department of Immunology and Molecular Biology,

Faculty of Veterinary Medicine, Free University,
Berlin, Germany
mveit@zedat.fu-berlin.de, schmidt.mfg@vetmed.fu-berlin.de

Synonyms

N-terminal myristoylation, N-acylation, palmitoylation, S-acylation

Definitions

Fatty acid acylation is the covalent attachment of fatty acids to proteins. This distinguishes acylation from other hydrophobic modifications such as “► [glypidation](#)”, ► [isoprenylation](#) and ► [cholesterol modification](#), where either the fatty acids are bound indirectly *via* a glycolipid anchor or where other lipid moieties (isoprenoids, cholesterol) are linked to the protein.

Characteristics

According to the linkage type and the fatty acid species involved in acylation two types of acylation are distinguished: N-terminal myristoylation (► [N-acylation](#)) and palmitoylation (► [S-acylation](#)).

Myristoylation

Myristoylation (1, 2) refers to the attachment of myristate, the 14-carbon saturated fatty acid, *via* an ► [amide-linkage](#) to the amino group of a glycine residue at the N-terminus of eukaryotic and viral proteins. N-terminal myristoylation is an irreversible process, which occurs co-translationally after the initiator methionine has been removed. With a few exceptions (proteins of the retina), fatty acids other than myristate are not transferred to proteins. Myristoylation strictly requires a N-terminal glycine, but not all proteins with such a glycine are acylated. Myristoylation also depends on downstream amino acids. Proteins with a serine or threonine at position six and basic amino acids at position seven are especially preferred substrates. Myristoylated proteins are found in diverse intracellular locations such as the cytosol, the nucleus, the endoplasmic reticulum and the plasma membrane.

Myristoylated proteins are often involved in signal transduction. Protein kinases such as the catalytic subunit of cAMP dependent protein kinase A and kinases of the src-family, the protein-phosphatase calcineurin and most α -subunits of heterotrimeric G-proteins are prominent examples. Myristoylated proteins of viruses are for example the gag-polyprotein precursor of retroviruses and the capsid proteins of

some papova- and picorna-viruses, such as polio- and foot and mouth disease-virus.

Palmitoylation

Palmitoylation (2–8) is the post-translational attachment of long chain fatty acids *via* a ►thioester linkage to cysteine residues. At least in some proteins, palmitoylation is reversible and has therefore the potential to regulate the subcellular localization and function of the

summarizes the main features of myristoylation and palmitoylation: In contrast to myristoylation the molecular signals that direct S-acylation have not been fully identified. However, by comparison of the hitherto identified palmitoylation sites the following conclusions can be drawn:

1. ►Transmembrane proteins are palmitoylated at cysteine residues located at the boundary between the transmembrane region and the cytoplasmic tail.

Protein-species	Membrane	Cytoplasm
Transferrin-Receptor	TGY C ISG S C	RKP KTVNAKTNDANEEDVALKM....NH ₂
CD-4	GIF F C V	R C RHRRRQAERMSQIKRLLSEKKTCQ..COOH
Synaptotagmin 1	VT C C F C V C	KK C LFKKKNNKKGKEKGKNA.....NH ₂
VSV G-protein	LIIGLFLVL	RVGIHL C IKLKRATKKRQIYTDIE....COOH
Influenza virus A Hemagglutinin	AMGLVF I C V	KNGNMR C T I C I -COOH
NDV fusion protein	LSLVLA C YL	MYKQKAQKQTLWLGNNLTLDQ.....COOH

(Palmitoylated cysteine residues are in bold and underlined)

protein. Palmitate (C 16) is often the predominant fatty acid, but other species such as myristate (C 14), stearate (C 18) and even arachidonic acid (C 22:4) have been

In some proteins, e.g. G-protein coupled receptors, the palmitoylated cysteine is found at more distal positions.

Protein-species	Membrane	Cytoplasm
Rhodopsin	NPVIYI	MMNKQFRNCMVTTL C C G KNPLGD...COOH
β2-Adrenalin receptor	NPLIYC	RSPDFRIAFQELL C LRRSSLKAYG.....COOH
Dopamine D1 receptor	NPIIYA	FNADFRKAFSTLL G C Y RL C PAT...COOH

identified in “palmitoylated” proteins. Therefore this modification is also referred to as S-acylation. Palmitoylated proteins are mostly integral or peripheral membrane-proteins present on the cell surface as well as on organelles of the secretory or endocytic pathway. A few palmitoylated proteins have been identified in other organelles such as mitochondria or are even secreted from the cell, for example the surfactant protein C present in alveoli of the lung. The following table

2. Intrinsically hydrophilic proteins, i.e. those without a hydrophobic transmembrane region are palmitoylated in three different regions:

At cysteine residues located near the N-terminus of the protein.

Palmitoylation in this region often occurs after myristoylation of the N-terminal glycine. Such proteins seem to have a characteristic MGC-motif at the N-terminus:

src-family of tyrosine-kinases	α-subunits of heterotrimeric G-proteins
Yes M G C IKSKEDKGPAMKY	α1 M G C TLSAEDKAAVERS
Fyn M G C VQ C KDKEATKLTE	αo M G C TLSAEERAALERS
Lyn M G C IKSKRKDNLDDE	αz M G C RQSSEEKEAARRS

In other myristoylated proteins the palmitoylated cysteine is present at a more distal location:

eNOS **MG**NLKS**VGQEPGPP****C**GLGLGLGLGL**C**GK
 AKAP18 **MG**QL**CC**FPFSRDEGK
 Vac8p **MG****SCCSC**

Other proteins are palmitoylated at cysteines near the N-terminus without prior myristoylation:

Gαq MTLESIM**ACCL**SEEAKEA
 Gα12 MSGVVRTL**SRCL**LPAEAG
 Gα13 MADFLPSRSVLS**VC**FPQ**C**VLN**GEAEQ**QRKSKEIDK**CL**S
 GAP-43 ML**CC**MRRTKQVEKNDD**DQ**KIEQDGI
 PSD-95 MD**CL**CIVTTKKYRYQDE**DTP**

Palmitoylation can also occur at a cluster of cysteine residues located in the middle of the protein.

SNAP-25B NH₂.....EKNLTDLGKF**CGLC****Vc**
 P**c**NKLKSSDA.....COOH
 cysteine string NH₂.....VV**CGLLTCCYCCCC****CCCC**FN
 protein **CCCGKCK**PKA.....COOH

Cysteine residues located near the C-terminus of the protein also serve as palmitoylation sites. In these proteins palmitoylation occurs after isoprenylation of the cysteine within the C-terminal CAAX-box, proteolysis of the AAX and methylation of the penultimate cysteine. This type of palmitoylation has been described for two forms of the ras-protein and the small G-protein rho-B.

H-Ras SGPG**C**MS**C**K**CVLS**-COOH
 N-Ras GTQ**GC**MGLP**CVVM**-COOH
 Rho-B YGSQNG**C**IN**C**KVL-COOH

The CAAX-box is marked in bold.

Enzymes for Acylation and Deacylation of Proteins

Myristoylation (1)

Myristoylation is catalyzed by a well-characterized enzyme, N-myristoyltransferase (NMT). Genetic studies have established that NMT is essential in *S. cerevisiae* and in *Drosophila melanogaster*. Mammals possess two NMTs, which are both ubiquitously expressed and show the same substrate specificity. Myristoylation is irreversible and thus no enzyme for demyristoylation has been described. In some proteins the myristate may be removed together with neighboring amino acids by proteolytic cleavage.

Palmitoylation (6)

Palmitoyl-acyl-transferases (PAT)

Enzymes that catalyze palmitoylation of proteins have not been identified with certainty. In the test tube some proteins can even be palmitoylated in the absence of any enzyme source with palmitoyl-CoA as lipid donor. This non-enzymatic acylation can occur in stoichiometric amounts and at authentic palmitoylation sites. However, non-enzymatic palmitoylation *in vitro* requires unphysiologically high concentrations of

►Pal-CoA and is therefore believed not to be responsible for palmitoylation of proteins *in vivo*.

Many attempts have failed to purify PAT with biochemical methods, but it is generally agreed that such an enzyme exists on cellular membranes. Enrichment of PAT in membranes such as the ER/Golgi intermediate compartment and the plasma membrane have been reported. Even synaptosomes prepared from neurons and the vacuole of yeast cells contain palmitoylating activity of a protein nature. It is therefore likely that a cell contains different PAT enzymes with different intracellular locations and different substrate specificities.

Using a genetic approach candidates for PAT have been identified in lower eukaryotes:

Skinny Hedgehog

In *Drosophila* the product of the *ski* gene is required for palmitoylation of hedgehog, a secreted protein with a key role in several developmental processes. However, acylation of hedgehog occurs through an amide bond by an unusual mechanism, which makes it unlikely that skinny hedgehog also catalyzes conventional palmitoylation by thioester bonding.

DHHC-CRD Proteins

In *S. cerevisiae* one candidate for PAT is a protein family containing a conserved Asp-His-His-Cys (DHHC) motif within a cysteine rich domain (CRD). One member of this DHHC-DRD protein family (erf2p) catalyzes palmitoylation of yeast ras. A second DHHC-CRD protein (akr1p) palmitoylates a casein kinase. It is not known whether all members of the DHHC-CRD protein family are palmitoyl-transferases with different substrate specificities and if their mammalian orthologues perform the same function.

Thioesterases

Palmitoyl-Protein Thioesterase 1

Palmitoyl-protein thioesterase 1 (PPT-1) has been purified with palmitoylated ras as a substrate. However, its location inside lysosomes suggests that PPT-1 is not involved in depalmitoylation of functional proteins, but is rather required for lysosomal degradation of palmitoylated proteins and peptides (see also Clinical Relevance).

Acyl-Protein Thioesterase -1

Using palmitoylated G-proteins as substrates, a cytosolic deacylating enzyme was purified and designated acyl-protein thioesterase (APT). Peptide sequencing revealed that APT is identical to a lysophospholipase, which is also known to hydrolyze Pal-CoA. APT can cleave fatty acids from several, but not all, palmitoylated proteins. Deletion of the only APT gene in yeast had no obvious effect on the viability of the yeast cells. Thus, the role of APT in the regulated turnover of fatty acids requires further investigation.

Function of Fatty Acylation

Binding of Acylated Proteins to Membranes and to Membrane Subdomains (2, 5)

Attachment of lipids to an otherwise hydrophilic protein increases its hydrophobicity and thereby could facilitate insertion into the lipid bilayer. This has been experimentally demonstrated for the attachment of palmitate to intrinsically hydrophilic proteins, which are thereby anchored to the inner leaflet of the lipid bilayer. In contrast, myristoylation alone can only transiently attach proteins to membranes, probably because myristate is less hydrophobic than palmitate. Therefore a second signal located in the vicinity of the myristoylated glycine is required for permanent membrane anchorage of proteins. This can be a polybasic cluster of amino acids that enhances electrostatic interactions with acidic phospholipids enriched in the inner leaflet of the bilayer. In other myristoylated proteins, palmitoylation of cysteine residues in the vicinity of the myristate group provides the second signal for stable membrane binding.

It is reasonable to assume that protein-bound fatty acids partition into the lipid bilayer for attaching the protein to the membrane. However, membrane targeting based on these lipid-lipid interactions alone would lead to a homogeneous distribution of the modified protein in all intracellular membranes. Therefore additional mechanism must secure the correct intracellular location for acylated proteins. According to the “membrane-trapping model” the location of PAT may determine targeting of a protein to specific membranes. N-terminal myristoylation can attach the protein transiently to various membranes, but only if a PAT is present

is the protein palmitoylated and permanently attached to this membrane. The protein can exit the membrane only by deacylation or by inclusion into budding vesicles mediating intracellular transport, for example from the Golgi to the plasma membrane.

Some palmitoylated proteins are not homogeneously distributed in the plasma membrane, but are enriched in specific subdomains called “lipid-rafts”, or “caveolae” if the palmitoylated protein caveolin is a major constituent. These lipid aggregates are insoluble in non-ionic detergent at low temperature and are characterized by their high cholesterol and sphingomyelin contents. The rafts serve as platforms where specific proteins are concentrated, especially those that participate in protein-protein interactions. The enrichment of palmitoylated proteins in rafts does not depend on the hydrophobicity of the lipid moiety alone, because myristoylated and isoprenylated proteins are not targeted to rafts. Likewise, not every palmitoylated protein is located in ▶“lipid-rafts”, demonstrating that additional signals must exist that regulate the protein composition of “lipid-rafts”.

Release of Acylated Proteins from the Membrane

Some acylated proteins cycle on and off the membrane. Palmitoylation is a reversible process and palmitoylated proteins can be released from the membrane simply by enzymatic cleavage of the fatty acid. This has been shown for ▶peripheral membrane proteins, but not for ▶integral membrane proteins that do not depend on palmitoylation for membrane attachment because their hydrophobic transmembrane region is sufficient for permanent anchorage.

One possibility for releasing a myristoylated protein from the membrane is to quench the second signal required for membrane binding. Phosphorylation of amino acids in the polybasic domain introduces negative charge thereby reducing the interaction with acidic phospholipids in the bilayer. Another possibility for regulating transient membrane interactions of myristoylated proteins is to regulate the surface exposure of the fatty acid. Some myristoylated proteins can occur in two conformations. In one conformation the myristate is sequestered in a hydrophobic pocket within the protein. In the second conformation the fatty acid is expelled and is then available to participate in membrane binding. This “myristoyl switch” is often triggered by binding of a ligand to the protein, for example calcium or GTP.

We have just described the general role of acylation for the attachment of proteins to membranes. In the remaining part of the article we will discuss acylation of specific proteins and their role in cellular functions such as signal transduction, cell transformation and neuronal development and secretion and also in the replication cycle of viruses.

Signal Transduction via G-protein Coupled Receptors, Heterotrimeric G-proteins and Kinases (7)

The role of acylation during signal transduction *via* G-protein coupled receptors (GPCRs), heterotrimeric G-proteins and kinases of the src-family has been intensively studied over the last few years. Myristoylation of the α -subunits of heterotrimeric G-proteins and of tyrosine kinases enhances their subsequent palmitoylation, probably by targeting the acylated α -subunit to a membrane-bound PAT. As already discussed above, palmitoylation then leads to permanent membrane binding and to targeting of the double acylated protein to lipid-rafts. Both myristate and palmitate increase the interaction of the acylated protein with other signaling molecules, in this case between the $G\alpha$ subunit and the $G\beta\gamma$ complex.

Palmitoylation and depalmitoylation are apparently regulated during signal transduction. When a specific signal transduction pathway is activated, increased turnover of protein-bound palmitate can occur both on the receptor (GPCR) and on the downstream G-protein. Deacylation of the G-protein is accelerated and the protein is released into the cytosol or is no longer attached to "lipid-rafts".

Palmitoylation of GPCRs is not essential for membrane binding, but palmitate is thought to insert into the membrane thereby forming an additional cytoplasmic loop. Despite the pronounced structural similarity between different GPCRs, the role of palmitate in the life cycle of GPCRs depends on the individual receptor. Some receptors require palmitoylation for their stable expression on the cell surface. Other receptors are unable to couple to G-proteins without covalently linked palmitate. In yet other receptors, palmitoylation has an effect on desensitization, phosphorylation, internalization and/or down-regulation. Thus, no generalizations can be made and the function of acylation has to be experimentally determined for each signal transduction pathway.

Neuronal Development and Neurosecretion (8)

"Lipid-rafts" are believed to mediate axon targeting of proteins as opposed to dendrites. Dynamic palmitoylation at two N-terminal cysteine residues is required for incorporation of GAP-43 into "lipid-rafts" and for its transport to the growth-cone of developing axons. There it modulates GPCR signaling cascades and regulates neurite outgrowth and path finding. Palmitoylation of GAP-43 decreases in mature axons and disrupting palmitoylation experimentally blocks axon outgrowth, indicating that membrane-bound, palmitoylated GAP 43 orchestrates the molecular machinery for axon extension.

However, not every palmitoylated protein is sorted to axons. PSD-95 is also palmitoylated at two N-terminal cysteine residues, but is transported to dendrites. At postsynaptic densities PSD-95 (together with other scaffolding proteins) regulates clustering of glutamate receptors of the AMPA- and NMDA-type. Certain patterns of synaptic activity reduce palmitoylation of PSD-95. The protein dissociates from postsynaptic sites and the number of glutamate receptors at the synapse is reduced.

Transformation of Cells by the ras Protein (2, 5, 7)

Mutations in the GTP-binding domain of the ras-protein can cause transformation of cells, because downstream signal cascades are permanently activated. To fulfill its function ras must be targeted to the inner leaflet of the plasma membrane. This is accomplished by protein modifications at the ►CAAX-box, isoprenylation, AAX-proteolysis, methylation and (in the case of N-ras and K-ras) palmitoylation. Isoprenylation occurs in the cytosol and targets ras-proteins to the endoplasmic reticulum, where proteolytic cleavage and methylation occur. Palmitoylation is then required for vesicular transport of ras to the plasma membrane and for its association with "lipid-rafts" or caveolae. Isoprenylation is carried out by the well-characterized enzyme farnesyl-transferase. Inhibitors of the enzyme can revert transformed cells to a normal phenotype and are currently being tested in clinical trials for use in cancer patients. Once the enzyme for the palmitoylation of ras has been unambiguously identified, it might also be a promising target for the development of cancer drugs.

Fatty Acid Acylation and the Viral Replication Cycle (3, 5)

Fatty acylation was originally discovered in viral proteins and many investigators have studied the role of myristoylation and palmitoylation in the replicative cycle of different viruses. Disappointingly, no generalization can be drawn from these numerous studies. Poliovirus is a non-enveloped virus that requires myristoylation of its VP4 protein for virus assembly. X-ray structures of whole virus particles show that myristate is an integral part of the virion subunit structure. This also demonstrates that a fatty acid need not necessarily be integrated into a lipid bilayer.

Nef and Gag, two proteins of the AIDS virus require myristoylation plus a polybasic motif for their membrane binding. Blocking myristoylation of the gag-polyprotein disrupts budding of AIDS-virus particles. Many spike glycoproteins of enveloped viruses, often those with membrane fusion activity, are palmitoylated

at cysteine residues in their transmembrane or cytoplasmic region. Although great progress has been made in elucidating the fusion mechanism of these glycoproteins, the role of palmitoylation in membrane fusion is probably quite subtle. The hemagglutinin of influenza virus and gp 160 of the AIDS-virus seem to require palmitoylation for the last step in membrane fusion, the opening or enlargement of the fusion pore. Palmitoylation of viral glycoproteins also effects the release of viruses from the infected cell. Palmitoylated viral glycoproteins as well as internal components of viral particles are often enriched in “lipid-rafts”. This concentration of viral components in membrane subdomains facilitates protein-protein interactions, assembly of virus particles and budding from the cell. Not every viral protein present in lipid-rafts is acylated, but those that are require palmitoylation for raft targeting.

Clinical Relevance

NMT as a Therapeutic Target (1)

Several human pathogens such as *Candida albicans* require a functional NMT for their survival. NMT from humans and from microorganisms both show the same preference for myristoyl-CoA, but have different peptide substrate specificities. Peptide derivatives have been developed that are species-specific inhibitors and show fungistatic or fungicidal activity.

Mutations in PPT Cause a Lysosomal Storage Disease (6)

Infantile neuronal ceroid lipofuscinosis, a lysosomal storage disease, is caused by genetic deficiencies of PPT-1. Pathologically, it is characterized by accumulation of autofluorescent material in the lysosomes of neurons and other cells. The disease is selectively manifested in the central nervous system, so that there is a progressive loss of neurons resulting in dementia and motor problems, epilepsy, blindness and, finally, early death. Cells from patients with the disease accumulate fatty acid thioester peptides derived from palmitoylated proteins.

References

1. Frazi TA, Waksman G, Gordon JI (2001) The biology and enzymology of protein N-myristoylation. *J Biol Chem* 276:39501–39504
2. Resh MD (1999) Fatty acylation of proteins: new insights into membrane targeting of myristoylated and palmitoylated proteins. *Biochem Biophys Acta* 1451:1–16
3. Schlesinger MJ, Veit M, Schmidt MFG (1993) Palmitoylation of cellular and viral proteins. In: Schlesinger MJ (ed) *Lipid Modifications of Proteins*. CRC-Press, Boca Raton, FL, USA, pp 1–19

4. Veit M, Ponimaskin E, Schmidt MFG (2002) Analysis of S-Acylation of Proteins. In: Kannicht C (ed) *Methods in Molecular Biology* 194, Posttranslational Modifications of Proteins: Tools for Functional Proteomics, Humana Press Inc., Totowa, NJ, pp 159–178
5. Bijlmarkers M-J, Marsh M (2003) The on-off story of protein palmitoylation. *Trends Cell Biol* 13:32–42
6. Linder M, Deschenes RJ (2003) New insights into the mechanism of protein palmitoylation. *Biochemistry* 42:4311–4320
7. Qanbar R, Bouvier M (2003) Role of palmitoylation/depalmitoylation reactions in G-protein-coupled receptor function. *Pharmacol Ther* 97:1–33
8. el-Husseini Ael-D, Bredt DS (2002) Protein palmitoylation: a regulator of neuronal development and function. *Nat Rev Neurosci* 3:791–802

FCMD

►Fukuyama Myopathy

Feedback

Definition

Feedback refers to a mechanism by which it responds to its output. There are two-type feedback mechanisms, positive and negative feedback. A positive feedback intensifies the response of the triggering signal, whereas a negative feedback is lowering the response of the signal.

►Axis Formation-Formation and Function of the Dorsal Organizer

FEP

►Free Energy Perturbation

Fermentation

Definition

Originally fermentation meant anaerobic respiration. In traditional biotechnological processes, only anaerobic processes were used. To date, the term fermentation is still used for microbiological production processes,

although many of them are performed under aerobic conditions.

► [Recombinant Protein Expression in Yeast](#)

Ferritin

Definition

Ferritin is an iron-storage protein consisting of an outer shell comprised of light and heavy ferritin chains. Iron is stored inside the shell.

► [Hemochromatosis](#)

Fertilization

Definition

The union of two germ cells, egg (female) and sperm (male), whereby the somatic chromosome number is restored and resulting offspring exhibit characteristics of their parents.

► [Mammalian Fertilization](#)

FEVR

► [Familial Exudative Vitreoretinopathy](#)

FG Repeats

Definition

FG repeats refers to a phenylalanine-glycine repeat motif found in the amino acid sequence of many nucleoporins, and is required for the interaction of nucleoporins with transport receptors.

► [Nuclear Pore Complex](#)

FGF

Definition

FGF stands for fibroblast growth factor. It belongs to a large family of growth factors that can bind to structural

proteins in tissues and is released in situations such as tissue injury. It stimulates proliferation of a large number of cell types, important for the formation of new blood vessels.

► [Growth Factors](#)

Fibrinolysis

Definition

Fibrinolysis is the dissolution of fibrin by enzymatic action.

► [Hereditary Hemostatic Defects and Recombinant Proteins for Treatment](#)

Fibroblast

GIULIO GABBANI

Department of Pathology, CMU-University of Geneva, Geneva, Switzerland

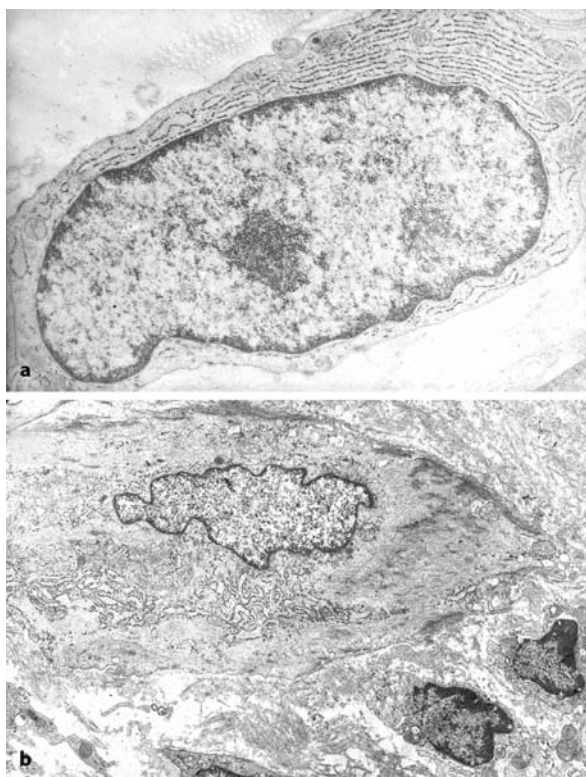
giulio.gabbiani@medecine.unige.ch

Definition

The fibroblast has been considered for a long time a relatively static cell fulfilling the function of synthesizing extracellular matrix components, such as collagen. However, during the last few years, new perspectives have opened concerning the biological features, the function and possibly the heterogeneity of this cell; we are now facing a period of intense research concerning new and potentially important contributions of the fibroblast to several physiological and pathological phenomena. In this article, we discuss the role of the fibroblast in contributing to the organization of ► [extracellular matrix](#) and, in particular, the aspects of its biology that open new perspectives in the understanding of connective tissue remodeling in normal and pathological situations.

Characteristics

The morphological features of fibroblasts under normal conditions are very similar in different organs. On electron microscopic examination, they exhibit an oval nucleus and a variably developed endoplasmic reticulum; this is in relation to their synthetic activity (Fig. 1a). The ► [intermediate filament](#) protein typical of fibroblastic cells vimentin.



Fibroblast. Figure 1 (a) Typical fibroblast of the human dermis showing an oval nucleus and a well developed endoplasmic reticulum. (b) typical myofibroblast from an experimental granulation tissue with an indented nucleus, a fairly developed endoplasmic reticulum and many microfilaments (Figure a, from Gabbiani G. et al. *The Journal of Experimental Medicine*, 1972, 135 (4):719–734 with copyright permission of The Rockefeller University Press and Figure B, from Gabbiani G. et al., *Experientia*, 1971, 27:549 with copyright permission of Birkhäuser Verlag AG).

Fibroblasts' Origin

It has been assumed for a long time that in adult mammals fibroblasts are resident mesenchymal cells and that they are renewed locally. This view has not always been predominant. For example, very early work by Connheim (1) suggested that fibroblasts could derive from circulating cells, macrophages in particular. Later, the experiments of R. Ross using parabiotic rats (two rats with the same genetic background, whose circulations have been connected) showed that when a wound was made in one of them, there was no evidence that the fibroblasts of the granulation tissue repairing the wound were derived from the other rat. This established the dogma that the origin of fibroblasts during tissue repair is local (1). However, in the last few years, several observations have challenged this view and convincingly suggested that, at least in part, fibroblasts derive from circulating cells such as

macrophages or more generally stem cells (2). These circulating fibroblasts or fibroblast precursors have been called (somewhat unfortunately from the etymological point of view) fibrocytes (2). Other experiments have shown that if a silicone tube is implanted into the peritoneum of an animal, e.g. a rabbit, the layer of cells surrounding this tube is composed first of macrophages and then of fibroblastic cells that probably derive from these macrophages. The fibroblasts making a capsule around the tube may even transdifferentiate into smooth muscle cells (3). These observations open an important perspective on the characterization of fibroblast precursors that could eventually be used therapeutically during tissue repair and/or tissue reconstruction.

Another interesting new approach concerning the origin of the fibroblast has been the observation that these cells can derive from epithelial cells, particularly in locations such as the kidney interstitium, where the main source of fibroblasts is the epithelial cell from tubules (4) and in the mesothelium of several cavities, where the precursor cell is the mesothelial cell (5). **►Epithelial-to-mesenchymal transition** is an important biological phenomenon for example during development. Taken together these observations show that fibroblasts are more heterogeneous than one would have previously assumed.

Fibroblasts' Heterogeneity

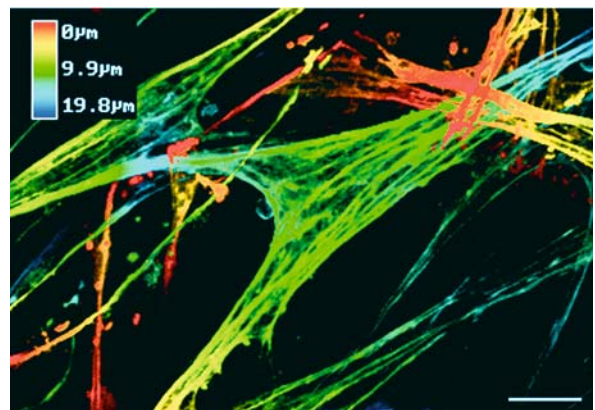
It is intuitively acceptable that fibroblasts present in different organs, although involved in similar functions, e.g. synthesis of extracellular matrix, show different biological features, since some organs are selectively rich in one of the collagen types such as Type I or Type III and may also show different proportions of accumulated connective tissue components. However, for a long time, differences among fibroblastic cells were somehow ignored. During recent years, culturing fibroblastic cells from different locations has indicated that they may behave differently e.g. in the synthesis of collagen types (6), complement fragments (reviewed in 7), cytokines (reviewed in 7) and structural proteins such as α -smooth muscle actin (6), the actin isoform typical of vascular smooth muscle cells that also participates in the force generation by myofibroblasts (see below) in order to produce wound contraction (see below). An interesting recent development is the description of discrimination between two fibroblastic phenotypes using the marker Thy-1, a cell surface protein whose function is at present not known. Using this experimental approach, it has been shown that Thy-1 positive fibroblasts are predisposed to differentiate toward the myofibroblastic phenotype (see below) whereas Thy-1 negative fibroblasts are predisposed to differentiate toward the adipocytic phenotype (7).

Regulatory Mechanisms

It appears more and more clear that fibroblasts can crosstalk with other cells of their microenvironment, through modulation of their phenotype or through production of ►cytokines.

Role of Fibroblasts in Wound Healing and Tissue Remodeling

Perhaps the most important advance in fibroblast biology of recent years is the finding that these cells participate in a crucial way in granulation tissue remodeling and wound contraction (reviewed in 6). When an open wound heals in the skin or in internal organs, the repair is essentially done by connective tissue cells rather than by epithelial cells, since in general mammalian organs do not possess or possess in a very low degree the capacity of regenerating. The process of wound healing starts with an extravasation of blood that coagulates; the fibrin coagulum contains platelets and white blood cells which liberate cytokines and growth factors that in turn play an important role at the beginning of the healing process. Among the actions of these substances, there is a) stimulation of the motility and contractility of fibroblasts and b) stimulation of ►angiogenesis. These actions allow the formation and accumulation of what is called granulation tissue, i.e. a tissue consisting essentially of small vessels, fibroblasts and extracellular matrix components that are synthesized by fibroblasts. Granulation tissue formation is followed by contraction, which is rather a retraction process than a real contraction (see below), which is important for reducing the size of the wound. This process, together with reconstitution of epithelial continuity, i.e. replication and movement of epithelial cells, is essential for the completion of wound healing. During the process of granulation tissue formation, fibroblasts acquire contractile capacity, which is reflected in their neoexpression of bundles of microfilaments (Fig. 1b) that contain contractile proteins such as actin and myosin, the classical molecules that generate force in organs such as cardiac and skeletal muscle. Actin and myosin are expressed as different isoforms (that means similar molecules with slightly different sequences) in different organs. In general fibroblasts do not express actin isoforms typical of muscle cells but granulation tissue fibroblasts, that have been called myofibroblasts, neoexpress α -smooth muscle actin, the actin isoform typical of vascular smooth muscle cells. Hence, fibroblasts not only assume a contractile function during the evolution of granulation tissue but also express proteins typical of a muscle cell. This expression corresponds morphologically to the appearance of cytoplasmic stress fibers (Fig. 2) i.e. bundles of filaments containing actin and myosin and capable of exerting a contractile activity. Stress fibers end at the cell membrane on adhesion



Fibroblast. Figure 2 Three-dimensional reconstruction by means of laserscan confocal microscopy of a myofibroblastic cell present in an epiretinal membrane from a patient suffering from proliferative vitreoretinopathy. The cells have been stained *in toto* by means of immunofluorescence with an antibody against α -smooth muscle actin. The colors change in function of the depth of the section. Myofibroblasts show typical stress fibers. Bar, 10 μ m. (from Serini G. and Gabbiani G., *Experimental Cell Research*, 1999, 250:273–283 with copyright permission of Elsevier Science and Academic Press)

complexes, which connect the cell periphery with the extracellular matrix, allowing the force exerted by stress fibers to be transmitted to the surrounding microenvironment (8). Using these organelles, myofibroblasts remodel the extracellular connective tissue in order to produce reduction of granulation tissue volume i.e. wound contraction. It should be noted that the contractile activity of myofibroblasts is not exerted by means of a mechanism similar to that of striated or smooth muscles, that is always followed by relaxation, but is more the consequence of a long lasting isometric tension that reduces the volume of the surrounding matrix (6). This change in volume is then stabilized by the synthesis of new extracellular components such as collagen molecules and the overall process results in a decrease in the wound area or volume. Recent work has shown that in myofibroblasts actin/myosin interaction in order to produce isometric force is regulated by Rho/Rho kinase mediated inhibition of myosin phosphorylation, a mechanism that is typical of primitive contractile phenomena (6).

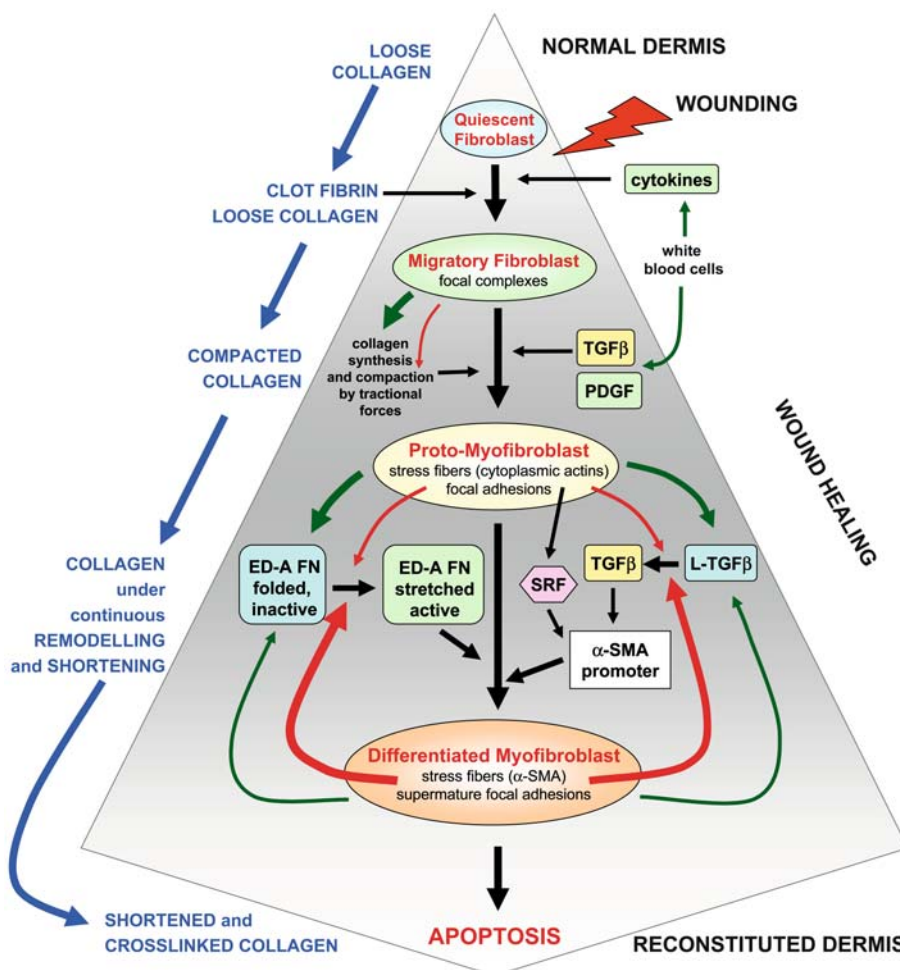
The fibroblast/myofibroblast transition depends on the concerted action of cytokines, such as transforming growth factors β (TGF β) and specific extracellular matrix proteins such as the cellular fibronectin splice-variant ED-A. Recently, acquisition and maintenance of the myofibroblast phenotype has been shown to depend on mechanical tension, which is regulated by the continuous interaction between the rigidity of the

extracellular matrix and the contractile activity of the myofibroblasts (6).

When a wound is reepithelialized and the continuity of the tissue is re-established, myofibroblasts undergo apoptotic phenomena that cause their death, thus provoking the evolution of granulation tissue into scar tissue that is relatively poor in cells. The mechanisms of myofibroblast apoptosis are not very clear but it is established that during pathological wound healing, e.g. hypertrophic scar development, these apoptotic changes do not take place.

Conclusions and Perspectives

The last ten years have seen a profound change in the paradigm of fibroblast biological function with a switch from a static role involving synthesis of extracellular matrix components to a more active participation in tissue repair, remodeling and differentiation. In particular the fibroblast/myofibroblast transition has been established as a key event in wound healing and fibrotic phenomena (Fig. 3). Among the processes that remain to be investigated are the transdifferentiation mechanisms of myofibroblasts from precursor cells located



Fibroblast. Figure 3 Scheme illustrating myofibroblast evolution during dermal wound healing. Stimulatory effects are depicted by black arrows; arrow thickness grossly corresponds to the level of activity. In the mechanically unloaded extracellular matrix (ECM) of normal dermis, fibroblasts are quiescent and do not develop stress fibers or cell-matrix contacts. Cytokines released by blood cells after dermal injury stimulate fibroblasts to form immature focal adhesions and to migrate into the provisional extracellular matrix, where they secrete and remodel collagen. Their tractional forces enhance matrix rigidity and lead to the development of the proto-myofibroblast, which is characterized by stress fibers and mature focal adhesions. Secretion and activation of TGFβ and ED-A fibronectin (ED-A FN) and increased matrix stress induce α-smooth muscle actin (α-SMA) accumulation and formation of the differentiated myofibroblast. When matrix reconstruction is completed, myofibroblasts are removed by apoptosis. PDGF, platelet-derived growth factor; L-TGF, latent transforming growth factor β; SRF, serum response factor (from Hinz B. and Gabbiani G., *Current Opinion in Biotechnology*, 2003, 14 :358–546 with copyright permission of Elsevier Science and Academic Press)

locally or immigrating from the circulation. In fibroblast/myofibroblast modulation, TGF β appears to be a crucial player together with ED-A cellular fibronectin. In addition, the role of mechanical factors such as tissue stiffness and tension is beginning to be elucidated. We are clearly at the beginning of an exciting era that will probably allow definition of the implications of the fibroblast in several important biological phenomena and allow one to influence these phenomena by strategies involving the mechanical or chemical factors modulating fibroblast biology.

Acknowledgements

This work has been supported by the Swiss National Foundation grant no 31-61.336.00. I thank The Rockefeller University Press, Birkhäuser Verlag AG, Elsevier Science and Academic Press for allowing the reproduction of Figs. 1, 2, 3. I am grateful to Dr. Marie-Luce Bochaton-Piallat for help in the organization of the manuscript and Mrs Evelyne Homberg for secretarial work.

- [Extracellular Matrix](#)
- [Senescence](#)
- [Wound Healing](#)

References

1. Gabbiani G, Montandon D (1977) Reparative processes in mammalian wound healing: the role of contractile phenomena. *Int Rev Cytol* 48:187–219
2. Chesney J, Bucala R (1997) Peripheral blood fibrocytes: Novel fibroblast-like cells that present antigen and mediate tissue repair. *Biochem Soc Trans* 25:520–524
3. Thomas AC, Campbell GR, Campbell JH (2003) Advances in vascular tissue engineering. *Cardiovasc Pathol* 12:271–276
4. Lan HY (2003) Tubular epithelial-myofibroblast transdifferentiation mechanisms in proximal tubule cells. *Curr Opin Nephrol Hypertens* 12:25–29
5. Yang AH, Chen JY, Lin JK (2003) Myofibroblastic conversion of mesothelial cells. *Kidney Int* 63:1530–1539
6. Tomasek JJ, Gabbiani G, Hinz B et al (2002) Myofibroblasts and mechano-regulation of connective tissue remodelling. *Nat Rev Mol Cell Biol* 3:349–363
7. Koumas L, Smith TJ, Feldon S et al (2003) Thy-1 expression in human fibroblast subsets defines myofibroblastic or lipofibroblastic phenotypes. *Am J Pathol* 163:1291–1300
8. Hinz B, Gabbiani G (2003) Mechanisms of force generation and transmission by myofibroblasts. *Curr Opin Biotechnol* 14:538–546

disulphide bonds, and containing about 5% carbohydrate. It occurs in insoluble fibrillar form in the extracellular matrix (ECM), and as a soluble form in plasma, the latter previously known as cold-insoluble globulin. Fibronectin is involved in many cellular processes like cell-cell adhesion, cell substratum adhesion, control of cell shape and induction of cell migration. Fibronectins have multiple domains that facilitate interaction with other components of the ECM, such as collagen, fibrin and heparin. Fibronectin can bind to integrins via the ► [RGD motif](#). There are several isoforms of fibronectin, all of which are the product of a single gene. The structure of these isoforms are made of three types of repeated internal regions called I, II and III, which exhibit different lengths and presence or absence of disulfide bonds.

- [Extra Cellular Matrix](#)
- [Integrin Mediated Signaling](#)

FID

Definition

FID refers to a signal that is detected at the end of a multidimensional NMR experiment. An initial state is created by the pulse sequence, and the decay of this initial state back to equilibrium is recorded through the induction of a current in a receiver coil.

- [Multidimensional NMR Spectroscopy](#)

FIH

Definition

FIH (factor inhibiting HIF-1) is a member of the 2-oxoglutarate dependent hydroxylase family, which hydroxylates an asparagine residue in the N-TAD region of HIF, resulting in reduced transcriptional activity of HIF by interfering with recruitment of the coactivator p300.

- [Hypoxia Inducible Factors](#)

Fibronectin

Definition

Fibronectin is high molecular weight glycoprotein, consisting of two chains (each of 250 kD) linked by

Filaggrin

Definition

Filaggrins are filament-associated proteins that interact with keratin intermediate filaments of terminally

differentiating mammalian epidermis via disulphide bond formation. The proteins are synthesised as large, insoluble, highly- phosphorylated precursors, containing multiple tandem repeats of 324 amino acids, which are not separated by a large linker. The precursor is deposited as keratohyalin granules. Aberrant expression of filaggrin has been implicated in a number of keratinising disorders.

► [Intermediate Filaments](#)

Filamentous Actin

► [F-Actin](#)

► [Actin Cytoskeleton](#)

Filopodia

Definition

Filopodia are finger-like extensions of the plasma membrane driven outwards by actin polymerization in which actin filaments are arranged in parallel.

► [Rho, Rac, Cdc42](#)

Filtering

Definition

In medical image processing, various filtering methods have been developed aimed at reducing the noise in the data set, while leaving the signal largely unchanged, and therefore increasing the signal-to-noise ratio.

► [EM Tomography](#)

Fingerprint/Fingerprinting

Definition

Fingerprinting refers to a high throughput method in which DNA (of large insert clones) is digested with restriction endonucleases, which results in a largely individual-specific series of bands. The measurement of relative mobilities of the fragments represents a

DNA fingerprint of each clone/individual. This technique enables the identification of other clones that share a large proportion of fragments with the same relative mobilities that represents an overlap of the different clones, thereby constructing a contig. In analogy to DNA fingerprinting, in proteomics the term fingerprint is defined as a series of conserved local sequence alignments, here called motifs, which have been iteratively refined from database searches. The pattern of fragments can be obtained when a protein is digested by a proteolytic enzyme, usually observed following two-dimensional separation by chromatography or electrophoresis.

DNA Fingerprinting has become very popular in forensic laboratories. Scientists have chosen repeating sequences in the DNA, which are present in all individuals on different chromosomes, and are known to vary from individual to individual except in identical twins. By this technique an individual can be identified at molecular level.

► [C. Elegans Genome, Comparative Sequencing](#)

► [Protein Databases](#)

FISH

Definition

FISH (Fluorescence in-situ hybridization) is a cytogenetic method of visualizing defined DNA sequences in metaphase or interphase chromosome spreads. It is useful for identifying chromosomes and parts thereof, deciphering chromosomal abnormalities and for gene mapping. A nucleic acid probe labelled with a fluorescent dye is hybridized to chromosomes of suitably prepared cells or histological sections, and the probe is then visualised by fluorescence microscopy and scored for the presence or absence of the signal.

► [Immunological Methods, the Use of Antibodies to Discover the Function of Novel Gene Products](#)

► [Large-Scale Homologous Recombination Approaches in Mice](#)

► [Mouse Genomics](#)

► [Nuclear Compartments](#)

Fission Yeast

Definition

This group of yeasts is characterised by cells that divide into two identical cells. This type of cell division also occurs in bacteria and in mitotic dividing mammalian

cells, and stands in contrast to “the budding yeast” *Saccharomyces cerevisiae*, where daughter cells are budded from one mother cell.

► [Recombinant Protein Expression in Yeast](#)

Fitness

Definition

Fitness is a measure of the survival and reproduction of different genotypes. It is usually used as a relative measure, e.g. individuals of genotype A are fitter than those of genotype *a* if they contribute proportionally more offspring to the next generation.

► [Gene Duplications](#)

Fixation

Definition

A mutation becomes fixed when it is the only allele present at a specific locus in a population.

► [Gene Duplications](#)

FK506 (Tacrolimus)

Definition

The immunosuppressive drug FK506 is a 23-member macrocyclic lactam which is isolated from *Streptomyces tsukubaensis*. Beside its immunosuppressive effects, it has neuroprotective properties and antiparasitic effects. It shows strong affinity for PPIases (peptidyl-prolyl cis/trans isomerase) of the FKBP (FK506-binding protein) family.

► [Nuclear Hormone Receptors: Regulators of Gene Transcription and Cellular Signaling](#)

► [Peptidyl Prolyl Cis/Trans Isomerases](#)

► [Receptor Serine/Threonine Kinase](#)

FKBP12

Definition

FKBP12 is a FK506-binding protein of 12 kD, which binds to the TGF- β Type I receptor and blocks its

kinase domain. It is released from the receptor during phosphorylation by the TGF- β Type II receptor. Blocking FKBP12/Type I receptor interaction with FK506 non-functional derivatives enhances the ligand activity, indicating that FKBP12 binding is inhibitory to the signaling pathways of the TGF- β family ligands.

► [Receptor Serine/Threonine Kinase](#)

FLAG-Tag

Definition

FLAG is a short, hydrophilic affinity tag. This peptide with the sequence DYKDDDDK, is recognized by the monoclonal mouse antibody M1 (anti-FLAG antibody), and can be used to facilitate superior detection and purification of recombinant fusion proteins. FLAG has proven utility in numerous applications such as Western blotting, immunocytochemistry, flow cytometry, protein purification, and in the study of protein-protein interactions and protein localization.

► [Affinity Chromatography and *In Vitro* Binding \(Beads\)](#)

► [Protein Tags](#)

FIAsh/ReAsH System

Definition

(Red) Fluorescent Arsenic Helix binder is an arsenic organic compound that fluoresce upon binding of tetracysteine motifs expressed in living cells. It is a method for the tagging and labeling of protein molecules, which allows correlative microscopy.

► [Electron Tomography](#)

Flexible Docking

Definition

Some or all of the possible conformational changes of the molecules upon binding are considered explicitly in the docking procedure. The search space for the docked complex is larger than for rigid-body docking, and

typically the searched degrees of freedom must be restricted to those considered most important. It is common, for example, to treat a protein target as rigid, and a low-molecular weight ligand as conformationally flexible.

► [Molecular Docking](#)

FLIM

► [Fluorescence Lifetime Imaging](#)

Flippase

Definition

Flippase is an integral membrane protein that facilitates either an energy-independent equilibration of lipids between the two membrane halves, or actively translocates lipids from one leaflet to the other (energy-dependent).

► [Biological Membranes](#)

Flow Cytometry

Definition

Flow cytometry (Cytofluorometry) is a method for quantitating components or structural features of cells, primarily by optical means. A suspension of cells or particles is passed through laser light focussed to a beam. As a cell or particle is struck by the beam, it emits scattered light. If the cells have been stained with fluorochromes, they will also emit fluorescence light at levels that directly corresponds to the density of fluorochrome inside or on the cells. The signals are picked up by detectors. Over 1000 cells per second can be distinguished by quantitating biophysical features (size, granularity) and biochemical properties (DNA content, expression of surface markers). These signals are then converted for computer storage and data analysis, and can provide information about various cellular properties. Flow cytometry is often used to characterize blood cells during infection or inflammation.

Many larger flow cytometers can also be used as ► [fluorescence-activated cell sorters](#) (FACS), which sort cells according to selected properties.

► [Immunochemical Methods, Monoclonal Antibodies](#)
 ► [Camel as a Model for Functional Genomics](#)

Floxed

Definition

Floxed designates a stretch of DNA that is flanked by *loxP* sites on either end.

► [Cre/loxP Strategies](#)
 ► [Mouse Genomics](#)

FLP/FRT System

Definition

The FLP/FRT system is a yeast site-specific recombination system applicable to many model organisms. The FLP/FRT system has been used in *Drosophila* and several other organisms to induce somatic recombination, generate chromosomal rearrangements, and as a basis for gene targeting. To induce mitotic recombination, the site-specific recombinase FLP (flippase) gene and the FRT target sequence have been introduced into flies. Induction of FLP expression (usually using heat-shock inducible FLP) catalyzes exchange of the homologous chromosome arms and the generation of somatic clones

► [Drosophila as a Model Organism for Functional Genomics](#)

FLT3

► [FMS-Like Tyrosine Kinase 3 Gene](#)

Fluorescence

Definition

Fluorescence denotes a process by which an atom or molecule is transiently excited by absorption of light at

the proper energy level (usually ultraviolet, visible or infrared light, excitation), and the absorbed energy is released as a photon having a wavelength longer than the absorbed energy (emission). The fluorescence excitation and emission processes usually occur in less than a nanosecond.

- [Fluorescence Correlation Spectroscopy](#)
- [Fluorescence Microscopy](#)
- [FRAP](#)
- [FRET](#)
- [In Vivo Imaging of Transgenic Mice with Fluorescent Protein Expression](#)
- [Protein Microarrays as Tool for Protein Profiling and Identification of Novel Protein-Protein Interactions](#)

Fluorescence Anisotropy

Definition

Similar to excitation, fluorescence emission by the excited fluorophore occurs in a plane oriented parallel to the emission dipole moment. The dipole moments for excitation and emission have a defined orientation with respect to the molecular axes, but can be separated from each other by an angle. During the excited state lifetime, rotation of the fluorophore will depolarize its emission with respect to the excitation polarization history. Fluorescence anisotropy is defined as the ratio of the difference between the emission intensity parallel, and perpendicular to the polarized electric vector of the exciting light in relation to its total emission intensity.

- [FRET](#)

Fluorescence Correlation Spectroscopy

PETRA SCHWILLE

Institute for Biophysics/BioTec, Dresden University of Technology, Dresden, Germany
 petra.schwille@biotec.tu-dresden.de

Definition

Fluorescence correlation spectroscopy (FCS) is a spectroscopic technique based on the laser-induced fluorescence of very small quantities of specifically

labeled biomolecules. Its primary measurement parameter is the spectrally resolved fluorescence intensity from a small spatial element, typically the diffraction-limited spot of a laser beam focused by a high-resolution microscope objective into a low concentration (100 nM and below) solution of fluorescent sample. In contrast to classical spectroscopy, special attention is not given to the average signal, but rather to the resolution of its temporal variations, induced by statistical fluctuations of measurement parameters such as particle number and brightness. Since these fluctuations are mediated by the thermal dynamics of single molecules, FCS is usually subsumed into the family of so-called single-molecule techniques although its standard measurement signal is ensemble-averaged due to comparably long (10–100 s) measurement times. Information that can be revealed by FCS on a single molecule level comprises local concentrations, mobility parameters, rate constants of association and dissociation reactions and time scales of internal fluctuations due to structural or photophysical dynamics. In combination with specific dyes, time scales of internal fluctuations can yield environmental parameters such as pH, ionic strength or oxygen levels. The best performance of FCS is obtained with aqueous solutions of 1–10 nM with average occupation numbers of 1–10 molecules simultaneously present in the measurement volume, which usually has the size of an *E. coli* bacterial cell, *ca.* 10^{-15} liters (femtoliters) and below. Cellular applications are possible, however particles must be sufficiently mobile in order to be successfully analyzed. To study molecular interactions between different molecular species, dual- or multi-color applications such as fluorescence cross-correlation spectroscopy are particularly powerful. During the past years, ► [two-photon excitation](#) has become an attractive alternative for cellular applications, due to the limited cumulative ► [photobleaching](#) and the possibility of exciting more than one fluorescent species with a single (infrared) laser line.

Description History

FCS was conceptually developed in the early 70s, as a special type of ► [relaxation technique](#) (1). The underlying idea is to follow the relaxation of a measurement parameter back to its equilibrium value after a small spontaneous deviation induced by thermal fluctuations. To resolve spontaneous, non-coordinated fluctuations, the size of the measurement system had to be reduced to the single molecule level. Background reduction and sufficient excitation of fluorophores were both optimized by the introduction of the laser-illuminated ► [confocal](#) setup that first allowed single molecule detection and triggered the development of modern

FCS (2) with its high sensitivity and manifold parameters to be analyzed. Since then, FCS has been commercialized and found widespread applications in analytical biochemistry, biotechnology and lately, also cell biology (3, 4, 5, 6).

Theory

► **Correlation analysis** is the mathematical formalism applied to simplify the evaluation of the extremely large numbers of single recorded fluctuation events during a data recording time of seconds to minutes. With a single measured signal, fluorescence ► **auto-correlation** analysis is performed. For two or more measured signals, auto- and cross-correlation functions can be determined. The theory for both modes is outlined in detail in (3). Autocorrelation analysis compares a fluctuating signal with itself at different times and thereby reveals characteristic repetitive time patterns, such as diffusion-mediated single molecule transits through the volume element or internal fluctuations during the residence time of a single molecule. Cross-correlation (4, 6) determines the relationship between different signals due to molecular interactions.

Setup and Data Recording

The setup that is nowadays employed in most FCS measurements is the confocal setup (2). Here, a microscope objective with a high numerical aperture (0.9 or larger) is illuminated through its back aperture by a parallel laser beam of large diameter (ideally the size of the back aperture). The laser is focused to a small diffraction-limited spot of less than 0.5 μm diameter in the sample. Fluorescence excited in this region is collected by the same objective, split from the excitation light by a dichroic mirror and imaged onto the surface of an extremely sensitive, single-photon counting avalanche photodiode. Resolution in the axial direction is enforced by a small (<100 μm) pinhole in front of the detector that can be replaced by an optical fiber of corresponding core diameter. The fluorescence recorded by the detector is coupled into a computer and correlated by hardware or software correlators quasi on-line to follow the correlation curves during the measurement on the monitor. This allows the tracking of artifacts of signal loss during data recording. Dual- or multi-color applications (5) can be employed by superimposing more than one excitation laser line in the focal volume element and by splitting the detection signal by additional dichroic mirrors before directing it to the different detectors. For ► **two-photon excitation**, high performance tunable pulsed laser systems (typically mode-locked titanium-sapphire systems of *ca.* 100 fs pulse width) are employed. In modern commercial systems, FCS units are often built into confocal scanning microscopes to address specific sites

for intracellular FCS applications. The cell is first imaged for identification of interesting measurement locations and the focal spot is then parked at a specific site for FCS recording. Mobile molecules in aqueous solution traverse the fixed volume element on time scales between microseconds and seconds, depending on their diffusional mobility. With measurement times in the seconds to minutes regime, thousands of single molecule transits are usually averaged, which ensures high statistical quality of the signal.

Labeling Dyes

The labeling dyes employed for FCS and other single molecule techniques must fulfill the same requirements as in standard fluorescence applications, such as high ► **absorption coefficients** and fluorescence ► **quantum yields**. Another property of key relevance for these ultrasensitive measurements is however a high photostability, i.e. the potential to withstand the large illumination intensity in the laser focus for a sufficiently long time (at least a single molecular transit time through the measurement volume). Photostability is usually quantified in FCS measurements by the maximum number of photons that can be harvested from a single molecule per unit time. Values of over 100,000 photons/sec per molecule in buffer solutions and at least 1,000 photons/sec per molecule in cellular applications have to be reached to guarantee reliable data evaluation in FCS. Frequently used dyes are several rhodamine derivatives (Rhodamine Green, TAMRA, Texas Red), cyanines (Cy-2, Cy-3, Cy-5), the Alexa dye family (from Alexa-488 in the blue to Alexa-647 in the red) and all ► **GFP** mutants and homologues with emissions larger than 500 nm (BFP and CFP are usually not photostable enough). EGFP and DsRed make very good pairs for intracellular dual-color cross-correlation applications. In buffer, Alexa-488 and Cy-5 are usually combined.

Concentration Range

Above 100 nM, fluctuations are hardly resolved and the signal-to-noise-level in FCS becomes too low. Below 1 nM, measurement artifacts due to fluorescent contaminations of the solvent can become substantial, and special care must be exerted to purify water and buffer components. In principle, there are no lower limits for the detection of fluorescent particles (single molecule aspect of FCS). However, long measurement times are required for subnanomolar concentrations in order to collect statistically relevant data.

Clinical Applications

FCS is a high-performance biophysical technique and has, due to its experimental challenges, not yet found widespread clinical applications, although several medical centers already possess a commercial FCS

instrument in combination with a confocal scanning microscope. The most obvious medical applications of FCS are in diagnostics, where its extreme sensitivity can be utilized. FCS can be employed to detect the specific recognition of ligands by receptors, antigens by antibodies and dimerization or aggregation of molecules at very low (nanomolar and below) concentrations in real time. If specific fluorescent probes are available, FCS can, in principle, facilitate pathogen detection in various body liquids down to pico- and femtomolar concentrations. However, this is only sensible at very low background levels and if additional steps for high sample throughput (beam-scanning or sample flow through a glass capillary) are taken. The most useful biological applications of FCS to date are probably

1. the determination of mobility parameters in living cells that allow probing of the organelle affiliation of proteins (e.g. cytosolic vs membrane-bound) or their specific binding to large or immobilized targets (3, 6)
2. the dynamic co-localization (co-diffusion) of biomolecules in solution, cells or tissue measured by dual-color cross-correlation. Co-diffusion is a valuable means for precisely determining fractions of bound and free partners in recognition and dissociation assays and a powerful complement to the ►FRET measurements that are presently often employed for these purposes (6).

References

1. Magde D, Elson EL, Webb WW (1972) Phys Rev Lett 29:705
2. Eigen M, Rigler R (1994) Sorting single molecules: application to diagnostics and evolutionary biotechnology. Proc Natl Acad Sci USA 91:5740–5747
3. Elson EL, Rigler R (2001) Fluorescence correlation spectroscopy. Theory and applications, Springer, Berlin Heidelberg New York
4. Schille P, Meyer-Almes F-J, Rigler R (1997) Dual-color fluorescence cross-correlation spectroscopy for multi-component diffusional analysis in solution. Biophys J 72:1878–1886
5. Schille P, Haupts U, Maiti S et al (1999) Molecular dynamics in living cells observed by fluorescence correlation spectroscopy with one- and two-photon excitation. Biophys J 77:2251–2265
6. Bacia K, Schille P (2003) A dynamic view of cellular processes by in vivo fluorescence auto- and cross-correlation spectroscopy. Methods 29:74–85

Fluorescence *In Situ* Hybridization

►FISH

Fluorescence Labelling

Definition

Proteins of two different biological conditions are labelled with two different fluorescent compounds, mixed and separated within one 2-DE gel (DIGE, fluorescence difference gel electrophoresis). The fluorescence of a spot is measured at two different wavelengths, and the intensity difference corresponds to the difference of the amount of a certain protein species between the two biological conditions under investigation.

►Two-Dimensional Gel Electrophoresis

Fluorescence Lifetime Imaging

Definition

Fluorescence Lifetime Imaging (FLIM) is a fluorescence microscopy technology, in which image contrast in the fluorescence microscope is derived from the fluorescence lifetime at each point in the image. Biological materials (such as proteins, lipids, nucleic acids, and ions) are labelled with fluorescent dyes. Fluorescence decay times (typically some nanoseconds after excitation) are then measured on a pixel by pixel basis. The lifetime of these fluorophores in the living cell is sensitive to the surrounding intracellular environment, namely ion intensity (e.g. hydrogen ion concentration (pH), Ca^{2+} , Mg^{2+} , Cl^-), hydrophobic properties, oxygen concentration, as well as to molecular binding and interaction. Fluorescence lifetimes are mostly independent of the local probe concentration and photobleaching. Thus, fluorescence lifetime microscopy allows to perform accurate ion concentration measurements. The binding or the interaction between molecules can be measured in combination with FRET (Fluorescence Resonance Energy Transfer).

►FRET

Fluorescence Microscopy

Definition

A common clinical and research technique in optical microscopy, which relies on excitation of fluorescent molecules with a specific wavelength region to produce

an image generated by fluorescence emission at longer wavelengths.

- [Fluorescence Microscopy: Single Particle Tracking](#)
- [Fluorescence Microscopy: Spectral Imaging vs. Filter-Based Imaging](#)
- [In Vivo Imaging of Transgenic Mice with Fluorescent Protein Expression](#)
- [Wide-Field Fluorescence Microscopy](#)

Fluorescence Microscopy: Single Particle Tracking

ULRICH KUBITSCHKE

Institute of Physical and Theoretical Chemistry,
Bonn, Germany
u.kubitschke@uni-bonn.de

Definition

In single-particle tracking (SPT) high-speed and high-sensitivity video microscopy is employed to trace the movements of single particles – in this context usually single lipid, protein or nucleic acid molecules or single molecular complexes composed of them – on membranes (artificial bilayers or cell membranes) or within the cellular interior. After data acquisition, the respective positions of the single particles are determined by image processing algorithms thus providing a time sequence of positions, which form a pathway or trajectory. Typically, a time resolution of milliseconds and a [localization precision](#) of a few tens of nanometers can be achieved.

Description

Single particle or single molecule tracking is based on the microscopic imaging of single molecules. The molecules in question must be labeled in a specific way. This is usually done by fluorescent tags. A wide range of fluorescent entities are in use, ranging from autofluorescent proteins, which can be genetically coupled to proteins, to fluorescent dyes, which can be coupled covalently to purified material, and to the recently developed quantum dots.

Highly diluted, singled-out fluorescent molecules produce diffraction-limited light spots in the image plane upon efficient excitation. This intensity pattern is formally described by an Airy function, which is very similar to a simple two-dimensional (2D) Gaussian function. The intensity distribution has a full width at half maximum of 250–300 nm, if green light and a high numerical aperture microscope objective lens is used for imaging. Thus, the shape of nanometer-sized objects cannot be recognized. However, the center of

mass of the intensity distribution, which corresponds to the position of the object, can be determined by fitting a two-dimensional Gaussian function to the observed intensity distribution with high precision. This precision depends on the signal-to-noise ratio (SNR) of the image, and the physical stability of the experimental setup employed. Under the sub-optimal conditions in a live cell experiment, where the SNR is hampered by background fluorescence, a localization precision of 20–50 nm can be achieved on a regular basis.

Instrumental preconditions for SPT are the optimization of light detection by carefully adjusting the optical filter systems to the employed fluorescence bandwidth, the utilization of strong excitation light sources usually in form of laser light, the use of [super-resolution](#) and last but not least the employment of sensitive, fast [CCD camera](#) systems. The digital images obtained must be processed after data acquisition by automatic pattern recognition methods and peak finding algorithms. These procedures initially allowed the mapping of single fluorescent molecules in lipid bilayers and later on the surface and within the interior of living cells (1, 2).

The observation of molecules within axially extended systems such as liquids or the interior of living cells is hampered by the problem of an enhanced background due to out-of-focus fluorescence. Only a single plane – the focal plane – is imaged sharply. All fluorescent structures or particles present in the spaces above and below the focal plane do not contribute to a sharp image, but produce an overlaying image blur, which reduces the SNR of the single molecule signals in focus. The general effect is well known from photography, but in microscopy it is more severe due to the small focal length of the microscope objective lens employed. Therefore, a fluctuating background is produced by fluorescent particles, which move around in the out-of-focus space. However, for low enough substrate concentrations the signals from the focal plane still rise high enough above this background to achieve a localization with nanometer precision.

By single particle observation in the cellular interior, the transport pathways of particles such as mobile proteins, molecular motors, molecular complexes or vesicles in intact biological systems can be observed. Conclusions with regard to the underlying cellular architecture can be drawn and mobility restrictions as well as binding processes can be measured. It is easier than with other techniques such as [fluorescence recovery after photobleaching](#) (FRAP) or [fluorescence correlation spectroscopy](#) (FCS) to discriminate different forms of mobility in heterogeneous systems. SPT is especially well suited to analyze slow dynamic processes characterized by diffusion coefficients below 20 $\mu\text{m}^2/\text{s}$. One could designate SPT as a complementary technique to FCS.

The acquisition of image series with high frame rates permits the tracing of single molecule pathways with the mentioned high precision, if the motion of the molecules during image acquisition is negligible. A further requirement is that the molecules remain and move in the focal region of the microscope objective lens. The signals of molecules moving away from the focal plane are quickly dominated by background noise. Hence, single-molecule imaging has an intrinsic three-dimensional (3D) resolution. For immobile or slowly moving molecules the inherent 3D resolution can be used to obtain time series of 3D image stacks (i.e. 4D image data), from which the 3D positions of the single molecules can be determined as a function of time without using confocal optics.

Experimental Setup

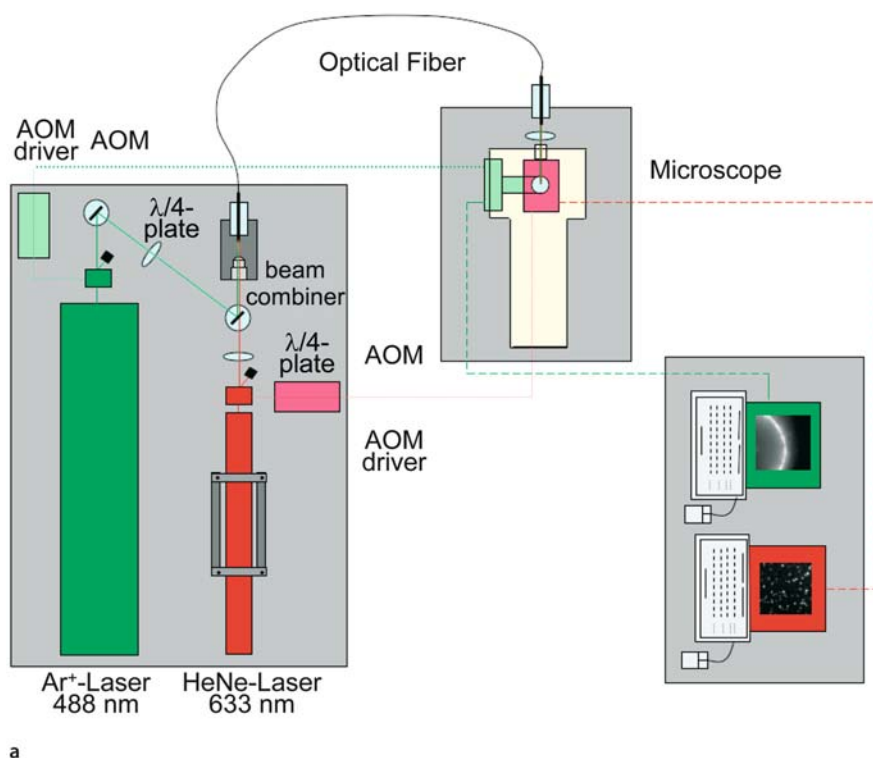
Microscopic imaging of single particles or even molecules requires a fluorescence video microscope of the highest optical sensitivity and physical stability. Although different experimental setups can be employed, several general features must be present (Fig. 1). Optimal fluorescence light detection efficiency is of greatest significance. Utilization of at least two fluorescence channels is beneficial. The core of the instrument is often a commercial inverse fluorescence microscope, because these allow a manipulation of biological samples during the experiment. The shape of the point spread function (PSF) produces the intrinsic 3D resolution in single particle imaging, if the probe concentration is low enough. In this context, low means that considerably less than one fluorescent particle should be present within one focal volume element. This corresponds to approximately 0.25 femtoliter when a high numerical aperture objective lens is used and results in a final particle concentration of less than 100 picomolar. Typically lasers are used for fluorescence excitation. The laser light is passed through acousto-optical modulators for rapid intensity switching and is circular polarized by $\lambda/4$ -plates. If several lasers are used, the beams are combined by appropriate beam splitters and passed through an optical mono-mode fiber to the microscope. A collimated light beam exits the fiber and illuminates with a Gaussian intensity distribution a small area of 10 by 10 μm^2 in the object field. The irradiance may conveniently be adjusted using analog acousto-optical modulators (AOMs) or acousto-optical tunable filters (AOTFs). They work as switches to turn on and off the laser irradiation, and define the transmitted power. The irradiance within the sample is thus set to 0.1–2 kW/cm^2 . The emitted fluorescence light is separated from the excitation light by a dichromatic beam splitter, which reflects the excitation lines onto the sample, while the respective fluorescence emission bands are transmitted to the detection system. The fast image acquisition of

scientific-grade CCD cameras is well suited for imaging and tracking of possibly fast moving molecules. If two laser lines are used for excitation of two separate fluorescent dyes, the two fluorescence bands must be separated by an additional dichromatic beam splitter in front of the detection cameras. This allows a simultaneous recording of two fluorescence channels. The insertion of a dichromatic wedge filter at the position of the final dichromatic beam splitter is another solution for obtaining dual or even triple channel detection. It projects two chromatically different and laterally offset images onto a single CCD. Different fluorescence images from an identical object region can be aligned to each other by calibration with images from multicolored microbeads. At the start and the end of the image integration time, the CCD cameras issue trigger signals, which are fed into the drivers of the acousto-optical devices and thereby turn the respective laser illumination on and off. This ensures that the illumination of the sample is kept to a minimum. In general, light integration over 1–50 ms is required to obtain images from single molecules with an SNR, which is sufficient for a nanolocalization with a precision in the nanometer range.

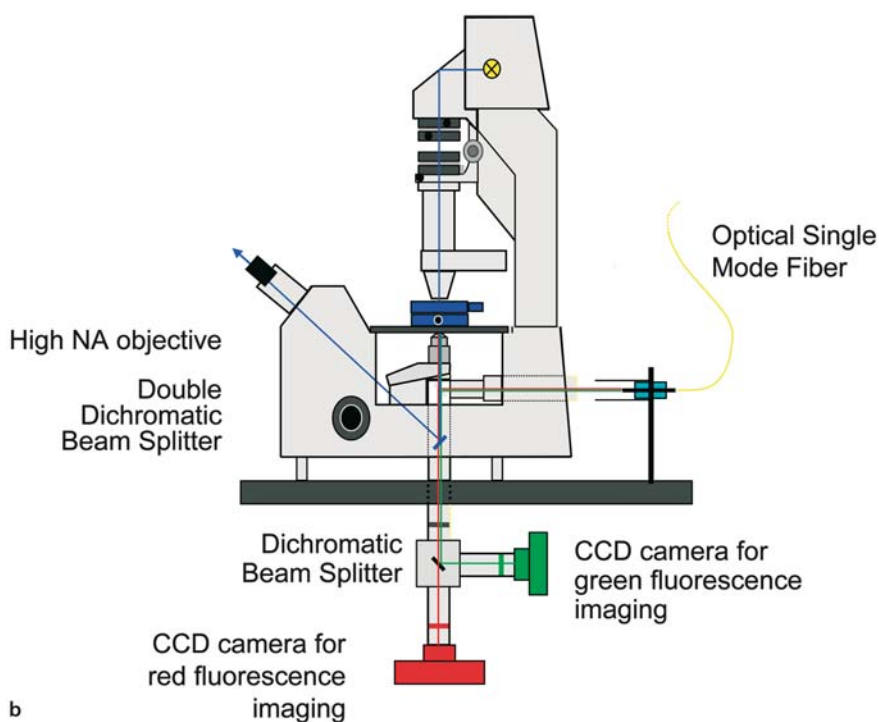
When imaging fast moving objects such as mobile single molecules, a high imaging frame rate is of great importance. Considering the small image integration time, frame rates up to 1 kHz should be achievable. However, in a scientific-grade CCD camera the image readout is rate limiting. For example, the time required to read-out a full image on a chip with 1000×1000 pixels is in the range of 1 s or longer, causing rather low frame rates. A simple method to accelerate the rate is to read out only a subregion of the full frame, which yields a frame rate in the range of 10/s. The frame rate may be further considerably increased by using high-speed imaging techniques. In one technique a mask shields a major part of the CCD from the illumination light. This region of the chip may be used as an intermediate image storage. By direct programming of the chip, several subframes can be acquired rapidly and transferred into this intermediate storage area with high speed before the image series is finally read out. This approach yields high frame rates – but only for about 5–20 images in a succession. A more direct approach to increasing the frame rate is to use a high-speed camera such as an [intensified CCD](#) or an [electron-multiplying CCD](#). These devices combine high sensitivity and high speed in an ideal way, and reach imaging rates that correspond directly to the inverse of the image integration time, i.e. close to 1000/s.

Particle Localization with Nanometer Precision

The image of a particle with a size significantly smaller than light microscopic resolution – like a single

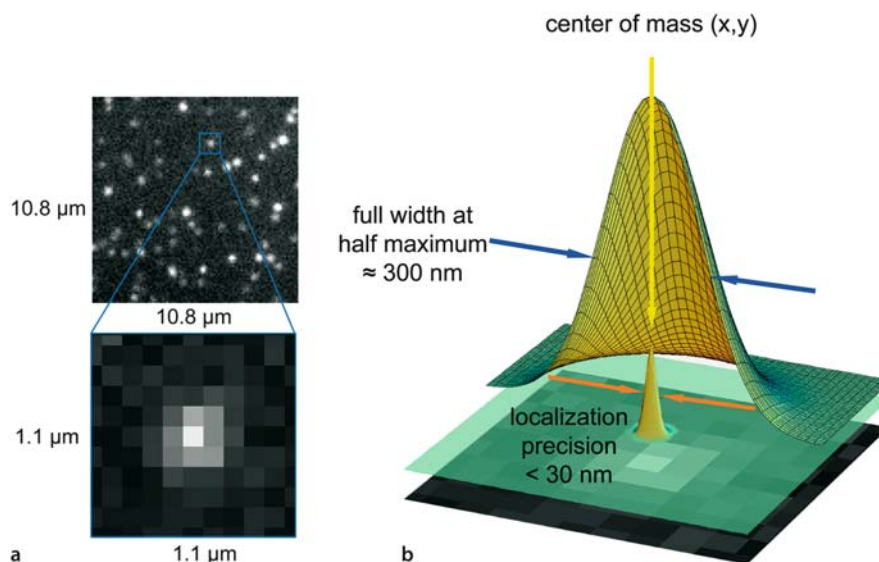


a



b

Fluorescence Microscopy: Single Particle Tracking. Figure 1 Instrument for single molecule tracking. Detailed views of the microscopic setup. (a) Overview and (b) closeup of the microscope. A detailed description of the setup is given in the text.



Fluorescence Microscopy: Single Particle Tracking. Figure 2 (a) Image of a single fluorescence labeled protein molecule (surface-attached BSA molecules labeled by Oyster565, excitation wavelength 633 nm, emission above 650 nm, irradiance 2 kW/cm²). (b) The 2D-Gaussian function (surface plot) approximating the intensity distribution (grayed image, below) yields the position of the center of mass with a precision (orange arrows), which is significantly smaller than the half-width of the diffraction limited spot size (arrows).

molecule – has a complex intensity profile, which is designated as the PSF. The PSF limits the spatial resolution of a microscope and can be measured by imaging a subresolution object under super-resolution conditions. The experimentally observed PSF can be well approximated by a 2D Gaussian function (Fig. 2). By fitting a 2D Gaussian to the observed intensity profile of a submicroscopic particle, the position of the object center can be determined with a very high precision, which mostly depends on the SNR of the image. This localization precision is usually 10–20× smaller than the width of the PSF. Image acquisition is a stochastic process due to various experimental noise sources. In single-molecule imaging, the inherent photon Poisson noise is inevitable and the camera readout noise is the second important noise source. Therefore the object position that is found by fitting algorithms is merely an estimator for the true object position.

The SNR of a particle image is determined by the brightness of the utilized probe and specific instrument parameters such as the image integration time, detector pixel size, image readout rate, photoelectron conversion factor and laser irradiance. The SNR of a single molecule signal is defined as:

$$\text{SNR} = \frac{I_0}{\sqrt{\sigma_{\text{bg}}^2 + \sigma_{I_0}^2}} \quad (1)$$

where I_0 designates the signal intensity of the maximum of the PSF above background, σ_{bg}^2 the

variance of the background noise and $\sigma_{I_0}^2$ the variance of the maximum signal intensity above background.

An analytical approximation for the localization precision of subresolution objects can be derived (3, 4, 5). An expectation value for the localization precision can thereby be calculated as a function of the SNR, which represents a limit for the experimentally achievable localization precision. For example, an SNR between 10 and 2 corresponds to a localization precision between 10 and 60 nm.

Clinical Applications

Single molecule techniques can be employed for studying basic cell biological systems and processes. Numerous biological processes are initiated by a very small number of effector molecules, e.g. receptor activation as the first step in signal transduction chains or DNA replication or transcription. A detection limit at the single molecule level allows the analysis of the start of such biochemical reaction cascades. Furthermore, important processes within living cells take place on length scales between molecular distances (0.1–1 nm) and the optical microscope resolution limit (≈ 250 nm). Such processes on an “intermediate-sized” length scale usually involve supramolecular complexes containing many nucleic acids and proteins. This makes their reconstitution and study *in vitro* difficult, and their size complicates direct observation and analysis with conventional microscopy. Due to the localization precision in the nanometer range, single molecule

microscopy is well suited for the study of such systems *in vivo*.

In the future a major application for single molecule detection techniques will be the molecular description of individual cells containing or expressing a distinctive collection of receptor molecules. SPT will have an application in all diagnostic situations, where diagnostic information may be derived from a few distinctive molecules on or within a small number of cells. Examples of this are the occurrence of graft *versus* host reactions or graft *versus* leukemia reactions after transplantation or the characterization of success after tumor therapy.

Therapeutic Consequences

In the ideal case, single particle tracking within living cells provides a view of cellular molecular dynamics by directly showing the pathways and activities of intracellular molecular components. It can provide information on molecular mobility thus giving insight into structural determinants of lipid or protein mobility. It reveals information on the assembly sites of protein complexes and ultimately even on interaction topology, loci and partners. A central topic in the current progress of medical research is the search for individual molecular patterns in a single patient's disease to find improved targets for individually fine-tuned therapeutic strategies. SPT provides the means to approach these questions.

References

1. Saxton MJ, Jacobson K (1997) Single-particle tracking: applications to membrane dynamics. *Annu Rev Biophys Biomol Struct* 26:373–399
2. Sako Y, Yanagida T (2003) Single-molecule visualization in cell biology. *Suppl. to Nature Rev Mol Cell Biol* 4: SS1–SS5
3. Kubitschek U, Kückmann O, Kues T et al (2000) Imaging and tracking of single GFP molecules in solution. *Biophys J* 78:2170–2179
4. Thompson RE, Larson DR, Webb WW (2002) Precise nanometer localization analysis for individual fluorescent probes. *Biophys J* 82:2775–2783
5. Ober RJ, Ram S, Ward ES (2004) Localization accuracy in single-molecule microscopy. *Biophys J* 86:1185–1200

Fluorescence Microscopy: Spectral Imaging vs. Filter-based Imaging

M. CRISTINA CARDOSO

Max Delbrueck Center for Molecular Medicine, Berlin, Germany
cardoso@mdc-berlin.de

Definition

In spectral imaging, the entire spectrum of emitted light is utilized by positioning a prism or optical grating device in the emitted light path, which diffracts the light into its spectral components. These can be acquired in a sequential mode or, more efficiently, in parallel detectors. In contrast, optical filters positioned in the light path transmit a selected range of wavelengths with high efficiency while rejecting, through reflection and destructive interference, all other wavelengths.

Description

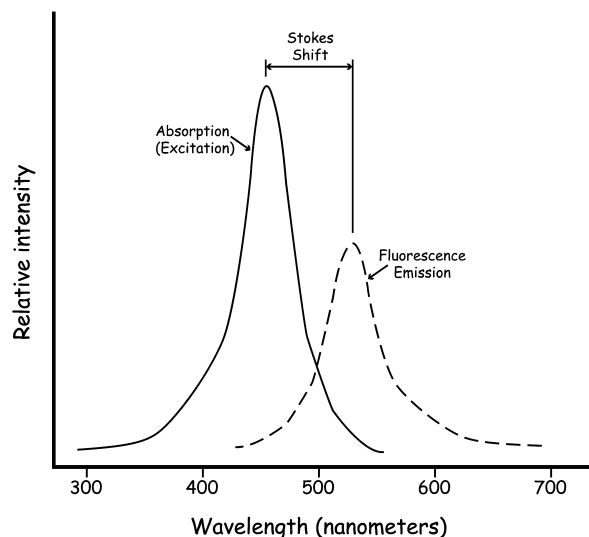
Light is an extremely versatile tool for studying biological systems and their functions in a non-invasive manner. When light shines upon an object, it can be reflected, transmitted, absorbed or re-emitted. Spectroscopy measures these phenomena to determine the physical properties of the substance being studied. The addition of imaging to spectroscopy aims to provide spatial information to the ability to determine what substances are contained within the object. Our own vision is a form of imaging spectroscopy. The combination of imaging and spectroscopy was pioneered on airborne satellite-based remote sensing, with the aim of determining from a distance the distribution of natural or man-made substances based on their spectral signatures.

Optical imaging combines a wide variety of light sources with non-invasiveness and high spatial and temporal resolution, allowing the investigation of structure-function from the subcellular level up to entire organisms and in real time. Fluorescence is the most rapidly expanding optical microscopy technique in both the medical and biological sciences, which has spurred the development of highly sophisticated fluorescence microscopes. In contrast to other modes of optical microscopy, fluorescence microscopy allows the detection of a single molecular species based on its light emission characteristics. ▶ **Fluorescence** is the process by which an atom or molecule absorbs light at a particular ▶ **wavelength** and subsequently emits light of a longer wavelength (the difference between excitation and emission wavelengths is termed the ▶ **Stokes shift**). Thus, using fluorescence microscopy one can determine with high spatial and temporal resolution, the precise subcellular localization of one or several fluorescently labeled molecules within a living cell. Furthermore, diffusion rates and interactions with other molecules as well as their changes in time and with pathological alterations can be determined. In addition, the fact that fluorochromes are susceptible to environmental influences (pH, oxygen content, ion concentrations, etc.) permits using them as a probe for local environmental characteristics.

The ever-expanding range of fluorescence microscopy applications requires the quantitative analysis of a wide number of fluorochromes at the same time and place within a sample. This requires methods that allow good discrimination of, often spectrally overlapping, fluorochromes. ►**Fluorochromes** (also called ►**fluorophores**) are substances, which fluoresce upon irradiation and are characterized by their excitation and emission spectra. The absorption or ►**excitation spectrum** is obtained by measuring the relative fluorescence intensity at a certain wavelength when the specimen is excited at varying wavelengths. The ►**emission spectrum** is red-shifted (Stokes shift) and results from excitation at a certain wavelength. Most excitation and emission curves overlap to a certain extent (Fig. 1).

The essential feature of a fluorescence microscope is to provide a way of exciting the specimen with a selected wavelength followed by isolating the much weaker fluorescence emission (three to six orders of magnitude less than the excitation) coming from the specimen. The majority of commonly used light sources (natural and artificial) are polychromatic, i.e., emit a broad range of wavelengths that cover the entire visible light spectrum and often extend into the ultraviolet and infrared regions. In order to select the wavelength of excitation from a polychromatic light source, specialized filters are normally used. In addition, filters are also used to discriminate between excitation light and the light emitted by the specimen. This discrimination is possible since the emitted light is of longer wavelength than the excitation light. The choice of filters is of paramount importance in fluorescence microscopy and depends on the spectral characteristics of the fluorochromes used. Filters are constructed in a wide variety of shapes and physical dimensions and can be employed to remove or pass wavelength bands ranging in size from hundreds of nanometers down to a single wavelength. Many filters work by absorbing light, while others reflect unwanted light, but pass a selected range of wavelengths. The characteristics of a filter are seen by its transmission curve. The main types of filters for fluorescence microscopy are interference filters and ►**dichromatic beamsplitters** (3).

►**Interference filters** are glass substrates onto which thin layers of metal salts are deposited. There are different types, named band-pass and short and long-pass interference filters. They operate by transmitting a selected wavelength region with high efficiency while rejecting, through reflection and destructive interference, all other wavelengths. They can be positioned in front of the excitation light source (commonly a mercury or xenon lamp) to select the excitation wavelength or before the detection system (most often



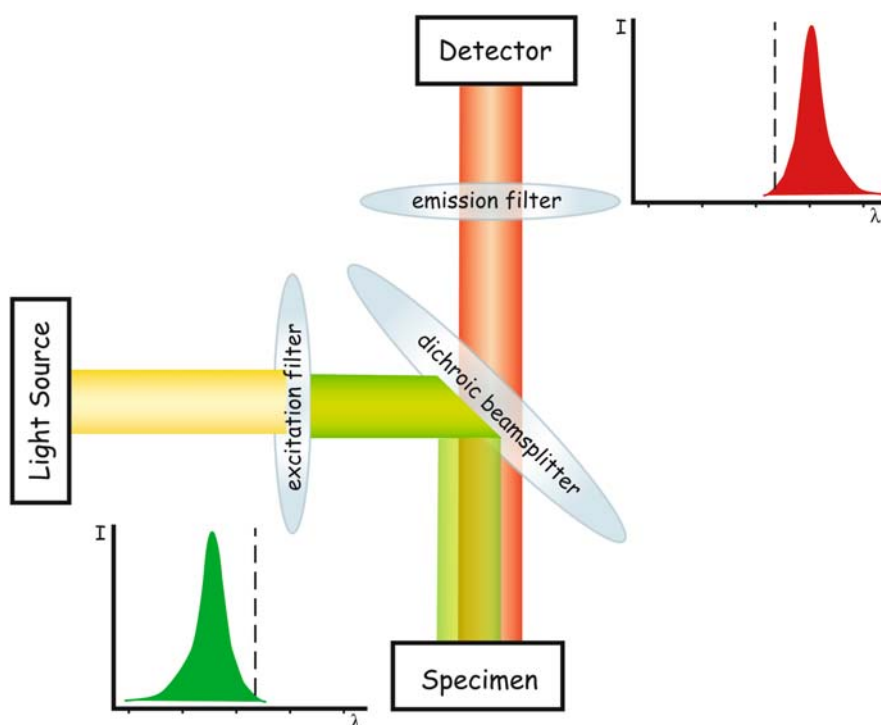
Fluorescence Microscopy: Spectral Imaging vs. Filter-based Imaging. Figure 1 Excitation (solid line) and emission (dashed line) spectra of a fluorochrome.

nowadays a CCD camera) to select a range of emitted wavelengths.

Dichromatic beamsplitters reflect light of wavelengths shorter than the specified wavelength and transmit light of longer wavelengths. The dichroic beamsplitter is positioned at a 45-degree angle to the optical axis of the illumination path and reflects the exciting rays into the objective in an epi-illumination setup where the objective serves as condenser. Light emitted by the sample is collected by the objective, traverses the dichroic filter and is further selected by the emission (interference) filter.

A typical filter set configuration for fluorescence microscopy consists of an excitation (interference) filter, a dichromatic beamsplitter and an emission (interference) filter mounted in a cube as shown in Fig. 2. The individual filters can also be positioned each separately into the light path as part of rotating filter wheels, which permit different combinations of the three basic components to be chosen in a more flexible way. The filters in a configuration should be chosen so that the wavelength range transmitted by the excitation interference filter matches the reflection range of the dichroic. In that way, the excitation light is effectively directed onto the sample. In addition the emitted light from the specimen must match a high transmission range from the dichroic in order to pass through to the detector. The emission filter, albeit less relevant, still imposes further constraints onto the light that reaches the detector.

Even in perfect combinations of filters there is always some crossover or bleedthrough of light between the

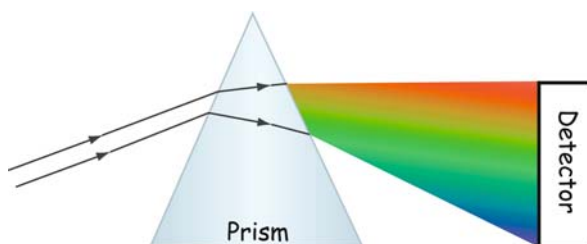


Fluorescence Microscopy: Spectral Imaging vs. Filter-based Imaging. Figure 2 Separation of excitation and emission using a classical filter set configuration. Polychromatic light (yellow light beam) is selected (green light) by the excitation interference filter and reflected onto the specimen by the dichroic beamsplitter. Fluorescence emission of longer wavelength from the sample (red light) is transmitted through the dichroic and further selected by the emission filter before it reaches the detector.

filters. Furthermore, the more discriminating a filter set is, the less light it passes and with increasing numbers of fluorochromes used in a single sample the more difficult it is to find filters which fit all the fluorochromes one needs. Moreover, the strongly overlapping emission spectra of commonly used fluorescent dyes and proteins severely limit their combined use in multicolor imaging using optical filters. A better solution is to have systems where the selection of wavelengths is flexible, in particular at the detection side or the entire range of emitted light (spectrum) is detected and resolved into the individual components. In recent years, several commercial manufacturers have started to implement both these solutions, in particular into their laser scanning confocal microscopes where the excitation light source is monochromatic.

Flexible selection of wavelength can be performed using tunable filters, in particular ▶[acousto-optical filters](#). These are electro-optical devices that allow the modulation of the light intensity as well as wavelength. Light penetrating the optical crystal is deflected depending on its wavelength and the wavelength of the ultrasonic field applied to the crystal. These filters do not have the mechanical limitations, speed

constraints and vibrations associated with the rotating filter wheels. Furthermore, they offer far greater flexibility and durability as compared to interference filters upon exposure to heat and high intensity light. In order to remove the constraints of filters on the detected range of wavelengths, spectral detectors can be used. The light emitted by the specimen consists of mixtures of wavelengths, due to autofluorescence and excitation crosstalk (excitation of multiple fluorochromes with overlapping absorption spectra) as well as the broad range of wavelengths emitted by single fluorochromes (represented in its emission spectrum). A prism or optical grating device positioned in the emitted light path diffracts this light into its spectral components, which can subsequently be acquired in a sequential mode or, more efficiently, in parallel detectors (Fig. 3). In order to select a particular range of the emission spectrum, a motorized slide aperture can be positioned before the detector, letting only a particular wavelength range of the spectrum pass. Alternatively, the entire spectrum can be detected and the contributions of the different fluorochromes calculated based on their spectral signatures. For this unmixing of the spectral signatures, the relative contributions of each of the fluorochromes into each



Fluorescence Microscopy: Spectral Imaging vs. Filter-based Imaging. Figure 3 Spectral detection of emitted light. Fluorescence emitted by the sample (arrows), consisting of a mixture of wavelengths, is dispersed *via* a prism and the entire emission spectrum gets projected onto the detection system.

detection channel are first recorded as reference spectra. The latter are used to build up a matrix of fluorophore specific weight for the calculation of an individual fluorophore's contribution to each pixel. By applying this matrix to the spectral image acquired, one can determine the contribution of the individual fluorophores at every pixel in the image and effectively separate their signals. This mathematical operation is called **linear unmixing**. In principle, any fluorescent contribution can be calculated back and this can even be used to subtract the contribution of background fluorescence from an image. The combination of spectral detection with linear unmixing analysis (1,4) allows the simultaneous use of highly overlapping fluorescent molecules, which has thus far been a major limitation in multicolor imaging. Moreover, since most of the emitted fluorescence is recorded, greatly increased sensitivity is attained as compared to classic filter-based systems.

The information summarized here focuses on commonly used fluorescence microscopy systems in biomedical research. Additional information on spectral imaging can be found elsewhere (2).

References

1. Dickinson ME, Bearman G, Tille S et al (2001) Multi-spectral imaging and linear unmixing add a whole new dimension to laser scanning fluorescence microscopy. *BioTechniques* 31:1272–1278
2. Farkas DL (2001) Spectral microscopy for quantitative cell and tissue imaging. In: Periasamy A (ed) *Methods in cellular imaging*. Oxford University Press, New York, pp 345–361
3. Ploem JS (1989) Fluorescence microscopy. In: Lacey AJ (ed) *Light microscopy in biology. A practical approach*. Oxford University Press, Oxford, pp 163–185
4. Zimmermann T, Rietdorf J, Pepperkok R (2003) Spectral imaging and its applications in live cell microscopy. *FEBS Letters* 546:87–92

Fluorescence Quantum Yield

Definition

Fluorescence quantum yield refers to the total number of photons emitted as fluorescence divided by the total number of photons absorbed by the molecule. The range of the fluorescence quantum yield is between 0 and 1.

► FRET

Fluorescence Recovery after Photobleaching

► FRAP and Other Photobleaching Methods

► Functional Imaging

Fluorescence Resonance Energy Transfer

► FRET

Fluorescence-Activated Cell Sorter

Definition

Fluorescence-activated cell sorter represents a highly specialized **flow cytometer** capable of detecting, separating and analyzing cells based on certain physical parameters, such as cell size, granularity, and the presence of cellular components, based on the presence of certain specifically bound fluorochromes.

► Flow Cytometry

► Immunochemical Methods, Localization

Fluorescent Proteins

Definition

Naturally occurring proteins with an intrinsic fluorophore generated by an autocyclization reaction of

neighboring aminoacids. The ►[green fluorescent protein](#) is derived from the jellyfish *Aequorea victoria*. Several variants with different spectral properties have been generated (enhanced green, cyan and yellow fluorescent proteins) by combinatorial mutagenesis. A novel class of fluorescent proteins has been isolated from reef corals (red FPs such as DsRed and HcRed). ►[In Vivo Imaging of Transgenic Mice with Fluorescent Protein Expression](#)

Fluorochrome

A natural or synthetic dye or molecule that is capable of exhibiting fluorescence. Fluorochromes (also termed fluorescent molecules, probes or fluorescent dyes) are usually polynuclear heterocyclic molecules containing nitrogen, sulfur, and/or oxygen with delocalized electron systems and reactive moieties that enable the compounds to be attached to a biological species.

►[Fluorescence Microscopy: Spectral Imaging vs. Filter-Based Imaging](#)

►[In Vivo Imaging of Transgenic Mice with Fluorescent Protein Expression](#)

5-Fluoroorotic Acid

Definition

5-Fluoroorotic acid (5-FOA) is a toxic analog of orotic acid, and is the most widely used negative selection marker in yeast. Yeast cells with the functional URA3 gene are sensitive to this drug, since they convert it into a product that is toxic to the cell. This negative selection system can efficiently be used to select for a specific protein-protein interaction in yeast, or to characterize drugs that specifically inhibit a protein-protein interaction.

►[Protein Interaction Analysis: Variations of the Yeast Two-Hybrid System](#)

5-FOA

►[5-Fluoroorotic Acid](#)

Fluorophore

Definition

The structural domain or specific region of a molecule that is capable of exhibiting fluorescence see also fluorophore.

►[In Vivo Imaging of Transgenic Mice with Fluorescent Protein Expression](#)

FlyBase

Definition

FlyBase (<http://flybase.bio.indiana.edu/>) is a comprehensive database for information on the genetics and molecular biology of *Drosophila*.

►[Drosophila as a Model Organism for Functional Genomics](#)

Fmoc Chemistry

Definition

FMOC chemistry describes a peptide synthesis strategy in which the 9-fluorenylmethoxycarbonyl group (Fmoc group) is used as a temporary protecting group for the N-terminus. The Fmoc group is cleaved by secondary amines like piperidine.

FMR Protein

►[FMR1](#)

FMR1 Knockout Mouse

Definition

Fmr1 knock-out mice are transgenic animal models for the human Fragile X syndrome. In these mice, the

► *fmr1* gene is inactive, and mice lack normal FMR1 protein. The knockout mice show macroorchidism, learning deficits, and hyperactivity. These knockout mice allow investigators to determine the role of the targeted gene, by observing the phenotype of individuals that lack the FMR protein completely.

► [Fragile X Syndrome](#)

► [Transgenic and Knockout Animals](#)

fmr1/FMR1

Definition

Fragile X mental retardation 1 (*fmr1*) gene is the gene that is mutated in fragile X mental retardation syndrome. *fmr1* encodes for the Fragile X mental retardation protein FMR1 (also known as FMRP), which can act as a translational suppressor in neuronal dendrites. Loss of function of the *fmr1* gene causes Fragile X syndrome.

► [Fragile X Syndrome](#)

► [Micro RNAs](#)

fMRI

► [Functional Magnetic Resonance Imaging](#)

FMRP

► [FMR Protein](#)

FMS-Like Tyrosine Kinase 3 Gene

Definition

The FMS-like tyrosine kinase 3 (FLT3) gene, also known as fetal liver kinase 2 (Flk2), encodes a membrane-bound receptor protein tyrosine kinase (RPTK), which belongs to the RTK subclass III family and has a crucial role in the proliferation, differentiation

and survival of normal haematopoietic cells. The human FLT3 gene is located on chromosome 13q12.

► [Leukemia](#)

Focal Adhesions

BERHARD WEHRLE-HALLER^{1,2}, BEAT A. IMHOF

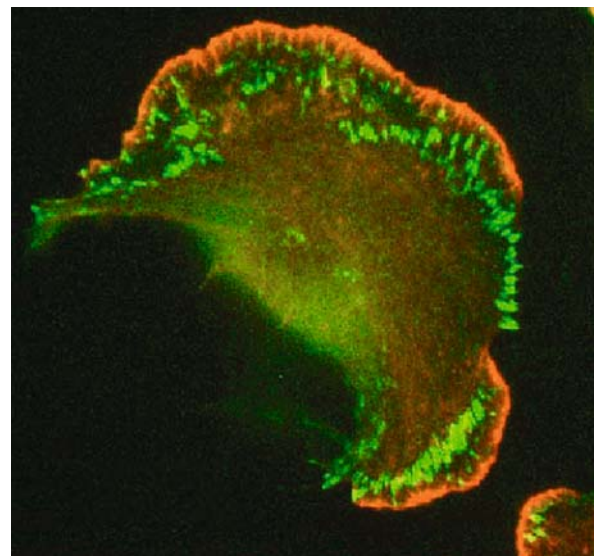
¹Department of Pathology and Immunology and

²Department of Cellular Physiology and Metabolism, University of Geneva, Geneva, Switzerland

beat.imhof@medecine.unige.ch

Definition

The integrity of an organism depends on adhesion of cells to their immediate vicinity formed by a network of extracellular matrix or a neighboring cell (Fig. 1). Cells produce adhesion molecules, the ► [integrins](#), which link the intracellular cytoskeleton to the extracellular environment (Fig. 1). Strength of cell adhesion can be achieved by grouping individual integrins in clusters,



Focal Adhesions. **Figure 1** Photograph of an adherent migrating cell. Integrins (β3 chain linked to EGFP shown in green) are organized in focal adhesions adhering to vitronectin. The actin-cytoskeleton is labeled with rhodamine in red. Ribs of bundled actin filaments ending in focal adhesions form the lamellipodium. The intracellular thin network of actin filaments exists but is difficult to visualize at the resolution shown.

the so-called focal adhesions. The formation of focal adhesions involves several steps, including integrin activation and recruitment of cytoskeletal adaptor proteins (1).

Characteristics Integrin Activation

Integrins are a family of heterodimeric receptors binding to ligands only when an activated conformation is formed (2). The ligand-binding domain of integrins consists of a headpiece in which α and β chains meet (Fig. 2). In its inactive conformation, this headpiece cannot bind to ligands and is bent towards the cell membrane through a domain called genu (hinge). Activation of the bent integrin involves a separation of the two integrin chains and opening of the molecule like a switchblade, making the headpiece available for ligand binding (3). This mechanism increases the affinity for binding to an individual ligand. However, integrins usually act in clusters and this increases the avidity for groups of integrin ligands such as extracellular matrix components. Clustering of integrins needs lateral mobility. A cascade of events and molecules induced by **▶receptor tyrosine kinases** (RTKs) or **▶G-protein coupled receptors** leads to unfolding of integrins and increased lateral mobility allowing clustering. Among the many molecules involved in integrin activation, a selected number of important ones is described here (Fig. 2).

Molecular Interactions Rap1 and Rap-Ligand

Rap1, Rap2 and their splice variants are four **▶small GTPases** of the Ras protein family that play a pivotal role in mediating inside-out signals to activate the integrins (4). They cycle between an inactive GDP-bound form and an active GTP-bound form. Guanine nucleotide exchange factors such as Epac and C3G regulate Rap1 conversion (Fig. 2). Cells transfected with constitutively active Rap1 adhere and migrate without requiring a G protein signal. Spa1 or RapGAPII are Rap specific GTPase-activation proteins that inactivate Rap1 by modifying the GTP bound into the GDP bound form. Transfection of Spa1/RapGAPII blocks integrin mediated cell adhesion and migration. In cells of hemopoietic origin, activated Rap1 binds the Rap1-ligand RAPL and this complex activates the integrin through binding to the **▶GFFKR domain** of the integrin α chain. Over-expression of RAPL can activate integrin mediated cell adhesion but expression of a RapL mutant, which is incapable of Rap1 binding, inhibits adhesion. So, cell surface receptor induced activation of Rap1 seems to be key in the signal transduction pathway that activates integrins and cell adhesion.

Talin

Talin is a cytoskeleton protein with a globular head and a rod-like domain. The head domain contains a FERM (ezrin, radixin, moesin) domain which binds to the non-phosphorylated **▶NPxY/F motif** of the β integrin chain and activates the integrin, probably by keeping the cytoplasmic tails of the α and β integrin chains at a distance allowing unfolding of the extracellular, ligand binding domains (5). The regulatory link of RTK and G protein mediated integrin activation through talin binding is not clear but talin is required for integrin activation downstream of several physiologically relevant signaling pathways. In a resting cell, talin is present in an inactive, folded conformation with masked integrin binding the FERM domain in its globular head. One way to regulate talin-integrin interactions occurs through binding of **▶PIP2** (phosphatidylinositol 4,5-bis-phosphate), which leads to a conformational change of the talin FERM domain and increases the binding affinity of talin for the integrin β chain. Interestingly the PIP2 producing enzyme type I γ -90 phosphoinositol phosphate kinase (PIPKI γ -90) can also bind to talin, which then produces its "own" localized PIP2. Unfortunately, PIPKI γ -90 binding to talin competes with talin binding to the integrin, which creates a negative loop of regulation. Another way of activating talin is its cleavage into a head and a rod domain by the protease calpain. The separated talin head domain has a six fold higher affinity for integrin β chain than does the intact molecule and hence it is a more efficient integrin activator than the intact talin. Interestingly, phosphorylating the tyrosine residue of the NPxY/F talin-binding motif of the integrin by Src-family kinases can block talin binding to the integrin. Thus, talin is a general regulatory molecule for integrin activation and there are several mechanisms to control talin binding to the integrin.

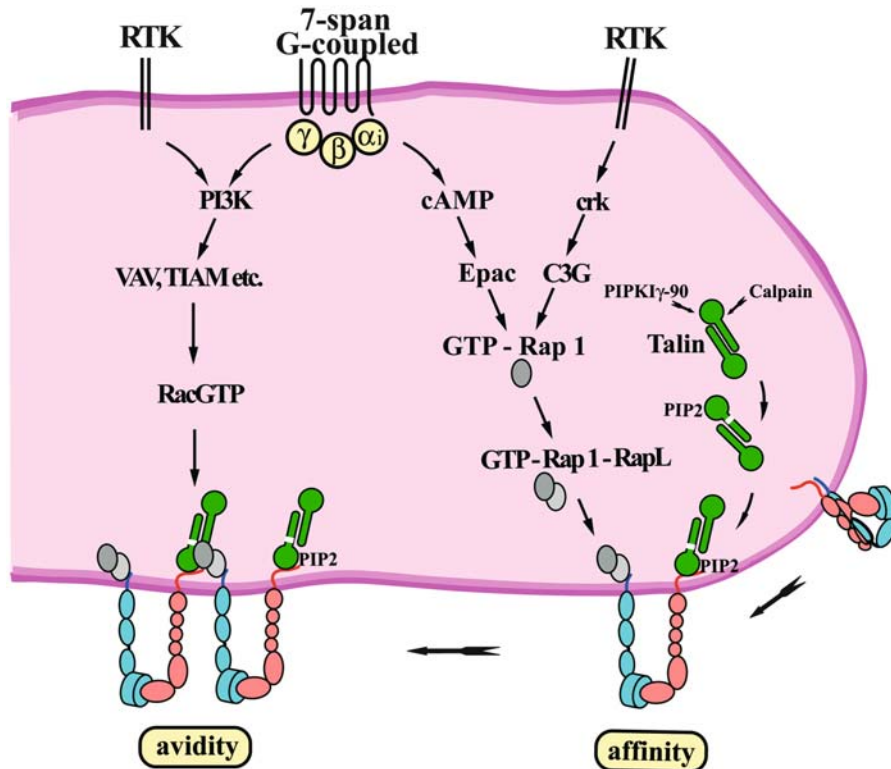
Atypical ζ PKC and Integrin Clustering

The atypical protein kinase C ζ (ζ PKC) is associated with a protein complex that regulates cell polarity. Recently it has been found that cell activation by G proteins activates the kinase activity of ζ PKC by its interaction with phosphatidyl inositol 3- phosphate kinase PI(3)K and brings the enzyme to the plasma membrane, where it leads to increased mobility of integrins. This effect is needed to allow clustering of activated integrins and achieve the highest level of integrin **▶avidity**.

Formation of Focal Adhesions

Activated integrins bind to their ligands. The intracellular tail of integrins then interacts with adaptor proteins and the actin cytoskeleton. Clustering of these complexes leads to the formation of focal adhesions (Fig. 3).

Integrin Activation



Focal Adhesions. Figure 2 Integrin activation. Receptor mediated signaling through receptor tyrosine kinases (RTK) and G-protein coupled receptors activates two major intracellular signaling pathways. First, Rap1 GTPase is activated by the cAMP dependent GTP exchange factor Epac or C3G and will then bind to RapL forming a complex that targets the cytoplasmic domain of the α subunit of integrins. At the same time the cytoskeletal linker protein talin is activated by calpain, which cleaves the talin head from its tail, or by binding of PIP2 (phosphatidylinositol 4,5-bis-phosphate) to the talin head domain. This results in recruitment of talin to the cytoplasmic domain of the β integrin subunit. Both Rap1/RapL and talin head domain binding will induce an allosteric switch in the integrin heterodimer creating high affinity binding to integrin ligands (affinity switch). The induction of the small GTPase Rac downstream of PI3 kinase leads to integrin clusters, which have an increased avidity for multimeric or clustered integrin ligands (avidity switch).

Rac1

One of the first steps in focal adhesion assembly is the activation of the small GTPase Rac1 (Fig. 2). GTP-bound Rac1 induces the polymerization of an actin network at the periphery of the cell, the ►**lamellipodium**. Within this network, the adaptor protein vinculin binds PIP2 and recruits it to the plasma membrane.

Vinculin

Vinculin activation by PIP2 leads to interaction with the Arp2/3 complex. This induces nucleation and polymerization of actin at the plasma membrane. The nascent fibers emanating from the vinculin/Arp2/3 complex can now interact with the major F-actin binding site of talin. It is thought that this interaction recruits activated integrin into nascent focal adhesions.

This complex can be reinforced by the integrin-linked kinase (ILK).

Integrin-linked Kinase

ILK interacts with two adaptor proteins PINCH and parvin and forms a trimeric linker between actin filaments in focal adhesions and the plasma membrane (6). The structural importance of ILK is best illustrated in ILK deficient animals, which suffer from severe cell adhesion defects. In *Drosophila*, ILK deletion results in an embryonic lethal muscle defect, disrupting the actin cytoskeleton from the integrin and the plasma membrane.

Conclusion

Two classes of focal adhesion adaptor proteins provide the physical link between the ECM and the actin

cytoskeleton. The first class consists of structural proteins such as talin, filamin and α -actinin, which crosslink the actin cytoskeleton to the cytoplasmic tail of integrins. The second class of adaptors is formed by regulatory proteins such as the trimeric ILK-PINCH-parvin complex (Fig. 3).

Regulatory Mechanisms

Dynamics and Regulation of Focal Adhesions

Focal adhesions can be distinguished according to their size, shape and location within the cell and have been named focal complexes or focal contacts respectively. During cell migration, small point-like focal complexes form in the leading lamellipodium whereas large, elongated focal contacts form in the central and posterior parts the cell. Focal complexes appear in response to Rac1 activity; focal contacts require RhoA signaling and evolve by maturation from focal complexes mediated by actomyosin tension (Fig. 4) (7).

A Role for RhoA and Rac1

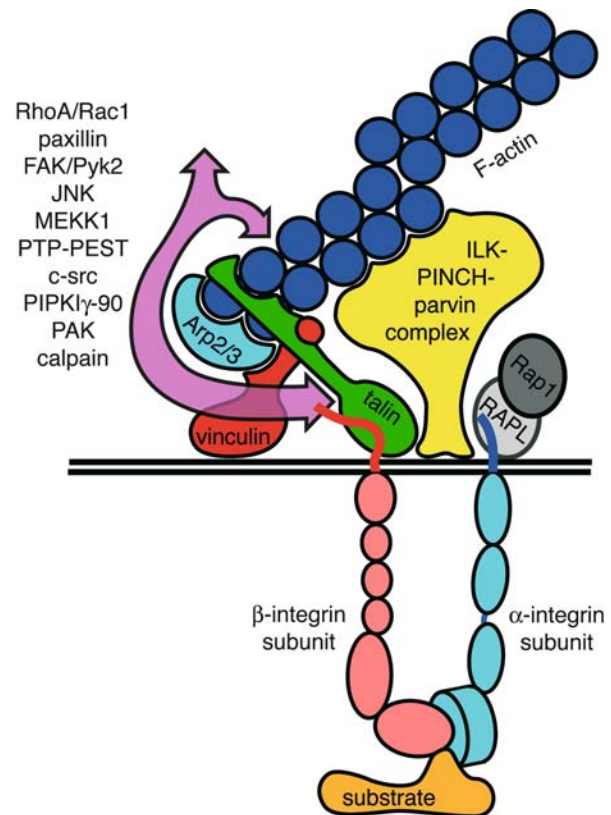
RhoA is a member of the small GTPase family inducing the formation of actin stress fibers and focal contacts. The two Rho effectors Citron and ROCK are kinases acting on the **actin cytoskeleton** by phosphorylation of myosin light chain. The phosphorylated **myosin light chain** then interacts with myosin and induces intracellular contractility and the formation of focal contacts. As another target, ROCK inactivates the myosin light chain phosphatase leading to the maintenance of focal contacts. Rac1, another GTPase family member counter-balances the effect of RhoA. It blocks myosin light chain kinase *via* its downstream effector PAK (p21-activated kinase).

FAK/Pyk2

The tyrosine kinases **FAK** (focal adhesion kinase) and Pyk2 are involved in G protein-induced cell adhesion and migration. These are proline-rich molecules with 45% homology to each other. FAK and Pyk2 bind to the cytoskeleton protein paxillin and are recruited to clusters of integrins. In FAK deficient cells, focal adhesions form more slowly but are more stable, resulting in an immobile phenotype. This suggests that FAK signaling is important for focal adhesion remodeling and cell migration. Similarly, cells from Pyk2 deficient mice fail to migrate in response to chemotactic signals.

Paxillin

Paxillin, a 68 kD focal adhesion adaptor protein exhibits multiple protein-recognition motifs and is recruited to nascent focal adhesions. It directly binds to the cytoplasmic tail of integrin and the structural focal adhesion protein vinculin. It also binds and recruits signal-transducing molecules, such as the kinases FAK,

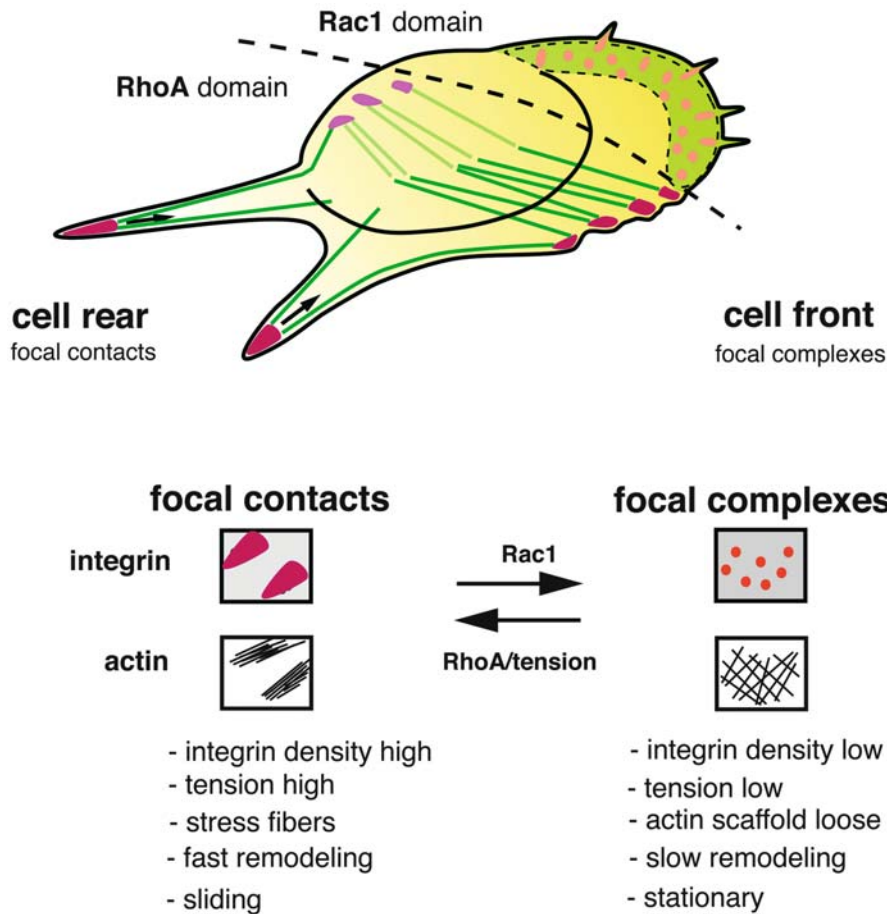


Focal Adhesions. Figure 3 The structural and regulatory link between the substrate and the actin cytoskeleton. Integrins are activated by intracellular adaptor molecules such as Rap1/RAPL or talin that bind to the α -subunit (blue) or the β -subunit (red) respectively. As soon as the high affinity form of integrin (open configuration) is attached to immobilized substrate (orange), the docking of the vinculin/Arp2/3 (red/blue) and ILK-PINCH-parvin (yellow) complexes stabilizes the initial integrin-substrate contact and reinforces the connection of the integrin to the actin cytoskeleton (dark-blue). This structural link now recruits focal adhesion adaptors (pink) that either serve to further stabilize focal adhesions, or in contrast, lead to their remodeling or dispersal (for explanation of the respective roles of the regulatory adaptors see text).

Src and ILK and the phosphatase PTP-PEST to focal adhesions. Vinculin regulates cell motility by reducing the interaction of paxillin with FAK and vinculin deficient cells are highly motile. Thus paxillin-FAK interaction controls cell adhesion and migration. Accordingly, paxillin null cells show reduced focal adhesion kinase (FAK) phosphorylation and delayed **spreading** and migration.

Protein-Tyrosine Phosphatase-PEST

PTP-PEST is a cytosolic protein-tyrosine phosphatase that dephosphorylates focal adhesion adaptor proteins



Focal Adhesions. Figure 4 Differential remodeling of focal adhesions during cell migration. Cell migration is driven by Rac1-dependent actin polymerization in the advancing lamellipodium. Integrins that are incorporated into nascent focal complexes exhibit low density and slow turnover rates, resulting in stationary focal adhesions that firmly anchor the actin network to the substrate. Under the influence of RhoA signaling, stationary focal complexes mature into focal contacts that are linked to actin stress fibers. The integrin cluster density increases with the applied intracellular actomyosin tension and the integrin exchange rate increases, leading to focal contacts, which can be remodeled. The plasticity of focal contacts due to intense remodeling is best illustrated during the retraction of the cell tail, where focal contacts undergo a polarized renewal, which results in sliding.

such as paxillin and signal adaptor proteins leading to activation of Rac1. Over-expression of PTP-PEST reduces cell spreading due to interference with Rac1 activity. In contrast, fibroblasts lacking PTP-PEST spread normally but form increased numbers of focal adhesions and do not migrate.

MEK Kinase 1

The activation of the dual specific MEK kinase 1 (MEKK1) is a critical event in the dynamic remodeling of focal contacts during cell migration. MEKK1 acts downstream of the receptor tyrosine kinase pathway and signals to Jun kinase (JNK) and stimulates the activity of the Ca^{2+} dependent intracellular protease calpain. Calpain cleaves various focal contact components such as talin, filamin and β -integrin tails and its

absence alters the actin cytoskeleton and reduces cell migration.

Jun Kinase

JNK stimulates the detachment of focal adhesions by the phosphorylation of serine 178 of paxillin. Small focal adhesions in fast migrating cells such as fish keratinocytes are transformed into larger immobile focal adhesions by the expression of the S178A paxillin mutant.

Visualizing the Formation of Focal Complexes and Focal Contacts

The attached movie summarizes integrin activation, the formation of focal complexes at the front of an adherent migrating cell, cytoskeleton and adaptor protein

organization, the building of focal contacts and their resolution at the tail of the migrating cell.

- [Integrin Signaling](#)
- [Rho, Rac, Cdc42](#)

References

1. Geiger B, Bershadsky A, Pankov R et al (2001) Transmembrane crosstalk between the extracellular matrix-cytoskeleton crosstalk. *Nat Rev Mol Cell Biol* 2:793–805
2. Hynes RO (2002) Integrins: bidirectional, allosteric signaling machines. *Cell* 110:673–687
3. Takagi J, Petre BM, Walz T et al (2002) Global conformational rearrangements in integrin extracellular domains in outside-in and inside-out signaling. *Cell* 110:599–611
4. Bos JL, Rooij J de, Reedquist KA (2001) Rap1 signalling: adhering to new models. *Nat Rev Mol Cell Biol* 2:369–377
5. Cram EJ, Schwarzbauer JE (2004) The talin wags the dog: new insights into integrin activation. *Trends Cell Biol* 14:55–57
6. Brakebusch C, Fassler R (2003) The integrin-actin connection, an eternal love affair. *Embo J* 22:2324–2333
7. Wehrle-Haller B, Imhof B (2002) The inner lives of focal adhesions. *Trends Cell Biol* 12:382–389

Focal Complexes/Focal Contacts

- [Focal Adhesion](#)

Folding Funnel

Definition

The folding funnel is a representation of the change in enthalpy (vertical axis) and conformational entropy (both horizontal axes) during protein folding.

- [Protein Folding](#)

Follicle

Definition

Follicle is the vesicle in the ovary that contains the egg.

- [SRY – Sex Reversal](#)

Follicle Stimulating Hormone

Definition

Follicle stimulating hormone (FSH) is the hormone produced by the anterior pituitary gland that stimulates follicular growth in the ovary, and stimulates aromatization of androgens in the granulosa cells of the ovary in females. In males, FSH stimulates estrogen formation from androgens in the Sertoli cells of the testes, and helps to stimulate androgen binding protein.

- [Hypothalamic and Pituitary Diseases Genetics](#)

Fölling Disease

- [Phenylketonuria](#)

Force Field

- [Molecular Mechanics Force Field](#)

Force Generation

Definition

Force generation is the fundamental process of muscle contraction produced by the sliding movement of thick and thin filaments in a sarcomere.

- [Familial Dilated Cardiomyopathy](#)
- [Muscle Contraction](#)

Force Transmission

Definition

Force transmission is the transduction of mechanical force produced by sarcomeres through the cytoskeleton to the sarcolemma and extracellular matrix.

- [Familial Dilated Cardiomyopathy](#)
- [Muscle Contraction](#)

Forkhead Genes or Proteins

►Winged Helix Transcription Factors

Förster Resonance Energy Transfer

MARTIN BEUTLER, RAINER HEINTZMANN
Randall Division of Cell and Molecular Biophysics,
King's College, London, UK
martin.beutler@kcl.ac.uk
rainer.heintzmann@kcl.ac.uk

Definition

Förster resonance energy transfer (FRET) is a distance-dependent quantum mechanical process that occurs between chromophores. In a FRET-pair, an excitation energy exchange of the electronic states takes place *via* a dipole-dipole coupling mechanism. The ►donor in a FRET pair is a fluorophore in its electronic excited state, passing the energy to another chromophore, the ►acceptor. This process, however, can only occur if the donor ►fluorescence emission and the acceptor excitation spectra show a spectral overlap. The efficiency of FRET is dependent upon the distance and relative orientation between donor and acceptor. Typically, quantification of FRET requires a mutual distance of fluorophores in the range of 1nm to 10 nm. This distance sensitivity makes it an important method for investigating a variety of biological phenomena showing changes in the proximity, mutual orientation or rigidity of proteins and other macromolecules. FRET is used in microscopy, flow cytometry and spectroscopy to analyze interaction between or conformational changes in donor and acceptor molecules (e.g. proteins, ligands, lipids) even in living cells. In practice, two molecules of interest are modified to fluoresce in order to investigate their interaction or even to determine the sub-cellular localization of their interaction.

Description Theory

FRET was first described theoretically by Theodor Förster in 1948 (1). Ever since, investigation of this phenomenon has been applied to establish now routinely employed methods in cell and protein

biology. This development has been reviewed and discussed in depth (2, 3), but methodological progress is still being made.

For FRET in a pair of molecules to occur, the donor first requires excitation, e.g. by absorption of a photon (Fig. 1). The excited donor can then de-excite thermally, emit a photon as fluorescence or transfer energy to the acceptor *via* electrical dipole-dipole coupling (in contrast to Dexter type energy transfer the electrons are not exchanged). For FRET, it is not necessary for the acceptor to be able to emit fluorescence, but in most FRET experiments fluorescent acceptors are nonetheless employed. However, for some applications dark acceptors are advantageous (2). The transfer efficiency is influenced by the spatial distance between the donor and acceptor molecules and is given by an inverse sixth power law in dependence on the donor acceptor distance (Equation 1 and Fig. 2). This fact is a consequence of the dipole-dipole interaction. FRET can therefore be employed as a molecular ruler (4).

$$E = \frac{1}{1 + (r/R_0)^6} = \frac{R_0^6}{R_0^6 + r^6} \quad (1)$$

with E being the energy transfer efficiency, R_0 the radius of 50% transfer efficiency (Förster radius) and r the donor-acceptor distance.

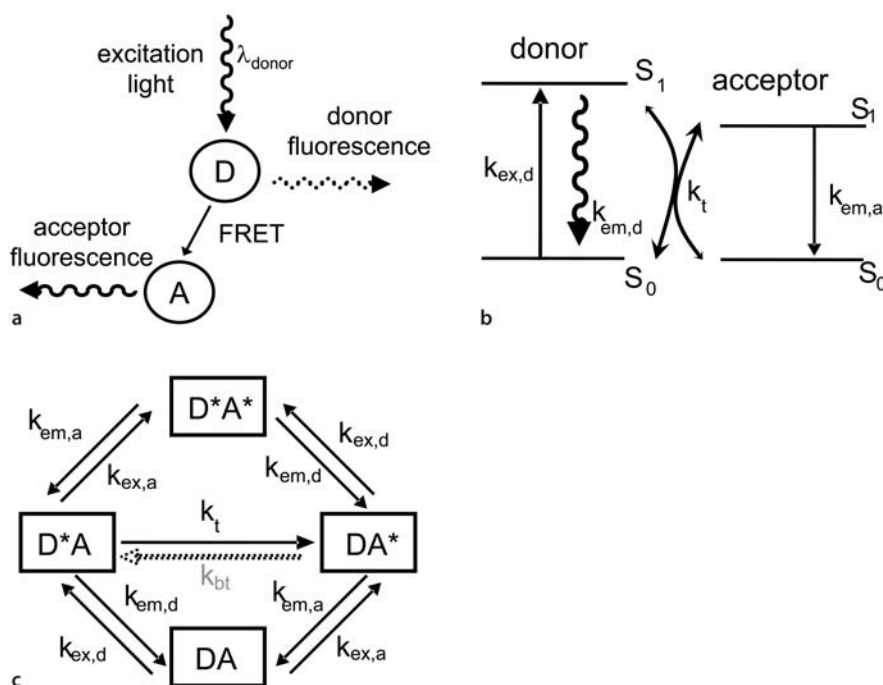
The Förster radius R_0 is proportional to the spectral overlap integral of the normalized donor emission spectrum and the acceptor absorption spectrum (J) (Equation 2) and depends on the relative orientation of the molecules and the refractive index (n). Figure 3 shows their spectra and their spectral overlap region. The Förster radius is computed as

$$R_0 = 0.211(\kappa^2 n^{-4} Q_D J(\lambda))^{1/6} \quad (2)$$

$$J(\lambda) = \int \varepsilon A(\lambda) \varepsilon m D(\lambda) \lambda^4 d\lambda$$

where κ is the orientation factor representing the directions of the emission dipole of the donor with respect to the acceptor excitation dipole. For randomly orientated molecules, κ^2 can be assumed as 2/3. n is the refractive index of the medium and Q_D is the quantum efficiency of the donor. The emission spectrum of the donor (with its integral normalized to one) is given by $\varepsilon m D(\lambda)$, and $\varepsilon A(\lambda)$ denotes the wavelength-dependent acceptor extinction coefficient.

In microscopy, flow cytometry or spectro-fluorometry, the most often desired parameters are the concentration of FRET pairs, the concentration of free donor ("free" means no acceptor in proximity to the donor) and free acceptor, as well as the distance (given by the FRET efficiency) between donors and acceptors in pairs or the



Förster Resonance Energy Transfer. Figure 1 (a) The donor molecule absorbs light energy. It can emit the energy by fluorescing (emission of a photon) or alternatively transfer the energy to the acceptor molecule, which, if fluorescent, could then itself emit a photon. (b) Energetic state model: before light absorption, the donor molecule is in its ground state denoted as S_0 . After the up-take of a photon ($k_{ex,d}$) it transits to the first excited singlet state S_1 . From here it can de-excite via light emission (fluorescence, $k_{em,d}$) or radiationless excitation (k_t) of the acceptor molecule (this process is, to some extent, reversible, especially when donor and acceptor are molecules of the same type). The acceptor changes from the ground state to the first excited state and de-excites ($k_{em,a}$) by light emission (if fluorescent) or heat emission. (c) Energy state diagram of the acceptor-donor system (DA) for defining differential equations. $*$ indicates an excited state of donor (D) or acceptor (A). Rate constants are the same as in (b) (10).

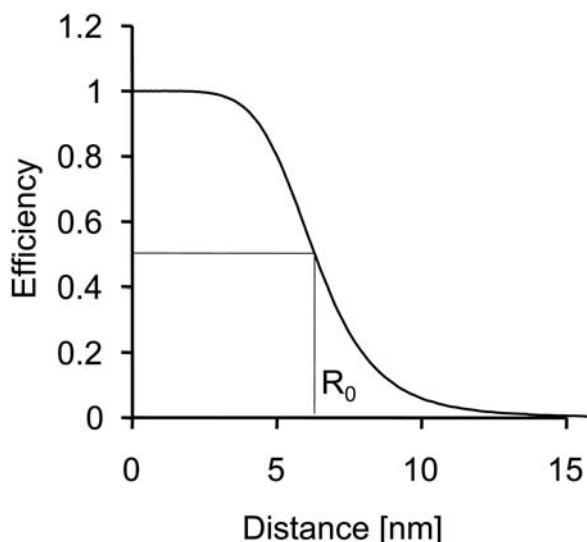
orientation of donor to acceptor. In most cases, not all of the parameters can be determined simultaneously. Often some of the desired parameters can only be retrieved by other calibration experiments or must be assumed as *a priori* known constants.

Organic dye molecules have been widely used both as acceptors and donors in FRET experiments. They are often covalently coupled to target molecules, exhibit a high ►fluorescence quantum yield (which means an easily detectable signal) and are small in size, a fact that diminishes perturbation of biological systems. Recently, quantum dots have been employed as donors in FRET experiments (5). These are minuscule semiconductors with a diameter of a mere few nanometers. Their properties are similar to single fluorophore molecules, but they offer a rather broad excitation spectrum (“free” choice in excitation wavelength), relatively monochromatic emission (width ~ 30 nm) and a high quantum yield. In addition, quantum dots are highly photo-stable, allowing for longer measuring periods than with ordinary dyes.

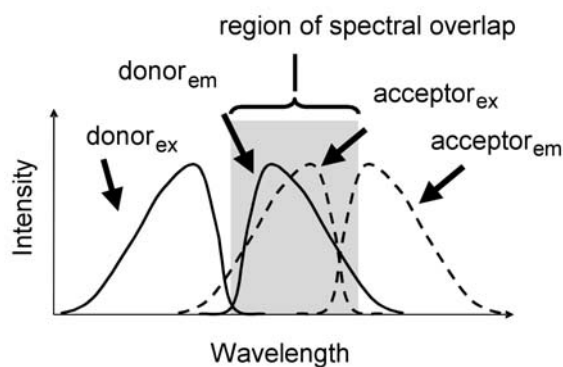
Methods

Sensitized Emission

Many FRET measurements are based on spectral methods, i.e. sensitized emission measurements (6). They can be used stand-alone, but may also be combined with any of the methods described below. A method-sensitized emission exploits the fact that donor and acceptor molecules absorb and emit light energy at different wavelength regimes (Fig. 3). Therefore, in donor and acceptor pairs that show very little spectral overlap, the fluorescence of donors can be detected by performing excitation at the donor excitation wavelengths and measuring emission at the donor fluorescence emission wavelengths (Fig. 3). Acceptor fluorescence can be detected in an analogous manner. The sensitized emission (sometimes termed “FRET channel”) can be measured by excitation at a wavelength range within the donor excitation spectrum and detection of the acceptor emission within its emission spectrum. Ideally, this signal depends only on the concentration of FRET pairs and the FRET



Förster Resonance Energy Transfer. Figure 2 FRET transfer efficiency in dependence on the donor-acceptor distance. 50% transfer efficiency is achieved at the Förster radius, R_0 . This example shows a Förster radius of 6.31 nm, calculated for the FRET pair Alexa 488-Alexa 546.



Förster Resonance Energy Transfer.

Figure 3 Normalized donor and acceptor excitation and emission spectra of a FRET pair. The dashed area marks the overlap between donor emission and acceptor excitation spectra.

efficiency. However, many FRET pairs that are suitable for biological experiments (e.g. green fluorescent protein and its spectral derivatives) show substantial spectral overlap (7). In other words the donor also fluoresces in the wavelength range of the acceptor emission channel (“donor crosstalk”) and the acceptor is directly excited at donor excitation wavelengths (“direct excitation”). These effects can be corrected mathematically (“spectral unmixing”) by careful calibration,

using the respective individual fluorophores (7). A problem occurs in FRET experiments if not all of the donors find a binding partner or the affinity of the binding is low. In these cases, many free donors and acceptors are present that do not show FRET. However, they exhibit fluorescence in the respective detection channels. In this case, additional measurements are required (see methods further below) in combination with sensitized emission to achieve reasonable estimates of the concentration of the two species and the FRET efficiency of the interaction. Depending on the investigated molecules, it may be debatable whether a fixed FRET efficiency can be assumed for bound pairs.

Acceptor ► Photobleaching

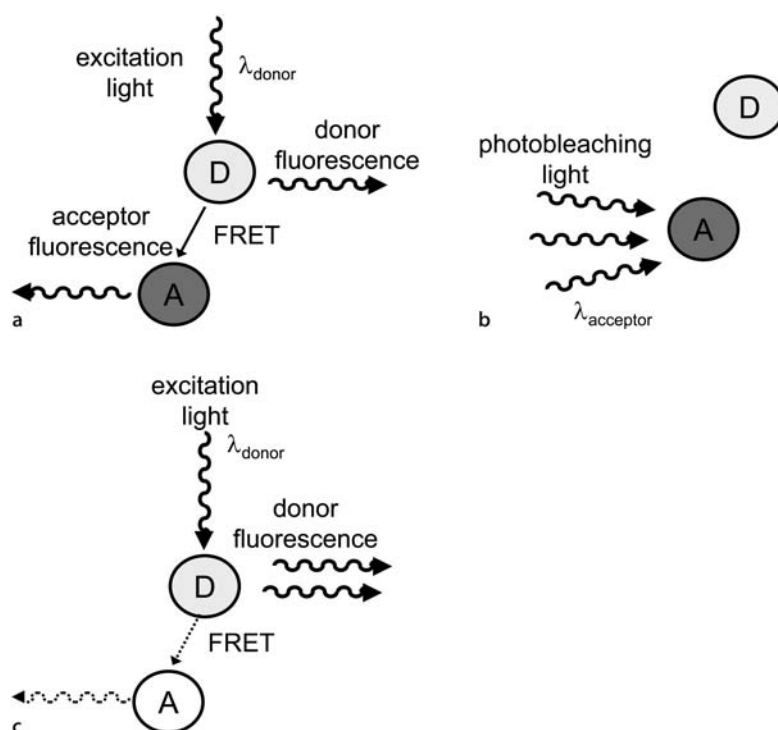
Photobleaching techniques are widely used because they can be performed easily and yield relatively robust estimates of FRET (8). In acceptor photobleaching approaches, the acceptor of a FRET pair is illuminated with intense light of wavelengths strongly absorbed by the acceptor but as little as possible by the donor. The acceptor then undergoes a photochemical destruction process with a substantial change in its absorption cross-section at the donor emission wavelength. After this, FRET cannot occur any longer and the donor fluorescence increases, as it is no longer quenched by energy transfer to the acceptor. The amount of donor fluorescence before and after the photobleaching procedure can be used to estimate the FRET efficiency (E):

$$E = 1 - I_{bb}/I_{ab} \quad (3)$$

with I_{bb} being the donor fluorescence intensity before and I_{ab} being its intensity after photobleaching

For the principle underlying the acceptor photobleaching measurement see Fig. 4. An experimental example obtained with a confocal microscope is shown in Fig. 5. Problematic cases for this method include incomplete bleaching of the acceptors, non-zero extinction coefficients of bleached acceptors in the donor emission spectrum and photodestruction of donors during acceptor depletion.

Alternatively, FRET efficiencies can be estimated by careful investigation of the time traces of bleaching experiments (2, 3) specifically targeted at bleaching the donor. A donor molecule displaying FRET exhibits an improved resistance to bleaching in comparison to “free” donors. From the analysis of fluorescence decay time constants, the FRET efficiency can then be calculated. One evident major problem in the above approaches based on photodestruction is that the experiment cannot be repeated after bleaching. These techniques are therefore not well suited to time-resolved FRET studies. In addition, excessive photobleaching is often lethal to live samples such as cells.



Förster Resonance Energy Transfer. Figure 4 Acceptor photobleaching approach. (a) An intact FRET pair is illuminated by excitation light—FRET does take place. (b) The sample is exposed to intense light at acceptor excitation frequency causing photodamage to the acceptor. (c) Measurement of a FRET pair with a photo-damaged acceptor. FRET transfer is decreased to a minimum and the donor fluorescence increased due to de-quenching of the donor.

► Fluorescence Lifetime Imaging (FLIM)

Each excited state of a molecule or fluorophore has a decay time called its fluorescence lifetime. The lifetime is usually stated as the time span a given excited state population requires to de-excite to $1/e$ (~ 0.37) of the original population. After exciting a population of fluorophores with a light pulse, the fluorescence emission intensity decays (often exponentially) as the number of excited molecules decreases over time (6). If a donor transfers its energy efficiently to an acceptor, the lifetime of the donor excited state is decreased when compared to the unperturbed donor. There are several ways to determine fluorescence lifetime experimentally (2, 6). FLIM is a rapid, robust and non-destructive method for FRET estimation. FRET efficiencies as well as the number of FRET pairs can be estimated, when combined with spectral analysis (9).

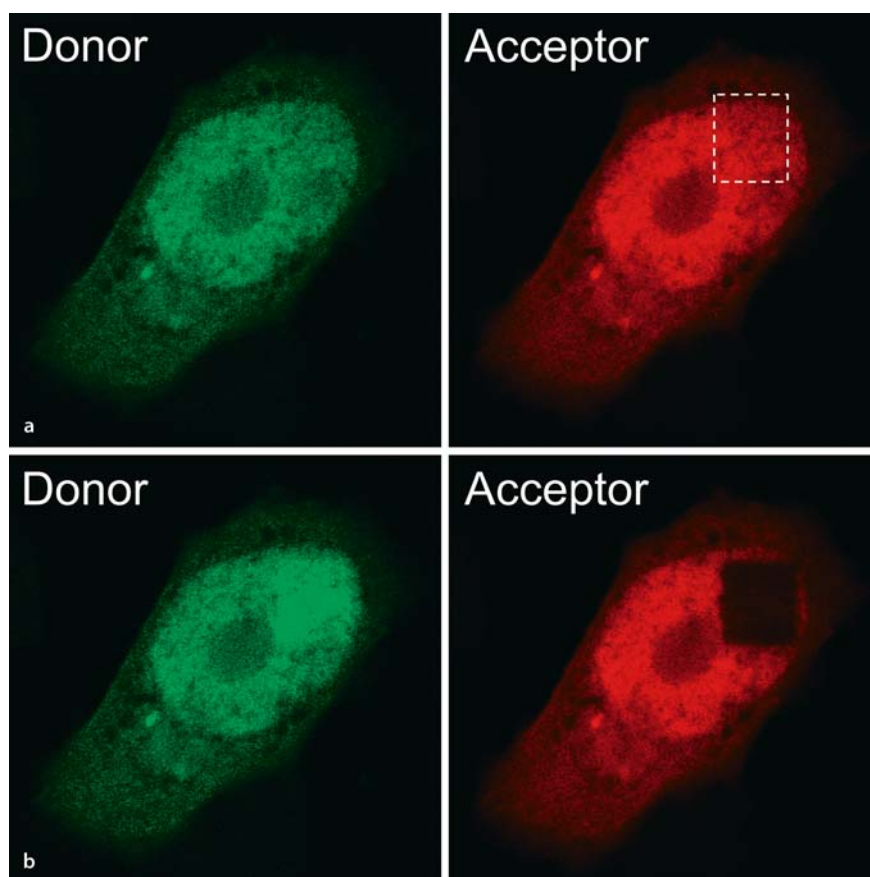
Fluorescence Anisotropy

In FRET measurements that employ ► fluorescence anisotropy or fluorescence polarization, the donor is excited with polarized light (2, 3). Polarized excitation predominantly excites molecules whose excitation dipole moment is collinear to the polarization of the

excitation light (“photoselection”). If the rotational correlation time of the donor is long in comparison to its fluorescence lifetime, the emission of the donor signal will be partially polarized. Light emission after energy transfer that originates from a randomly oriented acceptor will not be polarized. However, due to the dependence of κ on the mutual orientation of the molecules, the effect of this method is often only a further reduction in the degree of polarization. FRET can occur between two fluorophores of the same type (e.g. donor and acceptor are both GFP), providing they meet the FRET criteria outlined above. FRET between molecules of the same type can then be used to quantify the formation of homo-dimers and higher oligomers of the same protein. This method cannot predict a change in lifetime or a massive spectral change, since the acceptor signal is not spectrally separated from the donor signal and both molecules have similar properties.

Photochromic FRET (pcFRET)

In pcFRET, a recently applied technique, photoswitchable acceptors are employed (2). These acceptors can be switched between two states by ultra-violet illumination. One state is capable of absorbing the



Förster Resonance Energy Transfer. Figure 5 Donor (CFP) dequenching under acceptor (YFP) depletion. (a) Donor (left) and acceptor (right) before bleaching. (b) Donor and acceptor after intense illumination at acceptor wavelength of a sample-region (as indicated in a) leading to its photodestruction. These are CHO cells transfected with protein kinase C reporter (CKAR construct, kindly provided by Alexandra C. Newton) in the non-phosphorylated (closed) state. All images are acquired under illumination at donor excitation wavelength (458 nm) and detection with a bandpass (480/20 nm) for CFP and longpass (>545 nm, 545 dichroic and LP 505) for YFP. Note the increase in donor fluorescence (left b) after photo-destruction of the acceptor. The data were kindly provided by Elena Kardash.

donor fluorescence, whereas the extinction ratio of the other state is substantially reduced. This switching of states is reversible. In comparison to the photobleaching approach, which is not a reversible process, fewer quanta are necessary to cause the desired state transition, which has obvious advantages in speed and cell viability.

Outlook

In addition to various combinations of the approaches described, there are other promising approaches for FRET estimation, such as saturation approaches in which donor or acceptor molecules are driven into excited state saturation by intense excitation light (2, 10). At low light intensities, the fluorescence emission of a fluorophore population is proportional to the excitation intensity. When exposed to higher excitation light intensity, more and more fluorophores are driven

into the excited state. In extreme cases the fluorescence intensity finally reaches a limit as defined by the radiative rate. In this limiting case, all fluorophores are in the excited state since a fluorophore is immediately re-excited after fluorescence emission. At less extreme but high intensity, the proportionality between excitation intensity and fluorescence emission is violated, with many molecules being in the excited state at all times. In some systems, excited molecules are unsuitable for absorption in the wavelength range required for FRET. If the acceptor in a FRET pair is saturated by high light intensities within its direct absorption spectrum, the donor becomes de-quenched, since less acceptor molecules are now available in the singlet (non-excited) state. Related approaches are based on an increased excitation intensity required for saturation in the presence of the acceptor. When combined with spectral methods, this approach may

deliver FRET efficiencies as well as donor and acceptor quantities.

Acknowledgements

We thank Thomas Jovin and Stephan Hoepfner for critical reading of the manuscript. We also thank Elena Kardash for the supply of the data in Fig. 5 (all from the Department of Molecular Biology at the Max Planck Institute for Biophysical Chemistry, Göttingen). The CKAR construct (Fig. 5) was kindly provided by Alexandra C. Newton. M. Beutler was financed by DFG Project HE 3492/2-1, SPP 1128.

► [Protein-Protein Interaction](#)

► [Two-Hybrid System](#)

References

1. Förster T (1948) Intermolecular energy migration and fluorescence. *Ann Phys Leipzig* 2:55–75
2. Jares-Erijman EA, Jovin TM (2003) FRET imaging. *Nat Biotechnol* 21:1387–1395
3. Marriott G, Parker I (eds) (2003) *Biophotonics, Part A and B*. Academic Press, San Diego
4. Stryer L (1978) Fluorescence energy-transfer as a spectroscopic ruler. *Annu Rev Biochem* 47:819–846
5. Jovin TM (2003) Quantum dots finally come of age. *Nat Biotechnol* 21:32–33
6. Lakowicz JR (1999) *Principles of Fluorescence Spectroscopy*. Kluwer Academic/Plenum Publishers, New York
7. Zimmermann T, Rietdorf J, Girod A et al (2002) Spectral imaging and linear un-mixing enables improved FRET efficiency with a novel GFP2-YFP FRET pair. *FEBS Lett* 531:245–249
8. Bastiaens PIH, Jovin TM (1998) Fluorescence resonance energy transfer microscopy. In: Celis JE (ed) *Cell Biology: A Laboratory Handbook*. Academic Press, New York, pp 136–146
9. Hanley QS, Arndt J, Jovin DJ, Jovin TM (2002) Spectrally resolved fluorescence lifetime imaging microscopy. *Appl Spectrosc* 56:155–166
10. Hanninen PE, Lehtela L, Hell SW (1996) Two- and multiphoton excitation of conjugate-dyes using a continuous wave laser. *Opt Commun* 130:29–33

Forward Genetics

Definition

Forward genetics comprise of genetics techniques (e.g. knockout or transgenic studies) that aim to identify genes/mutations that produce a certain phenotype. A mutagen is very often used to accelerate this process. Once mutants have been isolated, the mutated gene can be molecularly identified. Once the sequence of the gene has been identified, the mutant phenotype can assign its function with nearly absolute certainty.

► [Morpholino Oligonucleotides, Functional Genomics by Gene ‘Knock-down’](#)

► [SLE Pathogenesis Genetic Dissection](#)

Fos

► [Jun/Fos](#)

Founder (Gene) Effect

Definition

Founder (gene) effect refers to the high frequency of a mutant gene in a population founded by a small ancestral group, when one or more of the founders were carriers of the mutant gene.

► [Familial Hypercholesterolemia](#)

► [Genetic Epidemiology](#)

Founder Mutation

Definition

Founder mutation describes a mutation that is present in different, seemingly unrelated, individuals of a given population, because it arose in an ancestor (the founder). Subsequently, this mutation is inherited by individuals in later generations (► [Identity-by-descent](#)).

► [Bloom Syndrome](#)

► [Genetic Epidemiology](#)

Fourier Shell Correlation

Definition

The FSC is the commonly accepted method to estimate the resolution limit (usually defined at a correlation of 0.5) of three-dimensional (3D) reconstructions from

► [Cryo Electron Microscopy](#)

► [Cryo-Electron Microscopy: Single-Particle Reconstruction](#)

Fourier Transform

Definition

The mathematical transformation of a function into its representation as a sum of a spectrum of wave components. As one of its myriad applications in science and engineering, the Fourier transform relates the electron density distribution to the waves diffracted from a crystal.

► [X-Ray Crystallography—Basic Principles](#)

Fourier Transform Ion Cyclotron Resonance Mass Spectrometer

Definition

Fourier transform ion cyclotron resonance mass spectrometer is an instrument to measure mass-to-charge ratios dependent on the principle of an ion orbiting in the presence of a strong magnetic field. While orbiting in this field, the ions are excited by a radio frequency signal, which produces a detectable image current. The time-dependent image current has to be subjected to Fourier transformation to obtain the component frequencies of the ions, which can then be correlated to their mass-to-charge ratios.

► [Mass Spectrometry: ESI](#)

Fovea

Definition

The fovea is the very center of the macula region, which is located roughly in the center of the retina. This area is responsible for our central, sharpest vision.

► [Retinitis Pigmentosa](#)

Fox Genes or Proteins

Definition

Fox genes stand for Forkhead box. This box comprises of genes encoding transcription factors that are related to the *Drosophila* gene forkhead. All Fox genes contain

a characteristic DNA binding motif, the winged helix domain.

► [Winged Helix Transcription Factors](#)

Fps

Definition

Fps is a member of a distinct subfamily of the non-receptor protein-tyrosine kinase family. Fps is involved in regulating cytoskeletal rearrangements and inside out signalling that accompany receptor ligand, cell matrix and cell cell interactions.

Fragile Site

Definition

Fragile sites on chromosomes are regions at which the chromosome is liable to break.

► [Fragile X Syndrome](#)

► [Marfan Syndrome](#)

► [Chromosomal Instability Syndromes](#)

Fragile X Mental Retardation Syndrome

PETER STEINBACH

Department of Human Genetics, University of Ulm, Ulm, Germany

peter.steinbach@medizin.uni-ulm.de

Synonyms

Fragile X Syndrome; Martin Bell Syndrome; Marker X Syndrome

Definition

► [Mental retardation](#) manifests during the developmental period and causes children to have educational problems. They learn and develop more slowly than a typical child of the same age and have certain limitations in intellectual functioning and adaptive behaviour, i.e. in skills such as communication, taking

care of him or herself and social performance. Intellectual functioning is measured by an ►IQ test. The average score is 100. People scoring below 70 are thought to be mentally retarded. Adaptive behaviour is measured by comparing what a child can do in comparison to other children of the same age. About 3% of all children in the population have mental retardation. The most common causes are problems during pregnancy (e.g. rubella infection, alcohol or drug abuse by the mother) or during birth (e.g. oxygen deficiency of the baby), health problems (e.g. infectious diseases such as measles or meningitis, extreme malnutrition, exposure to poisons), and genetic conditions. One of the most common forms of inherited mental retardation, besides ►Down syndrome and phenylketonuria, is fragile X syndrome.

Characteristics

The name of the syndrome derived from the “fragile ►X chromosome” that is seen in affected males and females (Fig. 1). Under normal circumstances each somatic cell contains twenty-three pairs of ►chromosomes harbouring the genetic material (►DNA). Twenty-two pairs are morphologically the same in males and females. The remaining two chromosomes (X and Y) are the sex chromosomes. They determine whether a person is male (XY) or female (XX). A fragile X chromosome carries a particular ►fragile site (designated ►FRAXA) that is prone to breakage when cells are cultured under certain conditions outside the body (*in vitro*). Thereby the fragile site becomes visible under the microscope. At the molecular level, i.e. in the DNA of the X chromosome, the FRAXA site represents an abnormally expanded segment of a particular gene, designated ►FMR1 (fragile X mental retardation 1).

Common Symptoms

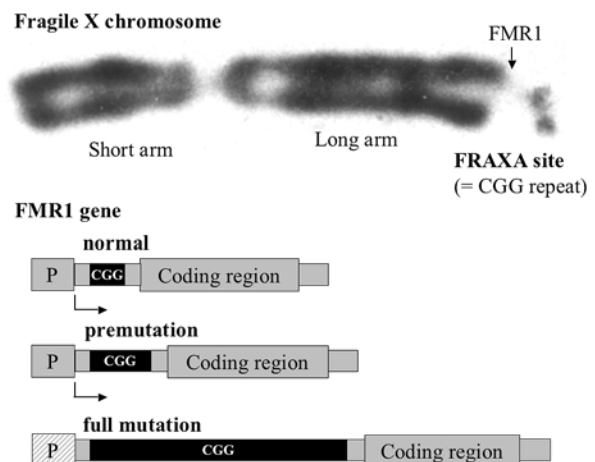
The common symptoms of fragile X syndrome include intellectual, physical and behavioural characteristics. The most significant feature is developmental delays ranging widely from ►learning disabilities to severe mental impairment. Learning disabilities are often accompanied by delays in speech and communication skills. There can also be delays in gross and fine motor skills. Some people with fragile X mental retardation have a long face, large ears that stick out from the head, connective tissue weakness leading to floppy muscle tone and flat feet, hyperflexible joints and large testicles ►Mitral Prolapse and High-pitched Speech. About 25% are affected by seizures. The behavioural characteristics include a tendency to anxiety and unstable mood, to trouble coping with sensory stimuli and to autistic like behaviours such as hand flapping, avoiding eye contact, saying or doing the same thing

over and over and disliking change in routine. Many have ►attention deficit hyperactivity disorder (ADHD) showing poor concentration and poor social skills. Some experience tantrums and emotional outbursts. Boys are typically more severely affected than girls. Most boys have mental retardation but only about one-half of girls have significant intellectual impairment while the rest share either normal IQ or learning disabilities. Emotional and behavioural problems are common in both sexes.

Treatment

There is currently no cure for fragile X syndrome. But specialist therapies, targeted education and medication can help maximise the potential of each child and to improve the quality of life. However, most boys and girls remain significantly affected throughout their lives. Because of speech and language delay and because of their poor concentration, the children may seem hard to educate. However, the child may have more ability than he or she can demonstrate, e.g. good copying skills, excellent long-term memories, good senses of humour, and often good vocabulary and reading skills.

Researchers are currently investigating two main strategies of causal treatment.



Fragile X Mental Retardation Syndrome. Figure 1

The fragile X chromosome at mitosis, and the FMR1 gene. The fragile site (FRAXA), located at the FMR1 locus (arrow), shows a break leading to dislocation of a small part of the chromosome's long arm. The chromosomal fragile site corresponds to the gene's CGG trinucleotide repeat (CGG) situated between the promoter (P) and the region encoding the FMR protein (coding region). Transcriptional activity (symbolised by angled arrow at the beginning of the transcribed gene region) is usually lost upon expansion of a premutation to a full mutation coinciding with methylation of the promoter region (hatched symbol). For further explanation see text.

- 1) Gene therapy by inserting functional copies of the *FMRI* gene into ►neurons of ►*FMRI* knockout mice to refine techniques for achieving consistent delivery of genetic material into humans.
- 2) Gene repair to restore the normal function of a mutated *FMRI* gene whose coding sequence is normal but which is not expressed because the gene is turned off.

Genetics

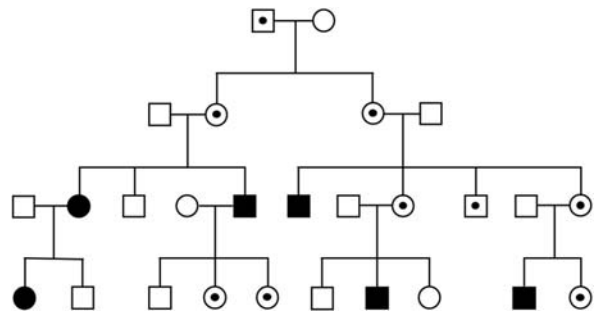
Fragile X mental retardation syndrome is caused by the absence of functional FMR protein (►FMRP), encoded by the fragile X mental retardation gene, *FMRI* (Fig. 1). This gene was discovered in 1991 when the FRAXA site was isolated by molecular cloning. The fragile site turned out to be a CGG ►trinucleotide repeat, i.e., a stretch of DNA containing a repeat sequence of C, G, and G nucleotides that reads CGG-CGG-CGG, etc. (1, 2). Fragile X syndrome is the first human disease identified to be caused by an unstable DNA sequence. The CGG repeat is situated in the *FMRI* gene but does not contribute to biosynthesis of the FMR protein. The repeat varies in length from one person to another, ordinarily falling within a “normal”-range (5 to 58 trinucleotides). In most individuals who have fragile X syndrome the stretch of CGGs is expanded to several hundreds of CGGs (>220). This large expansion is called a “►full mutation”. In this situation the gene usually becomes heavily modified by ►DNA methylation and is shut down.

Inheritance

Fragile X syndrome is inherited. The mutated gene is transmitted by female or male individuals who are “carriers” (Fig. 2). The fully mutated *FMRI* gene of affected children is always received from a carrier mother. She may carry a full mutation herself but mostly she has a “►premutation”, i.e. the length of her CGG repeat is between the normal and the full mutation range (59 to 200 CGGs). The majority of female premutation carriers did not receive the gene from their mothers but from the fathers. In this case, the father is a carrier male passing the premutation to all his daughters but none of his sons, who inherit the Y chromosome. Carriers of a premutation do not typically have symptoms of fragile X syndrome, but the premutated CGG repeat is “unstable” in that it is prone to further expansion when passed from a mother to her children. A full mutation male may also have children, but he only passes a premutation to his daughters.

Timing of Full Mutation

Expansion of a premutated CGG repeat to a full mutation allele probably occurs when the ►oogonia in the ovary of a female premutation carrier reproduce by



Fragile X Mental Retardation Syndrome. Figure 2 Fictitious pedigree of a family with fragile X mental retardation syndrome resulting from expansion of the CGG trinucleotide repeat in the *FMRI* gene. Females are symbolised by circles, males by squares. Filled symbols indicate affected individuals who have a full mutation. Male and female carriers of a premutation are indicated by a dotted symbol.

mitotic cell divisions before they enter ►meiosis to become egg cells that can be fertilised by sperm. The full mutation allele, that a fragile X embryo receives only through an oocyte, is probably not methylated initially, and is mitotically unstable in that cells with different lengths of the CGG repeat are generated upon continual mitotic activity. The unstable behaviour results in a mosaic of cells with different CGG repeats. Methylation of full mutation alleles only occurs at a later stage of development. The methylated CGG repeats then acquire mitotic stability by an unknown mechanism (3, 4).

In the testis of a fragile X male, premutated CGG repeats may also expand in dividing ►spermatogonia. But, in the absence of the normal *FMRI* gene in the male, cells with larger repeats produce less FMR protein and reproduce more slowly than cells with a short repeats. After a history of mitotic cell divisions that is much longer compared to that of oocytes, adult fragile X males, no matter if they are a carrier of a pre or a full mutation, have only premutations in their sperm.

Genetic Testing

The CGG repeat sequence is used to identify those who carry a pre or a full mutation and those who do not. According to the guidelines accepted by informed professionals, genetic testing of any individual requires genetic counselling and her or his informed consent. The DNA test for fragile X is offered by specialised laboratories and is quite accurate. It usually takes about two weeks to get the results.

Cellular and Molecular Regulation

Fragile X mental retardation syndrome is genetically caused by mutations that change the DNA sequence

of the *FMR1* gene. Most of these changes do not allow for the production of ►messenger RNA (►transcription) that is normally translated into protein (►translation). The normal *FMR1* gene is only transcribed when the ►promoter region, that is located “before” the CGG repeat (Fig. 1), is open to ►transcription factors, i.e. proteins that bind to the promoter DNA to initiate or regulate transcription.

In full mutation fragile X patients (98% of the cases) the promoter region is closed. The DNA of sufficiently large CGG repeats probably folds abnormally and thereby attracts an enzyme that methylates the triplet repeat and the nearby promoter region as well. Next, the methylated DNA binds to specialised proteins and enzymes that modify the ►chromatin of the *FMR1* promoter. The modified chromatin is remodelled and becomes so dense that the promoter is no longer accessible to transcription factors.

In about 1–2% of all fragile X cases, a portion of the *FMR1* gene including the promoter and the CGG repeat is lost. Changes of the genetic code may also be a genetic cause of fragile X syndrome but have been found only rarely. All but one of these changes generated one or more premature termination signals, because the genetic code was read through an altered frame, and thereby caused degradation of the aberrant messenger ►RNA by a cellular mRNA surveillance system.

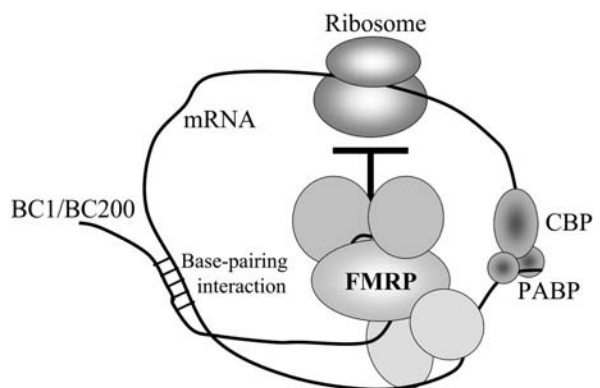
In one exceptional case, a severely affected male showed normal quantities of abnormal FMR protein with one of the amino acids exchanged by a “missense” mutation. This change probably interfered with the ability of the protein to interact with other molecules to form a messenger ribonucleoprotein particle (mRNP). These particles contain messenger RNA (mRNA) and a number of previously identified proteins interacting with normal FMRP and with each other. The FMR protein has RNA binding domains and selectively binds to its own and to several other messenger RNAs in a highly specific manner.

High quantities of FMR protein are produced in all cells when they reproduce by mitotic division and in some non-dividing cells including neurons in several regions of the brain. While the function of FMR protein in dividing cells is less well understood, the protein’s normal role in brain is to help shape the synaptic connections of neurons that underlie learning and memory.

The vast majority of ►synapses receiving signals from other neurons sit on the heads of tiny protrusions of the ►dendrites. These protrusions are known as spines. Tens of thousands of spines can coat a single neuron making the dendrite look spiny, like a rose. In the normal brain, FMR protein is also produced near synapses and its level increases when the synapse is stimulated by nerve impulses. High-frequency

stimulation of a synapse increases its sensitivity and the efficiency of synaptic transmission over a long period of time. This phenomenon is called ►long term potentiation (LTP). It contributes to the formation of memories on which learning depends. LTP coincides with a controlled synthesis and activation of “key-” proteins at the synapse.

In the absence of FMR protein, i.e., in fragile X patients and in knockout mice lacking a functional *FMR1* gene (5), the density and shape of dendritic spines is abnormal, resembling the immature state of the early period of brain development when the number of spines with immature synapses is highest, before experience acts to mould brain organisation by selecting which synapses mature and survive and which fail to mature and consequently degenerate (6). In the absence of FMR protein, the level of several synaptic key proteins is significantly higher compared to the normal brain, where the translation of the corresponding messenger mRNAs is down-regulated by FMR protein functioning as a specific repressor of translation (7). Together with those mRNAs that are regulated by FMRP the protein is bound to a synaptic repressor complex (Fig. 3). This particle differs both structurally and functionally from the FMRP-containing mRNP outside the brain.



Fragile X Mental Retardation Syndrome. Figure 3 Translational regulation by FMR protein (FMRP) at the synapse. Interacting with known and unknown protein partners (symbolised by elliptic figures), FMRP binds to BC1/BC200 in a specific messenger ribonucleoprotein particle (mRNP). Direct interaction between BC1/BC200 RNA and a regulated mRNA is thought to bring the mRNP complex in the vicinity of the site where translation is normally initiated by binding of a ribosome, but is blocked by the FMRP complex. CBP, “Cap-binding protein” bound to the first nucleotide base of the mRNA, interacting with “Poly(A) binding protein” (PABP) that coats the mRNA’s tail formed by a series of A nucleotides.

The specificity of the synaptic mRNP complex is probably determined by the “▶[small dendritic RNA](#)” (▶[BC1/BC200](#)) that is included in the complex. BC1/BC200 RNA, also known as brain cytoplasmic RNA 1 (▶[BCYRN1](#)), is localised at synapses, does not encode a protein, but has been suggested earlier to be engaged in the regulation of ▶[dendritic protein biosynthesis](#). BC1/BC200 binds directly to FMR protein, and is also able to undergo base-pairing interaction with the mRNAs that are regulated by FMRP. There is some evidence that the FMRP mRNP represses translation of its target mRNAs in the initiation phase by blocking their binding to the ▶[ribosome](#) (5).

Thus, fragile X mental retardation is probably caused by synaptic dysfunction due to loss of appropriate regulation of dendritic protein biosynthesis in the absence of FMR protein.

Clinical Relevance

The identification of expanded triplet repeat sequences as the underlying cause of genetic disease has been a fascinating new development in human genetics. In fragile X mental retardation, expansion of a premutated CGG trinucleotide repeat in the *FMR1* gene on the human X chromosome to a full mutation upon maternal transmission sets off a molecular pathway leading to functional loss of the FMR protein that eventually results in a synaptic dysfunction phenotype. Investigation of the pathogenetic mechanisms in fragile X syndrome has already rendered valuable insights into complex processes of molecular genetics and cell biology, such as DNA synthesis and repair, DNA methylation and gene regulation. Research on fragile X syndrome is now beginning to significantly improve our knowledge of translational regulation at synapses, one of the key processes in the biochemistry of learning and memory.

▶[Inherited Mental Retardation Syndromes](#)

References

1. Verkerk AJ, Pieretti M, Sutcliffe JS et al (1991) Identification of a gene (FMR1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* 65:905–914
2. Imbet G et al (1998) FMR1 and mutations in fragile X syndrome: molecular biology, biochemistry, and genetics. In: Wells RD, Warren ST (eds) *Genetic instabilities and hereditary neurological diseases*. Academic Press, San Diego, pp 27–54
3. Wohrle D, Salat U, Hameister H et al (2001) Demethylation, reactivation, and destabilization of human fragile X full-mutation alleles in mouse embryocarcinoma cells. *Am J Hum Genet* 69:504–515

4. Steinbach P et al (1998) Systems for the study of triplet repeat instability: cultured mammalian cells. In: Wells RD, Warren ST (eds) *Genetic instabilities and hereditary neurological diseases*. Academic Press, San Diego, pp 509–528
5. Bakker CE et al (1994) FMR1 knockout mice: a model to study fragile X mental retardation. *Cell* 78:23–33
6. Comery TA, Harris JB, Willems PJ et al (1997) Abnormal dendritic spines in fragile X knockout mice: maturation and pruning deficits. *Proc Natl Acad Sci USA* 94:5401–5404
7. Zalfa F, Giorgi M, Primerano B et al (2003) The fragile X syndrome protein FMRP associates with BC1 RNA and regulates the translation of specific mRNAs at synapses. *Cell* 112:317–327

Fragile X Syndrome

▶[Fragile X Mental Retardation Syndrome](#)

Fragile X(A) Syndrome

Definition

Fragile X(A) syndrome is an X-linked mental retardation syndrome associated with ‘fragile X phenotype’ (as detectable after specific cell culture treatment for chromosome preparation; nowadays molecular genetic diagnosis of repeat expansion mutations in the ▶[fmr1](#) gene). The syndrome is characterized by moderate to severe mental retardation, dysmorphic signs (large ears, increased size of testicles, prominent jaw, joint hypermobility, chest deformity, ▶[mitral valve prolapse](#)) and high-pitched speech.

▶[Marfan Syndrome](#)

▶[Repeat Expansion Diseases](#)

Fragment-Based Ligand Design

Definition

Fragment-based ligand design describes a method used to identify ligands for a target protein by building up potent ligands in a modular way, starting from small fragments.

▶[NMR-Based Screening](#)

Frame (+1 Frame/−1 Frame)

Definition

With respect to an open reading frame on an mRNA, +1 frame is the translational reading frame consisting of codons shifted one nucleotide downstream (toward the 3' end of the mRNA). −1 frame refers to a reading frame to which codons are shifted one nucleotide upstream (towards the 5' end of the mRNA).

► [Translational Frameshifting, Non-standard Reading of the Genetic Code](#)

Frameshift Suppression/Suppressor

Definition

Frameshift suppression is a mechanism of translational termination suppression, used by some retroviruses. The process allows ribosomes to read (at a low frequency) through a frameshift mutation, producing the normal protein despite the genetic alteration. Translation can proceed from ► [gag](#) or pro into ► [pol](#).

► [Retroviruses](#)

Frameshift (Mutation)

Definition

Frameshift means the programmed shift of the reading frame during translation, which leads to a protein that differs in sequence from the sequence predicted from the mRNA on the basis of the standard genetic code, and is required to obtain functional protein. In most cases, the reading frame is shifted by one base in a +1 or −1 direction. Frameshifting frequently depends on the presence of particular structural elements (hairpin, pseudoknot) in the mRNA downstream of the shift site. The nature of the shift sequence is such that it provides a sufficiently stable base pairing between codon and anticodon, after the anticodon has shifted by one base. Frameshift mutations alter the reading frames of the corresponding messenger RNAs, so that codons downstream from the mutation site are out of register and are not translated properly. If the resulting protein is expressed, a frameshift mutation generally results in a truncated protein with an abnormal C-terminus.

► [Hereditary Spastic Paraplegias](#)

► [Ribosomes](#)

FRAP

- [Fluorescence Recovery after Photobleaching](#)
- [FRAP and other Photobleaching Methods](#)

FRAP and Other Photobleaching Methods

ERIK LEE SNAPP

Department of Anatomy and Structural Biology, Albert Einstein College of Medicine, Bronx, NY, USA
esnapp@aecom.yu.edu

Definition

► [Photobleaching](#) is the irreversible photo-induced destruction of the fluorescent properties of a fluorophore. While the loss of ► [fluorescence](#) can be problematic in fluorescence microscopy imaging experiments, photobleaching can be exploited in live cell imaging experiments to reveal information concerning the kinetic properties and environment of fluorescently labeled molecules in cells. Four different photobleaching applications (FRAP, FLIP, selective photobleaching and IFRAP) are described below. The different methodologies are illustrated in Fig. 1.

FRAP (Fluorescence Recovery After Photobleaching)

Fluorescent molecules in a small ► [region of interest](#) (ROI) of the cell are rapidly and irreversibly photobleached using a ► [confocal laser scanning microscope](#) (CLSM) with a high-powered laser beam. Subsequent

Frameshift Stimulator Sequences

Definition

Frameshift stimulator sequences refer to sequences found near to sites of programmed translational frameshifting, which have the effect of increasing the frequency of frameshifting.

► [Translational Frameshifting, Non-standard Reading of the Genetic Code](#)

movement of surrounding nonbleached fluorescent molecules into the photobleached ROI is imaged and quantitated using an attenuated light source (Fig. 1A). FRAP can reveal information concerning a protein's mobility and association with complexes and the protein's environment.

Selective Photobleaching

This method is similar to FRAP, but is generally used to characterize and quantitate nondiffusive movement of fluorescently labeled molecules. Specifically, a fluorescently labeled compartment or ROI is discretely photobleached using a CLSM to visualize and measure nondiffusive transport into and between discontinuous

compartments (e.g. trafficking between the endoplasmic reticulum to the Golgi complex).

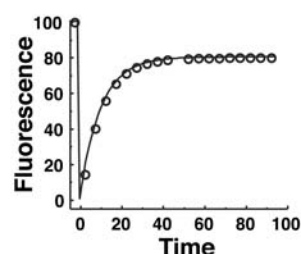
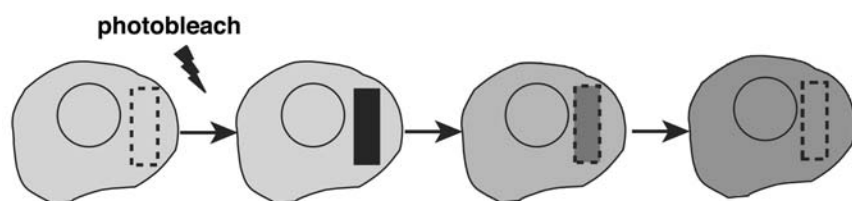
IFRAP (Inverse FRAP)

Everything outside a fluorescently labeled compartment or ROI is selectively photobleached and a CLSM is used to follow loss of fluorescence from the ROI (Fig. 1B). IFRAP complements selective photobleaching experiments and can potentially be used to study bind and release kinetics of proteins.

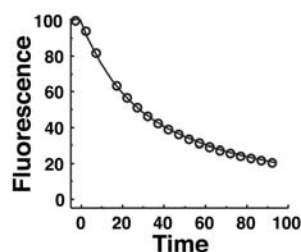
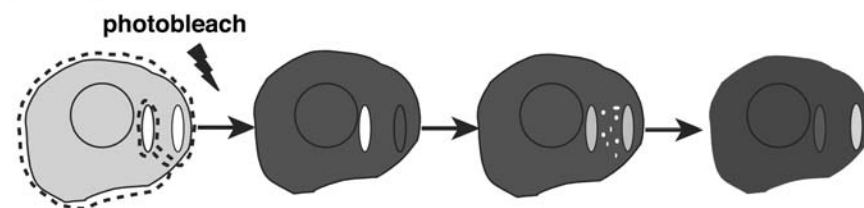
FLIP (Fluorescence Loss in Photobleaching)

Fluorescence in a discrete region of the cell is repeatedly bleached while images of the entire cell

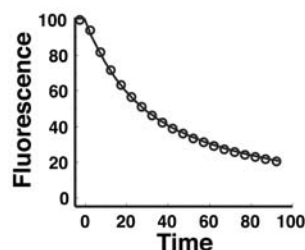
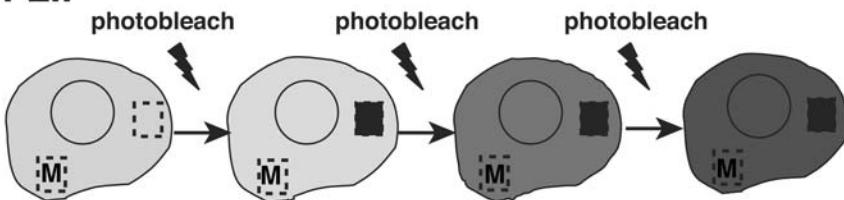
FRAP



IFRAP



FLIP



FRAP and Other Photobleaching Methods. Figure 1 Kinetic microscopy techniques. (a) In FRAP, a region of the cell (black dashed line box) is selectively and intensely irradiated to photobleach fluorescent molecules. The return of fluorescent molecules into that region is assessed quantitatively to determine diffusion coefficients and mobile fractions. (b) In IFRAP, the entire cell is photobleached except for one of two discrete structures (black dashed outline). The exit of fluorescent molecules or carriers from the unbleached region is assessed quantitatively to measure transport kinetics or bind and release kinetics. In addition, dim fluorescent carriers and their trafficking to other structures can be better visualized. In the example, fluorescent molecules have trafficked in vesicles from the unbleached structure to the photobleached structure. (c) In FLIP, a region of the cell (black dashed line box) is repeatedly photobleached. Movement of fluorescent molecules into the photobleached ROI results in loss of fluorescence from areas outside the box and can be used to determine the boundaries for a protein's diffusional movement within a cell. Loss of cellular intensity can be monitored in a separate region of interest, marked "M."

are monitored (Fig. 1C). In this photobleaching technique, the rate of loss of fluorescence rather than fluorescence recovery is monitored. FLIP can probe the extent of continuity of various intracellular membrane systems.

Description

Cellular proteins can freely diffuse, be immobilized to a scaffold or be actively transported. These dynamic properties influence what function a protein serves within the cell. Photobleaching can reveal a protein's diffusional properties, movement between compartments within cells and life history.

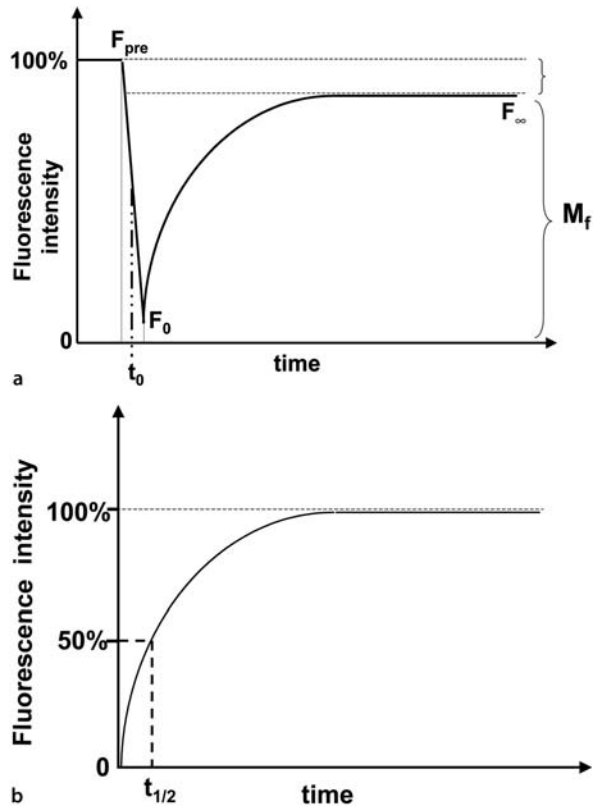
Photobleaching techniques depend on the availability of fluorescently labeled molecules such as fluorescent dyes and their labeled antibodies. ►GFP chimeras have become a popular method of fluorescently labeling proteins in living cells. An important property of GFP is that it can be photobleached without detectable damage to the surrounding environment. This is presumably because the GFP's cage-like structure, which surrounds a small cyclic peptide fluorophore (5), shields the external environment from any damaging effects caused by reactive photobleaching intermediates. These characteristics of GFP-chimeras make them ideal reagents in photobleaching experiments.

Fluorescence Recovery after Photobleaching

Several forms of FRAP or FPR (fluorescence photobleaching recovery) have been described. Popular methods include spot photobleaching, pattern photobleaching and strip bleaching. Spot photobleaching (1) can be performed on custom microscopes and some commercially available CLSMs. The strip photobleach method is suitable for commercially available CLSMs. Strip photobleaching permits imaging of the whole cell, while monitoring the fluorescence recovery into the photobleached ROI. This allows information concerning the changes in the spatio-temporal distribution of fluorescence to be readily determined.

The kinetic parameters that can be determined from FRAP experiments are the ►mobile fraction (M_f) of the fluorescent fusion protein and its ►diffusion coefficient (D). M_f and D are distinct parameters. D is a characteristic of the mobile pool of fluorescent proteins. A typical FRAP curve, which provides this information is shown in Fig. 2a. The M_f is the fraction of fluorescent proteins capable of diffusing into the photobleached region during the time course of the experiment. M_f is determined by calculating the ratio of the final to the initial fluorescence intensity in the bleached region, corrected for the amount of fluorescence removed during photobleaching.

D reflects the mean squared displacement that a protein explores through a random walk over time and has



FRAP and Other Photobleaching Methods. Figure 2 Plot of fluorescence intensity in a ROI versus time after photobleaching a fluorescent protein. (a) The prebleach (F_{pre}) is compared with the asymptote of the recovery (F_{∞}) to calculate the mobile and immobile fractions. Information from the recovery curve (from F_0 to F_{∞}) can be used to determine D . (b) The data plotted in part A can be transformed into a new scale in which F_{∞} represents 100% of the recovery and F_0 is 0% recovery. The time at which 50% of fluorescence has recovered into the photobleach ROI is the $t_{1/2}$. This analysis is useful for quantifying complex diffusion and nondiffusive processes.

units of area per time (usually $\text{cm}^2 \text{s}^{-1}$ or $\mu\text{m}^2 \text{s}^{-1}$). All proteins undergo this type of diffusive movement if they are not immobilized or experiencing active transport. The diffusion constant, D , is obtained by plotting the recovery of relative fluorescence intensity within the bleached region as a function of time and fitting this recovery curve with various equations. Several equations and simulations are available for calculating D for spot and strip photobleaching. For more information concerning FRAP theory and techniques see (3, 6).

Changes in M_f can reveal new information. A decrease in M_f indicates that the protein could be binding to fixed molecules or forming immobile aggregates or that

the protein is confined to a compartment and cannot contribute to fluorescence recovery in a separate disconnected compartment. When the M_f increases, the protein has been released from either a restricted compartment or a fixed macromolecular complex.

Changes in D and deviations from the predicted D value can reveal important information about the environment and size of the fluorescently labeled molecule. For example, a D significantly lower than a predicted value (indicating slower diffusion) suggests that a fluorescent protein could be incorporated into an aggregate or a large complex, because D is inversely proportional to protein size. Alternatively, the environment of a protein could be notably more viscous than expected or the protein could be interacting transiently with large or fixed molecules. By contrast, if D is significantly higher than predicted (indicating faster diffusion), the protein might be showing nondiffusive behavior such as ►flow or directed movement by motor proteins or the viscosity of the environment might be decreased. Given that there are many ways to interpret the D and M_f values for different proteins, combining FRAP data with biochemistry and cell biology is essential for determining the correct interpretation of the data.

The theoretical D for a protein is related to the molecular radius of the protein, the viscosity of the medium within which it is diffusing and whether the protein is soluble or integrated into a membrane. Membranes have a much higher viscosity than cytoplasm, so the lateral diffusion of an integrated membrane protein is considerably slower than that of a soluble protein and this is reflected in a lower D value. For a soluble spherical protein, an eightfold increase in size will lead to a twofold decrease in D . The relative insensitivity of D to small changes in molecular radius makes FRAP a poor method for detecting dimerization or other subtle changes in size.

The mobilities of several proteins are complicated by behaviors such as protein-protein interactions or binding to a matrix that might slow or immobilize a protein or populations with multiple D s. Such conditions complicate the analysis of D and may prevent fitting of data by traditional diffusion analysis. When a diffusion equation or simulation is unable to fit the fluorescence recovery data, the $t_{1/2}$ measure can be used to compare relative recovery rates between samples. The $t_{1/2}$ is the time required for the fluorescence intensity in the photobleach ROI to recover to 50% of the recovery asymptote fluorescence intensity (Fig. 2B). Other processes that lead to recovery into the bleached area (including vesicle transport or flow based movement) also may be quantitated and studied by this method. Though $t_{1/2}$ can be a useful tool, it is only relevant for the user's system and conditions.

Selective Photobleaching

Within cells, some proteins are transported between compartments within membrane-bounded vesicles moving along cytoskeletal fibers. It is difficult to gain insight into the rate of this movement in cells by only observing GFP chimeras at steady state. By selectively photobleaching the donor but not the acceptor compartment (or *vice versa*), vesicle transport of GFP-tagged cargo can be monitored by recording the rate of fluorescence recovery in the photobleached compartment. Data obtained from this type of experiment can be used to model kinetically the flux or cycling rates of the protein between compartments or to determine rate constants and residency times for proteins in distinct compartments. An additional application of selective photobleaching is to enhance the imaging of dim structures in cells or in areas of the cell next to very bright objects (i.e. photobleaching of Golgi fluorescence to visualize cargo delivery to this organelle).

Inverse FRAP

To probe the behavior and rates of trafficking of fluorescent proteins moving out from a structure, IFRAP can be performed. IFRAP essentially follows the FRAP protocol. However, instead of a strip, the investigator creates a photobleach ROI including the area of the entire cell, but excluding the organelle or structure of interest. For example, the Golgi complex (4) can be photobleached and then fluorescence recovery of either fluorescence into the photobleached organelle or fluorescence trafficking out of the unbleached organelle into the photobleached area surrounding the organelle can be imaged for both qualitative and quantitative analysis.

FLIP (Fluorescence Loss in Photobleaching)

FLIP is similar to FRAP in that an ROI is photobleached with a high power laser, but in FLIP the ROI is repeatedly photobleached over time to examine the behavior of the entire fluorescent pool. If fluorescent molecules within a cell are mobile and can diffuse into the photobleach ROI, the entire fluorescent pool of a compartment will be photobleached. Results from FLIP experiments can provide information about the connectedness of structures and organelles. For example, GFP-fusion proteins localized in the ER membrane show complete loss of fluorescence upon FLIP of a small area in the ER indicating that the luminal spaces of the ER are normally continuous throughout the cell (2). In addition, FLIP experiments can address whether a protein can diffuse uniformly across a compartment or whether there are regions of restricted mobility.

References

1. Axelrod D, Koppel DE, Schlessinger J et al (1976) Mobility measurement by analysis of fluorescence photobleaching recovery kinetics. *Biophys J* 16:1055–1069
2. Cole NB, Smith CL, Sicaky N et al (1996) Diffusional mobility of Golgi proteins in membranes of living cells. *Science* 273:797–801
3. Lippincott-Schwartz J, Snapp EL, Kenworthy A (2001) Studying protein dynamics in living cells. *Nat Rev Mol Cell Biol* 2:444–456
4. Nichols BJ, Kenworthy A, Polishchuk RS et al (2001) Rapid cycling of lipid raft markers between the cell surface and Golgi complex. *J Cell Biol* 153:529–542
5. Prendergast FG (1999) Biophysics of the green fluorescent protein. *Methods Cell Biol* 58:1–18
6. Snapp E, Altan N, Lippincott-Schwartz J (2003) Measuring protein mobility by photobleaching GFP-chimeras in living cells. Unit 21.1: In: Bonifacino J, Dasso M, Harford J (eds) *Current Protocols in Cell Biology*. John Wiley & Sons, Inc. New York

Frasier Syndrome

Definition

Frasier syndrome describes a rare disorder which is characterized by the presence of an XY sex chromosome constitution, and undermasculinized external genitalia that may range from ambiguous in appearance to normal-looking female genitalia. There is also kidney disease (glomerulosclerosis) and an increased risk of gonadal tumors (gonadoblastoma). Mutations in the donor splice site in intron 9 of the Wilms' tumor (WT1) gene have been shown to cause Frasier syndrome

► [Splicing](#)

FRAXA

Definition

FRAXA refers to the chromosomal fragile site located in the long arm of the X chromosome, which can be induced to break in cultured cells of people with a full mutation in the ► *fmr1* gene.

► [Fragile X Syndrome](#)

Free Energy

Definition

Free energy defines a fundamental thermodynamic state function taking into account both energy and entropy. The free energy change for a reaction (or a transition between two states) is a measure of how far the reaction is from equilibrium. It determines whether and in which direction the reaction can occur.

► [Molecular Dynamics Simulation in Drug Design](#)

Free Energy Change (dG)

Definition

Free energy change (dG) describes a thermodynamic state function that represents the work, at constant pressure, necessary to achieve a given state from the reference state. The Gibbs free energy change (dG) provides a measure of the stability of the thermodynamic state.

► [Thermodynamic Properties of DNA](#)

Free Energy Perturbation

Definition

Free energy perturbation describes a statistical thermodynamical method to calculate the free energy difference between two systems A and B by generating an ensemble using one of the two potential functions and evaluating an ensemble-average expression that depends on the difference between the two potential functions. The two systems should be sufficiently similar, so that one may be regarded as a “perturbation” of the other.

► [Molecular Dynamics Simulation in Drug Design](#)

Free Induction Decay

► [FID](#)

Free Radicals

LEOPOLD FLOHÉ, HEIKE BUDDE
MOLISA GmbH, Magdeburg, Germany
l.flohe@t-online.de

Definition

In chemical terms, free radicals are molecules with an uneven number of electrons. More precisely, free radicals are molecules that contain one or more unpaired electrons and thus are paramagnetic. This physicochemical definition includes a large number of highly diversified natural compounds, such as enzymes containing transition metals or catalyzing univalent redox-transitions, cofactors of the respiratory chain like ubiquinone, intermediates in various catalytic processes, and finally low molecular weight compounds such as molecular dioxygen (O_2), endothelium-derived relaxing factor (EDRF; $\cdot NO$) and the superoxide anion ($\cdot O_2^-$). In cell biology, however, the term “free radical” is used in a narrower definition to describe small, highly reactive oxygen-, nitrogen-, sulfur-, or carbon-centered radicals that are implicated in oxidative or **▶nitrosative stress**, respectively.

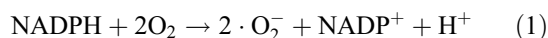
This article will focus on radicals produced enzymatically such as $\cdot O_2^-$ and $\cdot NO$ and “reactive oxygen species (**▶ROS**)” derived therefrom. It has, however, to be stressed that the terms “oxygen-centered free radical” and “ROS” are not identical. The latter includes aggressive species such as H_2O_2 , alkylhydroperoxide, singlet oxygen, hypochlorite and peroxy-nitrite that are not radicals but are formed from, or may initiate, free radical reactions.

Characteristics Chain Reactions

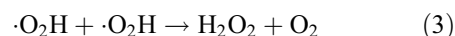
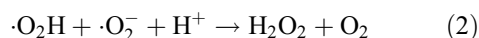
A common feature of free radicals is their ability to initiate, propagate or terminate free radical chain reactions (1). If a free radical is formed (chain initiation), it tends to react with any kind of electronically saturated molecules. The product thus formed is inevitably a radical that may react with another molecule to create a new radical (chain propagation). Only if the reaction partner of a radical is a radical itself, can the free radical chain reaction be terminated by forming a product with an even number of electrons (chain termination). In such free radical chain reactions, labile intermediates are often formed that decompose with formation of two radicals, each being able to initiate a new chain reaction (chain branching). If not balanced by free radical scavengers, this leads to an uncontrolled situation, known by the term “**▶oxidative stress**” (1).

Phagocytic Killing, a Superoxide-driven Free Radical Chain (1, 2)

A typical free radical process in biology is the “killing reaction” of phagocytes, which is an indispensable component of the innate immune response but, if unbalanced, causes oxidative tissue damage as, e.g., in septicemia. Herein, a free radical chain is initiated by **▶NADPH oxidase**, a transmembranal cytochrome b-containing flavoprotein ($gp91^{phox}$). It separates the electron pair of NADPH to form two $\cdot O_2^-$ radicals.



The superoxide radical anion, $\cdot O_2^-$, is not a particular aggressive radical. It is a weak base that associates to the superoxide radical ($\cdot O_2H$) with a pK of 5. At the low pH of the phagocytic vacuole it dismutates spontaneously according to Eq. 2 or 3:

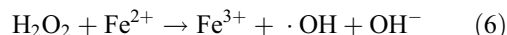


At neutral pH, the $\cdot O_2^-$ anions require enzymatic catalysis to dismutate. These reactions seemingly terminate the radical chain, since O_2 , although a biradical, is less reactive than $\cdot O_2^-$ and H_2O_2 is no longer a radical. H_2O_2 and other hydroperoxides ($ROOH$), however, can re-initiate radical chains in different ways. They may decompose according to 4 and 5 to form hydroxy ($\cdot OH$) or alkoxy radicals ($RO\cdot$).

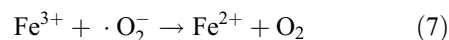


According to Eq. 4 the chain termination according to (2) and (3) leads to chain branching. A variety of other chain branchings and/or propagations, which differ between tissues, appear to be physiologically more important.

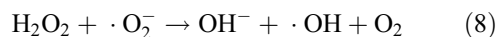
In general, H_2O_2 decomposes in the presence of Fe^{2+} ions with formation of $\cdot OH$, as has been known for a long time from Fenton chemistry (Eq. 6).



The Fe^{2+} thereby consumed can be regenerated by $\cdot O_2^-$ (or other reductants; Eq. 7):

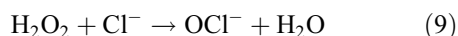


Summing up reaction (6) and (7) reveals a continuous production of $\cdot OH$ from $\cdot O_2^-$ and H_2O_2 as long as catalytic amounts of transition metals are available:

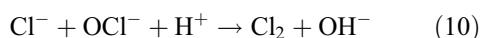


This special case of Fenton chemistry is also called the Haber-Weiss reaction. It is a major source of the hydroxyl radical. The latter reacts with all kind of SH groups, polyalcohols and aromatic compounds with diffusion-limited rate constants near $2 \times 10^{10} \text{ M}^{-1} \text{ s}^{-1}$ and accounts for most of the toxicity attributed to superoxide-driven free radical chains.

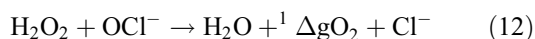
In activated Polymorphic Clear Lencocytes (PMNs), the myeloperoxidase uses the ►respiratory burst-derived H_2O_2 to produce hypochlorite, a non-radical but equally aggressive oxidant (Eq. 9):



Hypochlorite according to Eq. 10 and 11 is in equilibrium with chlorine molecules and chlorine atoms. The latter, as radicals, re-initiate and branch the radical chain:



Hypochlorite and H_2O_2 further generate singlet oxygen ($^1\Delta_g \text{O}_2$), a non-radical excited form of molecular dioxygen.

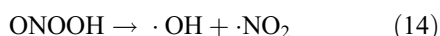


$^1\Delta_g \text{O}_2$, commonly obtained by UV light, is the reactive form of O_2 that, e.g., readily oxidizes unsaturated lipids.

Particularly in macrophages, the production of $\cdot\text{O}_2^-$ may occur simultaneously with that of $\cdot\text{NO}$. These two radicals instantly react with each other to peroxynitrite:



Again a non-radical but highly aggressive oxidant is generated, which is implicated in bacterial killing, protein nitration and nitrosation. Moreover, peroxynitrite under acid conditions decomposes with formation of two radicals $\cdot\text{OH}$ and $\cdot\text{NO}_2$, and thus re-initiates two radical chains:



While $\cdot\text{NO}$ itself is a fairly inert and, in physiological terms, often beneficial radical, the peroxynitrite-mediated reactions tend to account for pathological phenomena subsumed under the term nitrosative stress. Any of the $\cdot\text{O}_2^-$ derived radicals (Eq. 1–13) may initiate chain reactions in lipid phases by hydrogen abstraction from unsaturated lipids. Carbon-centered radicals thus formed add O_2 to yield lipidperoxy radicals, $\text{ROO}\cdot$ that are reduced by lipid-soluble antioxidants, e.g. tocopherols, to lipidhydroperoxides. If not enzymatically

reduced by any of the ►glutathione peroxidases, these may, in analogy to Eq. 4–6, decompose to $\cdot\text{OH}$ and/or alkoxy radicals ($\text{RO}\cdot$), which sustain the radicalic “►lipid peroxidation”.

Molecular Interactions (1, 3, 4)

Apart from the nonspecific interactions of radicals with each other and other susceptible molecules implicated in oxidative stress, both primary biological radicals, $\cdot\text{O}_2^-$ and $\cdot\text{NO}$, have their specific targets that mediate either their decomposition or specific cellular responses.

$\cdot\text{O}_2^-$ and $\cdot\text{NO}$ have pronounced affinity to heme proteins. The reduction of cytochrome c by $\cdot\text{O}_2^-$, which is used for routine $\cdot\text{O}_2^-$ determination, may be quoted as an example. Similarly the formation of nitroso hemoglobin by $\cdot\text{NO}$ can be used analytically and is a physiological scavenging mechanism. Moreover, $\cdot\text{NO}$ binding to heme proteins is the chemical basis of its role as signaling molecule (see below). Both radicals modify protein thiols. The relative specificity of thiol modifications is due to the molecular environment of the SH groups, which have to be exposed and dissociated. The chemistry of such protein modification is complex and in most cases the radicals are not considered to be the proximal reactants. Oxidation of protein SH to SOH followed by disulfide formation is rather due to H_2O_2 or ONOO^- than to $\cdot\text{O}_2^-$. S-Nitrosation, i.e. the formation of R-SNO, is also not directly achieved by $\cdot\text{NO}$ but by the nitroxyl cation, NO^+ , which is formed from $\cdot\text{NO}$ and O_2 via N_2O_3 . Since nitrosothiols readily react with free thiols, nitrosation represents an alternative route to protein mixed disulfides.

Nitration of proteins at tyrosine residues under conditions of $\cdot\text{NO}$ generation often occurs in a rather specific manner. Again, not $\cdot\text{NO}$ itself, but $\cdot\text{NO}_2$ that is generated from peroxynitrite is the immediate reactant. While specific enzymatic routes to degrade $\cdot\text{NO}$ are unknown, $\cdot\text{O}_2^-$ is eliminated at almost diffusion-limited rates by ►superoxide dismutases (SOD), which are metalloenzymes of three distinct protein families, containing Fe, Mn or Cu as catalytic transition metal. The SODs acting with copper contain Zn as additional metal cofactor for structural stabilization. The unusually high rate constants of the Cu/Zn SODs ($2 \times 10^9 \text{ M}^{-1} \text{ s}^{-1}$) are explained by an electrostatic gradient that draws the negatively charged $\cdot\text{O}_2^-$ into a tunnel of increasing positive charge with the complexed Cu^{2+} ion at its bottom. The catalytic mechanism is otherwise simple. A first $\cdot\text{O}_2^-$ delivers its unpaired electron to the Cu^{2+} , a second one picks it up again. An analogous redox shuttling of the transition metal takes place in the Fe- and Mn-SODs.

Regulatory Mechanisms

Regulation of Radical Production (2, 4, 5)

For obvious reasons the formation of free radicals is strictly controlled. $\cdot\text{O}_2^-$ production by phagocytes is triggered by a receptor-mediated phosphorylation cascade that leads to an association of the cytosolic components p47^{phox} , p40^{phox} , p67^{phox} and the GTP binding proteins Rac 1/2 to the catalytic transmembranal flavocytochrome $\text{gp91}^{\text{phox}}$ and to p22^{phox} (2). Formation of the active complex is associated with a massive oxygen consumption ("respiratory burst") and release of $\cdot\text{O}_2^-$ into the phagosome or the extracellular space. Typical activators of the respiratory burst are microbial structures (e.g. formyl peptides) and numerous inflammatory mediators such as anaphylatoxins, platelet activating factor, TNF α , leukotriene B $_4$, and interleukins. The ability of calcium ionophores and phorbol esters to induce a respiratory burst suggests an involvement of PKC-type enzymes in the initial steps of signaling. The relevance of the system to host defense is evident from mutations in any of the system components that all result in defective bacterial killing (Chronic Granulomatous Disease).

Many variations of this theme have meanwhile been detected in other cell types (3). Either the activating components or the catalytic entity itself may be changed. The flavocytochrome $\text{gp91}^{\text{phox}}$ (Nox 2) may be replaced by any of the homologous Nox proteins. In contrast to Nox 1–4, Nox 5 has a calcium-binding domain attached; the Duox proteins, in addition, contain an intracellular peroxidase domain. Although less efficient, the $\cdot\text{O}_2^-$ producing systems of non-phagocytic cells are also discussed in the context of innate immune response. In particular cases, alternative roles are obvious. In the thyroid gland, e.g., a Duox 1/2 system, *via* $\cdot\text{O}_2^-$ production, probably provides the H_2O_2 required for thyronine synthesis. In nematodes, Duox systems are implicated in oxidative cross-linking of cuticular proteins *via* dimerisation or trimerisation of tyrosine residues. It remains to be worked out which of the system variants is responsible for growth factor-induced ROS formation, which is also believed to result from an initial $\cdot\text{O}_2^-$ synthesis by any of the $\text{gp91}^{\text{phox}}$ -related proteins.

$\cdot\text{NO}$ is generated in mammals by three distinct, though homologous, NO synthases (NOS) from arginine, NADPH and oxygen (3, 4). The neuronal form, nNOS, which is predominantly, but not exclusively formed in the nervous system, and the endothelial form, eNOS, are low output enzymes that are primarily regulated at the protein level by the Ca^{2+} /calmodulin system. They require free Ca^{2+} for activity, while their activity is reduced by serine/threonine phosphorylation by Ca^{2+} /calmodulin-dependent PK II, PKA, C or G, and other protein kinases. Of the NOS-bound cofactors, which

are FAD, FMN, heme and tetrahydrobiopterin (BH_4), the last deserves particular interest, since suboptimum saturation changes the NOS reaction specificity towards that of an $\cdot\text{O}_2^-$ producing system. Regulation of nNOS and eNOS at the transcriptional level, e.g. by sex hormones and corticosteroids, has also been observed but only modulates the, in essence, constitutive expression (for more details see 2). In contrast, the high-output "inducible NOS", iNOS, is primarily regulated at the expression level (3), while little is known about regulation of the enzyme itself apart from inhibition of (activating) dimerisation by kalirin in the nervous system and by NAP110 in macrophages. The pattern of inducers/inhibitors, the signaling pathways and affected promoter sequences vary substantially between cell types (3). Most consistently, the iNOS promoter is positively affected by NF κB , octamer factor (Oct), interferon regulator factor-1 (IRF-1), C/EBP, HIF-1, estrogen receptor- β , pregnane X receptor (PXR), antioxidant-responsive element binding protein, nuclear factor- Jl6 (NF- Jl6) and Krüppel-like factor 6 (KLF6). Interestingly, there is a considerable, although not strictly conserved, overlap between the cytokines and signaling pathways leading to iNOS induction and those triggering oxidative burst. This implies that iNOS induction can in most cases be rated as an inflammatory or stress response and that, under these conditions, excessive amounts of $\cdot\text{NO}$ and $\cdot\text{O}_2^-$ are produced simultaneously, a situation that inevitably results in peroxynitrite formation.

Regulation by Free Radicals

$\cdot\text{O}_2^-$ formation upon receptor stimulation by, e.g., EGF, PDGF or insulin is considered to positively influence signal transduction and, accordingly, the radical is often described as a second messenger (3, 4). In view of the fast enzymatic dismutation and the multiple possible interactions of the radical with other cellular components it is, however, almost impossible to attribute the physiological consequences to $\cdot\text{O}_2^-$ itself. Cautious interpretation attributes the observation to "ROS" (4) and there is ample evidence that $\cdot\text{O}_2^-$ -dependent modulation of signaling is achieved through protein thiol modification by H_2O_2 . In addition, alkylhydroperoxides directly formed by lipoxygenases may contribute to enhancement of signaling. In line with this view, exposure to cell-permeable thiols and overexpression of GPx- or Prx-type peroxidases usually dampen or even block such signaling cascades. In most cases of redox-regulated signaling cascades, however, the molecular targets of ROS still remain as unclear as the reacting oxidant species and the nature of target modification.

Much better defined is the regulatory role of $\cdot\text{NO}$. By binding to the heme moiety of soluble guanylyl cyclases

it enhances the formation of cyclic GMP. This appears to account for many of the specific $\cdot\text{NO}$ effects such as modulation of cGMP-regulated ion channels, phosphodiesterases and protein kinases. A direct binding of $\cdot\text{NO}$ to heme moieties is also discussed as a mechanism to induce $\cdot\text{O}_2^-$ formation in the mitochondrial respiratory chain and is implicated in triggering apoptosis. The effect of $\cdot\text{NO}$ on GTP-binding proteins involves nitrosation of cysteine residues and thus is not a direct effect of $\cdot\text{NO}$. Similarly, the effects of $\cdot\text{NO}$ on MAP kinase and other signaling cascades are often not mimicked by cGMP. They rather resemble the effects of ROS and might involve peroxynitrite-dependent oxidations or nitrosations.

► [Molecular Aging Research](#)

References

1. Sies H (1985) Oxidative stress. Acad Press, London, p 1507
2. Babior BM (1999) NADPH oxidase: an update. Blood 93:1464–1476
3. Forman HJ, Fukuto J, Torres M (2003) Signal transduction by reactive oxygen and nitrogen species. Kluwer Acad Publisher Dordrecht, The Netherlands, pp 1–411
4. Dröge W (2002) Free Radicals in the control of cellular function. Physiol Rev 82:47–95
5. Kleinert H, Schwarz PM, Förstermann U (2003) Regulation of the expression of inducible nitric oxide synthase. Biol Chem 384:1343–1364

FRET

► [Forsker Resonance Energy Transfer](#)

Frictional Ratio

Definition

Frictional ratio designates the ratio of the frictional coefficient of a molecule to the frictional coefficient of a compact sphere with the same mass and density. This number, usually denoted f/f_0 is a measure of the shape asymmetry of a molecule.

► [Analytical Ultracentrifugation](#)

Friedreich Ataxia

Definition

Friedreich ataxia is an autosomal recessively inherited disease with sclerosis in the dorsal and lateral columns of the spinal cord, usually beginning in childhood or youth, caused by mutations with expansion of the trinucleotide-repeat GAA in the Frataxin gene on chromosome 9q13.

► [Repeat Expansion Diseases](#)

Frizzled

Definition

Frizzled proteins belong to a family of 7 transmembrane spanning serpentine receptors, which function as receptors for Wnts and have been shown to interact with Wnts via their cysteine rich domain (CRD). Drosophila has 4 Frizzleds, while most vertebrates have 10 or more.

► [Wnt/Beta-Catenin Signaling Pathway](#)

FSC

► [Fourier Shell Correlation](#)

FSH

► [Follicle Stimulating Hormone](#)

FTI

Farnesyl transferase inhibitor

► [Protein Prenylation](#)

FT-ICR

Fourier transform ion cyclotron resonance

► [Proteomics in Cancer](#)

Fukutin

Definition

Fukutin belongs to a family of glycosylation enzymes in bacteria and yeast. Mutations in the ► [fukutin](#) gene lead to ► [Fukuyama myopathy](#) (FCMD).

► [Limb Girdle Muscular Dystrophies](#)

Fukuyama Myopathy

Definition

Fukuyama myopathy or Fukuyama-type congenital muscular dystrophy (FCMD OMIM: 253800) is an autosomal recessive or transmitted congenital muscular dystrophy that is highly prevalent in Japan (approximately 1 in 10,000 births). The disease is associated with mental retardation, seizures, and an unusual brain abnormality. It is due to a mutation in the fukutin gene.

► [Limb Girdle Muscular Dystrophies](#)

Full Mutation

Definition

Full mutation refers to a disease causing expansion of the CGG trinucleotide repeat in the ► [fmr1](#) gene, responsible for the fragile X mental retardation syndrome.

► [Fragile X Syndrome](#)

Full-Length cDNA

Definition

Full-length cDNA is defined as a cDNA that contains the sequence information of the whole original mRNA it derives from. However, cDNA is frequently defined as being full-length if it covers the complete protein

coding part of the original mRNA (some sequence of the 5' UTR may not be present). While the latter cDNA is still useful to express the encoded protein, the identification of the transcription start site and all sequence motifs that had been present in the 5' UTR of the mRNA is not possible.

► [Full Length cDNA Sequencing](#)

Full-Length cDNA Sequencing

STEFAN WIEMANN, REGINA ALBERT,
PETRA MOOSMAYER, INGO SCHUPP
Molecular Genome Analysis, DKFZ, Heidelberg,
Germany
s.wiemann@dkfz.de

Definition

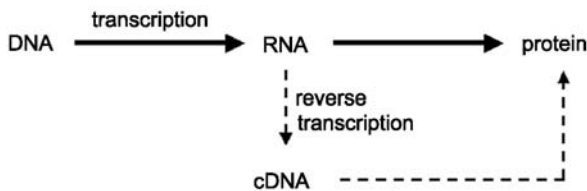
► [cDNA](#) is a 'complementary' or 'copy-DNA' that is generated *in vitro* from cellular ► [mRNA](#) (messenger ► [RNA](#)). ► [Full-length cDNAs](#) contain the complete sequence information of their respective mRNA templates. This information is determined by full-length cDNA sequencing and provides knowledge of the protein encoded by the original mRNA and, in addition, of the untranslated regions (► [UTR](#)) of this mRNA. The so called open reading frame (► [ORF](#)) specifies the protein sequence that is encrypted in the mRNA while the UTR frequently contain sequence elements, e.g. regulating mRNA stability or localization.

Characteristics

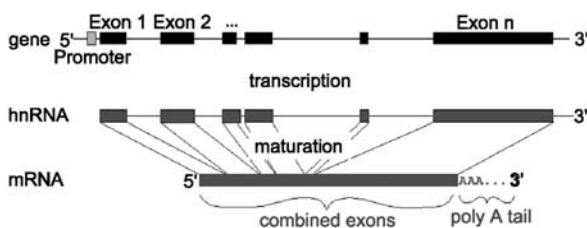
RNA is an Intermediate between DNA and Protein

mRNA is an intermediate between DNA (the gene) and the encoded protein (Fig. 1) and is one of several RNA species that are present in cells. DNA stores the genetic information, and proteins are the molecules executing most cellular processes. Most genes are present in one or two copies in the chromosomes and alterations in the copy number mostly result in disease (e.g. trisomy 21 causes ► [Down syndrome](#)) or death. The genome needs to be stable over generations. In contrast, the amount of mRNA that cells produce from these genes can vary over several orders of magnitude to allow for a rapid and fine-tuned expression of these mRNAs and the proteins. RNA is generated exclusively as required and quickly degraded when no longer necessary. The expression of mRNA is often tissue specific or occurs only in defined developmental stages.

In addition, genes in eukaryotes are mostly structured in ► [exons](#) and ► [introns](#). To produce mRNA, introns



Full-Length cDNA Sequencing. Figure 1 From DNA to protein. *In vivo*, protein is generated from the DNA (genes) with the help of an intermediate messenger or mRNA. RNA can be isolated and copied *in vitro* into cDNA with the help of the enzyme reverse transcriptase (dotted line to from mRNA to cDNA). By determining the sequence of cDNA, the gene and the sequence of the encoded protein are identified. A cDNA can also be used to translate the encoded protein *in vitro* (dotted line from cDNA to protein).



Full-Length cDNA Sequencing. Figure 2 From the gene to the mRNA. The genome (on top) contains genes, which have a promoter region and other elements regulating the amount of RNA that is transcribed. In eukaryotes, most genes consist of exons (boxes) and introns (thin lines). The exons contain the information needed to make protein or a functional RNA. The functions of introns are not completely clear yet; however, they are frequently involved in the regulation of transcription, in alternative splicing and in evolutionary processes. First the gene is transcribed by DNA-dependent RNA-polymerase (transcription) into a perfect copy (hnRNA = heterogeneous nuclear RNA - center) of the gene, although omitting the promoter region. The following maturation of the RNA consists of processes during which a stretch of A-residues is polymerized at the 3' end (poly A tail) and the introns are removed by splicing. In the end, the exons are lined up in a non-interrupted order. All these processes, starting from transcription and ending with the mature mRNA happen in the nucleus.

are removed in a natural process named **▶splicing** to yield the protein coding part in a non-interrupted fashion (Fig. 2). The mature mRNA is transported into the cytoplasm of cells, where the encoded protein is translated with help of ribosomes.

cDNA is Synthesized from RNA

Unfortunately mRNA degrades very easily and cannot be manipulated e.g. cut with restriction enzymes.

Consequently, the mRNA must be converted into more stable DNA before analysis. This reaction, which is called reverse **▶transcription**, is catalyzed by the enzyme reverse transcriptase (1) and produces cDNA (Fig. 1). Since cDNA normally descends from mature mRNA, it contains the protein encoding part of the original gene in non-interrupted order. cDNAs are also a means for the identification of alternative splice forms (**▶Alternative splicing**), when different sets of exons are used *in vivo* to assemble the template mRNA. The cDNA can additionally be used to express the encoded protein *in vitro* (Fig. 1).

This DNA is cloned into vectors that can be propagated and amplified in bacteria to produce cDNA libraries (2). In analogy to a book library, which consists of a large collection of different books, cDNA libraries should contain representatives of, ideally, all mRNAs that were expressed in the tissue from which the RNA was isolated.

When cDNAs are generated and cloned to make cDNA libraries, mRNAs expressed at low levels are usually underrepresented, whereas highly expressed mRNAs lead to high numbers of cDNA clones in these libraries (Table 1). The different abundances of cDNA species impose difficulties, both in gene discovery and in full-length sequencing projects. New genes or targets for full-length sequencing are commonly identified by random sequencing of cDNAs on a large scale. mRNAs present in high and medium abundance are quickly identified by this approach. However, mRNAs expressed at low levels often escape identification. One effect of this problem is that the complete number of human genes is still unknown, despite the genome sequence being fully determined.

Full-Length Cloning and Sequencing vs. Gene Discovery

1. The sequencing of cDNAs can be carried out to obtain two levels of information (Fig. 3). The full-length sequencing of cDNAs is performed to determine the complete sequence of the protein coding as well as the non-coding parts of the mRNA in question. Splice variants of mRNAs often lead to variant proteins with different functions or activities (see clinical relevance) and are also identified by full-length sequencing of cDNAs. In addition, the cDNAs are physical resources, i.e. DNA clones, suitable for experimental exploitation (e.g. expression of proteins).
2. The sequencing of **▶ESTs** (expressed sequence tags) has proved valuable in the discovery of genes. ESTs are sequence information generated from cDNA clones; however, only the ends of the clone-sequences are determined (Fig. 3). This short stretch (tag) of sequence is still sufficient to identify the gene this cDNA represents (3). Note that the

Full-Length cDNA Sequencing. Table 1 mRNA abundance classes. The amount of specific mRNAs in cells is highly variable. While some genes are expressed at high levels, resulting in large amounts of a specific mRNA (copies of mRNA per cell), other mRNAs are expressed in very small quantities. Most mRNAs are present in low or intermediate abundance.

Abundance Class	Copies per Cell	Abundance of each mRNA
Low	5–15	<0.004%
Intermediate	200–400	<0.1%
High	12,000	3%

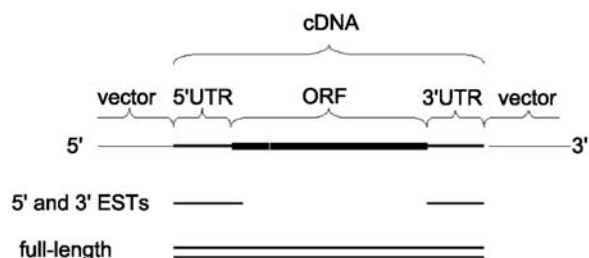
terms “cDNA” and “EST” need to be clearly distinguished. While a cDNA is a physical resource (clone), an EST is a piece of sequence information that is generated from a cDNA.

Large-scale projects aiming at the full-length sequencing of cDNAs were initiated considerably after the production of ESTs had become fashionable. Most (human) genes are believed to be represented by one or more EST-sequences in public databases, however, a significant fraction of the respective genes is not (yet) covered by full-length cDNA clones and sequences.

Model Organisms (Human, Mouse, Rat, Others)

With improved sequencing and ►cDNA library technologies available, the number of full-length cDNA sequencing projects and the organisms analyzed has grown over the years. The first human full-length cDNA analysis project was established in 1994 in Japan (4). Long human cDNAs were generated and sequenced at the Kazusa DNA Research Institute. Since then, a small number of large-scale projects have evolved in Japan (5), Germany (6) and in the USA (7) with the focus on human and mouse cDNAs. The overall goal of these projects has been to generate full-length cDNAs for all genes contained and expressed in the respective species. For human over 50% of the expected genes are available as full-length sequences, for most of them the cDNA clones are also made available.

The number of species analyzed by full-length cDNA sequencing is low compared to the large number of organisms that have been sequenced at the genome level. This is partly due to the high variability of mRNA levels in cells and tissues (Table 1) and also because of the different sequencing set-up necessary to analyze cDNAs, compared to genome sequencing. An up-to-date list of individual projects is available at the National Center for Biotechnology Information (NCBI) in the USA (► <http://www.ncbi.nlm.nih.gov/genome/flcdna/>). Most organisms that are currently analyzed in full-length cDNA sequencing projects are important in



Full-Length cDNA Sequencing. Figure 3 EST vs. Full-length sequencing. A cDNA is cloned into a plasmid vector, amplified in bacteria and then purified. Using vector primers, the terminal sequences of the cDNA insert are determined at the 5' and 3' ends (in the direction of the original mRNA), which produces ESTs (expressed sequence tags). These sequences are error-prone since they have been determined analyzing only one of the two DNA strands (ESTs drawn as single lines). In full-length sequencing, the sequence of the cDNA is determined by analyzing both DNA strands (drawn as a double line). An increase in accuracy is reached by analysis of the two DNA strands compared to the relatively poor quality obtained in EST sequencing. A cDNA ideally contains the 5' UTR and 3' UTR (un-translated region), and also, in the case of protein-coding cDNAs, the non-interrupted ORF (open reading frame), where the protein is encoded.

cell biology (e.g. *Drosophila*, *Xenopus*) and medicine (e.g. human, mouse, zebrafish).

Clinical Relevance

Full-length sequences and clones serve at least two functions with clinical relevance.

1. Identification of the full-length mRNAs and encoded proteins is essential in disease gene identification and for the identification of splice variants associated with disease (clotting factor VIII in haemophilia A, presenilin in Alzheimer). The full-length sequences then form the basis for the functional characterization of the proteins and for the development of diagnostic tools.
2. Full-length cDNAs can be used to produce the proteins encoded by the cDNA. For example, human clotting factor VIII is commonly expressed in bacteria using a full-length cDNA as starting material. The factor is purified and used therapeutically in the treatment of haemophilia A patients. Thus, full-length cDNA can also serve as a basis for therapeutics.

With the genome being sequenced, and growing information on genes and gene products becoming available, the importance of full-length cDNAs in functional genomics, proteomics and medical sciences will further increase.

References

1. Baltimore D (1970) RNA-dependent DNA polymerase in virions of RNA tumour viruses. *Nature* 226:1209–1211
2. Gubler U, Hoffman BJ (1983) A simple and very efficient method for generating cDNA libraries. *Gene* 25:263–269
3. Venter JC (1993) Identification of new human receptor and transporter genes by high throughput cDNA (EST) sequencing. *J Pharm Pharmacol* 45:355–360
4. Nomura N, Miyajima N, Sazuka T, Tanaka A, Kawarabayashi Y, Sato S, Nagase T, Seki N, Ishikawa K, Tabata S (1994) Prediction of the coding sequences of unidentified human genes. I. The coding sequences of 40 new genes (KIAA0001-KIAA0040) deduced by analysis of randomly sampled cDNA clones from human immature myeloid cell line KG-1. *DNA Res* 1:47–56
5. Okazaki Y et al (2002) Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* 420:563–573
6. Wiemann S et al (2001) Toward a Catalog of Human Genes and Proteins: Sequencing and Analysis of 500 Novel Complete Protein Coding Human cDNAs. *Genome Res* 11:422–435
7. Strausberg RL et al (2002) Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proc Natl Acad Sci USA* 99:16899–16903

Full-Size cDNA

Definition

Full-size (or full-length) cDNA designates a complete cDNA copy of mRNA (complementary DNA, ►cDNA) extending from the 5' capping site to the 3' polyA region.

►RNA Capping

►Full Length cDNA Sequencing

Functional Assays

JEREMY C. SIMPSON

Cell Biology and Biophysics Programme, European Molecular Biology Laboratory (EMBL), Heidelberg, Germany
simpson@embl-heidelberg.de

Synonyms

Functional assays are encompassed within the terms high-throughput screening and high-content screening.

Definition

One goal of current proteomic approaches is to provide a listing of all the proteins present in either an

organelle, cell type or whole tissue at a particular point in time. Although such data are extremely valuable, this is only one step towards our understanding of global protein function. Once a particular set of proteins has been identified, methodologies must be applied to screen these proteins for their cellular role. In this regard, functional assays can be defined as systematic *in vivo* experiments that are designed to determine the involvement of each protein in a particular cellular pathway or biological process. This approach provides both the first clues to the cellular role of individual proteins and a means to classify the proteome into functionally related groups. In keeping with this definition, the term functional assays is not used here to describe general enzymatic activities such as phosphorylation or glycosylation. The readout from functional assays can be obtained in a variety of ways, but the description given here will use the paradigm of microscopy-based methods to identify proteins that are involved in intracellular transport.

Characteristics

In multicellular organisms, biological function is ultimately determined at the cellular level. Although the human body is composed of at least 250 cell types, some of which perform specialised functions, a large number of biological processes are common to nearly all cells. These include cell division, the ►secretion of various proteins or macromolecules, the uptake of nutrients and the cell's response to external stimuli. A large part of the proteome in any cell type is dedicated to these processes and the challenge is to correctly assign proteins to these as a first step towards a full understanding of their individual role. To reconcile such large lists of proteins with a putative function therefore requires the application of systematic and directed assays, which can be characterised by certain features:

1. methodology that allows testing of individual proteins in turn,
2. criteria to choose which proteins are appropriate for which assay,
3. clear and reliable readouts from the experiment, preferably quantitative,
4. the potential for scaling up to thousands of samples,
5. easy adaptability as requirements change.

The purification of every protein from the human proteome is still technically unachievable and furthermore, since it requires their removal from the normal cellular environment, it may be considered an inappropriate means to determine their *in vivo* function. Consequently, functional assays must adopt an approach whereby each protein under investigation can be easily identified while still in its cellular context. This has now become possible due to the complete

sequencing of the genomes of a number of model organisms. Such sequencing projects not only provide a means of identifying all the possible proteins in an organism, they also provide a physical resource of many thousand clones and ►[open reading frames](#) (ORFs) that encode the proteins (1). The availability of large clone collections provides an excellent tool for functional assays, as these can be taken, genetically tagged and then reintroduced into cells in appropriate ►[expression vectors](#). The tag enables the protein to be specifically localised within the cell and the effect of the over-expression of individual proteins on a given process can be monitored. Since the introduction of a number of recombination-based cloning systems, many thousand ORFs can be subcloned rapidly into these expression vectors without having to take account of their individual restriction enzyme sites. The advantage of this whole approach is that tagged ORFs can be easily introduced into model cell systems using either plasmid or viral vectors. Simple plasmid vectors are easy to prepare and propagate and indeed a wide variety are available for use with recombination cloning strategies. However, when used in transient ►[transfection](#) experiments, it is inevitable that not all of the cells receive the DNA for over-expression. This is generally not a problem with viral vectors, but their preparation is more labour intensive and the viral infection itself can induce cellular changes. The design of functional assays around any type of expression vector does however have one major caveat; in order to see a phenotypic effect, it is assumed that the over-expressed protein of interest is present in sufficient amounts to titrate out its interacting partners. In this way, over-expression experiments attempt to override the normal operation of a particular process by upsetting the equilibrium of the endogenous factors involved.

The choice of tag is important when designing functional assays. Many ►[epitope tags](#) are now available, such as haemagglutinin, hexahistidine, V5, FLAG and so on, and these have the advantage of being relatively small (a few amino acid residues) compared with the size of the final expressed protein. This means that generally, the function of the tagged protein and its interactions are preserved. However, the disadvantage of these epitope tags is that they can only be visualised (for example using antibodies) after the cells have been fixed. Another type of tag makes use of entire proteins that have fluorescent properties. The most frequently used of these is the ►[green fluorescent protein](#) (GFP). As GFP fluorescence still occurs when appended to another protein and without the need for co-factors, it is a marker that can be utilised in living cells from many organisms. Although it is significantly larger than an epitope tag, it folds into a compact barrel-shaped structure and studies indicate that the functionality of

the protein to which it is attached is retained in the majority of cases. In addition, GFP-tagging allows the protein under study to be visualised in real time in living cells, potentially allowing more information to be retrieved from the functional assay.

Having generated a comprehensive set of appropriately tagged ORFs encoding proteins to be assayed, it may be necessary to first choose appropriate subsets of proteins for screening. There are a number of possibilities here, again depending on the assay in question. For example, since bioinformatic analysis is available for every predicted ORF, the assay could be designed around a subset of proteins that contain a particular ►[protein domain](#). Work by Cesareni and colleagues provides such useful information for the SH3 protein recognition domain (2). Their analysis has highlighted the presence of over 1500 SH3 domain-containing proteins, with functions in totally diverse processes. They have further suggested a sub-classification of these domains, thereby providing a potential starting point from which to screen a realistic number of proteins in any functional assay of interest. Such a strategy is potentially applicable to proteins containing any domain, such as those annotated in the ►[SMART databases](#) and ►[PFAM databases](#).

Another possibility for identifying an interesting subset of proteins for a functional assay is to screen and classify them according to their subcellular localisation. Indeed, researchers are making significant efforts to determine this information for every protein. Most recently, this has been achieved on a genome-wide scale in the yeast *Saccharomyces cerevisiae* (3) and it is inevitable that this will be extended to mammalian systems. Numerous projects have localised the products of hundreds of human ORFs and this information is already proving useful in determining subsequent strategies for characterising the proteins (4). Pre-screening proteins according to their localisation undoubtedly helps to make any functional assay more efficient. For example, if one were interested in identifying all the proteins involved in the ►[endocytosis](#) of nutrients from outside the cell, it makes more sense to assay those proteins that localise to the plasma membrane than those that localise to the nucleus. As more localisation projects come to fruition, functional assays will become increasingly focussed according to these results, in such a way that candidate proteins for the process of interest can be more rapidly identified. A functional assay in which individual, tagged proteins are over-expressed should not only provide clear, interpretable results, but it should also be possible to distinguish the results obtained from individual cells. Such a readout on a cell-by-cell basis is crucial if one is using transfection techniques to over-express a protein of interest. In this scenario, not all the cells in a well or dish will be transfected and correspondingly will not all

display a phenotypic effect in the assay. Therefore, taking a measurement from the entire well or dish will result in an erroneously high 'background' reading from the non-transfected cells, which could mask the phenotype of the cells that have been transfected. In order to make the data acquisition more reliable, it is therefore preferable to take measurements from individual cells. This has the advantage that the results will be more accurate and furthermore, allows the direct comparison of transfected and non-transfected cells under the same conditions. Microscopy-based data acquisition systems are ideally suited to this use because the two cell types can be easily distinguished and quantitative measurements of fluorescence can be made. Again, if we take the example of a functional assay designed to assess the endocytosis of a ligand into a cell, the protein under test could be tagged with GFP and the ligand, such as transferrin, could be labelled with a red fluorophore such as rhodamine. On addition to the culture medium, the labelled transferrin would be internalised by the cells and over a period of time degraded. Such an assay could therefore provide information about the uptake and degradation (loss of rhodamine fluorescence) of the ligand at various time points and allow a comparison of these measurements between transfected and non-transfected cells to be made.

Such cell-based functional assays can therefore be envisaged for any cellular activity of interest and could be applied to any fraction of the proteome. However, in order to do this for a large set of molecules, a certain degree of automation is required. Such automation can be at the level of sample preparation – namely the use of a pipetting robot – but it can also be at the level of data acquisition and analysis. For example, we have recently developed a screening platform, the central component of which is an automated microscope that is capable of acquiring images from multiple colour channels and from a variety of plate formats, including 96-well plates (5). As this system also contains an autofocus feature, an entire 96-well plate can be assayed without manual intervention. To demonstrate this platform, a functional assay to assess the secretion of a transmembrane marker protein was carried out. Such systems will inevitably become more widespread and more sophisticated, so that functional assays will become possible on a proteome-wide scale.

The approaches described here largely relate to experiments that utilise protein over-expression as a means to screen for function. However, as a result of the relatively recent discovery and application of **small interfering RNAs** (siRNAs) to effectively down-regulate proteins in mammalian cell systems, functional assays can now be used to screen cells that are depleted in specific proteins. Furthermore, if fluorescently labelled siRNA molecules are used, it is

again relatively easy to adapt the screening platform to specifically detect those cells that received the oligonucleotides. Although not all siRNA oligonucleotides are effective at down-regulating their target protein, combinations of siRNAs have been shown to have a greater efficacy and may ultimately provide the possibility of genome-wide knockdown screens.

Finally, it is worth considering the practicalities of routinely performing functional assays on the entire proteome of an organism. In order to undertake such a task, it is necessary to go beyond assays in plates and instead carry them out in a microarray format. ORFs and siRNAs can be spotted on to glass slides, then overlaid with cells that subsequently take up the nucleic acids in the presence of appropriate transfection reagent (6). The resulting highly localised transfected areas within a lawn of cells mean that functional assays on hundreds of proteins can be carried out on a single glass slide. Furthermore, since such nucleic acid spotting requires only small amounts of material to create many identical arrays, the assays should be highly reproducible. Arrays representing the human genome are already commercially available, and so it is only a matter of time before this is also the case for arrays of tagged ORFs, which are more suited for functional assays.

Clinical Relevance

Currently the application of functional assays to biological problems remains in the research laboratory. This is primarily because at this time most of the proteome is still uncharacterised and also because the development of robust large-scale functional assays is still in its infancy. However, as more diseases become associated with specific protein dysfunction, the need for established functional assays to assist clinical work will increase. In the field of intracellular protein transport, which encompasses both secretion and endocytosis, a whole variety of diseases can now be attributed to defects in a variety of transport machinery molecules (7). Assays that address such fundamental biological processes should prove essential to reconcile the results from proteomic projects with further understanding of the molecular biology of disease. Ultimately this is the purpose of primary biomedical research.

References

1. Strausberg RL, Feingold EA, Grouse LH et al (2002) Generation and initial analysis of more than 15,000 full-length human and mouse cDNA sequences. *Proc Nat Acad Sci USA* 99:16899–16903
2. Cesareni G, Panni S, Nardelli G et al (2002) Can we infer peptide recognition specificity mediated by SH3 domains? *FEBS Lett* 513:38–44
3. Huh WK, Falvo JV, Gerke LC et al (2003) Global analysis of protein localization in budding yeast. *Nature* 425:686–691

4. Simpson JC, Wellenreuther R, Poustka A et al (2000) Systematic subcellular localisation of novel proteins identified by large-scale cDNA sequencing. *EMBO Rep* 1:287–292
5. Liebel U, Starkuviene V, Erfle H et al (2003) A microscope-based screening platform for large scale functional analysis in intact cells. *FEBS Lett* 554:394–398
6. Ziauddin J, Sabatini DM (2001) Microarrays of cells expressing defined cDNAs. *Nature* 411:107–110
7. Aridor M, Hannan LA (2002) Traffic jams II: an update of diseases on intracellular transport. *Traffic* 3:781–790

Functional Group

Definition

Functional group refers to a defined arrangement of atoms that recursively occurs in molecules and stands for particular properties of these molecules. For example, carbonic acids possess a carboxylate group as functional group ($-\text{COOH}$) that can release a proton and form polar or hydrogen-bonding interactions.

► [Structure-Based Drug Design](#)

Functional Imaging

BURKHARD WIESNER

Research Institute for Molecular Pharmacology, Berlin, Germany

wiesner@fmp-berlin.de

Definition

Functional imaging represents a range of microscopic techniques in which the aim is to extract quantitative information about structure and physiological function from image based fluorescence data. Using ► [confocal laser scanning microscopy](#) (CLSM) the dynamics of molecular mobility, protein protein interactions, processes of signal transduction and structural characteristics of cell surfaces can be examined.

Description

Use Of Caged Compounds

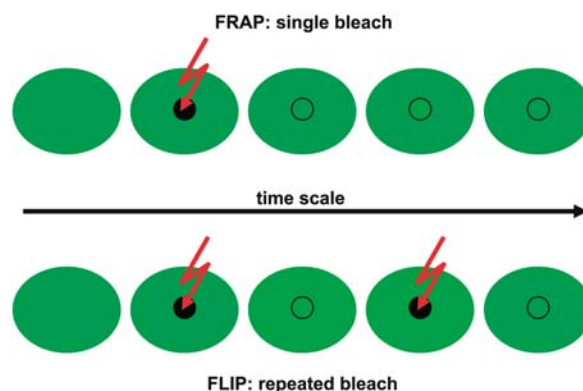
Caged compounds are an elegant means of producing rapid jumps in the concentration of chemical messenger molecules inside cells. These compounds are photolabile, inactive derivatives of biologically active molecules, whose biological activity has been disabled by a photoprotecting or caging group. Caged compounds can be applied to cells under steady-state conditions without evoking biological responses. The biologically active substance is rapidly released by a

photochemical reaction. Flash photolysis using near-UV light cleaves the modifying group and rapidly generates the biologically active molecule (1). For many applications, it is useful to determine quantitatively the degree of photolysis of the caged compound inside cells. In a few cases this has been possible by rather complicated calibration of the cellular reaction. With novel caged compounds this quantitative determination is feasible by measurements of fluorescence (2).

Fluorescence Photobleaching Techniques

Fluorescence recovery after photobleaching (FRAP) and the related FLIP technique (fluorescence loss in photobleaching) are popular methods that utilize changes in the recovery of fluorescence after local bleaching events to measure the dynamics of 2D or 3D molecular mobility, e.g. diffusion or transport of fluorescently labeled molecules in membranes or inside living cells (3). Fig. 1 shows the schematic of both procedures. FRAP enables the measurement of the relative mobility and the fraction of mobile fluorophores. With FLIP it is possible to measure the continuity and transport between two populations of fluorophores. Fig. 2 shows the relative fluorescence intensity within the bleach region plotted as a function of time. Some methods calculate the effective diffusion coefficient D_{eff} directly from the time ($t_{1/2}$) to reach half final intensity I_H . But to calculate D_{eff} as accurately as possible, I must be corrected for the background intensity and the amount of total fluorescence removed by the bleach.

Very recently, another FRAP related method was described (4) and named FLAP (fluorescence localization after photobleaching). In this type of experiment, the molecule of interest is labeled with two fluorophores, e.g. CFP and YFP, one to be bleached and the other to act as the reference label. This combination permits the monitoring of the localization of the bleached molecules themselves as the FLAP signal is



Functional Imaging. Figure 1 Scheme of the techniques FRAP (top) and FLIP (bottom).

Functional Genomics, the Systematic Analysis of the Function of All Genes and Gene Products in Parallel

H. LEHRACH

Max Planck Institute for Molecular Genetics, Berlin, Germany

lehrach@molgen.mpg.de

Definition

Functional genomics is the systematic study of gene products and their functions in biological processes. Functional genomics deals with gene knockouts and their phenotypic characterization, gene expression profiles (mRNA or protein levels) in different tissues of an organism, the sub-cellular location of proteins, three-dimensional protein structure, post-translational modification and processing of proteins, protein-protein interactions, and many more functional aspects of gene products that can be assayed systematically and at high throughput.

► **Structural Genomics: Structure-to-Function Approaches**

Introduction

In a sense, life can be considered a ‘computational’ process: The organism ‘computes’ its phenotype from the information in its genome, modulated by the environment as well as by some random components. This ‘computation’ is carried out by a complex network of processes, involving many of the genes and gene products of the organism. These networks have evolved over billions of years, translating the information in the genome into the phenotype. Selection of phenotypes therefore selects the genome, driving evolution.

Diseases are disturbances in these networks, usually caused by some combination of genetic and environmental factors. Understanding the disease mechanism will therefore require an understanding of both the underlying network in its normal, healthy state and the exact disturbance leading to the disease. To understand such complex processes, we have, in the past, tried to subdivide the processes into individual segments, which could then be analyzed by small labs working in (relative) isolation. This approach has clearly been very successful in identifying many of the basic biological mechanisms, but has given us less than sparkling progress in solving other areas of major importance to us all. Little progress has, for example,

been made in improving the rates of cure of many common cancers.

The Limits of Classical Biology

This slow progress in many important areas could very well be due to very basic limitations in our current approaches to understanding biology. It is, for example, not proven (and in my view not even likely), that all important problems in biology can be effectively approached by the divide and conquer strategy we have been using up to now. Biological processes, like neural networks, have evolved without any regard to the ease with which they can be subdivided and therefore effectively studied by our current strategies. Neither neuronal networks, nor many problems in mathematics (e.g. the traveling salesman problem), can be effectively subdivided in this way. Are we reaching the limits of our current strategy (as well as the organizational structure science has developed to implement this strategy)?

Functional Genomics, a Major Step Forward

In functional genomics, the systematic analysis of the functions and interactions of the genes and gene products encoded by the genome, we do have a key new tool in hand to put biological research on a new basis. Functional genomics is however also a key to progress in many of the real world problems we are facing. Like the situation in an analysis of a neural network, we will probably have to identify and analyze essentially all components of the complex processes we are investigating, rather than being able to concentrate on a few ‘interesting genes’ supposed to carry out the key steps of a process.

To try to understand the functions of all the genes of the organism, we have to be able to analyze the entire flow of information from the gene to the effect changes in the gene have on the phenotype of cells and ultimately of the organism, covering essentially all the types of analysis that have been traditionally used to try to understand the function of a single gene in classical (molecular) biology.

Since the function of a gene cannot be separated from the analysis of its expression pattern (transcriptomics) or the analysis of the structure and function of its protein product (proteomics) or even the analysis of the metabolic changes caused by e.g. the inactivation of a gene (metabolomics), these (and other -omics approaches) are all legitimate parts of functional genomics (proteomics is however covered in a separate article in this book and will therefore not be treated in this review).

Such an identification and analysis of thousands or tens of thousands of components (genes, gene products, metabolites etc.) can be carried out most easily by

systematic, highly automated procedures, incorporating mechanisms to control the processes carried out, to track samples and results and to eliminate errors as much as possible. Since functional genomics relies much more on systematic, automated data production, it is in many respects more efficient than the normal, manual strategy of producing data in biological research. Systematic data generation makes it much easier to standardize conditions and materials used and, since automated procedures are typically easier to control, data will often be of higher quality and will generally be more easily accessible, since they can be more easily submitted to the corresponding databases. The function of each gene can however be defined on many (at least three) levels. It can for example be defined in molecular terms (e.g. protein kinase), in the context of the cell (e.g. cell-cycle regulator) or in the context of the entire organism (e.g. recessive embryonic lethal). It is therefore very difficult to define the function of a gene in a straightforward fashion. To make the situation even more complex, it is highly likely that each gene has multiple functions on each of these levels, making it even more difficult to assign a clear function for each gene. In a sense, we might therefore most easily describe the function of any gene implicitly by computer models (of a complexity approaching the complexity of the organism itself (systems biology) of the processes in which it (and its gene products) are involved. The ultimate goal of functional genomics can therefore be defined as being able to compute the phenotype of organisms from their genomic sequence (as well their environment). In this sense, functional genomics will be a core research activity for many decades (if not centuries). Only after this goal has been achieved we can really argue that we have entered the 'post genomic' phase of biological research.

Genomic Tools

The primary information about the genes in an organism is provided by the sequence of its genome. If this information is not available, the analysis of expressed sequences can provide information on the most abundantly expressed genes. Gene identification can be carried out by a number of techniques.

The first phase of this work, the determination of the genomic sequence of many of the key organisms, provides an essential background to any of this work. The sequence of the human genome (1), as well as of the genomes of many other organisms including mouse (2), rat (3), chicken, *Fugu* (4), *Ciona* (5), *C. elegans* (6), *Drosophila* (7), *Arabidopsis* (8), *S. cerevisiae* (9) and *S. pombe* (10) to name a few (see also ►<http://www.ensembl.org/> or ►<http://genome.ucsc.edu/>), have been determined. In addition, a

number of other systematic analyses have been carried out at the transcript level [German cDNA sequencing consortium (11), FANTOM encyclopedia of mouse genes (12), IMAGE consortium (13), TIGR gene indices (14)]. These analyses have given us information on many of the basic components (genes, transcripts, proteins) of a large number of organisms.

Gene Predictions

To identify the functions of all genes, we first have to be able to identify them within the sequence of the genome. This problem is fairly straightforward for a subset of the genes, namely those coding for proteins. This is partly due to the improvements in informatics tools available, with Genscan being probably the most widely used program (15). Basically, all prediction methods are combinatorial approaches based on, e.g. splice site detection, multimer and motif statistics etc. together with artificial intelligence (AI) techniques like neural networks or HMMs.

Important roles are however also played by the results of sequence comparisons and the large amount of information from the analyses of expressed sequence tags (ESTs) available for a large number of organisms. Based on different types of information, our estimates for the number of protein coding genes in the human genome has continued to decrease, from initial estimates of 100,000 or higher to current estimates of between 20,000 and 25,000. Information on the current best estimates of the identity and structure of these genes is available for many organisms with completely sequenced genomes from the ENSEMBL project (www.ensembl.org) and the UCSC bioinformatics center (►<http://genome.ucsc.edu>), both based on an automated analysis pipeline. For selected organisms (*D. melanogaster*, *C. elegans*, yeast etc.) species specific databases, which tend to be hand curated, are generating additional information (www.flybase.org, www.wormbase.org), which in turn enters the automated annotation systems.

Recently, a very interesting experimental approach to gene prediction was introduced. A high-resolution oligo chip covering specific regions of the human genome was shown to be able to detect many novel loci with potential transcriptional activity (16).

Evolutionary Analysis

One of the most important sources of information on the functional importance of specific segments of a gene or of a specific element in the genome is provided by its evolution. Regions that are functionally important tend to be more conserved than those that have no clear function. On the other hand, genes that are under strong positive selection can be evolving faster than regions without any discernable function. In

protein coding genes, this can, for example, be recognized by comparing the rate of evolution of nucleotide positions that do not lead to a change in the amino acid of the protein (e.g. many changes in the third base of a codon), kS, with the rate of evolution of those that do change the amino acid, kN, of the encoded protein (kS/kN). Especially for closely related species like human and chimp, a detailed analysis by genome-genome comparison can deliver deep insights into the evolutionary dynamics (17), e.g. the unexpectedly high mutation rates in 5' UTRs of protein coding genes in humans compared to the chimp. Evolutionary analyses are getting more straightforward, as more and more genomes are completely sequenced and analyzed and can provide insights into the mechanisms of evolution by comparisons even between distant genomes (e.g. *C. elegans*, *D. melanogaster*, and vertebrates like mouse, man, etc.). Based on the genome and transcript/protein sequences of several invertebrate and vertebrate species, two independent studies (18,19) recently provided new arguments for a genome-wide duplication event at the transition from invertebrates to vertebrates, which could be dated to around 600 million years ago (under the assumption of a molecular clock model).

Identification and Analysis of Transcripts

Transcript Identification

A major effort has been invested world wide in cDNA analysis projects in order to gain an overview of a major fraction of the genes of an organism, to analyze the structure of the transcript or transcripts produced from each gene and to gain some insights into the abundance of transcripts in different tissues and at different stages of development. ► **Expressed sequence tags** (ESTs) have been generated for many organisms, many tissues and many stages of development. For instance, there are 149 species for which EST projects have been initiated, for which dbEST, the data base of expressed sequence tags (► www.ncbi.nlm.nih.gov/dbEST), contains over 10,000 entries per species (Release 092404; Sept. 24th, 2004). The six top species (number of ESTs in brackets) are human (5.9 Mio), mouse (4.2 Mio), *Ciona intestinalis* and rat (684000 each), zebrafish (575000) and wheat (561000). To shortcut the procedure of generating comprehensive gene catalogues for selected species, alternative techniques have been developed and used in an attempt to generate sufficient sequence information to recognize each transcript at lower cost than *via* the sequencing strategy.

As an example, oligonucleotide fingerprinting has been used in a number of projects to identify new genes as a powerful means to reduce complexity prior to sequencing (20), e.g. in *Amphioxus* (21), sea

urchin (22), zebrafish (23), mouse, medaka and sugar beet (24). Since the information generated covers the entire transcript, this procedure should, in principle, also be better at identifying specific splice forms than the EST strategy.

Other procedures have attempted to identify short sequence elements (sequence tags) within each transcript in order to be able to quantify the abundance of transcripts in different materials. Examples of such approaches are ► **serial analysis of gene expression** (SAGE) (25) and massively parallel signature sequencing technology (MPSS) that involves microbead arrays (26).

Alternative Splice Products

Alternative splicing contributes significantly to human proteome complexity and allows large proteomic complexity from a limited number of genes. It explains to some extent the difference between the low number of human protein-coding genes and the number of human proteins. Annotation of the human genome shows that alternative splicing is a major cause of transcriptome diversification.

The bulk of intron-containing transcripts are alternatively spliced. A single primary transcript yields different mature RNAs leading to the production of proteins with diverse functions.

More than 45% of the genes in the human genome have alternative transcripts ranging from two to 40. 50% of these genes have alternative promoters. The human genome is currently annotated with 23,245 gene loci (NCBI Build 34). For these loci over 43,000 transcripts are known. Despite an early report that most alternative splices occur within the 5'untranslated region, recent studies indicate that 70–88% of alternative splices change the protein product. The majority of these changes appear to be functionally interesting, such as replacement of the amino or carboxy terminus, or in-frame addition and removal of a functional unit.

Bioinformatics studies have reported that about 25% of genes have alternative polyadenylation forms (mRNAs that are cleaved and polyadenylated at different sites). In addition 6,418 of the annotated loci have two or more promoters. These numbers show that most of the genes are highly flexible entities with regard to transcripts and regulation. It has been estimated, that 15% of the point mutations that cause human genetic disease affect splicing (27). Alternatively spliced exons are typically identified by aligning EST clusters to reference mRNAs or genomic DNA. A high rate of false negatives is the greatest disadvantage of methods that require mapping ESTs to the genome sequence. This approach is not useful for genomes that lack robust EST coverage.

Non-messenger RNAs

For a long time non-messenger RNAs (nmRNA) were regarded as a somewhat exotic class of RNAs with very few members known. The most prominent ones were structural RNAs (tRNA, ribosomal RNA, snRNA) and a handful of classic examples like Xist (28) (X inactivation) or H19 (29) (tumor suppressor RNA). Only recently has the detection of the micro-RNAs and small interfering RNAs (siRNAs) (30) produced remarkable scientific and technical interest in the fundamental roles of nmRNAs, which is, for example, reflected in the repeated election of small nmRNAs as “Molecule of the Year” or runner-up in Science Magazine (31).

In the last few years, several large-scale studies gave rise to the question as to whether the number of nmRNAs active in regulation, suppression and modification of mRNAs and their protein products hadn't been notoriously underestimated. Recently, Affymetrix scientists (16) tested the transcriptional activity in the human genome. A novel high-resolution chip covering the entire human chromosomes 21 and 22 with evenly spaced oligonucleotides revealed a ~10-fold higher proportion of non-genic transcribed regions, many of which could be verified by subsequent experimentation. Analysis of the mouse FANTOM cDNA collection published by members of the RIKEN consortium resulted in the surprising detection of at least several thousand large RNAs that exhibit features of mRNAs, but do not contain any significant open reading frames (32).

It may be speculated that the increased complexity of higher vertebrates as compared to invertebrates (e.g. *C. elegans*) cannot be explained by a mere 1.5-fold increase in protein-coding genes, but might be based on a complex network of non-protein coding sequences (33).

Analysis of Transcript Levels

Sequence identifiers are used to analyze transcript abundance in ‘electronic Northern’. As data continue to increase, the reliability of the information is similarly increased. A significant difficulty in establishing these ‘electronic Northern’ is, however, the accurate description of the material used to make the library, the need to eliminate normalized libraries from the analysis and the complex relationships between the different descriptions.

The use of DNA or oligonucleotide arrays to probe the abundance of transcripts has constituted a revolution, since it allows the large-scale analysis of the transcriptome level of tens of thousands of genes in parallel within one single hybridization experiment. Originally based on the DNA and clone array format developed by us in 1987 (22, 34, 35), it has

been extended later to a number of other formats. The main differences in the formats arise from the nature of the DNA probes, the labeling procedures and the material of the solid array surface. A particularly useful variant has been oligonucleotide chips (36, 37, 38) with (typically short) oligonucleotides either synthesized *in situ* by a variety of techniques or (longer oligonucleotides, e.g. 70 mers) spotted after synthesis. A number of successful commercial formats are available (for example Affymetrix, Agilent). Furthermore, more and more chromosomes are mapped to so-called tiling arrays that represent the entire (non-repetitive) genomic sequence of the chromosome by oligonucleotides, which has contributed additional information on transcripts outside the canonical gene set (16, 39).

For all these analyses significant amounts of material (typically in the range of micrograms of total RNA) have to be generated. The development of amplification procedures has contributed to a decrease in the requirement for material and has, for example, allowed the analysis of material microdissected from tissue sections. This application is particularly important in the analysis of cancer tissue containing many different cell types (40). Though these types of microdissection and similar procedures have contributed to increasing the anatomical resolution achieved, this is nevertheless the weak point in most of the techniques that analyze the expression levels of many genes in parallel.

In situ hybridization techniques, in contrast, are typically only able to analyze the expression level of a single gene at a time, but can achieve very high anatomical resolution. They are therefore much better in the identification of spatial expression patterns. Such analyses have, for example, been carried out systematically in the mouse using the whole mount *in-situ* hybridization technique (41, 42). Even higher anatomical resolution, down to the single cell level, can be achieved using *in situ* hybridization to frozen sections (43).

A weak point in all these analytical techniques is still the identification of the expression patterns of different splice forms, since this type of hybridization will typically score the overall abundance of different exons, which can, however, be parts of alternatively spliced transcripts. For this, additional techniques will have to be developed.

The chip-based analyses are still error prone and the correlation between different platforms is rather low. Thus, they require verification with independent techniques at the single gene level, mostly performed by RT-PCR or Taqman analysis.

Other techniques, which have been quite popular for some time, are now being used less often. An example

of this is the technique of differential display, based on the selective amplification of sequences by short oligonucleotides. Similarly, differential cloning is now being used rarely in those systems for which an abundance of information is available (man, mouse, rat). It is, however, still useful for organisms for which gene sets are not well defined and genomic sequence is not available.

Proteomics

Since the function of many genes is to code for proteins, proteomics, the systematic analysis of the structure and function of proteins, is an essential component of the functional genomics of protein coding genes. This topic is covered in detail in [►Proteomics – from Proteins to Disease Mechanisms and the Development of Novel Therapeutic Strategies](#).

Analysis of Gene and Protein Function in Cells

A number of aspects of gene (and protein) function can be most easily studied in cell systems. Examples are localization in the cell, the effect of protein over-expression or elimination of a protein by [►RNA interference](#) (RNAi) on a cell phenotype and the analysis of protein-protein interactions in mammalian cells. The analysis can be carried out on the microtiter plate scale, using automated equipment to handle the large number of plates that have to be analyzed in any realistic experiment. Some types of analysis can also be carried out on a smaller scale (cell chips), considerably reducing the effort involved in the analysis of many samples in parallel.

Protein Localization

The intracellular localization of the gene product can provide important information about its function, due to the high degree of compartmentalization of the cell. Most protein activities can be assigned to particular cellular compartments. Moreover the interactions of the proteins within regulatory networks rely to a large extent on proper localization of their components. Together with other functional data such as the sequence of the gene, its expression profile and the expression, modification and interaction of its product with other proteins, information about the distribution of the protein within the cells contributes significantly to integrative functional characterization of the gene. Until recently protein localization studies have been limited to a particular gene of interest. Accomplishment of the mouse and human genome project resulted in a vast number of sequences of unknown function. As a result the demand for a high throughput functional genomics solution has been growing. Hitherto functional analyses of genes have been performed on an individual basis, thus creating a

bottleneck for analysis of gene expression data. Recently several attempts at genome-wide functional analysis have been made in budding yeast (44, 45) and, to a lesser extent, man (46).

In addition to micro-well based assays, transfected cell arrays offer a robust platform for protein localization studies, allowing the simultaneous screening of a large number of novel genes, thereby reducing the amount of reagents (antibodies and dyes for counterstaining for particular organelles) and cells per assay (see section Cell Arrays below).

RNAi, Morpholinos, Antisense Constructs

RNA interference (RNAi) is the post-transcriptional silencing of gene expression that occurs in response to the introduction of double stranded RNA into cells. Application of RNAi in experimental systems has provided a quantum leap in the elucidation of gene functions.

In this approach, short, double stranded RNA molecules are used to trigger a cellular response, leading to the destruction of the corresponding transcripts. For short lived proteins and for rapidly dividing cells, this will lead to a reduction in the concentration of the corresponding protein, reaching, in some cases, an almost complete elimination. A number of strategies have been developed for use in mammalian cells, using short chemically synthesized oligonucleotides, short double stranded RNA products generated *in vitro* from longer ds RNA generated from appropriate DNA constructs *in vitro* or vector systems able to produce short double stranded RNAs *in vivo*.

To facilitate large-scale functional genomics studies using RNAi, several high throughput approaches have been developed based on microarray or micro-well assays. The recent establishment of large libraries of RNAi reagents combined with a variety of detection assays further opened the door for genome-wide screens of gene function in mammalian cells. For many systems this new strategy has replaced similar strategies developed earlier (the use of morpholino containing oligonucleotides to block translation, the use of antisense oligos or RNAs).

Protein Over-expression

The influence of specific genes on cellular phenotypes can be most easily studied by either reducing or eliminating the protein (e.g. by RNAi) or by increasing its level (e.g. by transfecting the coding sequence under the control of a strong promoter). The effects of over-expressing a gene on the phenotype of a cell can be analyzed by a range of cell biological assays. Changes in cellular localization patterns as determined by immunofluorescence techniques or the

effect of each gene on, for example, cell-cycle or apoptosis can be determined in micro-well based assays or in high throughput assays such as transfected cell arrays.

Cell Arrays

A recently described, novel cell-based microarray system, called the transfected cell array (TCA) (47) paved a way for high throughput gene analysis in the field of functional genomics. The principle of the TCA technique is based on the transfection of DNA or RNA molecules immobilized on a solid surface into mammalian cells with subsequent detection of the physiological effects caused by the introduction of the foreign nucleic acid in these cells. Briefly, full-length open reading frames of genes inserted in expression vectors are printed at a high density on a glass slide along with a lipid transfection reagent using a robotic arrayer. When the microarray of DNA constructs is covered with a monolayer of adherent cells, only the cells growing on top of the DNA spots become transfected, resulting in the expression of specific proteins in spatially distinct groups of cells. The phenotypic effects of this reverse transfection of hundreds or thousands of genes can be detected using specific cell-based bioassays. Cell arrays can be used for the large-scale analysis of protein-protein interactions (mammalian two-hybrid), gene silencing (siRNA, double stranded "decoy" oligonucleotides, morpholinos), promoter activity and the subcellular localization of proteins.

Analysis of cell arrays is restricted to those phenotypes that can be easily scored under a microscope and can be analyzed effectively in spite of the fact that typically only a fraction of cells take up the DNA or siRNA (fluorescent assays, other easily detectable phenotypes like apoptosis etc). However there are a number of advantages in using cell arrays as compared to functional assays performed on mammalian cells in micro-well plate format. The array approach requires fewer cells per number of genes tested. This is especially relevant for human primary cells (cells recently isolated from organs or tissues) since only small numbers of these cells can be isolated out of a tissue and the *in vitro* expansion of these cells is rather limited. Although primary cells are notorious for their low capacity to be transfected, reverse transfection of these cells is possible, but it requires careful optimization of the experimental conditions for each cell type tested (48). Transfected cell arrays also require far less DNA/RNA as well as transfection and signal development reagents as compared to assays performed in microwell plate format. Since in the case of high throughput technologies the cost of the single sample analysis must be reduced to an

absolute minimum, cell arrays are at the moment the most effective functional genomic tool available.

Cell arrays (47) offer a particularly efficient route to analyzing the effects of many genes on specific phenotypes of the cell. In this approach, DNA constructs, siRNAs or other effectors are spotted on glass arrays in some matrix (e.g. gelatin plus lipofection reagents). The chips are then covered with appropriate indicator cells and the effect of the reagent on the phenotype of the cell is observed. Since thousands of constructs can be analyzed on a single chip using this approach, this procedure offers a chance to analyze the effects of most or all of the genes of the genome on a particular phenotype. The analysis is however restricted to a number of phenotypes that can be scored easily under a microscope and can be scored effectively in spite of the fact that typically only a fraction of the cells take up the DNA or siRNA (fluorescent assays, other easily scoreable phenotypes like apoptosis etc). Other phenotypes will therefore have to be converted into phenotypes scoreable in this format, e.g. by introducing an appropriate indicator construct that translates the primary phenotype into a fluorescence signal.

The transfected cell array can be regarded as a sophisticated protein expression system where location of particular proteins on the slide can be identified by the coordinates of the arrayed cDNA. Cell arrays can, for example, be used for identification of disease related proteins by analyzing the binding of patient sera on arrays of expressed cDNA libraries, leading to identification of molecular diagnostic markers. Furthermore, the cell arrays are also well suited for screening of small molecules such as G-protein-coupled receptors for their ability to interact specifically with protein targets. Again, cell microarrays present a more advanced alternative to standard protein arrays, since in TCA the proteins are synthesized *in situ* in the physiological environment of the cells, thus enabling proper post-translational protein folding and glycosylation. Furthermore, cell arrays represent a powerful functional genomics tool for target gene identification. Through the accurate selection of cell line properties and signal detection assays, usage of cell arrays on a genome-wide scale can rapidly lead to identification of target genes related to particular diseases.

Analysis of Gene Function in the Organism

Ultimately the function of each gene has also to be studied in the organism itself, work that is typically being carried out in model organisms. The mouse especially has been an incredibly effective system for this type of analysis, due to the genetic resources and the highly developed tools to manipulate the mouse

genome. We will particularly focus here on approaches to identifying the phenotype associated with a specific gene or a specific set of genes (gene driven approaches). In addition, there are many types of analysis that start from a phenotype, aiming to identify the gene or genes involved (phenotype driven approaches). While in man for obvious reasons only phenotype driven analyses are possible, model organisms typically allow both gene driven and phenotype driven approaches.

Transgenic Animals

For a number of different model organisms (mouse, rat, zebrafish, ►*C. elegans*, ►*D. melanogaster*) genes can be efficiently introduced into the germ line, allowing detailed analysis of the function of a single gene. Use of tissue specific or inducible promoters allows the analysis of gene function in only a single tissue or after addition of an appropriate inducer (or withdrawal of an inhibitor).

Homologous Recombination

For the analysis of the function of a single gene, homologous recombination offers the possibility of introducing very specific changes in the genome. This technique works exceedingly well for a number of lower eukaryotes (e.g. *S. cerevisiae*, *Physcomitrella*), but has also become the main tool in analyzing gene function in the mouse. Mutations are introduced by homologous recombination into the genome of an embryonic stem (ES) cell and then introduced into the germ line (49,50), allowing detailed functional analyses by standard mouse genetic techniques. Changes introduced can range from the exchange of a single base pair to the introduction of large-scale rearrangements or deletions in the genome.

Gene Traps

A high throughput alternative for the analysis of gene function in the ES cell or the organism is provided by the gene trap system, based on the random insertion of DNA sequences into the genome of a mouse ES cell (51). The gene trap vector is designed so that a reporter/selection cassette (b-Geo) can be spliced onto an endogenous gene transcript if the construct inserts into an intron of an actively transcribed gene in the appropriate orientation. An internal ribosome entry site (IRES) allows expression of the reporter/selection cassette independently of the reading frame of the endogenous transcript. This comparatively rare event can therefore be selected (for neo/G418 resistance) and analyzed further to determine in which gene and in which position of this gene the insertion has taken place. Insertions can either be obtained by transfection of DNA fragments or by the use of appropriate

retroviral constructs. A number of efforts to construct complete gene trap libraries have been performed and have resulted in a large library of ES cell lines carrying insertions in defined genes (52). Several additional modifications of this principle have been designed (poly A trap, trap vectors enriching for secreted proteins, insertions generating conditional mutations etc.).

Chemical Mutagenesis, Tilling

An additional approach to generate mutations in the genome, particularly useful for organisms, for which no equivalent to the mouse ES cell system exists, takes advantage of the high mutagenic potential of some chemicals (e.g. ethylnitrosourea, ENU). After treatment of the male with high doses of ENU, a high frequency of mutations can be generated. In combination with high throughput screening systems able to identify point mutants or deletions within a specific gene, this can be used to identify a set of animals carrying mutations within the gene of interest or in a specific gene region (53). A related strategy has, for example, been used to generate the equivalent to a knockout mutation in the rat, in which the standard knockout procedure is not possible (54). Other successful examples are archives of small genomic deletions in *C. elegans* (55, 56), and phenotype-based mutagenesis screens in mouse (57, 58), zebrafish (59) and medaka (60). Chemical mutagenesis can also be applied to mouse ES cells (61, 62), holding great promise for the construction of allelic series within the genes of interest (63).

RNAi Constructs

RNAi has been used very successfully to degrade specific transcripts in a number of organisms. While the largest screens have been carried out in *C. elegans*, using a number of different approaches, an analogous strategy is increasingly being used in mouse ES cells or mouse embryos. In this case, specific vectors have to be constructed that are able to produce a primary transcript that can be processed to give the appropriate interfering RNA *in vivo*. Compared to most of the other systems (gene traps, knock-outs, ENU mutagenesis), this strategy has the enormous advantage of being able to produce dominant mutants, since a single copy of the construct will typically be able to inactivate the transcripts from both chromosomes. Most vector systems used contain a RNA polymerase III promoter driving the synthesis of short hairpin RNAs (shRNA) in mammalian cells. These are then processed intracellularly into siRNA-like molecules. ShRNAs turned out to be as effective as siRNA and moreover can exert their silencing effect over longer periods of time. Alternatively, two tandem pol III promoters have

been used to obtain sense and antisense siRNA sequences from expression constructs (convergent transcription). The silencing effect can be further extended using adenoviral, retroviral or lentiviral RNA expression vectors. Greater versatility and more refined applications will become possible with the development of cell- and tissue-specific RNAi expression constructs. Although most of the RNAi inducing systems were used for the specific silencing of a limited number of genes, large siRNA expression libraries in high throughput RNAi screens would allow for the identification of new genes involved in different cellular processes. In fact, efforts are underway to develop larger siRNA expression libraries and use of these libraries in cell-based RNAi screens is already generating important results.

Databases, Systems Biology

Databases

The collection and use of information from functional genomics approaches represents a major challenge to bioinformatics/computational biology. In contrast to sequence databases, which are restricted to a single data class, data from functional genomics experiments can be extremely complex and are typically only useful if a large amount of information on the experiment, the materials used, the exact experimental conditions etc is provided. In addition, many experiments are technically only feasible (or at least possible) in particular species. It is therefore essential to combine functional genomics information across many organisms, based on the orthology between the genes in the different species.

A number of relevant databases have been developed. Expression data, for example, have dedicated databases (ArrayExpress ►<http://www.ebi.ac.uk/arrayexpress/>, Gene Expression Omnibus ►<http://www.ncbi.nlm.nih.gov/geo/>, GenomeCubeTM Express ►<http://www.rzpd.de/asearch/>), in which expression data following the MIAME standard (Minimum Information About a Microarray Experiment, ►<http://www.mged.org>) can be deposited. There are a number of databases for protein interactions (Biomolecular Interaction Network Database (BIND) ►<http://www.bind.ca/>, Database of Interacting Proteins ►<http://dip.doe-mbi.ucla.edu/>, Grid ►<http://biodata.mshri.on.ca/grid/>, Protein-Protein-Interaction (PPI) ►<http://mips.gsf.de/proj/yeast/CYGD/interaction/>, Worm Interactome (64) ►<http://vidal.dfci.harvard.edu/>), protein structure (Protein Data Bank (PDB) ►<http://www.rcsb.org/pdb/>), gene traps (German Genetrap Consortium (GGTC) ►<http://tikus.gsf.de/>, Sanger Institute Gene Trap Resource (SIGTR) ►<http://www.sanger.ac.uk/PostGenomics/genetrap/>, BayGenomics ►<http://baygenomics.ucsf.edu/>, Soriano gene trap lines ►<http://www.fhcrc.org/labs/soriano/trap.html>)

and many other aspects of functional genomics. In addition to these databases centered around the type of information they store, there are many organism specific databases, (Mouse Genome Informatics (MGI) ►<http://www.informatics.jax.org/>, Rat Genome Database (RGD) ►<http://rgd.mcw.edu/>, WormBase ►<http://www.wormbase.org>, FlyBase ►<http://flybase.bio.indiana.edu>, the Zebrafish Information Network (ZFIN) ►<http://zfin.org>, etc.) in which all data on the genes of a specific organism are usually stored, typically in curated form. Access to this information is however typically focused on gene-by-gene queries and is therefore not optimal for the display of information on thousands of genes at a time, typical of functional genomics problems.

This has been the main driving force behind the development of GenomeMatrix (www.genome-matrix.org), a database/data base interface system developed by us. In this, information on the different genes in the different organisms is displayed in the form of a matrix of colored rectangles, with each column representing one gene (and its orthologs in the other organisms displayed), while each row corresponds to one particular type of information relevant to understanding the function of the genes.

As far as possible, the information on one particular data type for one particular gene is represented by the color of the box (usually possibly for scalar information). The color of the box can for example identify the function of the gene (GO annotation), represent a quantitative measure (normalized expression levels for a gene displayed on a red-green or yellow-blue scale) or simply indicate that there is information of one particular type for this particular gene available (e.g. an X ray structure of a protein, a picture showing a developmental phenotype of a mouse knockout strain, a RNAi phenotype in *C. elegans* etc.). Clicking on the boxes then displays the information available, stored either locally or in a database or website anywhere in the world. Since this system has a very low overhead, it is easy to adapt to changes in the many databases holding the primary information, a feature required for any realistic system attempting to keep track of the multitude of relevant data produced worldwide.

Systems Biology

Ultimately any display of information will be useful to identify unexpected correlations, formulate hypotheses and plan new experiments. It will, in itself, not be sufficient to be able to use the information to make quantitative predictions from the enormous amount of data available (in the future possibly even on every single patient). This is the domain of systems biology. Here we do not only have the possibility of representing our entire knowledge of biological

objects, their interactions, their function etc. in a formalized language, these systems also allow us to make quantitative predictions, which can be verified or refined by further experiments in an iterative process and which can be used in many practical applications to guide e.g. the treatment of individual patients. This approach offers the possibility of testing a hypothesis in the computer before expensive developments, e.g. the development of a new drug. In a sense, functional genomics, the systematic generation of data on the function of all the genes of an organism only covers the aspect of producing the information required. Systems biology is the component that is essential to make this information useful and to enable us to use functional genomics to solve the many problems, which affect our daily lives.

Acknowledgements

The concepts expressed in this review have been the result of many discussions with many people. I would particularly like to thank Bernhard Herrmann, Marie-Laure Yaspo, Steffen Hennig, Patricia Ruiz, Claudia Falter, Ralf Herwig, Heinz Himmelbauer, Marc Sultan, Zoltan Konthur, Dominique Vanhecke, Michal Janitz, Ilaria Piccini, Sylvia Krobisch, Günther Zehetner, Andreas Hewelt and Christoph Wierling for contributions and discussion.

References

1. Lander ES, Linton LM, Birren B et al (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860–921
2. Waterston RH, Lindblad-Toh K, Birney E et al (2002) Initial sequencing and comparative analysis of the mouse genome. *Nature* 420:520–562
3. Gibbs RA, Weinstock GM, Metzker ML et al (2004) Genome sequence of the Brown Norway rat yields insights into mammalian evolution. *Nature* 428:493–521
4. Aparicio S, Chapman J, Stupka E et al (2002) Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. *Science* 297:1301–1310
5. Dehal P, Satou Y, Campbell RK et al (2002) The draft genome of *Ciona intestinalis*: insights into chordate and vertebrate origins. *Science* 298:2157–2167
6. The *C. elegans* sequencing consortium (1998) Genome sequence of the nematode *C. elegans*: a platform for investigating biology. *Science* 282:2012–2018
7. Adams MD, Celniker SE, Holt RA et al (2000) The genome sequence of *Drosophila melanogaster*. *Science* 287:2185–2195
8. *Arabidopsis* Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796–815
9. Goffeau A, Barrell BG, Bussey H et al (1996) Life with 6000 Genes. *Science* 274:546, 563–567
10. Wood V, Gwilliam R, Rajandream MA et al (2002) The genome sequence of *Schizosaccharomyces pombe*. *Nature* 415:871–80
11. Wiemann S, Weil B, Wellenreuther R et al (2001) Toward a catalog of human genes and proteins: sequencing and analysis of 500 novel complete protein coding human cDNAs. *Genome Res* 11:422–435
12. FANTOM Consortium; RIKEN Genome Exploration Research Group Phase I & II Team (2002) Analysis of the mouse transcriptome based on functional annotation of 60,770 full-length cDNAs. *Nature* 420:563–573
13. Lennon G, Auffray C, Polymeropoulos M et al (1996) The I.M.A.G.E. Consortium: an integrated molecular analysis of genomes and their expression. *Genomics* 33:151–152
14. Quackenbush J, Liang F, Holt I et al (2000) The TIGR gene indices: reconstruction and representation of expressed gene sequences. *Nucleic Acids Res* 28:141–145
15. Burge C, Karlin S (1997) Prediction of complete gene structures in human genomic DNA. *J Mol Biol* 268:78–94
16. Kapranov P, Cawley SE, Drenkow J et al (2002) Large-scale transcriptional activity in chromosomes 21 and 22. *Science* 296:916–919
17. Hellmann I, Zollner S, Enard W et al (2003) Selection on human genes as revealed by comparisons to chimpanzee cDNA. *Genome Res* 13:831–837
18. Panopoulou G, Hennig S, Groth D et al (2003) New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal genomes. *Genome Res* 13:1056–1066
19. McLysaght A, Hokamp K, Wolfe KH (2002) Extensive genomic duplication during early chordate evolution. *Nat Genet* 31:200–204
20. Meier-Ewert S, Lange J, Gerst H et al (1998) Comparative gene expression profiling by oligonucleotide fingerprinting. *Nucleic Acids Res* 26:2216–2223
21. Panopoulou G, Hennig S, Groth D et al (2003) New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal genomes. *Genome Res* 13:1056–1066
22. Poustka AJ, Groth D, Hennig S et al (2003) Generation, annotation, evolutionary analysis, and database integration of 20,000 unique sea urchin EST clusters. *Genome Res* 13:2736–2746
23. Clark MD, Hennig S, Herwig R et al (2001) An oligonucleotide fingerprint normalized and expressed sequence tag characterized zebrafish cDNA library. *Genome Res* 11:1594–1602
24. Herwig R, Schulz B, Weisshaar B et al (2002) Construction of a ‘unigene’ cDNA clone set by oligonucleotide fingerprinting allows access to 25 000 potential sugar beet genes. *Plant J* 32:845–857
25. Velculescu VE, Zhang L, Vogelstein B et al (1995) Serial analysis of gene expression. *Science* 270:484–487
26. Brenner S, Johnson M, Bridgham J et al (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nat Biotechnol* 18:630–634
27. Modrek B, Lee C (2003) A genomic view of alternative splicing. *Nat Genet* 30:13–19

28. Brown CJ, Hendrich BD, Rupert JL et al (1992) The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. *Cell* 71:527–542
29. Juan V, Crain C, Wilson C (2000) Evidence for evolutionarily conserved secondary structure in the H19 tumor suppressor RNA. *Nucleic Acids Res* 28:1221–1227
30. Dykxhoorn DM, Novina CD, Sharp PA (2003) Killing the messenger: short RNAs that silence gene expression. *Nat Rev Mol Cell Biol* 4:457–467
31. Couzin J (2002) Breakthrough of the year. Small RNAs make big splash. *Science* 298:2296–2297
32. Numata K, Kanai A, Saito R et al (2003) Identification of putative noncoding RNAs among the RIKEN mouse full-length cDNA collection. *Genome Res* 13:1301–1306
33. Mattick JS, Gagen MJ (2001) The evolution of controlled multitasked gene networks: the role of introns and other noncoding RNAs in the development of complex organisms. *Mol Biol Evol* 18:1611–1630
34. H Lehrach, R Drmanac, J Hoheisel et al (1990) Hybridization fingerprinting in genome mapping and sequencing. *Genome Analysis 1: Genetic and Physical Mapping*, Cold Spring Harbor Laboratory Press, 39–81
35. Lennon GG, Lehrach H (1991) Hybridization analyses of arrayed cDNA libraries. *Trends Genet* 7:314–317
36. Lockhart DJ, Dong H, Byrne MC et al (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* 14:1675–1680
37. Wodicka L, Dong H, Mittmann M et al (1997) Genome-wide expression monitoring in *Saccharomyces cerevisiae*. *Nat Biotechnol* 15:1359–1367
38. Hughes TR, Mao M, Jones AR et al (2001) Expression profiling using microarrays fabricated by an ink-jet Oligonucleotide synthesizer. *Nat Biotechnol* 19:342–347
39. Kampa D, Cheng J, Kapranov P et al (2004) Novel RNAs identified from an in-depth analysis of the transcriptome of human chromosomes 21 and 22. *Genome Res* 14:331–342
40. Gillespie JW, Ahram M, Best CJ et al (2001) The role of tissue microdissection in cancer research. *Cancer J* 7:32–39
41. Neidhardt L, Gasca S, Wertz K et al (2000) Large-scale screen for genes controlling mammalian embryogenesis, using high-throughput gene expression analysis in mouse embryos. *Mech Dev* 98:77–94
42. Gitton Y, Dahmane N, Baik S et al (2002) HSA21 expression map initiative. A gene expression map of human chromosome 21 orthologues in the mouse. *Nature* 420:586–590
43. Carsona JP, Thaller C, Eichele G (2002) A transcriptome atlas of the mouse brain at cellular resolution. *Current Opinion in Neurobiology* 12:562–565
44. Kumar A, Harrison PM, Cheung KH et al (2002) An integrated approach for finding overlooked genes in yeast. *Nat Biotechnol* 20:58–63
45. Huh WK, Falvo JV, Gerke LC et al (2003) Global analysis of protein localization in budding yeast. *Nature* 425:686–691
46. Simpson JC, Wellenreuther R, Poustka A, Pepperkok R, Wiemann S. (2000) Systematic subcellular localization of novel proteins identified by large-scale cDNA sequencing. *EMBO Rep.* 1:287–92.
47. Ziauddin Sabatini (2001) Microarrays of cells expressing defined cDNAs. *Nature* 411:107–110
48. Yoshikawa T, Uchimura E, Kishi M et al (2004) Transfection microarray of human mesenchymal stem cells and on-chip siRNA gene knockdown. *J Control Release* 96:227–232
49. Doetschman T, Gregg RG, Maeda N et al (1987) Targetted correction of a mutant HPRT gene in mouse embryonic stem cells. *Nature* 330:576–578
50. Thomas KR, Capecchi MR (1987) Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell* 51:503–512
51. Gossler A, Joyner AL, Rossant J et al (1989) Mouse embryonic stem cells and reporter constructs to detect developmentally regulated genes. *Science* 244:463–465
52. Skarnes WC, von Melchner H, Wurst W et al (2004) A public gene trap resource for mouse functional genomics. *Nat Genet* 36:543–544
53. Till BJ, Reynolds SH, Greene EA et al (2003) Large-scale discovery of induced point mutations with high-throughput TILLING. *Genome Res* 13:524–530
54. Zan Y, Haag JD, Chen KS et al (2003) Production of knockout rats using ENU mutagenesis and a yeast-based screening assay. *Nat Biotechnol* 21:645–651
55. Yandell MD, Edgar LG, Wood WB (1994) Trimethylp-soralen induces small deletion mutations in *Caenorhabditis elegans*. *Proc Natl Acad Sci USA* 91:1381–1385
56. Jansen G, Hazendonk E, Thijssen KL et al (1997) Reverse genetics by chemical mutagenesis in *Caenorhabditis elegans*. *Nat Genet* 17:119–121
57. Hrabe de Angelis MH, Flaswinkel H, Fuchs H et al (2000) Genome-wide, large-scale production of mutant mice by ENU mutagenesis. *Nat Genet* 25:444–447
58. Nolan PM, Peters J, Strivens M et al (2000) A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. *Nat Genet* 25:440–443
59. Haffter P, Granato M, Brand M et al (1996) The identification of genes with unique and essential functions in the development of the zebrafish, *Danio rerio*. *Development* 123:1–36
60. Furutani-Seiki M, Sasado T, Morinaga C et al (2004) A systematic genome-wide screen for mutations affecting organogenesis in Medaka, *Oryzias latipes*. *Mech Dev* 121:647–658
61. Chen Y, Yee D, Dains K et al (2000) Genotype-based screen for ENU-induced mutations in mouse embryonic stem cells. *Nat Genet* 24:314–317
62. Munroe RJ, Bergstrom RA, Zheng QY et al (2000) Mouse mutants from chemically mutagenized embryonic stem cells. *Nat Genet* 24:318–321
63. Vivian JL, Chen Y, Yee D et al (2002) An allelic series of mutations in Smad2 and Smad4 identified in a genotype-based screen of N-ethyl-N-nitrosourea-mutagenized mouse embryonic stem cells. *Proc Natl Acad Sci USA* 99:15542–15547

represented by a simple difference image (unbleached fluorophore – bleached fluorophore). FLAP is therefore comparable with methods of photoactivation or uncaging fluorescent probes, but avoids exposing the living cells to harmful UV irradiation and takes advantage of the fluorescent proteins that can be expressed in living cells.

GFP (green fluorescent protein) fusion proteins are excellent reagents for use in photobleaching studies, because GFP inflicts much less photodamage on its surrounding environment than do other fluorophores (5). Either technique benefits from excellent confocal microscopy technology, which allows the monitoring of the dynamics not only of one fluorescent tag, but also of two or even more spectrally overlapping labels.

Fluorescence Resonance Energy Transfer

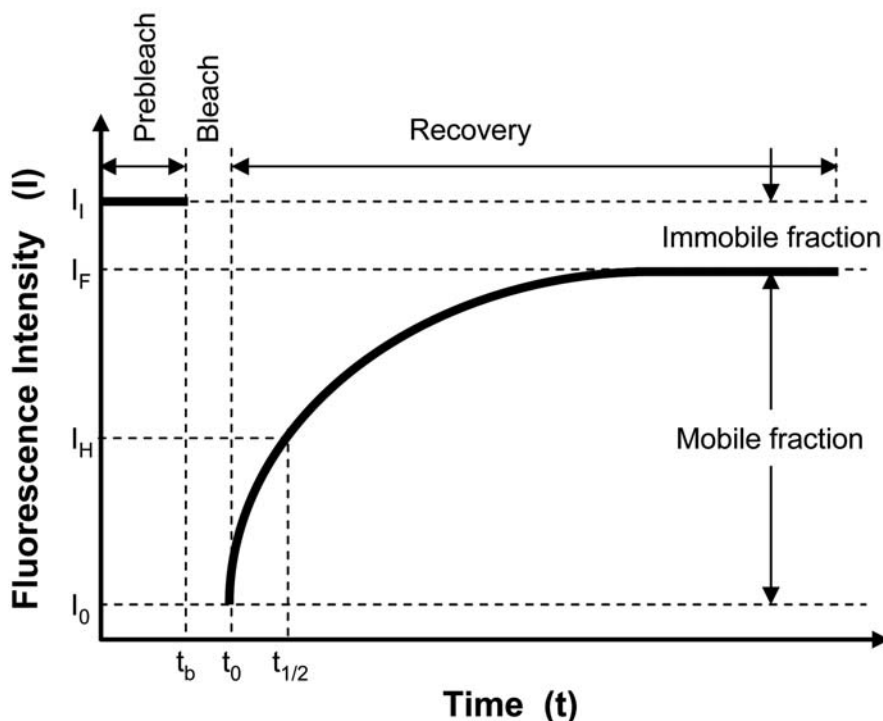
Fluorescence resonance energy transfer (FRET) is the non-radiative transfer of photon energy from a donor fluorophore to an acceptor fluorophore when both are located within close proximity (1–10 nm). Fig. 3 shows the scheme for this process. FRET can be used to reveal interactions between proteins or structural changes within a molecule (6). The method is very sensitive to

the distance between the fluorophores and thus can serve as an indicator of close proximity.

There are only certain pairs of fluorophores suitable for FRET experiments, since besides other prerequisites the donor emission has to overlap the excitation spectrum of the acceptor. Fluorescent proteins such as the green fluorescent protein (GFP) are very suitable for FRET experiments. They can be genetically fused to proteins of interest and expressed in cells, making them an excellent reporter system for gene expression and protein localization in living cells. Several enhanced fluorescent protein variants with different spectral properties are available. The cyan-colored CFP as donor and the yellow-colored YFP as acceptor are best suited for FRET experiments in living cells.

Fluorescence Correlation Spectroscopy

► **Fluorescence correlations spectroscopy (FCS)** is based on measurements of the fluorescence emitted from a small open observation subregion of the sample (7). This method uses statistical fluctuations in the intensity of the small illuminated sample volume to obtain information about the process that provokes these fluctuations. FCS is difficult to apply to cells but can provide important complementary information in

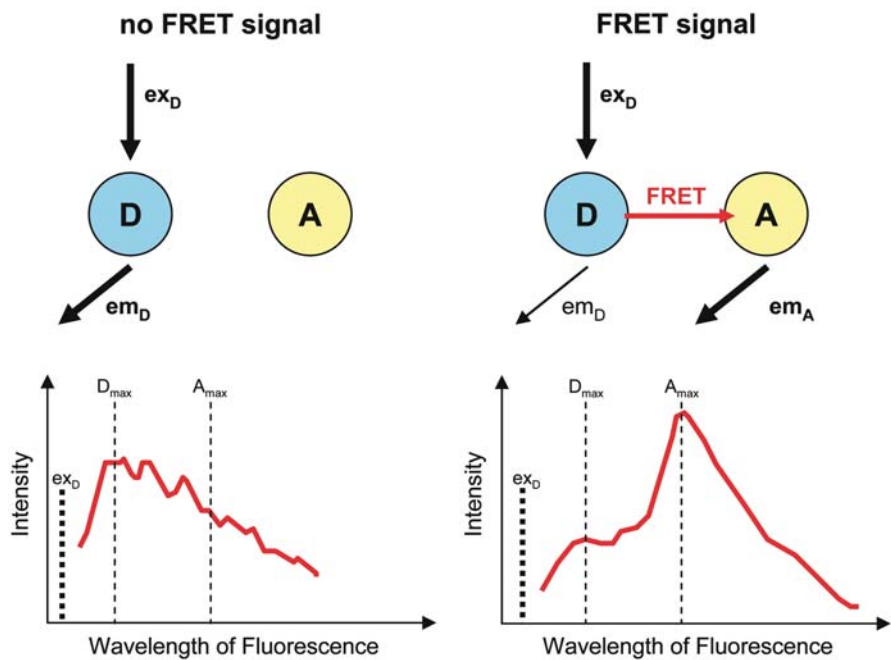


Functional Genomics, the Systematic Analysis of the Function of All Genes and Gene Products in Parallel.

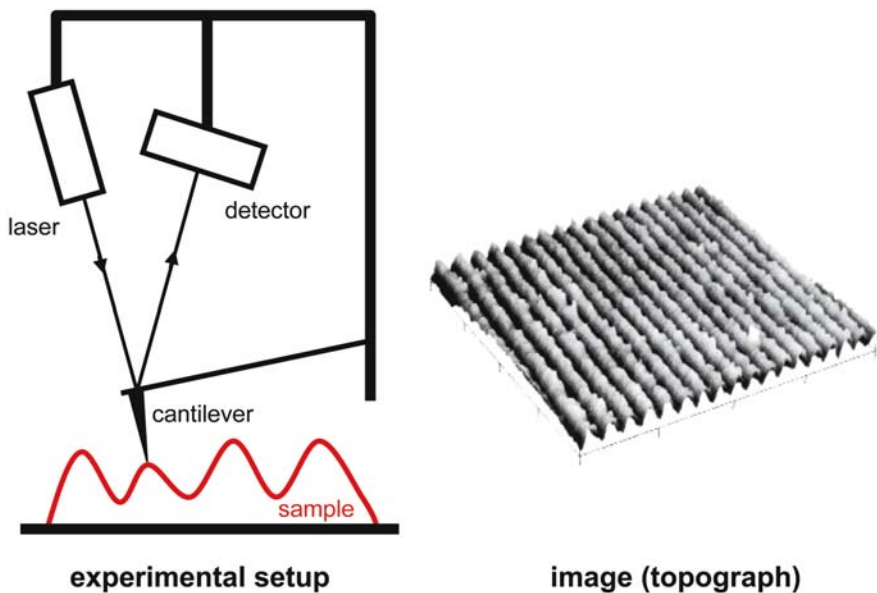
Figure 2 The idealized plot of fluorescence intensity (I) as a function of time (t) shows the parameters of a quantitative FRAP experiment. The bleach region is monitored during a pre-bleach period to determine the initial intensity I_i . The region is bleached using high intensity illumination from time t_b to t_0 , and the recovery is started at t_0 with the intensity I_0 . The fluorescence intensity increases during recovery to the intensity I_F . I_H is the half final intensity $((I_F - I_0)/2)$, and $t_{1/2}$ the half life.

some circumstances. The measured quantities include translational diffusion, rotational motion, chemical kinetics, the fluorescence lifetime of the excited state, diffusion and interactions between molecules, and aggregation and self-association. So FCS can be used, e.g. to study interactions of ligand with its target

molecules/receptor, antibody/antigen interactions on the surfaces of prokaryotic or eukaryotic cells, interactions of antisense oligonucleotides or regulatory proteins with target DNA or RNA or enzymatic activity. To reduce background noise, two-photon excitation has been applied to FCS (8).



Functional Genomics, the Systematic Analysis of the Function of All Genes and Gene Products in Parallel.
Figure 3 The principle of fluorescence resonance energy transfer (FRET). ex_D —excitation of the donor; em_D —emission of the donor; em_A —emission of the acceptor; D_{max} —maximum of the fluorescence of the donor; A_{max} —maximum of the fluorescence of the acceptor.



Functional Genomics, the Systematic Analysis of the Function of All Genes and Gene Products in Parallel.
Figure 4 The principle of the atomic force microscope.

Fluorescence Lifetime Imaging Microscopy

In the fluorescence lifetime imaging microscopy (FLIM) method, image contrast in the fluorescence microscope is derived from the fluorescence lifetime at each point in the image. FLIM has the advantage of providing 2-dimensional maps of the fluorescence lifetime with high spatial and temporal resolution. Importantly, FLIM measurements are independent of probe concentration and photobleaching. FLIM also provides an opportunity to study the dynamics of the environment surrounding the fluorophores, using time resolved emission anisotropy. Fluorescence lifetime probes already exist for the measurement of e.g. oxygen concentration, Ca^{2+} , Mg^{2+} , Cl^- , pH and K^+ .

Scanning Force Microscopy

Scanning force microscopy (SFM), or atomic force microscopy (AFM), represents a unique approach to detecting structures and forces in the nanoscale range (9). Its major advantage is that it can produce high-resolution topographic images of the surfaces of biological samples in aqueous and physiologically relevant environments without the need to stain the specimen. The principle of this method is shown in Fig. 4. The potential to measure ultralow forces enabled detection of specific interaction forces on the single molecule level. Molecular recognition of a single ligand-receptor pair can be observed in its native environment. The interaction forces between the ligand-receptor combinations biotin-avidin, antibody-antigen, sense-antisense DNA, cell recognition proteins and NTA (nitrilotriacetate) -His6 (histidine 6) were thus quantified in recent years. In combination with a confocal microscope, it is possible to examine the cellular topography in contact imaging mode simultaneously with fluorescence detection (e.g. for the cytoskeleton) in living cells.

Reference

1. Wiesner B, Hagen V (1999) Measurement of intracellular Ca^{2+} changes using novel caged cyclic nucleotides and confocal laser scanning microscopy. *J Photochem Photobiol B Biol* 49:122–119
2. Hagen V, Frings S, Bendig J et al (2002) Fluorescence spectroscopic quantification of the release of cyclic nucleotides from photocleavable [Bis(carboxymethoxy) coumarin-4-yl]methyl esters inside cells. *Angew Chem Int Ed* 41:3625–3628
3. White J, Stelzer E (1999) Photobleaching GFP reveals protein dynamics inside live cells. *Trends Cell Biol* 9:61–65
4. Dunn GA, Dobbie IM, Monypenny et al (2002) Fluorescence localization after photobleaching (FLAP): a new method for studying protein dynamics in living cells. *J Microsc* 205:109–112
5. Patterson GH, Knobel SM, Sharif WD et al (1997) Use of the green fluorescent protein and its mutants in quantitative fluorescence microscopy. *J Biophys* 73:2782–2790

6. Hovius R, Vallotton P, Wohland T et al (2000) Fluorescence techniques: shedding light on ligand-receptor interaction. *TiPS* 21:266–273
7. Madge D, Elson EL, Webb WW (1972) Thermodynamic fluctuations in a reacting system – measurement by fluorescence correlation spectroscopy. *Phys Rev Lett* 29:705–708
8. Berland KM, So PTC, Gratton E (1995) Two-photon fluorescence correlation spectroscopy: Method and application to the intracellular environment. *Biophys. J* 68:694–701
9. Lee GU, Kidwell DA, Colton RJ (1994) sensing discrete streptavidin-biotin-interactions with atomic force microscopy. *Langmuir* 10:354–361

Functional Magnetic Resonance Imaging

Definition

Functional magnetic resonance imaging (fMRI) is a non-invasive technique used to visualize brain function. Changes in chemical composition of brain areas or changes in the flow of fluids that occur over timespans of seconds to minutes can be monitored.

► [In Vivo Imaging of Transgenic Mice with Fluorescent Protein Expression](#)

Functional Proteomics

Definition

Proteomics, the large scale identification of proteins by 1-D and 2-D gel-electrophoresis and mass spectrometry, is traditionally divided into two main areas: expression proteomics and functional proteomics. Functional proteomics comprises the analysis of the function of each protein encoded by the genome.

► [Protein Microarrays as Tool for Protein Profiling and Identification of Novel Protein-Protein Interactions](#)

► [Proteomics in Cardiovascular Disease](#)

Functional Studies

Definition

An all-encompassing study of the function of an organism's genes and gene products, including transcriptional regulation, post-transcriptional processing, translational regulation, protein modification, protein interactions, etc. to understand basic mechanisms of gene regulation, structure and function. Functional studies often involve analyzing the effects of the

perturbation of normal gene/protein function (Mutational studies, overexpression/misexpression, knock-down/knockout studies, etc.).

► **ES Cell Differentiation as a Model System for Functional**

► **High-Throughput Approaches to the Analysis of Gene Function in Mammalian Cells**

► **Morpholino Oligonucleotides, Functional Genomics by Gene ‘Knock-down’**

► **Rheumatism Related Genes; Identification**

► **Xenopus as a Model Organism for Functional Genomics**

Furin

Definition

Furin is a membrane-associated, calcium-dependent, serine protease that belongs to the subtilisin-like prohormone convertase (PC) family. Members of this family of cellular enzymes cleave most prohormones and neuropeptide precursors. Numerous other cellular proteins, some viral proteins, and bacterial toxins that are transported by the constitutive secretory pathway are also targeted for maturation by PCs. Retroviruses use furin to process the mature Env protein complex (by cleavage of SU from TM).

► **Retroviruses**

Fusion

Definition

Fusion characterizes a process which describes the merging of two originally separated membranes. Fusion plays an essential role in intracellular vesicular transport between organelles. Fusion also occurs at the level of cell-cell interaction. For example, during myogenesis myoblasts fuse to form myotubuli. Release of the genetic information of enveloped viruses such as influenza virus and Human Immunodeficiency Virus (HIV) into the host cells, requires the merger of the viral envelope membrane with the host target membrane. Fusion of biological membranes is a local process regulated by specific proteins.

► **Biological Membranes**

Fusion Domain

Definition

Fusion domain describes a hydrophobic region near or at the N-terminus of the retrovirus ► **TM** (transmem-

brane) protein, which is exposed to interact with the target cell membrane following interaction of the retroviral ► **SU (surface)** protein with the receptor.

► **Retroviruses**

Fusion Proteins

Definition

Fusion proteins are two proteins or protein fragments which are fused into one by means of recombinant DNA techniques.

► **Protein Tags**

► **Two-Hybrid System**

Fusion Tag

Definition

Fusion tag is a peptide or protein, capable of being detected experimentally, which is fused to one end of a protein of interest, via expression from recombinant DNA vectors, in order to analyze the expression of the protein of interest.

► **Immunochemical Methods, Localization**

Fyn

Definition

Fyn is a Src-family kinases and is implicated in the control of cell growth.

► **Signal Transduction: Integrin-Mediated Pathways**

FzB

► **Secreted Frizzled Related Proteins**

G4 DNA

- G-Quartet DNA

GABA

Definition

GABA (gamma aminobutyric acid) acts as an inhibitory neurotransmitter in the central nervous system. GABA-ergic neurotransmission via (inhibitory) interneurons is involved in most brain functions.

- [Addiction, Molecular Biology](#)

Gag

Definition

Gag is a retrovirus gene that encodes the retroviral internal structural proteins ► [MA](#), CA, and ► [NC](#), and some others.

- [Retroviruses](#)

Gain-of-Function Mutations

Gain of function mutations describes mutations that result in new (protein) functions in heterozygous individuals.

- [Huntington's Disease](#)
- [Loss of Function Mutation](#)
- [Mendelian Forms of Human Hypertension and Mechanisms of Disease](#)
- [Mouse Genomics](#)
- [Tumor Suppressor Genes](#)

Gain-of-Function Screens

Definition

Gain-of-function screens are mutational screens based on the over- or misexpression of genes, aimed at detecting genes that may show no loss-of-function effect. Ectopic expression of an endogenous transcription unit adjacent to the 5' end of the randomly integrated P element is dependent on the expression of the GAL4 gene (*Drosophila melanogaster*), and thus may be spatially and temporally controlled.

- [Drosophila as a Model Organism for Functional Genomics](#)

GAL4–Expression System

Definition

The GAL4–expression system is a method for directed gene expression that can be used to misexpress genes in specific cell types, or tissues, at different times of development. This system relies on the generation of transgenic lines that carry “activator” or “effector” constructs. Activator lines express the yeast transcription factor, Gal4, under the control of a desired promoter, whereas effector lines contain DNA binding motifs for Gal4–(UAS) linked to the gene of interest.

- [Drosophila as a Model Organism for Functional Genomics](#)

Gametes

- [Germ Cells](#)

Gamma Rays

Definition

Gamma rays are an energetic form of electromagnetic radiation produced by radioactivity or other nuclear or subatomic processes such as electron-positron annihilation. Gamma rays are often defined to begin at an energy of 10 keV, although electromagnetic radiation from around 10 keV to several hundred keV is also referred to as hard X-rays. Gamma rays are a form of ionizing radiation; they are more penetrating than either alpha or beta radiation (neither of which is electromagnetic radiation), but less ionizing.

► [Molecular Imaging](#)

Gammaretroviruses

Definition

Gammaretroviruses refers to the genus of simple retroviruses that includes murine and feline leukemia viruses. Many species contain oncogenes and cause leukemias and sarcomas.

► [Retroviruses](#)

Gammaretroviruses IN (Integrase)

Definition

Gammaretroviruses IN (integrase) is the retrovirus virion enzyme, and is a product of the pol gene.

► [Retroviruses](#)

Gamma-Secretase (Complex)

The γ -secretase (complex) is a multimeric complex that is composed of at least four different transmembrane proteins (presenilin 1 or presenilin 2 (► [PSEN1/PSEN2](#)); (► [APH-1](#)); nicastrin; PEN-2). The last step of ► [amyloid](#) generation from amyloid precursor protein (► [APP](#)) is performed by the γ -secretase complex. The γ -secretase is also important in other pathways, as for example in the cleavage of ErbB4, intra cellular domains of Notch, and similar types of proteins.

► [Alzheimer's Disease](#)

Gap

Definition

A space introduced into an alignment to compensate for insertions or deletions in one sequence relative to another.

► [Protein Databases](#)

Gap Junctions

W. HOWARD EVANS

Medical Biochemistry and Immunology, Cardiff University School of Medicine, Cardiff, Wales, UK
wmbwhe@cf.ac.uk

Definition

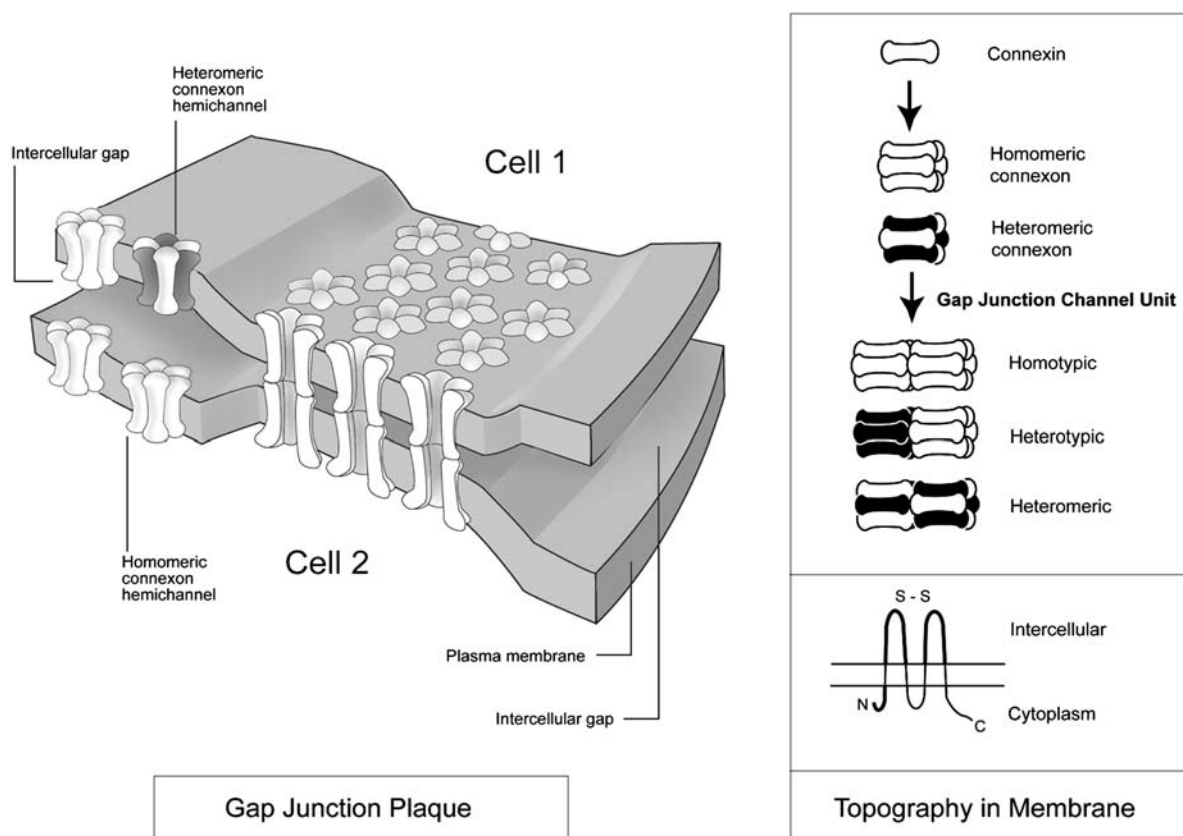
Gap junctions comprise minute regions of the cell's plasma (surface) membrane containing arrays of closely packed membrane channels. These channels traversing two aligned membranes separated by a 2–3 nm intercellular gap provide a communication pathway that directly connects cell interiors (Fig. 1). Communication across gap junctions enables single cells to co-ordinate, integrate and summate their metabolic and electrical interactive activities but also results in some loss of independence. Co-operation and harmonisation of cellular activities in tissues and organs is vividly illustrated by the ordered contraction of heart muscle, made possible because the beating of individual cells is synchronised and summated to the organ level by interactions facilitated by gap junctions. Indeed, each cardiac myocyte can communicate with about a dozen or more surrounding partner cells. In the brain, gap junctions enable the synchronisation of the electrical coupling of neuronal cell networks. In non-excitabile cells, small signalling molecules below 1200 daltons are exchanged across gap junctions. All animal cells communicate directly with each other across gap junctions, except striated muscle where the cells have fused, spermatozoa and non-nucleated erythrocytes and platelets (1).

Characteristics

Gap junctions are built from one of three biochemically different families of proteins.

► [Connexins](#)

In vertebrates, gap junctions are constructed from ► [connexin](#) protein units (Fig. 1). Over 20 different



Gap Junctions. Figure 1 A gap junction plaque showing diagrammatically the arrangement of paired connexon hemichannels in the membrane. These dock to generate a channel directly joining two cells. Connexons attach to the edges of plaques, but they may also function as hemichannels in their own right. In the top box, the oligomerisation of connexins into hexameric connexons and the various types of gap junctions formed are illustrated. In the lower box, the topography of a generalised connexin in the lipid bilayer is shown. Bold line indicates sequences with high homology.

connexins have been found in humans and rodents and various tissues synthesise different and often overlapping types of connexin (Table 1). They constitute a family of proteins displaying about 40% overall amino acid sequence identity. A widely adopted method of naming individual connexins uses the abbreviation Cx followed by the predicted molecular mass in kD. Increasingly, a prefix may be added to indicate the species e.g. h, human; m, mouse; zf, zebra fish, etc. When the molecular mass of the connexins is similar, a decimal point is added e.g. mCx 30.2, mCx 30.3 and mCx 31.1. However, although there is often little size variation between species, e.g. mCx36 and fish (skate) Cx35, there is sometimes greater disparity in the molecular sizes of functionally similar connexins, e.g. rat Cx46, bovine Cx44 and chicken Cx56 and clearly this nomenclature is not entirely satisfactory. Mouse and human connexins have been classified phylogenetically as follows: Group 1 or beta: Cx26, Cx30.3, Cx31, Cx31.1, Cx32; Group 2 or alpha: Cx33, Cx37,

Cx38, Cx40, Cx42, Cx43, Cx50, Cx56; Group3 or gamma: Cx45, Cx47 and an uncategorised group: Cx25, Cx29, Cx30.2, Cx36, Cx39, Cx40.1 and Cx58. Cx43 is by far the most widely distributed connexin (1). In heart ventricles, myocytes express Cx43 exclusively but myocytes in the atrium and Purkinje fibres also express Cx40. Endothelial cells lining the vascular wall express Cx37 in addition to Cx43, which is also expressed by the smooth muscle cells controlling the contraction of arteries. Cx36 is expressed by neurons in the retina, whereas in the brain astrocytes express mainly Cx43 and oligodendrocytes Cx32 and Cx47. Cells comprising the various layers in the skin express a range of connexins, especially Cx26, Cx30, Cx31, Cx32, Cx40, Cx43 and Cx45. In liver, hepatocytes express Cx26 and Cx32, with their relative abundance varying between species. Lens fibre cells express Cx46 and Cx50, but epithelial cells enveloping the lens capsule express Cx43. Up to 196 permutations of homologous and heterologous connexin based gap

junction interactions are possible when cells express two or more different types of connexins (Fig. 1). The relative amounts of each connexin in cells also varies during embryonic and tissue development.

Connexins span the membrane lipid bilayer four times and in an alpha helical conformation (Fig. 1). The protein's amino terminus (a highly conserved sequence of 21 amino acids) and carboxyl terminus (a highly variable number of amino acids as described below) and a single intracellular loop face the inside of the cell (Fig. 1). Two highly conserved amino acid sequences forming a loop (twenty in the first and forty amino acids in the second) project in a beta sheet conformation from the cell surface into the extracellular space; these loops contain three cysteine residues that are invariable in all connexins discovered and they are linked to each other by intramolecular disulphide bonds. These extracellular loop interactions are crucial for they provide a scaffold that aligns and extends the transmembrane [hemichannel](#) domains across the intercellular gap. The amino acid sequences in these two loops also set the rules governing the compatibility of interactions between various connexins. For example, cells with gap junction channels made from Cx32 and Cx26 communicate but cells making Cx32 do not communicate with those making Cx43. The types of connexin synthesised by cells thus dictate whether homo- or heterophilic gap junction channels are generated, with different channel pore characteristics and molecular selectivities. The third transmembrane domain and parts of the first and second transmembrane domains contribute the

wall of the overall channel. The greatest differences in amino acid sequences between connexins are found in the intracellular loop projecting into the cytoplasm and consisting of 30 amino acids in the smaller connexins, 50–55 amino acids in Cx33, Cx37, Cx40, Cx43, Cx46, Cx50 and Cx57, 80 amino acids in Cx45 and 100 amino acids in Cx36. The carboxyl terminal tail, often referred to as the regulatory domain, varies in length from 16 amino acids in Cx26 to 75 in Cx32, 156 in Cx43 and 275 in Cx57. Larger connexins are post-translationally modified by phosphorylation of several serine and threonine residues on the carboxyl tail. The functional consequences of phosphorylation of connexins remain to be defined as well as the key protein kinases and phosphatases involved in regulating gap junction channels. Connexins are not glycosylated, but are ubiquitinated and acylated.

The genes encoding mouse and human connexins are located on different chromosomes (Table 1). The general organisation of connexin genes is similar, and connexin phylogenetic trees indicate that connexins are likely to have arisen by gene duplication. Most connexin genes have a first exon containing 5' untranslated sequences and a large second exon containing the complete coding region as well as untranslated sequences. In contrast, the Cx32 gene contains two alternative first exons and their expression is tissue specific. Connexin 36 contains two exons both with translated and untranslated sequences with the coding region interrupted by an intron. The Cx45 gene has three exons, with most transcripts containing only

Gap Junctions. Table 1 Properties of some major mouse/human connexins

Connexin type	Chromosome mouse	mRNA (kb)	Cells/tissues where present	Disease involvement
Cx26	14	2.4	Mammary gland, liver, skin, cochlea	Deafness, palmoplantar hyperkeratosis
Cx30	14	20.23	Skin, cochlea	Deafness, ectodermal dysplasia, hair loss
Cx31	4	1.9, 2.3	Skin, cochlea, uterus	Deafness erythrokeratoderma variabilis
Cx32	X	1.6	Liver, oligodendrocytes	Peripheral neuropathy
Cx36	2	2.9	Neurons, retina	Visual defects
Cx37	4	1.7	Endothelium ovaries	bleeding
Cx40	3	3.5	Heart, endothelium	Atrial arrhythmia
Cx43	10	3.0	Diverse	Cardiac malformations
Cx45	11	2.2	Heart, brain thalamus, bladder	
Cx46	14	2.8	lens	cataract
Cx50	3	8.5	lens	cataract

exons 2 and 3. Cx40 also has three exons with the third containing the complete coding sequence. mRNA encoding connexins is subject to transcriptional control and tissue specific promoters (acting *via* multiple exons) account for hormonal and pharmacological regulation of expression. This is important, for example, in the contraction of uterine muscle during birth when gap junction numbers composed of Cx43 are regulated by oestrogens whereas gap junctions in the heart are not. Post-transcriptional regulation of mRNA levels is especially evident with some connexins, for example Cx32 and Cx26 in liver.

How are connexins organised into gap junctions? A hexagonal arrangement in which six connexins interact and surround a central pore of 2 nm was suggested by the regular packing of the presumed channel units in liver gap junctions stained with heavy metals as well as by atomic force microscopy. The most up to date structure is based on a three-dimensional electron crystallographic analysis of gap junctions prepared from cultured cells over-expressing recombinant Cx43. This model (2) shows the alignment of two composite hexameric connexin hemichannels at a resolution of 0.7 nm in the membrane plane and 2.1 nm in the vertical plane and it confirms independent biochemical and physical chemical studies showing the presence of 24 transmembrane alpha helices in each connexon hemichannel unit. The hemichannel unit has a dumb-bell shape and is 7 nm wide at the cytoplasmic aspect, narrowing to 5 nm at the extracellular aspect. The aqueous channel narrows from 4 nm diameter at the cytoplasmic entry point to 1.5–2.5 nm depending on the calcium concentration at the extracellular region where it becomes continuous with the partner hemichannel of a neighbouring cell. The arrangement of the connexin subunits in the gap junction ensures that the intercellular channel is completely insulated from the extracellular 'gap'.

► Innexins and ► Pannexins

Arthropod and vertebrate gap junctions differ in their overall thickness. Invertebrate gap junctions are constructed of a biochemically unrelated class of proteins called innexins. There is no amino acid sequence homology with connexins, but innexins adopt a similar topography in the membrane with four transmembrane domains and cytoplasmically oriented amino and carboxyl termini. Innexin transcripts expressed in *Xenopus* oocytes form functional gap junctions displaying typical channel gating characteristics. The *Caenorhabditis elegans* genome contains as many as 25 innexin genes, with single cells in this worm expressing more than one innexin transcript. Innexins are also present in *Drosophila* fruit flies where they feature in the development of the nervous system and intestine. Innexins have also been identified

recently in *Annelida*. The biochemical characterisation of innexins is awaited.

Pannexins are electrical junctions in vertebrates that are related to innexins. Pannexin 1 and 2 genes are abundantly expressed in the central nervous system, especially the hippocampus, olfactory bulb and cerebellum; pannexin 1 is confined to white matter. As with innexins, expression of pannexin transcripts generates intercellular channels with similar electrical characteristics to vertebrate gap junctions constructed of connexins. Their biochemical characterisation is awaited.

Assembly and Breakdown of Gap Junctions

Connexins are rapidly degraded; the half-life in heart and other cells is around two to four hours, a figure about 10–20× faster than most membrane proteins. Connexins are co-translationally inserted directly from ribosomes into the endoplasmic reticulum, threading four times into the membrane bilayer to achieve the typical connexin topography (Fig. 1). Connexins show a proclivity to oligomerise into hexameric hemichannels that exit the endoplasmic reticulum and enter the Golgi apparatus. The hemichannels are maintained in a closed configuration to limit continuity between cytoplasmic and luminal environments since ionic gradients allow cell-signalling responses. The hemichannels are then trafficked in membrane vesicles to the cell's plasma membrane and attach to the periphery of pre-existing gap junction plaques, a process that occurs simultaneously with their docking and alignment with hemichannels in the adjacent cell (Fig. 1). Connexins can be tagged at the carboxyl terminus with auto-fluorescent proteins such as green fluorescent protein or short tetracysteine-containing amino acid motifs to which arsenic-containing chemicals that fluoresce at different wavelengths will bind. These approaches have elegantly demonstrated in living cells how connexin hemichannels are moved to the plasma membrane and accrete into gap junction plaques and how gap junction units are internalised (3). Connexins are transported on 0.5 µm vesicles to the plasma membrane guided by a microtubular scaffolding and are removed from the central area of gap junction plaques as larger vesicles that correspond to annular gap junctions observed in the electron microscope. Gap junction plaques are built from up to thousands of paired hemichannel units that cannot be peeled apart once formed. Therefore, plaques are internalised into partner cells as complete units and are ultimately degraded in phagosomes or lysosomes. In contrast, incorrectly folded or mutated connexins are transferred directly from the endoplasmic reticulum for degradation in proteasomes.

Gap junctions are frequently located next to other adhesive junctions and it is not surprising that their

assembly by cells is closely co-ordinated. Connexins interact *via* their carboxyl tails with zona occludens 1 and occludin, two proteins associated with tight ►intercellular junctions, with tubulin, the major constituent of microtubules and with catenins, a further class of proteins associated with adhesive junctions. Specific connexins are also detected in ►lipid raft membrane microdomains where they associate with caveolins.

Regulatory Mechanisms

Gap Junction Channels

Measurements of electrical currents across gap junctions in paired *Xenopus* oocytes synthesising various recombinant connexins have shown that opening and closing (gating) of gap junction channels is determined mainly by a voltage difference between the paired cells and features amino acid sequences in the first to second transmembrane regions. In larger connexins, a second mechanism is identified involving a “ball and chain” type interaction between amino acids in the intracellular loop and the carboxyl tail. This chemical gating is mainly a function of pH with channels closing as pH drops (4). Gap junctions constructed of different connexins vary in their gating characteristics. Calcium ions also regulate gap junctional communication. Elevation of calcium in cells generally closes gap junctions. Cell signalling responses such as an increase in cytoplasmic calcium induced by mechanical aggravation of cells or release of inositol phosphates are propagated across gap junctions to neighbouring cells thus generating a calcium wave. The biochemical nature of the signal transmitted is not known for sure but one suggestion is that intercellular transmission of calcium waves involves the passage of inositol triphosphate through the gap junction channel. Calcium waves are also propagated between cells across connexin hemichannels. Here, ATP is released across the hemichannel and then binds in a paracrine fashion to purinergic receptors on neighbouring cells. In contrast, progress in establishing the biochemical nature of molecules/ions transmitted across gap junctions has been slow, probably because its direct nature has made their interception difficult. All that can be said is such entities are likely to be 1.2 Kd or less and with an ionic radius of below 1 nm. Gap junctions should not be regarded as nonselective pores joining cells, for the molecular selectivity for a range of permeants in the context of charge and size is a function of the connexin makeup of the channel. For example, gap junction channels distinguish between the passage of cAMP and cGMP.

Hemichannels

Unopposed connexin hemichannels in non-junctional regions of the plasma membrane (Fig. 1) were

previously thought of as biogenetic precursors of gap junctions. However, evidence is building up that they are regulatable entities in their own right, with functions influenced for example by calcium levels inside and outside the cell. These unpaired connexin hemichannels were first observed in the horizontal cells of catfish retina and their operation has now been studied extensively (5). Their presence in mammalian cells was demonstrated on the basis of passage of small dyes across the channels or the detection of electrical currents crossing them. Cells tolerate small numbers of open connexin hemichannels on the plasma membrane but the presence of larger numbers is often a pathological consequence of a metabolic insult such as in ischaemia, when cells release ATP across the open hemichannels in cardiac myocytes or glutamate in astrocytes. Importantly, hemichannels provide a second mechanism for connexin-dependent intercellular propagation of calcium waves that complements the calcium signalling occurring directly across gap junctions.

Modifications in Disease

Changes in gap junctional communication and especially mutations in connexin genes have been shown to correlate with a number of human diseases. Mutations in Cx43 are extremely rare and diseases in tissues expressing this connexin mainly involve modifications in the abundance of functional gap junction channels and the remodelling of pre-existing gap junctions as seen in cardiac hypertrophy and infarction and in endothelial dysfunction in arteries. In contrast, about 200 Cx32 mutations have been detected in the X-linked form of ►Charcot-Marie-Tooth Disease, a demyelinating syndrome that leads to degeneration of peripheral nerves. The channels constructed of Cx32 are located mainly in the paranodal loops and Schmidt-Landermann incisures of myelinating Schwann cells and provide a direct radial diffusion pathway that is about 300-fold shorter in distance than a circumferential route. Many missense, frameshift, deletion and nonsense mutations lead to a loss of function, many caused by failure of Cx32 to oligomerise correctly into hemichannels and to be targeted in a precise manner to the plasma membrane. Sometimes, gap junctions are formed by these mutated connexins but their operation is faulty. Gap junctions constructed from Cx32 in the liver and pancreatic acinar cells (Table 1) are unaffected. Surprisingly, myelination by oligodendrocytes in the central nervous system is unaffected. Mutations in Cx26 and Cx30 in the human inner ear are associated with congenital deafness, a disorder present in 1 in 1,000 births. Over 50 Cx26 mutations are known, with one common recessive mutation resulting in a severely truncated Cx26 protein. Many other site-specific mutations result in trafficking/►channel

assembly deficiencies. Gap junctions in the inner ear function in K^+ circulation from the interstitial space and through the cochlear supporting cells, a mechanism confirmed in mice in which the Cx26 gene was inactivated. Mutations in Cx26, Cx30 and Cx31.1 are also associated with disorders of the skin.

Mutations in Cx50 and Cx46 genes are associated with lens transparency and are prevalent in humans with inherited zonular pulverulent cataract. As with Cx43 in the heart, gap junctional communication is important in cell migration during organ growth and development. Knowledge gained by studying genetic mutations in the laboratory by functional expression in model cell systems has been extended and complemented in transgenic mice with single or double gene connexin “knockouts” and “knockins”. These approaches have added important inputs to our understanding of the physiological, pathological and developmental roles of individual connexins. The knowledge base will expand with the application of RNA interference techniques to silence connexin genes simultaneously *in vivo*. Further research aims will be to identify functions that can be attributed to individual connexins or specific combinations of connexins assembled into heteromeric channels. Higher definition molecular models will help explain how gap junction channels operate. The molecular basis of the specificity of transfer of signals between cells across gap junctions and the way they are interpreted will continue to be major foci for research.

►Neuron

References

1. Evans WH, Martin PE (2002) Gap Junctions; structure and function. *Mol Membr Biol* 19:121–36
2. Unger VM, Kumar NM, Gilula NB et al (1999) Three dimensional structure of a recombinant gap junction membrane channel. *Science* 238:1176–1180
3. Gaietta G, Deerinc TJ, Adams SR et al (2002) Multicolor electron microscopy imaging of connexin trafficking. *Science* 29:503–507
4. Harris AL (2001) Emerging issues of connexin channels: biophysics fills the gap. *Q Rev Biophys* 34:352–472
5. Paul DL, Goodenough DA (2003) Beyond the gap: functions of unpaired connexon channels. *Nat Cell Biol* 4:285–294
6. White TW, Paul DL (1999) Genetic diseases and gene knockouts reveal diverse connexin function. *Annu Rev Physiol* 61:283–310

GAPs

►GTPase-Activating Proteins

GAR Motif

Definition

A GAR motif is a glycine and arginine-rich region of a protein in which the arginine residues are often methylated.

►Methylation of Proteins

Gas Chromatography (GC)

Definition

Gas chromatography is an analytical method by which gaseous compounds are separated based on their volatility (boiling point) and interaction with a static liquid phase.

►Mass Spectrometry: Quantitation

Gastrulation

Definition

Gastrulation is a morphogenetic process in embryonic development, during which the three germ layers ectoderm, endoderm and mesoderm form. Mesoderm is formed from ectoderm by ►epithelial-to-mesenchymal conversion. The primitive ectoderm divides into surface ectoderm and neuroectoderm.

►Neural Development

Gatekeeper Genes

Definition

Gatekeeper genes are involved in the control of cell cycle progression, lifespan of a cell or cell death. They are often a target for mutations during cancer development. The ►APC gene is the major gatekeeper of the colon.

►Cell Cycle - Overview

►DNA-Repair Mechanisms

►Tumor Suppressor Genes

Gating

Definition

Gating refers to the conformational changes of ion channels during the opening and closing of the permeation pathway.

► [Ion Channels/Excitable Membranes](#)

GBP/FRAT

Definition

GBP (GSK-3 binding protein), and its mammalian homologue FRAT, bind to the serine-threonine kinase glycogen synthase kinase 3 (GSK3), and inhibit its phosphorylation of non-primed GSK-3 substrates. It has tumor promoting activity in lymphocyte.

► [Wnt/Beta-Catenin Signaling Pathway](#)

GDI

► [Guanine Nucleotide Dissociation Inhibitors](#)

GEF

► [Guanine Nucleotide Exchange Factor](#)

2D-Gel Electrophoresis

► [Two-Dimensional Gel Electrophoresis](#)

Gelsolin

Definition

Gelsolin and gelsolin-like proteins are ubiquitous F-actin fragmenting proteins. They perform three major functions: (1) They sever actin filaments to smaller fragments; (2) They cap free barbed ends thus inhibiting elongation; and (3) They nucleate actin polymerisation by stabilising dimers and trimers. Many of these proteins are regulated by Ca^{2+} and the phosphoinositide PIP_2 .

► [Actin Cytoskeleton](#)

Gemins

Definition

Gemins are proteins that associate with the ► [SMN \(survival of motoneuron\) complex](#) in a stable and stoichiometric manner. As these proteins and SMN colocalize in nuclear structures called “gems”, they are referred to as “Gemins”.

► [Spinal Muscular Atrophy](#)

GenBank

Definition

GenBank, the NIH genetic sequence database, is an annotated collection of all publicly available DNA sequences, located at <http://www.ncbi.nlm.nih.gov>. It is part of the International Nucleotide Sequence Database Collaboration, which is comprised of the DNA DataBank of Japan (DDBJ), the European Molecular Biology Laboratory (EMBL), and GenBank at NCBI.

► [Protein Databases](#)

Gene

Definition

Gene refers to a segment of DNA coding for a polypeptide chain of amino acids. A gene includes untranslated regions before and after the coding region

and intervening (intronic) sequences. Nomenclature: Human genes are abbreviated using italic capital letters and numbers without spaces, dashes etc. according to the approved gene symbols listed in the Human Gene Nomenclature Database (<http://www.gene.ucl.ac.uk/cgi-bin/nomenclature/searchgenes.pl>). Nomenclature for murine gene symbols underlies the same rules, but uses lower case letters after the first letter. Example: *PSEN1* = human gene for PS1; *Psen1* = murine gene for PS1.

Gene Annotation in Plants

S. ROMBAUTS, Y. VAN DE PEER, PIERRE ROUZÉ
Department of Plant Systems Biology, University of
Ghent, Ghent, Belgium
pierre.rouze@psb.ugent.be

Synonyms

Just as for gene and genome annotation in any other organism, gene annotation in plants makes use of algorithmic approaches based on statistics, artificial intelligence and machine learning, recently complemented with homology based approaches depending on the

availability of sequence data, whether related genomes, complete cDNAs or expressed sequence tags (ESTs).

Definition

All prokaryotic and eukaryotic organisms encode their genetic information in their genome, built up from DNA. The process of genome annotation consists of finding and decoding the information encrypted on the DNA molecules into known conceptual objects related to biological entities and functions. In general, annotation is mostly focused on finding genes, defining their structure and assigning a function to the product and the process resulting from the expression of each gene. But annotation is not restricted to genes, as non-coding RNAs, transposable elements, promoters and enhancers also make up the genome and are essential to an understanding of the organization and the various functions encoded by or embedded in genomes.

Characteristics

Gene annotation in plants is in essence not different from gene annotation in human or mouse, except that each genome although constituted by the same DNA has its own style that needs to be captured by the models used by the different prediction programs. Anticipating which are the genome specific characteristics, e.g. codon usage, gene density, length and composition of introns and intergenic sequences, as well as conservation of signals such as splice sites (Table 1) are all important for building adequate algorithms and for their proper training.

Gene Annotation in Plants. Table 1 Splice site prediction programs

Program	Organism	Method
GeneSplicer (152)	<i>Arabidopsis</i> , human	HMM + MDD
NETPLANTGENE (42) (http://www.cbs.dtu.dk/services/NetPGENE/)	<i>Arabidopsis</i>	NN
NETGENE2 (43) (http://www.cbs.dtu.dk/services/NetGene2/)	Human, <i>C.elegans</i> , <i>Arabidopsis</i>	NN + HMM
SPLICEVIEW (39) (http://l25.itba.mi.cnr.it/~webgene/wwwspliceview.html)	Eukaryotes	Score with consensus
NNSPLICE0.9 (44) (http://www.fruitfly.org/seq_tools/splice.html)	<i>Drosophila</i> , human or other	NN
SPLICEPREDICTOR (40,153) (http://bioinformatics.iastate.edu/cgi-bin/sp.cgi)	<i>Arabidopsis</i> , maize	Logitlinear models: (i) score with consensus; (ii) local composition
BCM-SPL (http://www.softberry.com/berry.phtml ; http://genomic.sanger.ac.uk/gf/gf.html)	Human, <i>Drosophila</i> , <i>C.elegans</i> , yeast, plant	Linear discriminant analysis

HMM, hidden MM; MDD, maximal dependence decomposition; NN, neural networks

The effort undertaken by several laboratories worldwide to sequence the *Arabidopsis thaliana* genome led, by the end of 2000, to the first full catalogue of genes present in a plant. This work also revealed that plants have about the same number of protein gene loci as vertebrate genomes (27,000 for *Arabidopsis*) but also that about 5,000 genes remained unknown, showing no homology to any known sequence in databases (1). These genes could only be predicted using *ab initio* prediction programs and need further experiment to prove that they truly exist, which has since been done for a number of them.

Genome annotation can be subdivided in two steps; structural annotation will provide gene structures and functional annotation as an as accurate as possible prediction of the function of each gene. To acquire structural annotation two main approaches are combined, intrinsic and extrinsic methods. Intrinsic (or *ab initio*) methods are all those methods that decipher genome content based solely on statistical/lexical models built by using human-curated data sets of sequences from the organism under investigation. The algorithms used for intrinsic approaches can be subdivided into signal sensors or content sensors. Signal sensors are algorithms that focus on the retrieval and identification of functional sites such as e.g. splice and translation initiation sites, transcription and polyadenylation signals and include methods such as position weight matrices, neural networks and support vector machines. Content sensors recognize regions along a sequence that have local characteristics differing from the surrounding sequence, include mainly methods such as Markov models and all their variants and are broadly used to distinguish coding from non-coding regions.

The purpose of the multiplicity of methods used is to decompose the problem of gene annotation into key components and achieve the best results on each component, which can thereafter be reassembled towards a whole gene structure (2).

Extrinsic or comparative approaches on the contrary rely on the availability of sequences from other genomes and proteins, where regions that show enough similarity between genomes are believed to have the same biological meaning, as a result of common ancestry. The advantage of comparative methods to predict genes is that it allows the revealing of small and novel genes without ambiguities and, more importantly, enables the detection of non-coding features that can hardly be detected otherwise. One needs nevertheless to be careful as to the choice of organisms used for comparative methods to achieve good results. Too closely related genomes might not reveal the information hidden in the genomes, as regions larger than coding sequences will remain conserved, making the delineation of gene structures impossible. If on the

other hand genomes are too distantly related, only a few genes will keep significant similarity to be correctly modeled or even be found. For plants, genome sequences of one dicot, *Arabidopsis*, and one monocot, rice are presently available. Unfortunately they diverged some 200 million years ago and are therefore too divergent for comparative annotation. Other genomes that are currently being sequenced, poplar, *Medicago* and *Lotus* for dicots, maize for monocots will soon fill this gap.

Besides experimental approaches, functional annotation can only be achieved through comparative methods where the knowledge of genes from one organism can be transposed to the genes and the genome of the organism concerned. To achieve this, homology searches and alignment programs are the main algorithms used (Table 2). The quality of the databases providing the data is of primary importance and is currently the main limitation. The homology searches are performed against protein, domain and motif databases using BLAST or FastA or by using hidden Markov profiles that show a better sensitivity and specificity. One point that needs to be stressed is the importance of consistency in genome annotation across species. The gene ontology (GO) project is an attempt to achieve this goal, through a hierarchical description of the genes in a genome according to the functions of their products at the various biological levels, from the molecules to the biological processes and cellular components with which they are associated. The classification and standardized terminology of GO (gene ontology) has been initiated in the annotation of the *D. melanogaster* genome and it is hoped that GO will become a community-curated entity, providing a central frame and vocabulary for annotation.

Many gene prediction programs are publicly available; many of them are referred to on the web site maintained by W. Li (<http://linkage.rockefeller.edu/wli/gene/programs.html>). Several reviews are available on this topic too, among which that of Mathé et al. (2) is more specifically oriented towards plants. The large variety of gene prediction programs have the drawback that one does not necessarily know which program to use in which situation and which performs best depending on the organism of interest. The issues of specificity, or the ability to predict only real genes and sensitivity, or the ability to predict all the genes present in a sequence, have been addressed by Burset and Guigo (3), Rogic et al. (4) and in the specific case of plants by Pavy et al. (5). In those publications it has been made clear that programs rarely performed well enough to be able to predict all the genes and that 30–40% of the genes were likely to be wrongly predicted even by the best intrinsic methods. In addition, it is clear that the more extrinsic data (of high quality) become available, the better the annotation will be.

Gene Annotation in Plants. Table 2 Homology-based gene prediction programs

Program		Organism	Databank or required input	Alignment	Gene reconstruction
AAT (66)	(► http://genome.cs.mtu.edu/aat.html)	Primates, rodents, other	cDNA, protein	DDS (improved BLASTX), DPS (improved BLASTN)	NAP, GAP2
ALN (62)			Protein	Tron code, PAM 250	
CEM (81)		Two genomic sequence	BLASTX output, WMM for sites	DP	
EbEST (73)	(► http://ares.ifrc.mcw.edu/EBEST/ebest.html)	Human, other	dbEST BLASTN, EST clustering, Smith ± Waterman-based gapped alignment	3'-UTR detection, assembly of EST-tagged exons	
Est2genome (74)		EST or cDNA, preferably BLASTN output	Modi/Eed Smith ± Waterman Needleman-Wunsh algorithm	No	
GeneSequer (67,68)	(► http://bioinformatics.iastate.edu/cgi-bin/gs.cgi)	<i>Arabidopsis</i> , maize, generic plant	dbEST or EST database or proteins	Spliced alignment, splice recognition with SplicePredictor if missing EST match	Yes
GeneWise (60)	(► http://www.sanger.ac.uk/Software/Wise2/genewiseform.shtml)	Human One protein or a HMM profile	Global alignment translated ORF/protein	DP (dynamite)	
GENQUEST	(► http://compbio.ornl.gov/Grail-bin/EmptyGenquestForm)	dbEST, SwissProt, Prosite, BLOCKS, GSDB	Smith-Waterman, Blast, Fasta		
ICE (64)	(► http://theory.lcs.mit.edu/ice)	dbEST, OWL	Look-up	DP	
INFO (63)		Nr	25mer look-up table, protein/protein alignments scored with PAM 40, PAM 120, PAM 250, BLO62	No	
ORFgene2 (61)	(► http://l25.itba.mi.cnr.it/~webgene/wwworfgene2.html)	Human, mouse, <i>Drosophila</i> , <i>Aspergillus</i> , <i>Arabidopsis</i> , <i>Caenorhabditis</i>	SwissProt	BlastP, WAM for splice sites, identity score on frequencies of dipeptides	Compatibility graph, DP
PredictGenes	(► http://cbrg.inf.ethz.ch/Server/subsection3_1_8.html)	Invertebrates, vertebrates, prokaryotes, plants	SwissProt	PAM 250	DP

Gene Annotation in Plants. Table 2 Homology-based gene prediction programs (Continued)

Program		Organism	Databank or required input	Alignment	Gene reconstruction
PRO-CRUSTES (59)	(► http://www.hto.usc.edu/software/procrustes/wwwserv.html)	Vertebrates	One homologous protein	Protein/protein alignments scored with PAM 120	DP
Pro-Gen (83)	(► http://www.anchorgen.com/pro_gen/pro_gen.html)		Two genomic sequences Alignment of translated sequences scored with PAM 120	DP	
ROSETTA (80)	(► http://crossspecies.lcs.mit.edu/)	Human, mouse	Two genomic sequences	GLASS (global alignment system), PAM 20, Genscan method for splice sites	DP
SGP-1 (82)	(► http://soft.ice.mpg.de/sgp-1)	Vertebrates, angiosperms	Two genomic sequences or a pairwise local alignment output	Local alignment	DP
SIM4 (69)		All eukaryotes	cDNA/genomic	HSP from Blast	No
SLAM (85)	(► http://baboon.math.berkeley.edu/~syntenic/slam.html)	Human, mouse	Two genomic sequences	Generalized pair HMM	DP
Spidey (70)	(► http://www.ncbi.nlm.nih.gov/IEB/Research/Ostell/Spidey/index.html)	Vertebrates, <i>Drosophila</i> , <i>C. elegans</i> , plant	One genomic sequence/set of mRNAs	Two Blasts: high stringency and low stringency	
SYNCOD (72)	(► http://l25.itba.mi.cnr.it/~webgene/wwwsyncod.html)	Human, mouse, <i>Drosophila</i> , <i>Arabidopsis</i> , <i>Aspergillus</i> , <i>Caenorhabditis</i>	BLASTN output	Silent/replacement ratio, Monte Carlo simulations	No
TAP (75)	(► http://sapiens.wustl.edu/~zkan/TAP/)	Human, mouse, <i>Drosophila</i>	dbEST	WU-BLASTN, SIM4	Yes
Utopia (84)		All eukaryotes	Two genomic sequences	Local alignment	Yes

The above described programs are different from annotation platforms, which do not attempt to make predictions themselves, but present the results from different prediction programs graphically and thus have to be seen as complementary tools that facilitate human driven annotation.

Clinical Relevance

The use of plants for medical purposes dates back thousands of years and was part of the magic exerted by medicine man, sorcerers and druids. Even now many pharmaceutical compounds that are commonly used are or were once extracted from plants. Genetic engineering

Gene Annotation in Plants. Table 3 Ab initio gene prediction programs (possibly with homology integration)

Program		Organism	Gene elements	Gene model	Homology
DAGGER (91)			Site scores	Directed acyclic graphs	
EuGène (31)	(► http://www.inra.fr/bia/T/EuGene)	<i>Arabidopsis</i>	Three-periodic IMM for exons, one IMM for introns, one for intergenic regions, one for UTR. NetGene2/SplicePredictor for splice sites	DP	<i>EST/cDNA, protein</i>
GeneId3 (89)	(► http://www1.imim.es/geneid.html)	Vertebrates, plants	Rule-based method; WAM, discriminant analysis.	DP	<i>EST</i>
GENEFINDER (28): FGENE, FEX	(► http://genomic.sanger.ac.uk/gf/gf.html ; ► http://www.softberry.com/berry.phtml)	Human, mouse, <i>Drosophila</i> , <i>Caenorhabditis elegans</i> , yeast, dicots, monocots, <i>Schizosaccharomyces pombe</i> , <i>Neurospora crassa</i>	Linear discriminant analysis	DP	<i>Protein</i>
GENEFINDER (Green)			Log likelihood ratio score matrix on MM	DP	
GeneGenerator (92)		Maize	Logitlinear models for splice sites, start; 3rd to 5th order MM for exons and introns	DP	
GeneMark (29)	(► http://opal.biology.gatech.edu/GeneMark/genemark24.cgi)	Prokaryotes, eukaryotes	5th order MM (homogeneous for introns, three-periodic for exons)	No	
GeneMark.hmm (35)	(► http://opal.biology.gatech.edu/GeneMark/eukhmm.cgi)	Human, mouse, <i>Drosophila</i> , <i>Gallus gallus</i> , <i>Arabidopsis</i> , rice, maize, <i>Chlamydomonas reinhardtii</i> , <i>C. elegans</i> , <i>Hordeum vulgare</i> , <i>Triticum aestivum</i>	5th order MM (homogeneous for introns, three-periodic for exons)	GHMM, DP	<i>Under development</i>
GeneModeler (57)	(► ftp://ftp.tigr.org/pub/software/gm/)	Eukaryotes	Nucleotide and dinucleotide composition, consensus for splice sites	Rule-based method	
GeneParser (27)	(► http://beagle.colorado.edu/~eesnyder/GeneParser.html)	Vertebrates	NN	DP	<i>EST</i>

Gene Annotation in Plants. Table 3 Ab initio gene prediction programs (possibly with homology integration) (Continued)

Program		Organism	Gene elements	Gene model	Homology
Genie (44,96)	(▶http://www.fruitoy.org/seq_tools/genie.html)	<i>Drosophila</i> , human, other	NN	GHMM, DP	<i>Protein</i>
GenLang (154)	(▶http://www.cbil.upenn.edu/genlang/genlang_home.html)	Vertebrates, <i>Drosophila</i> , dicots Grammar rules, WAM, hextuple frequencies	Chart parsing, DP		
GenomeScan		Vertebrates	Genscan method, BLASTP or BLASTX	GHMM, DP	<i>Protein</i>
GENSCAN (30)	(▶http://genes.mit.edu/GENSCAN.html)	Vertebrates, <i>Arabidopsis</i> , maize	WAM for acceptor; MDD for donor; 5th order MM (homogeneous for introns, three-periodic for exons)	GHMM, DP	<i>Protein</i> , <i>GenomeScan</i> (99)
GENVIEW2 (25)	(▶http://l25.itba.mi.cnr.it/~webgene/wwwgene.html)	Human, mouse, Diptera	Linear combination, dicodon statistic	DP	
GlimmerM (33,34)	(▶salzberg@cs.jhu.edu)	Small eukaryotes, <i>Arabidopsis</i> , rice	Three-periodic IMM for exons (order 0-8), IMM for introns, 2nd order MM for splice sites	DP	
GRAIL/GAP3/GrailEXP (90,155)	(▶http://compbio.ornl.gov/public/tools/)	Human, mouse, <i>Arabidopsis</i> , <i>Drosophila</i>	NN	DP	<i>EST</i> , <i>cDNA</i>
GRPL (97)		Human, <i>Drosophila</i> , <i>Arabidopsis</i>	Reference point logistic for splice sites, 5th order MM (homogeneous for introns, three-periodic for exons)	GHMM, DP	<i>Protein</i>
HMMgene (98)	(▶http://www.cbs.dtu.dk/services/HMMgene/)	Vertebrates, <i>C. elegans</i>	Three-periodic 4th order MM for exons, 3rd order MM for introns	CHMM	
MORGAN (48)	(▶http://www.cs.jhu.edu/labs/compbio/morgan.html)	Vertebrates	Decision tree system	DP	
MZEF (26)	(▶http://argon.cshl.org/geneænder/)	Human, mouse, <i>Arabidopsis</i> , fission yeast	Quadratic discriminant analysis	No	

Gene Annotation in Plants. Table 3 Ab initio gene prediction programs (possibly with homology integration) (Continued)

Program	Organism	Gene elements	Gene model	Homology
SORFIND (24)		Matrix method for start and splice sites, hexamer usage (Fourier measure)	No	
Twinscan (100)	Mouse, human	Genscan method; 5th order MM for UTR and intergenic, WAM for acceptor sites	GHMM	<i>Genomic sequence</i>
VEIL (47)	(► http://www.cs.jhu.edu/labs/compbio/veil.html)	Vertebrates	HMM	DP
Xpound (156)	(► http://bioweb.pasteur.fr/seqanal/interfaces/xpound-simple.html)	Human	Three-periodic 1st order MM for exons, 1st order MM for introns and intergenic	HMM

of plants to produce biopharmaceuticals is much more recent. Identification and mining of genes involved in the metabolism of such compounds in medicinal plants and in model systems is then an important issue. Gene annotation is the basic step for it, and its high quality can speed up and better focus the research on genes of potential pharmaceutical importance, especially the ones involved in secondary metabolism.

Besides being a source of pharmaceuticals, plants are also seen as a convenient and inexpensive way to produce proteins and other molecules of medicinal interest, a practice often referred as “pharming”. Most genes can indeed be expressed in a wide range of organisms. Therefore expression systems need to be tested for efficiency, cost and the biological activity of the products as the demand for high production levels at low cost is important to make modern medicine available for an ever expanding world population. Modified mammalian cells are in that respect valuable as far as the biological activity is concerned but their use is limited because of expensive culturing and difficult scaling up. The advantage of microbial organisms is that as well as the easier modification of the organisms, larger quantities can be manufactured using industrial bioreactors. Their disadvantage is that proteins do not become correctly glycosylated for usage in humans and that some proteins lack the proper folding and disulfide bridges. Plants on the other hand,

have many potential advantages for the production of recombinant proteins and the engineering of pharmaceuticals. First, growing plants is more economical than industrial facilities with bioreactors. Second, starting from plant material is already documented. Third, purification is not required when the therapeutic product can be administrated as food. Fourth, plants can be directed to target proteins into organs and intracellular compartments that confer better and more stable conservation (e.g. seeds). Fifth, the production levels that can be reached using modified plants approaches industrial scales. Last, the risk of human health threats due to contamination is reduced to a minimum.

The first recombinant therapeutic proteins, successfully produced in tobacco plants were different forms and parts of immunoglobulin (Ig), making plants virtually unlimited sources of inexpensive monoclonal antibodies. Ig produced in plants can effectively prevent infectious diseases and cancers in the mouse model or be used for *in vivo* tumor imaging. None are available commercially yet.

Glycosylation of proteins though remains a potential problem, as N-glycans in plants are structurally more diverse with a majority of oligosaccharides having β -(1,2)-xylose and α -(1,3)-fucose linked to the Man₃-GlcNAc₂ core. These are not found in mammalian N-glycans, while plant engineered proteins are lacking the sialic acid that represents 10% of the mouse sugar

content. These differences appear to have no influence on the binding affinities of Ig *in vitro*, however, there is some concern about the immunogenicity and allergenicity of plant engineered proteins. To circumvent this problem, proteins have been made edible, suppressing the problem since plant materials are ubiquitous in the human diet, the rationale being that plant engineered proteins can trigger a mucosal immunity. A high dosage is needed to maximize the chances of the protein reaching the gut intact. Therefore high expression levels in the plant tissues must be achieved. If annotation proper is not giving an answer to this issue, taking into account the features selected for making efficient, *ab initio* prediction software can be used as a guide. This tells us what makes a gene most adapted to its genome style and in turn most expressed. The latest approach to achieving high expressions levels was to express the proteins in the chloroplast. Expressing proteins in chloroplasts is an alternative for high expression with some advantages such as the absence of silencing of the engineered gene and the environmentally friendliness of the process, as spreading of chloroplast-targeted genes to other plants becomes very unlikely (6, 7).

References

1. The Arabidopsis Genome Initiative (2000) Analysis of the genome sequence of the flowering plant *Arabidopsis thaliana*. *Nature* 408:796–815
2. Mathé C, Sagot M-F, Shiex T et al (2002) Current methods of gene prediction, their strengths and weaknesses. *Nucl Acids Res* 19:4103–4117
3. Burset M, Guigo R (1996) Evaluation of gene structure prediction programs. *Genomics* 34:353–367
4. Rogic S, Mackworth A, Ouellette F (2001) Evaluation of genefinding programs on mammalian sequences. *Genome Res* 11:817–832
5. Pavy N, Rombauts S, Déhais P et al (1999) Evaluation of gene prediction software using a genomic data set: application to *Arabidopsis thaliana* sequences. *Bioinformatics* 15:887–899
6. Daniell H, Streatfield SJ, Wycoff K (2001) Medical molecular farming: production of antibodies, biopharmaceuticals and edible vaccines in plants. *Trends Plant Sci* 6:219–226
7. Price B (2003) Conference on Plant-Made Pharmaceuticals. *IDrugs* 6:442–445 <http://www.cmp2003.org/>

Gene Chip Technology

- ▶ Biochip Technology
- ▶ DNA Chip Technology
- ▶ Gene Chip Technology
- ▶ Microarray Technology

Gene Chip Technology and Its Application to Molecular Medicine

HEIKE ZIMDAHL, NORBERT HÜBNER

Max Delbrück Center for Molecular Medicine (MDC),
Berlin, Germany
nhuebner@mdc-berlin.de

Definition

DNA ▶**microarrays**, or DNA chips consist of thousands of individual DNA sequences arrayed at a high density on a single matrix, usually glass slides or quartz wafers, but sometimes on nylon substrates. Probes with known identity are used to determine complementary binding, thus allowing the analysis of ▶**gene expression**, DNA sequence variation or protein levels in a parallel format.

Description

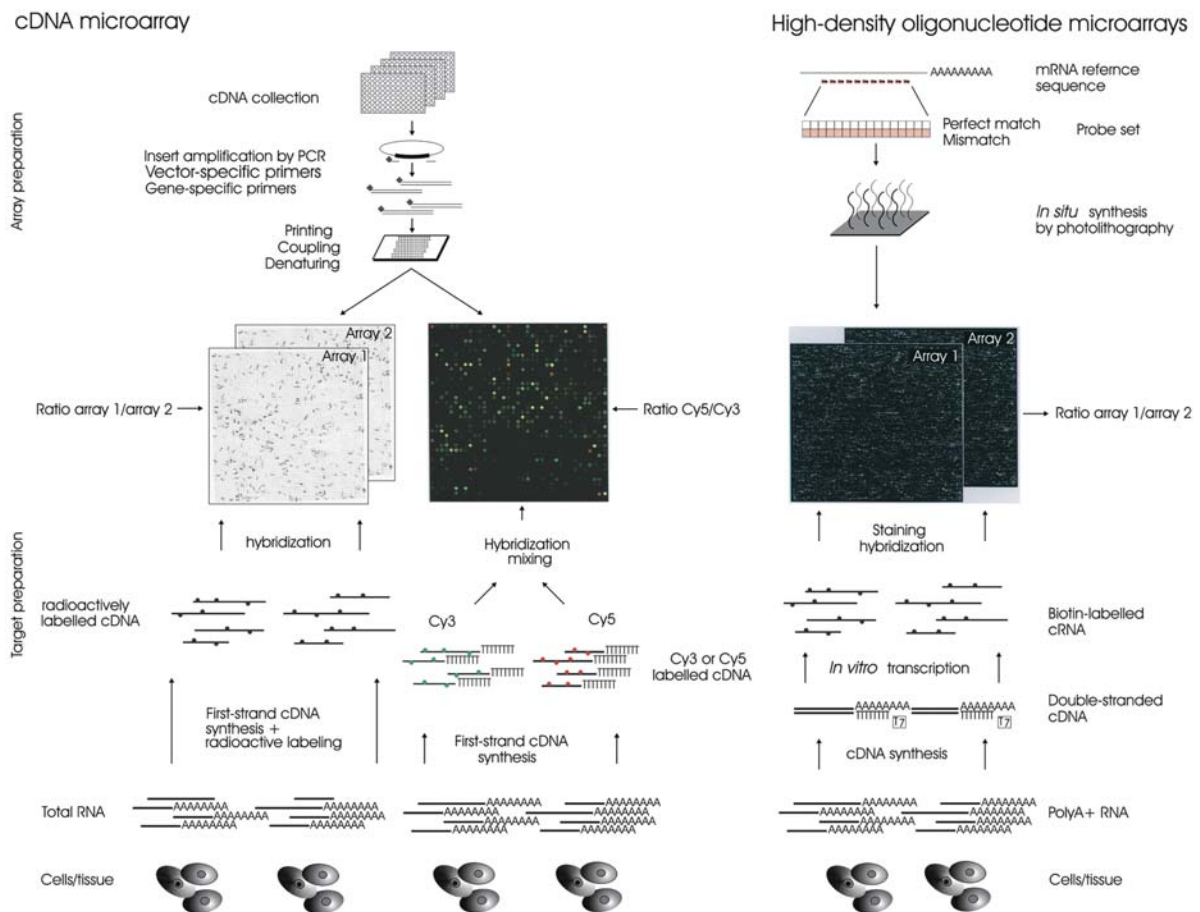
The analysis of gene expression levels in a certain tissue, cell type or stage of development provides essential information for any attempt to understand biological processes and to isolate differently regulated genes associated with a disease phenotype. Gene expression profiling may allow linkage of specific genes to disease susceptibility and drug response and the identification of genes and components of complex molecular pathways.

Various methods are available for detecting and quantitating gene expression levels, e.g. Northern blot, RNase protection assay, differential display, representational difference analysis (RDA) and serial analysis of gene expression (SAGE). For the high throughput analysis of gene expression on a global level, the microarray has emerged.

Microarrays usually contain thousands of arrayed ▶**cDNAs** or ▶**oligonucleotides**.

After completion of the human genome sequencing, the estimated number of human genes is ~30,000, but there are more ▶**mRNA** species that arise e.g. by ▶**alternative splicing**. Most investigators use in their experiments arrays comprising a substantial subset of the entire transcriptome. Gene expression analysis arrays are commercially available for a number of organisms, e.g. *H. sapiens*, *M. musculus*, *R. norvegicus*, *Arabidopsis*, *C. elegans* and *Drosophila*.

Expression profiling using microarrays represents the most efficient way of measuring gene expression – alterations in transcript levels in tissues or entire genomes can be simultaneously assayed (1-3). In addition to the capacity for parallel analysis of thousands of genes with a minimal amount of sample,



Gene Chip Technology and Its Application to Molecular Medicine. Figure 1 Schematic representation of the experimental strategy of cDNA and oligonucleotide arrays (modified according to Schulze A, Downward J, Nat Cell Biol 2001; 3:E190-E195(36)). cDNA arrays: For the array preparation, inserts from cDNA clones (libraries) are amplified and PCR products are spotted onto glass slides or nylon membranes at specific positions using arraying robots. Target preparation: RNA from the two tissues or cell populations under comparison is used to synthesize cDNA in the presence of either radioactively labeled nucleotides or nucleotides labeled with two different fluorescent dyes, Cy3 and Cy5, during cDNA synthesis. Samples labeled with two different fluorescent dyes are mixed and are hybridized to the array, whereas samples labeled radioactively (or with one fluorescent dye) are hybridized to separate arrays. Signal intensity ratios are obtained by comparing either two different signals (Cy5/Cy3) on one array or by comparing signals of genes represented on two arrays (array 1/array 2). High-density oligonucleotide arrays: For the array preparation sequences of 16–20 short oligonucleotides (typically 25mers) are chosen from the mRNA reference sequence of each gene. Light-directed, *in situ* oligonucleotide synthesis is used to generate high-density probe arrays, usually containing over 300,000 individual elements. Target preparation: polyA⁺ RNA is prepared from different tissues and used to generate double-stranded cDNA carrying a transcriptional start site for T7 DNA polymerase. During *in vitro* transcription, biotin-labeled nucleotides are incorporated into synthesized cRNA molecules. Each cRNA sample hybridizes separately to the array. Target binding is detected by staining with a fluorescent dye coupled to streptavidin. Signal intensities on different arrays are used to calculate relative mRNA abundance for genes represented on the array.

other advantages include the high sensitivity, the economy of size (miniaturization) and the use of non-toxic chemicals.

Despite the enormous potential of the technology a number of issues attenuate the power of microarrays, e.g. the control for biological and environmental factors and fluctuations, the validation of data, the need for

computational tools to analyze the vast amount of data and to enable comparisons between arrays and finally the high costs of commercial microarrays.

Conceptually different approaches to the development of microarray technology have resulted in the generation of two different array formats, oligonucleotide and cDNA ('targets') arrays (Fig. 1). The 'target'

cDNAs (or oligonucleotides) are immobilized on nylon membranes or glass slides. cDNA arrays are generated by arraying PCR products of ►cDNA libraries or clone collections usually onto glass or nylon substrates. cDNA arrays offer flexibility in the choice of arrayed elements and lower costs, particularly for the preparation of smaller, customized arrays for specific investigations of a small number of genes. In addition, arraying of unsequenced clones from cDNA libraries can be useful for gene discovery.

The advantage of the *in situ* synthesized, high-density oligonucleotide arrays (Affymetrix, www.affymetrix.com) is the high reproducibility of *in situ* synthesis on oligonucleotide chips, allowing an accurate comparison of signals generated by samples hybridized to separate arrays.

Microarray Fabrication

The concept of being able to characterize large numbers of clones by ►hybridization analyses of high-density arrayed cDNA libraries was established more than a decade ago (1).

In general, arrays are described as macroarrays or microarrays, the difference being the size of the sample spots and size of the array. Macroarrays are usually printed on nylon membranes, contain sample spot sizes of about 250 microns or larger and can easily be imaged by existing gel and blot scanners. The sample spot sizes in microarrays are typically less than 200 microns in diameter and probes are attached onto glass-like substrates. Depending on the arrayed material, the most commonly used array platforms are cDNA arrays and oligonucleotide arrays (►<http://www.gene-chips.com/>). For the generation of a cDNA microarray (4), cDNA clones representing as many unique transcripts as possible are either selected within ►EST data (►<http://www.ncbi.nlm.nih.gov/> UniGene and ►<http://www.tigr.org/tdb/tgi/hgi>) or tissue-specific cDNA libraries are constructed or ordered (www.rzpd.de). The cDNA clone inserts are PCR amplified from plasmid DNA or amplified from bacterial cultures. In high-throughput applications the amplification of clones in cultures stored in 384-well plates is more cost efficient and less labor intensive than amplification from plasmid DNA. PCR products can be purified to remove unincorporated nucleotides and primers, e.g. by filtration using silica systems. Amplified PCR products are spotted, usually in denaturing or high-salt buffer, onto glass slides or nylon membranes using robotic systems. Spots are typically 100–300 µm in size and are spaced about the same distance apart. Using this technique, arrays consisting of more than 30,000 cDNAs can be fitted onto the surface of a conventional microscope slide.

Commercially available cDNA arrays consisting of thousands of distinct sequence-verified genes are

provided by companies, like Clontech, Agilent, Incyte Genomics, Invitrogen (Research Genetics) or non-profit, public, limited institutions like the German Resource Center (►www.rzpd.de).

For oligonucleotide arrays, short oligonucleotides from 20–25 mers (Affymetrix) up to 60mers (Agilent Technologies) are usually synthesized *in situ*, either by photolithography onto silicon wafers (high-density oligonucleotide arrays from Affymetrix) or spotted by ink-jet technology (e.g. Agilent Technologies). Alternatively, presynthesized oligonucleotides can be spotted onto glass slides or glass-like matrices. Oligonucleotide arrays have certain advantages over cDNA arrays, namely high reproducibility and the facts that the sequence information (usually EST sequences) alone is sufficient to generate the DNA to be arrayed. Furthermore the oligonucleotide arrays can be designed to allow both ►SNP and alternative splicing analysis and do not require amplification and purification of cDNA fragments (5).

Since short oligonucleotides may result in less specific hybridization and reduced sensitivity, the arraying of presynthesized longer oligonucleotides (50–100 mers) has recently been developed to counteract these disadvantages. However, the high costs of commercially available, *in situ*-synthesized oligonucleotide arrays and their accessories (spotters, scanners, analysis software) as well as the time-consuming design of oligonucleotide sets may limit their use for academic laboratories.

Affymetrix has pioneered the oligonucleotide array technology and generated a number of different commercially available arrays for various organisms (►www.affymetrix.com).

The Microarray Experiment

Careful experimental design of the microarray will ensure the maximal potential gain in efficiency and is particularly important if the resulting experiment is to be maximally informative, given the effort and the resources.

There are many protocols and different types of systems available; the basic procedure for a large-scale measurement of gene expression involves the preparation of total or mRNA from the biological sample(s) under investigation (e.g. ‘candidate’ organ) and the hybridization of copied ‘labeled’ RNA to the array (Fig. 1).

In most cases, the extracted mRNA is converted to cDNA (►reverse transcription), labeled and hybridized to the DNA elements on the array surface of the array. In some cases (e.g. hybridization of Affymetrix chips) the cDNA is labeled during *in vitro* transcription and ►cRNA is hybridized. To ensure a high reproducibility, fluctuations in sample preparation and hybridization need to be reduced to a minimum. Major sources of

random fluctuations to be expected are in probe, target and array preparation, e.g. in mRNA preparation, reverse transcription, labeling, target volume, hybridization parameters, overshining effects, non-specific background, variations in pin geometry during spotting of cDNA, slide inhomogeneities and image analysis (6). Replicates of each experiment should be used in order to reduce variability and to differentiate between experimental variation and real expression differences. Suitable internal controls ensure quality control measurements for samples and array. After the hybridization process, intensity signals from the hybridized RNA samples are detected by phosphoimaging or fluorescence scanning and independent images are generated.

Analysis and Data Management

Microarray experiments generate large and complex data sets, e.g. lists of spot intensities and intensity ratios. Basically, the data obtained from microarray experiments provide information on the relative expression of genes corresponding to the mRNA sample of interest. Computational and statistical tools are required to analyze the large amount of data in order to address biological questions.

Once images of hybridized microarrays are processed, arrayed spots are identified, relative signal intensities for each spot are measured and background intensity is subtracted. Signal intensities are usually normalized to compensate for experimental variability and to 'balance' the signals from the two samples being compared (7). All normalization techniques assume that all or a subset of spots (e.g. genes) on the array have an average expression ratio equal to one. The normalization factor is then used to adjust the data (signal intensities) from the two samples and to ensure that the total quantity of RNA hybridized to the array is the same. Finally mean spot or transcript intensities are calculated and ratios of intensities are used to account for relative expression differences. In a simple pairwise comparison of gene expression between two samples, the results can be shown in plots of the intensities or the log of the intensity ratios. Scatter plots are widely used to make the observed differential expression visible.

A variety of software tools utilizing different mathematical algorithms to perform microarray image analysis are available; a detailed discussion of the analytical tools is beyond the scope of this article (see reviews 8, 9).

The first information obtained after data analysis and extraction of gene expression analysis is identification of those genes with significant differential expression in two samples or in a time series after a given treatment. To address the full potential of genome-scale experiments a ►cluster analysis is performed to analyze the entire repertoire of transcripts. Basically,

cluster analysis uses a standard statistical algorithm to arrange and organize genes according similar patterns of gene expression (10).

The data management is of particular importance for further downstream analyses. Databases are an important resource for storing and retrieving the vast amount of data generated in a microarray experiment. A number of gene expression databases have been generated and are accessible to the public (e.g. Gene Expression Omnibus at the National Center for Biotechnology Information, www.ncbi.nlm.nih.gov/geo and ArrayExpress at the European Bioinformatics Institute, EMBL-EBI www.ebi.ac.uk/arrayexpress/). Due to the experimental variation inherent in microarray experiments, the validation of the results by alternative techniques, such as quantitative real-time PCR or Northern blotting is advisable.

The most challenging part is "making sense" of the complex data retrieved, including to distinguish, whether the gene's expression change is part of the etiology of the disease or part of the pathology of the disease. The first task is the identification of the relevant gene(s), the prediction of protein function and the identification of the genetic variation that modulates gene expression, including the number of loci involved, the effect of each single locus and the interaction between loci.

Clinical Applications

Microarrays certainly have multiple applications many of which will develop and evolve over time. Although the first application of microarrays was in monitoring gene expression, the strategy of using arrayed biomolecules to examine a biological sample is generally applicable, e.g. for mutation screening, ►polymorphism analysis, mapping and other applications. An increasing number of human diseases result from alterations in DNA sequence and/or altered gene expression patterns. Therefore, information about up- and down-regulation of multiple genes is important for identification of disease genes, understanding of gene function and for potential therapeutic and/or diagnostic applications. The first clinical application may be the use of microarrays for the molecular classification of cancer (see disease diagnosis).

Although gene expression analysis is a powerful approach to identify characteristics of disease states or signaling pathways, it should be noted that gene expression levels often represent complex, quantitative phenotypes, influenced by environmental and genetic factors and the regulation of mRNA levels is only one aspect of biological control. Protein levels are also controlled at several post-transcriptional steps and protein activity is controlled by post-translational modification. A complete picture may be obtained by studying the global level of cellular proteins by proteomics (e.g. protein microarrays).

Disease Diagnosis

The real promise of the examination of gene expression using microarrays is to identify genes, which are consistently up- or down-regulated and play significant roles in the development and progression of disease. For example, the over-expression of certain genes is correlated with a certain type of cancer. By monitoring expression a new generic approach for cancer classification has been established and a comprehensive and commonly accessible catalog of gene expression profiles will make an accurate multiclass cancer classification feasible. Improvement in precise, objective and systematic tumor classification at the molecular level will advance cancer treatment (3, 11).

Pharmacogenomics

The information about gene expression and sequence variation will also have an impact on many aspects of the drug discovery process and on drug efficacy (12). Use for prognostic markers or to identify therapeutic targets has great potential.

Microarrays are used at specified steps in the process of drug development:

- Target discovery, to identify genes or pathways with altered expression in diseased human tissues or in animal models of disease.
- Target validation, to determine that a gene product is causative of disease symptoms or that activation of the target protein ameliorates disease symptoms. An agonist/activator or an inhibitor, which may be therapeutic, could be identified using microarrays.
- Compound optimization, to screen a series of therapeutic drug candidates to find the compounds that are most specific for the target protein and those that cause unintended effects.
- Drug metabolism, to predict whether a drug candidate will cause drug–drug interactions.
- Drug efficacy, to identify the individual mode of action or adverse effects of a given drug.
- Microarrays should therefore prove useful in understanding the mechanistic basis of action of many drugs.

Toxicological research: Toxicogenomics

Toxicogenomics is concerned with the identification of potential human and environmental toxicants and with finding correlations between toxic responses to toxicants and changes in the genetic profiles of the objects exposed to such toxicants through the use of genomics resources. DNA microarrays allow the identification of highly sensitive and informative markers for toxicity, by monitoring the expression levels of thousands of genes simultaneously.

Therapeutic Consequences

It has been suggested that microarrays will be routinely used not only in the selection, assessment and quality control of the best drugs for pharmaceutical development and for disease diagnosis, but also for monitoring disease status and the outcomes of therapeutic interventions (11–15). Microarrays may be of great benefit with respect to personalized medicine, i.e. to select the most likely effective drug, to individualize dosing and to minimize the safety risk for cost effective health care management. Microarrays may also be useful to identify bacterial strains that are resistant to known antibiotics to avoid a lack of response in certain patients.

One important aspect is to find a correlation between therapeutic responses to drugs and the genetic profiles of patients to address questions, such as why do some drugs work better in some patients than in others and why some drugs may even be highly toxic to certain patients?

Additionally to gene expression profiling, microarrays are especially suited for high-throughput genotyping to examine sequence variation, i.e. to screen SNPs (single nucleotide polymorphisms). An individual's genotype or genetic profile may allow 1. determination of disease association and assessment of disease risk and 2. determination of dosage and type of drug. This will enable the selection of the most appropriate and efficient drug and reduce chances of an adverse drug reaction (16, 17).

To realize the potential of microarrays, it will be a challenge to develop a cooperative framework that meets the requirements of basic research and clinical medicine.

► Microarray Data Analysis

References

1. Lennon GG, Lehrach H (1991) Hybridization analyses of arrayed cDNA libraries. *Trends Genet* 7:314
2. Schena M, Shalon D, Heller R et al (1996), Parallel human genome analysis: microarray-based expression monitoring of 1000 genes. *Proc Natl Acad Sci USA* 93:10614
3. De Risi J, Penland L, Brown PO et al (1996) of a cDNA microarray to analyse gene expression patterns in human cancer. *Nat Genet* 14:457
4. Schena M, Shalon D, Davis RW et al (1995) Quantitative monitoring of gene expression patterns with a complementary DNA microarray. *Science* 270:467
5. Lockhart DJ, Dong H, Byrne MC et al (1996) Expression monitoring by hybridization to high-density oligonucleotide arrays. *Nat Biotechnol* 14:1675
6. Lee ML, Kuo FC, Whitmore GA et al (2000) Importance of replication in microarray gene expression studies: statistical methods and evidence from repetitive cDNA hybridizations. *Proc Natl Acad Sci USA* 97:9834

7. Schadt EE, Li C, Ellis B et al (2001) Feature extraction and normalization algorithms for high-density oligonucleotide gene expression array data. *J Cell Biochem* 37:120
8. Butte A (2002) The use and analysis of microarray data. *Nat Rev Drug Discov* 1:951
9. Quackenbush J (2001) Computational analysis of microarray data. *Nat Rev Genet* 2:418
10. Eisen MB, Spellman PT, Brown PO (1998) Cluster analysis and display of genome-wide expression patterns. *Proc Natl Acad Sci USA* 95:14863
11. Ramaswamy S, Ross KN, Lander ES et al (2003) A molecular signature of metastasis in primary solid tumors. *Nat Genet* 33:49
12. Gerhold DL, Jensen RV, Gullans SR (2002) Better therapeutics through microarrays. *Nat Genet* 32:547
13. Golub TR, Slonim DK, Tamayo P et al (1999) Molecular classification of cancer: class discovery and class prediction by gene expression monitoring. *Science* 286:531
14. Jain KK (2000) Applications of biochip and microarray systems in pharmacogenomics. *Pharmacogenomics* 1:289
15. Marton MJ, De Risi JL, Bennett HA et al (1998) Drug target validation and identification of secondary drug target effects using DNA microarrays. *Nat Med* 4:1293
16. Shi MM (2002) Technologies for individual genotyping: detection of genetic polymorphisms in drug targets and disease genes. *Am J Pharmacogenomics* 2:197
17. Guzey C, Spigset O (2002) Genotyping of drug targets: a method to predict adverse drug reactions? *Drug Saf* 25:553

Gene Cluster

Definition

Gene cluster refers to a set of co-regulated genes, presumably belonging to the same functional class.

► [RNAi Interference in Mammalian Cells](#)

Gene Conversion

Definition

Gene conversion refers to a non-reciprocal, limited transfer of sequence information from one chromosome to another, or from one chromosome region to another region, of the same chromosome. The donating sequence is copied to the receiving chromosome, which is thus “converted”. The donating sequence remains unchanged.

► [Chromosomal Instability Syndromes](#)

► [Spinal Muscular Atrophy](#)

Gene Dosage Analysis

Definition

Gene dosage analysis is aimed at quantifying the copy number of a gene in an individual.

► [Spinal Muscular Atrophy](#)

Gene Duplications

SEBASTIAN M. SHIMELD

Department of Zoology, University of Oxford,
Oxford, UK

sebastian.shimeld@zoo.ox.ac.uk

Synonyms

segment duplication, chromosome duplication, ► [trisomy](#) and polyploidy. Sister genes separated by duplication are referred to as paralogues.

Definition

The term gene duplication describes a situation where a single ancestral gene has been copied such that two (or more) copies of that gene now exist in a single genome. Gene duplications are saltatory events in genome evolution in that the change occurs in one generation, with the parents having one copy while the offspring have two copies. Therefore, as with other genetic changes, gene duplications typically first exist in one individual. Whether they spread to become fixed in a population or are purged will primarily depend upon the effect they have on the organism's ► [fitness](#). At a practical level, gene duplications that have become fixed within a taxon are recognised retrospectively by the presence of homologous genes within one genome. These are best known in the context of gene families, groups of genes of related sequence inferred to have originated by repeated gene duplication from a single ancestral gene.

Characteristics

The Origin of Duplicated Genes

The prevalence of gene families in eukaryotic and prokaryotic genomes suggests gene duplication has played a prominent role in evolution. How do gene duplications occur? Probably the most common event is tandem duplication, as evidenced by the prevalence of homologous genes situated in tandem in most genomes. Tandem duplication probably occurs most frequently due to mis-recombination at ► [meiosis](#),

although other factors might also be important. Repeated tandem duplication without purging of duplicated genes results in ►**gene clusters**, that is a group of related genes situated next to each other in the genome. Such gene clusters can be evolutionarily ancient (for instance the ►**Hox gene cluster**), but are more commonly of relatively recent origin, for example the cluster of zinc finger genes on chromosome 19q12. Genes can also duplicate in other ways. Single genes can be duplicated elsewhere in the genome by several potential mechanisms, including *via* piggybacking on adjacent active retrotransposons or insertion of a reverse transcribed mRNA into the genome by a viral reverse transcriptase. Several mechanisms allow multiple genes to be duplicated simultaneously. Sections of DNA containing multiple genes can duplicate either in tandem or elsewhere in the genome (for example by transposition). Evidence suggests that recently duplicated segments form an estimated 5% of the human genome, suggesting this is a frequent and ongoing process (1). These are also known as segmental or block duplications. Duplication of an entire chromosome with all its incumbent genes can also occur. Although this is strongly selected against in humans and typically unviable (an exception is trisomy of chromosome 21), some other lineages appear to be able to tolerate such duplications more easily and aberrant chromosome numbers are especially common in flowering plants. Finally, there is evidence that whole genome duplication has occurred in several lineages, including in yeast, plants and salmonid fish (2). There is also evidence that genome duplication occurred early in vertebrate evolution and therefore that all vertebrates including humans are ancestrally polyploid. Duplication of an entire genome includes duplication of all its incumbent genes.

The Fate of Duplicated Genes

When examining the genomes of living organisms, we infer the presence of previous gene duplications by the existence of genes with homologous sequences. By their nature, these represent gene duplications that have become fixed within a population such that all members of a species (or higher taxonomic division) possess both copies. The route from the initial gene duplication to ►**fixation** or loss is affected by several factors. Some gene duplications confer an obvious ►**selective advantage** and spread rapidly through a population, for example the tandem amplification of esterase genes in mosquitoes can confer a degree of resistance to specific pesticides. Conversely, fixation of a duplicated gene within a population is by no means certain. Many gene products are required at a specific dosage and duplication disrupts this balance. This is likely to reduce the fitness of the individual carrying the gene duplication, leading to it being rapidly purged from the population.

For many genes, however, it is likely that duplication has little or no effect on fitness and is effectively neutral. A common way to view the fate of duplicated genes in this context was as a race between the accumulation of mutations that eventually silence one gene (turning it into a ►**pseudogene** which would accumulate more mutations over evolutionary time and eventually become unrecognisable) and the acquisition of divergent functions by the two duplicates and hence the necessity for an organism to maintain both. Empirical data does not support this model, however, suggesting a high rate of retention of duplicate genes (3).

More recently, a modified form of this model has been proposed. Many genes, particularly in multicellular organisms, are multifunctional, either at the biochemical level or in terms of regulated expression in different organs or cell types. This creates the possibility of duplicated genes diverging by subfunctionalisation, which implies that duplicate genes diverge by maintaining separate functions that were all initially maintained by the single gene ancestor. An organism must maintain both copies in its genome or lose one of the functions, providing a selective reason for gene duplicates to be retained in a population. Genes whose expression in multiple tissues is regulated by separate ►**enhancers** may be particularly prone to subfunctionalisation, as mutations in different enhancers of two duplicate genes would necessitate the retention of both, even if the encoded amino acid sequences were identical. This provides a potential explanation for the apparently high retention rate of duplicate genes, at least in multicellular organisms (4).

A corollary of duplicate gene divergence is redundancy and compensation. Mutational analysis of families of homologous genes often reveals a degree of redundancy, such that the phenotype of a ►**double mutant** is more severe than the sum of the phenotypes of each single mutant. This implies the genes are partially redundant and that one gene can in part compensate for the lack of the other, due to co-expression in the same tissue and a similar biochemical function.

Evolutionary Implications of Gene Duplication

Gene duplication has the potential to provide a lineage with 'new' genetic material, in that one copy of the gene is, in principle, free to evolve new functions, while the other maintains existing functions. This concept has led to the suggestion that gene duplication and entire genome duplication may play an important role in evolutionary innovation. However direct experimental support for this suggestion is lacking and some evidence suggests that, contrary to the above, both duplicate copies experience purifying ►**selection** following duplication (5). Nevertheless, it is commonplace to find gene family members that have undoubtedly evolved by duplication playing different roles in

the development, biochemistry or physiology of an organism. This implies that some degree of evolutionary innovation frequently follows gene duplication. A second possible evolutionary consequence of gene duplication is the disruption of successful interbreeding between populations in which different genes have duplicated. This suggests gene duplication might be a powerful force behind the evolution of reproductively isolated populations, and therefore of new species (6).

Clinical Relevance

The phenotypic effect of chromosomal aberrations involving duplication of segments of DNA or the possession of extra chromosomes is due to the imbalance of gene products deriving from the duplicated genes. ►[Down syndrome](#) (trisomy of chromosome 21) is probably the best-known case involving an entire chromosome, as, unlike most trisomies in humans, embryos carrying an extra chromosome 21 are viable. There are numerous other syndromes involving the duplication of specific chromosomal regions (7). Incomplete gene duplications may also result in the fusion of two genes at the boundary between the original and duplicated DNA. This can result in novel gene products with deleterious properties.

References

1. Eichler EE (2001) Recent duplication, domain accretion and the dynamic mutation of the human genome. *Trends Genet* 17:661–669
2. Skrabanek L, Wolfe KH (1998) Eukaryotic genome duplication – where's the evidence? *Curr Opin Genet Dev* 8:694–700
3. Hughes MK, Hughes AL (1993) Evolution of duplicate genes in a tetraploid animal, *Xenopus laevis*. *Mol Biol Evol* 10:1360–1369
4. Lynch M, Force A (2000) The probability of duplicate gene preservation by subfunctionalisation. *Genetics* 154:459–473
5. Wagner A (2002) Selection and gene duplication: a view from the genome. *Genome Biol* 3:1012.1–1012.3
6. Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290:1151–1155
7. Inoue K, Lupski JR (2002) Molecular mechanisms for genomic disorders. *Annu Rev Genom Hum Genet* 3:199–242

Gene Expression

Definition

Gene expression describes the process by which a gene's coded information is converted into the

structures present and operating in the cell. Expressed genes include those that are transcribed into mRNA and then translated into protein, and those that are transcribed into RNA but not translated into protein (e.g. transfer and ribosomal RNAs).

►[Microarrays in Colorectal Cancer](#)

Gene Expression Data Analysis: Classification

Definition

For classification, algorithms such as weighted-voting, k -nearest-neighbor classifiers, support vector machines and artificial neural networks can be applied to the set of genes selected using supervised analysis of gene expression to build models capable of predicting the class of a particular sample. To test the robustness of classification, these methods are often coupled with a leave-one-out cross-validation analysis, in which one of the samples from the original 'training' set is withheld and a class prediction is made on the withheld sample. For complete validation, gene lists should be tested on a second 'test' set of samples that were not used to derive the discriminatory gene list.

►[DNA Chips](#)

►[Gene Chip Technology and Its Application in Molecular Medicine](#)

►[Microarrays in Colorectal Cancer](#)

Gene Expression Data Analysis: Supervised Analysis

Definition

Using this method, one searches for genes whose expression patterns correlate with an external parameter. The most commonly used 'supervising' parameters are clinical features such as survival, presence of metastases and response to therapy. Many statistical metrics have been used successfully in 'supervised' analyses, including the standard t -test, permutation-based tests, and signal-to-noise ratios.

►[Microarrays in Colorectal Cancer](#)

Gene Expression Data Analysis: Unsupervised Analysis

Definition

No external feature is used to guide the analysis process of gene expression. The data are used to search for patterns without any *a priori* expectation concerning the number or type of groups that are present. The most common ‘unsupervised’ analysis method is hierarchical cluster analysis, based on similarity metrics.

► [Microarrays in Colorectal Cancer](#)

Gene Expression Data Matrix

Definition

Gene expression data matrix refers to a table where each row represents a gene, each column represents a particular sample, or a particular experimental condition, and each position contains a number or a set of numbers characterising the expression level of the particular gene under the particular experimental condition.

► [Microarray Data Analysis](#)

Gene Expression Profile

► [Expression profile](#)

Gene Gun

Definition

A gene gun is a tool for *in vivo* transformation of cells or organisms (e.g. gene delivery; DNA vaccination). The gun is loaded with DNA- or RNA-coated gold particles that are injected into cells or tissues using a helium pressure pulse.

► [DNA-based Vaccination](#)

Gene Mapping

► [Genetic Epidemiology](#)

Gene Ontology

Definition

Gene Ontology is a project to produce a controlled vocabulary of terms relating to molecular function, biological process, or cellular components, developed by the Gene Ontology Consortium. Such controlled vocabulary allows consistent use of terminology when describing the roles of genes and protein in cells.

► [Protein Databases](#)

Gene Silencing

Definition

Gene silencing refers to repression of genes by the formation of a specialized chromatin structure (heterochromatin). Silenced genomic regions carry specific histone modification patterns (hypoacetylation, H3 K9 methylation in some organisms) and are bound by heterochromatic proteins.

► [Chromatin Acetylation](#)

Gene Silencing by Double-Stranded RNA

► [RNA Interference in Mammalian Cells](#)

Gene Targeting

Definition

Gene-targeting describes the production of a modified allele of a gene by the process of homologous recombination between a modified exogenous DNA

molecule with its endogenous counterpart (target). In transgenic technology, it is used to introduce a targeted mutation into the genome of mouse embryonic stem cells (►ES-cells). When injected into blastocysts, mutant ES-cells contribute to all tissues of the embryo, including germ cells. Once the mutant gene has entered the germ line, mouse strains heterozygous and homozygous for the mutated gene can be bred and analysed.

- Jun/Fos
- Morpholino Oligonucleotides, Functional Genomics by Gene 'Knock-down'
- Mouse Genomics
- Transgenic and Knockout Animals

Gene Therapy

- Clinical Gene Transfer

Gene Trapping

Definition

Gene trapping comprises of a strategy by which the insertion of a targeting vector into a gene leads to reporter gene activation. The inserted vector sequence acts as a tag that facilitates rapid cloning of the trapped gene.

- Large-Scale ENU Mutagenesis in Mice
- Medaka as a Model Organism for Functional Genomics
- Mutagenesis Approaches in the Zebrafish

Gene-Based Therapies

Definition

Gene-based therapies involve the transfer of genetic material into a host with the hope of ameliorating or curing a disease.

- Limb Girdle Muscular Dystrophies

Gene-Environment Interaction

Definition

Gene-environment interaction describes an interplay between genetic variation and environmental triggers (e.g. smoke exposure, infections), which influences susceptibility to a certain disease.

- Atopy Genetics

Gene-Gene Interaction

Definition

Gene-gene interaction describes an interplay between variations in two different genes that influence(s) susceptibility.

- Atopy Genetics

General Transcription Factors

Definition

General transcription factors describes a set of common transcription factors that, together with RNA polymerase, are necessary and sufficient to direct accurate transcription initiation from a core promoter *in vitro*.

- Core Promoters

Genetic Algorithm

Definition

Genetic algorithm is a statistical/mathematical term/method that is aimed at finding the optimal solution to a question. The process is the same as natural selection: (1) selection of the strongest individuals/solutions, (2) production of new organisms/new solutions, by mixing the previously selected elements, and (3) mutations, which are accidental changes in the organisms/solutions. The process is reiterated until no more improvement can be done. The last solution is taken as the final one.

- EST Mining for Expression Analysis

Genetic Background

ROBERT GERLAI

Psychology Department, University of Toronto at Mississauga, Mississauga, Canada
robert_gerlai@yahoo.com

Synonyms

Host genotype; host genome

Definition

Genetic background represents all the genes in the ►genome. The effects of “background” genes are often considered with regard to their ability to influence or modify the effects of ►mutations artificially generated by the experimenter in model organisms such as the mouse. Mutations may be induced or introduced using targeted mutagenesis (►Targeted Gene Disruption), for example by ►homologous recombination in ►embryonic stem cells in the mouse or by using random mutagenesis with chemical mutagens including ethyl nitroso urea ►ENU. These mutations interact with the genetic background and will express their effects at the phenotypical level in a manner modulated and modified by the genetic background, a phenomenon called ►epistasis. The present essay will focus on the genetic background from this perspective. It will explain the complications associated with epistasis and with genetic linkage using examples of ►knock out (or ►null mutant) mice developed for the analysis of molecular mechanisms of mammalian brain function and behavior. The issues raised in this essay, however, are general and will be valid for several other genetic manipulation approaches and for all fields of biology.

Characteristics

►Gene targeting allows to create null mutations in mice and to analyze how the mutant organism responds to the lack of the product of a ►single gene. This has facilitated the molecular dissection of such complex traits as mammalian brain function and behavior. However, numerous problems have been pointed out (2, 4, 5, 6) with regard to the interpretation of the phenotypical changes observed in null mutant mice. Briefly, the most controversial issues stem from the fact that scientists overlooked the influence of the genetic background. These issues can be divided into two main categories, both of which have general importance in genomics. The first is a cluster of problems associated with compensatory mechanisms. This problem is rather difficult as there is no general solution to avoid it. The second problem is associated with genetic linkage (the

so called flanking region problem). Practical solutions to this problem exist and will be presented. While the examples will be drawn mostly from the field of behavioral neuroscience, the points illustrated are valid for any biological trait.

The Complicating Effect of Compensatory Mechanisms

With gene targeting one can knock out a gene *in vivo* and create a mutant organism that lacks the gene product. The promise of gene targeting has been to reveal the *in vivo* function of the gene of interest. However, the functional relevance of gene targeting has been questioned [reviewed in (3)] because the mutation may lead to an avalanche of compensatory processes (e.g. up- or down-regulation of other genes) and resulting secondary phenotypical changes. Compensation may be due to ►genetic redundancy. Genetic redundancy in this context means that putative “helper” genes take over the function of the targeted one, e.g. become up-regulated and compensate for the absence of the product of the targeted gene. Although labor intensive, proper analysis of compensatory changes may allow one to reveal how biochemical pathways interact. For example, Chen et al. (1) showed that although a null mutation in protein kinase C γ subtype (PKC γ) in mice resulted in an apparently normal long-term depression (LTD) in the cerebellum of the mutants, LTD could be blocked by a PKC inhibitor only in control mice but not in the null mutants, suggesting that LTD was mediated, at least partly, by non-PKC dependent processes in the null mutant mice. “Compensatory” processes, however, can also induce phenotypical changes. For example, assume gene α serves hypothetical function ‘A’. Also assume that targeted gene α is compensated for by gene β in the genetic background; gene β becomes up-regulated in response to the absence of α gene product. The excess of gene β product is able to compensate for the lack of gene α product and no change is observed in function ‘A’ at the phenotypical level. However, over-expression of gene β may have some pleiotropic effects, i.e. may affect functions other than ‘A’, similarly to the way over-expression of genes alters the phenotype of transgenic mice. These functional alterations when observed by the investigator will be assigned to gene α . Although they are indeed due to the introduced mutation of gene α , they need not reveal the function of this gene *per se* because they are related to this gene only indirectly. This example is not hypothetical. Empirical evidence for similar situations can be found in the literature (2).

Teasing out the direct and indirect effects of the mutation is not trivial and dissection of the molecular mechanisms underlying complex phenotypical traits will require meticulous studies in which numerous factors need to be controlled. Lathe (6) discusses

several confounding factors. Instead of reiterating his arguments, the reader's attention is now drawn to what is proposed to be the main issue, the necessity of a "systemic approach" view in gene targeting (5) (also see ► [Phenomics](#)).

The Need for a Systemic View

The principal problem in genomics is a systemic one that concerns biological organization and the functional units of this organization. From a geneticist's viewpoint the units of biological organization are the genes and their function is to encode particular proteins. However, when it comes to the question of phenotypical effects, genes may not be the units and the definition of their function may be complicated. Clusters of genes in the genetic background defined by higher organizational level phenomena, including developmental, physiological or even behavioral, may represent the functionally relevant unit. Disruption of a single gene may alter a biochemical cascade within the functional gene cluster. Expression levels of the genes belonging to a functional cluster may change in concert. Investigation of such changes may reveal the biological organization of the organism. The boundaries of putative gene clusters may not be sharp. Some genes may belong more, others less, to a specific functional gene group. This also implies that the gene group organization may not be orthogonal, i.e. some genes may belong to more than one functional group. Functional groups may be hierarchically organized. A smaller number of genes may define subgroups that may make up groups that in turn may be organized into super-groups, etc. Disrupting single genes will perturb the organism and will force it to respond in a way inherent in its biological organization. The phenotypical changes one observes are the reflection of this organization. Instead of looking for the function of single genes, it is proposed that investigators should take a systemic organizational view into consideration, an approach conceptually similar to metabolic control analysis employed in biochemistry.

In summary, the effects of genetic manipulation must be investigated at all practically feasible levels of biological organization, including gene expression patterns, protein-protein interactions and a broad spectrum of phenotypical traits that may be affected by the direct and indirect effects of the mutation.

Polymorphism in the Genetic Background May Make the Results of Gene Targeting Studies Difficult to Interpret

Assume that knock out of gene α leads to differential expression of alleles b vs. B of gene β , and a regulatory change of gene β leads to different phenotypical effects depending on which allele (b or B) is present in the α null mutant mouse. Consequently, ► [polymorphism](#) in

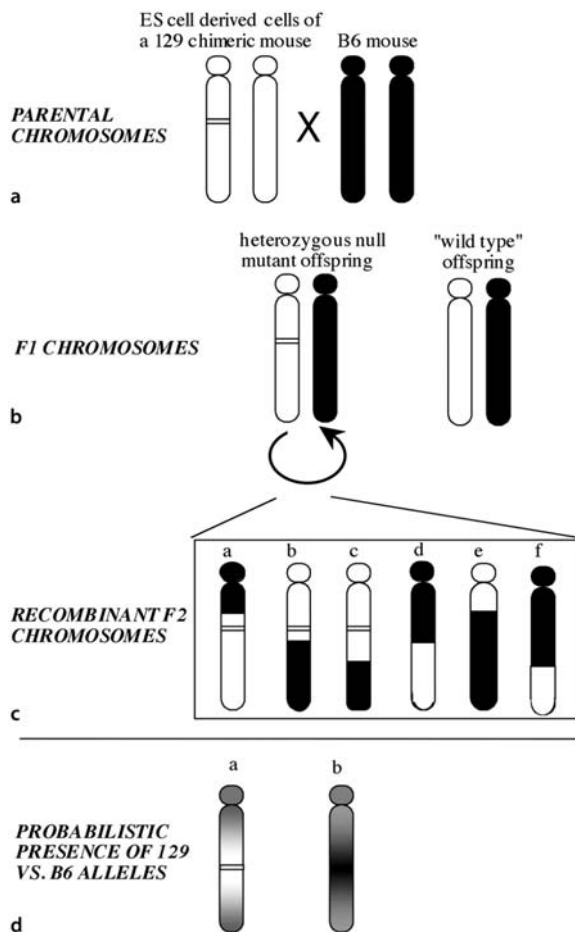
the genetic background will not allow one to conclude with certainty that a particular phenotypical change observed in the null mutant was due to the null mutation or to the genetic background. This issue is especially problematic if the genetic background of the null mutant animals is different from that of their wild type control counterparts, which is a typical problem in a large number of gene targeting studies.

Null Mutant Mice of Gene Targeting Studies Are Often the F2 Offspring of Two Mouse Strains

The genetics of the breeding strategy usually employed to generate null mutant mice is explained in Fig. 1. The figure also depicts why the hybrid genetic background leads to complications.

The Flanking Allele Problem

The F2 population is a segregating population in which mice have recombinant genotypes derived from the two parental mouse strains (Fig. 1). The difficulties arising from this are threefold. First, the recombination pattern, i.e. which locus contains strain 129 and which B6 alleles and whether in a homozygous or heterozygous form, may be different between littermates. Thus not even wild type littermates of their mutant counterparts represent an appropriate control, since their alleles could be different from those of the mutants not only at the locus of the gene of interest but also at other loci. This may lead to false positive results. Second, due to the genetic variation resulting from the hybrid segregating background, detecting significant effects of the mutated gene may be difficult, which leads to false negative results. These two problems can be alleviated by measuring larger numbers of animals, i.e. by increasing the power of statistical comparisons and decreasing the possibility of sampling error. Increasing the sample size, however, will not solve the third problem, which is associated with genetic linkage. If the targeted mutagenesis is conducted in ES cells from strain 129, the chromosome with the targeted locus will carry alleles of genes of 129-type. The probability of genetic recombination is generally inversely related to the distance between the loci of the genes (linkage). Thus, the 129-type alleles of the genes whose loci are close to the locus of the mutated gene will remain together with the mutated allele of the gene of interest (Fig. 1). In other words, any time the mutation is detected in a mouse, e.g. by ► [Southern blotting](#) or PCR, that particular animal will also carry the linked 129-type genes with high probability. Conversely, a non-mutant control, animal will not carry these 129-type alleles and will have B6 alleles with high probability if the 129-ES cell chimera was crossed to B6. In effect, the mutation can be seen as a ► [marker](#) for the 129-type genes linked to the locus of the targeted gene. Consequently, any phenotypical differences



Genetic Background. Figure 1 The genetic background of mice generated by gene targeting. Most gene targeting is carried out in cultured embryonic stem (ES) cells derived from one of the substrains of mouse strain '129'. The 129-type ES cells carrying the targeted mutation are introduced into a blastocyst and the surviving chimeric embryos develop to term, are raised to adulthood and are mated to "wild type", i.e. non mutated, mice. ES cells originating from mouse strain 129 carry one *chromosome* (white) with the disrupted allele (double lines) of the targeted gene. If these ES cells populate the germ-line in the chimeric mice, the mutation will be transmitted when the chimera is mated. A cross between a germline transmitting chimera and a C57BL/6 (B6) mouse (black chromosomes (a) will produce an F1 population (b) in which 50 % of the animals will have one copy of the mutant allele (heterozygous mutants) and 50% of them will have no mutant allele (wild type animals) at the targeted locus. Using Southern blotting or PCR (*polymerase chain reaction*) one can detect the presence of the mutant allele and identify the heterozygous mutant animals. If these animals are mated with each other, according to Mendel's law, homozygous mutant (two mutant alleles), heterozygous mutant (one mutant and one wild type allele) and wild type (two wild type alleles) animals will be obtained. It is also important to remember, however,

observed between mutant and control littermates of the hybrid genetic origin may be due either to the introduced null mutation or to the background genes linked to the targeted locus. Thus, one may find false positive results [for experimental examples see (3)].

Solutions to the Flanking Allele Problem

In order to decrease the probability of contribution of variable background genes, one could backcross the mutant hybrid animals for several generations to the strain of choice, e.g. to B6, and create a **congenic strain** that carries the mutation on the desired genetic background. However, complete elimination of 129-type genes that surround the locus of the gene of interest is not practically possible. For example, even with 12 generations of backcrossing (approximately 2 years of breeding) to B6, the length of the 129-type chromosome segment introduced to the B6 genome

how genes at loci other than the targeted one will be inherited. Crossover events during the meiotic process of gametogenesis will "shuffle" the alleles of these background genes and will create recombinant chromosomes (c), which will characterize the genotype of the sperm and the egg of the F1 mice. The genotype of an F2 individual, therefore, will be represented by a pair of such recombinant chromosomes. For example, a homozygous null mutant mouse may have chromosomes a and b, a and c, or b and c; a heterozygous mouse may have one of the recombinant chromosomes with the null mutation (a, b, or c) and another without the null mutation (d, e, or f); whereas a wild type control mouse may have chromosomes d and e, d and f, or e and f. Panel C shows that the null mutant allele of the targeted gene will be surrounded by 129-type genes, however, the wild type allele of the gene will be surrounded by B6 type genes. This *linkage disequilibrium* is simply due to the fact that the null mutant allele came from a strain 129 genetic background. In an animal produced from mating such F2 mice (F3 or the following generations), the null mutant allele could be surrounded by B6 genes only if, during the meiotic processes of gametogenesis, crossovers occurred precisely flanking both sides of the targeted gene, events whose combined probability is infinitesimally small. (d) shows the probabilistic distribution of 129 (white) vs. B6 (black) alleles in an F2 segregating population. The depicted chromosomes thus represent the genotype "average" of the F2 population. Note that in mice carrying the null mutation (chromosome 'a'), the probability of finding 129 alleles on the mutant chromosome increases the closer a given locus is to the locus of the targeted gene. However, in mice carrying the wild type allele of the targeted gene (chromosome 'b'), the probability of finding 129 alleles decreases the closer a given locus is to the locus of the targeted gene. Also note that as the distance increases from the locus of the targeted gene, the probability of the presence of 129 vs. B6 allele approaches 50-50%. (Modified from (3))

would be, on average, about 16 centiMorgans (cM) representing about 1% of the genome, i.e. about 3–400 genes. But if no alternatives are available, backcrossing is recommended because it stabilizes the genetic background of the mutant line by reducing the variation in recombination patterns across generations. Other, more complicated breeding strategies have also been proposed [reviewed in (3)]. They represent better solutions than simple backcrossing but they do not completely eliminate the flanking allele problem. Additional suggestions have also been made [for review see (3)]; rescue experiments may rule out the potential effects of linked 129-type genes and generation of “knock in” mice may be a control for the null mutant gene “knock out” animals. Furthermore, inducible knock out techniques, e.g. tetracycline transactivator systems, are recommended because comparison of pre- and post-induction phenotypes represents appropriate within subject control. Finally, generating null mutant mice with a pure genetic background is known to be the perfect solution for the problems associated with the genetic background, but up to date, this has not been preferred because most ES cells are derived from 129 type mice and these animals are not good breeders.

Clinical Relevance

Numerous genetic models of human diseases have been and will be created. However, the genetic manipulation is expected to lead to complex phenotypical changes that are influenced by a large number of genes in the genetic background as well as by environmental factors. Understanding compensatory mechanisms and systemic responses to the induced mutation and eliminating or addressing the confounding effects of background genes are important steps forward that will facilitate the modeling and ultimately the understanding of human diseases.

References

1. Chen C, Kano M, Abeliovich A et al (1995) Impaired motor coordination correlates with persistent multiple climbing fiber innervation in PKC gamma mutant mice. *Cell* 83:1233–1242
2. Crusio WE (1996) Gene-targeting studies: new methods, old problems. *Trends Neurosci* 19:186–187
3. Gerlai R (2001) Gene targeting: Technical confounds and potential solutions for the behavioral neuroscientist. *Behav Brain Res* 125:13–21
4. Gerlai R (1996) Gene targeting studies of mammalian behavior: Is it the mutation or the background genotype? *Trends Neurosci* 19:177–181
5. Gerlai R (1996) Gene targeting in Neuroscience: The systemic approach. *Trends Neurosci* 19:188–189
6. Lathe R (1996) Mice, gene targeting and behaviour: more than just genetic background. *Trends Neurosci* 19:183–186

Genetic Code

JEAN LEHMANN

Theoretical Biophysics, Department of Physics,
Royal Institute of Technology, Stockholm, Sweden
jean@theophys.kth.se

Definition

The genetic code is the set of correspondences between the codons of mRNAs and the amino acids of the proteins produced on the ribosomes through the ►translation of these mRNAs. This set also includes “stop” codons without an amino-acid assignment, which specify the termination of translation.

Codons are the units of information in mRNA. They are made up of three consecutive ►nucleotides, each of which can be one of the four bases U, C, G or A. Sixty-four different codons result from all the combinations. Twenty amino acids are coded through the phenomenon of translation. A twenty-first, selenocysteine, is coded in some organisms in particular situations (1, 2) and a still very rare twenty-second (pyrrolysine) has recently been discovered (2).

The set of correspondences between the codons and the amino acids is highly conserved in all organisms. The most common one is called the canonical genetic code, previously also called the universal genetic code before deviations were found in different ►taxa. It is usually depicted in a table with three entries (Fig. 1).

Since the sequences of mRNAs can be different from those of the genes for proteins from which they are transcribed, the coding rules are generally only valid for the mRNAs prior to translation. These differences arise from modifications on mRNAs through ►maturation processes occurring before translation (observed however almost only in Eukaryotes).

Characteristics

Main Molecular Processes Involved in Decoding

The coding rules are the outcome of extremely sophisticated molecular processes. Basically, the translation system converts genetic information contained in genes into functional proteins. The information is always processed in the form of an mRNA. In Eukaryotes, the latter can be the outcome of a complex maturation process from its primary transcript(s), which can involve several genes. In Eubacteria and Archebacteria, it is generally translated during or immediately after its ►transcription from a gene without any rearrangement. It may also come from the genome of an invasive entity (virus).

A start codon upstream of the mRNA signals the place of the beginning of translation. This codon is almost always AUG, but is also sometimes GUG (very rarely

1st pos. (5')	2nd pos.				3rd pos. (3')
	U	C	G	A	
U	phe	ser	cys	tyr	U
	phe	ser	cys	tyr	C
	leu	ser	trp	stop	G
	leu	ser	stop	stop	A
C	leu	pro	arg	his	U
	leu	pro	arg	his	C
	leu	pro	arg	gln	G
	leu	pro	arg	gln	A
G	val	ala	gly	asp	U
	val	ala	gly	asp	C
	val	ala	gly	glu	G
	val	ala	gly	glu	A
A	ile	thr	ser	asn	U
	ile	thr	ser	asn	C
	met	thr	arg	lys	G
	ile	thr	arg	lys	A

Genetic Code. Figure 1 The canonical genetic code table. The codonic positions are arranged from left to right of the table, corresponding to the reading direction of the mRNAs (5'–3'). Each of the sixty-four possible codons code for a particular amino acid or the stop function, which binds a release factor RF1 or RF2 at the end of translation. The amino acids are abbreviated to their three-letter notation.

UUG or CUG). Any of these codons can bind the initiator tRNA coding a modified methionine, N-formylmethionine. The start codon also positions the sequence into the correct **▶reading frame**. The process generally ends at a stop codon (UAA, UAG or UGA in the canonical genetic code), which binds a release factor RF1 (UAA, UAG) or RF2 (UAA, UGA).

The path leading a free amino acid to incorporation into a coded protein involves three main steps (Fig. 2):

1. **▶Activation** of the amino acid by the corresponding aminoacyl-tRNA synthetase (aaRS) through hydrolysis of a molecule of ATP.
2. The aaRS subsequently binds the amino acid to a cognate tRNA. Prior to translation, a complex made up of an **▶elongation factor** Tu (EF-Tu) and GTP binds the amino acid at the 3' end of the tRNA, giving it a much higher affinity for the ribosome.

3. During translation, occurring on the ribosomes, the successive codons of a particular mRNA are tested by the incoming tRNAs through anticodon-codon associations. A proofreading mechanism is preceded by hydrolysis of the GTP of the complex at the 3' end of the tRNA. Anticodon-codon complementarity is the decisive factor leading to the formation of the peptide bond between the last amino acid of the nascent protein and the amino acid carried by the incoming tRNA.

Accuracy of the Coding Rules

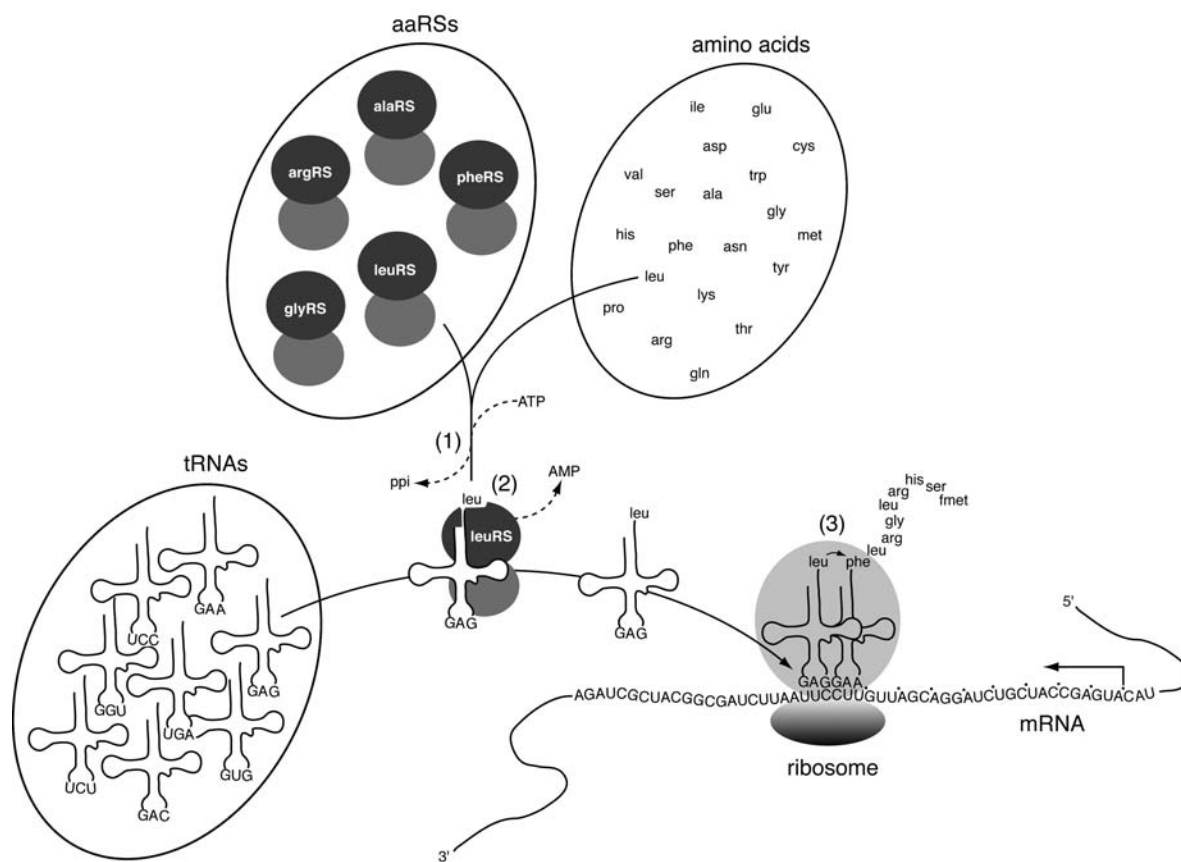
The reliability of the coding rules thus depends on the accuracy of two major processes of molecular recognition:

- a) Binding of both an amino acid and a tRNA by an aaRS (steps 1–2).
- b) Binding of the anticodon of a tRNA by a codon of an mRNA through base pairing (step 3).

The first process (a) constitutes a textbook case of RNA-protein recognition (3) and has been referred to as the second genetic code; as soon as a particular aaRS has bound an amino acid to a cognate tRNA, the code is virtually already established since the anticodon-codon associations occurring in (b) follow relatively strict base-pairing rules.

Each of the (generally) twenty different aaRSs thus recognizes both a tRNA and an amino acid. The recognition elements on the tRNAs are mainly situated in the acceptor stem and anticodon domains, which constitute the two major spatial regions of that molecule. Both positive and negative elements are used by the aaRSs to identify their associated tRNAs, and differentiation between two tRNAs can sometimes depend on the identity of only a few bases (3). Since more than one tRNA molecule is often necessary for the translation of the whole set of codons belonging to a particular amino acid, in several cases an aaRS must be capable of recognizing different tRNAs, which implies that the anticodon is sometimes not used as a discriminatory element (e.g. for ser-, leu- and ala-accepting tRNAs). The aaRSs are among the most studied enzymes, and their structural analyses have given important clues on the evolution of the genetic system (4).

The second process (b) involves an initial selection and a proofreading mechanism that are used by the ribosomes to discriminate between correct and incorrect anticodon-codon associations (5). These coupled mechanisms, separated by GTP hydrolysis, ensure a very low error frequency which in *E. coli* has been estimated as being between 6×10^{-4} and 5×10^{-3} . The initial selection step occurs when an incoming tRNA charged with its cognate amino acid (combined with EF-Tu and GTP) binds the ribosome and tests the



Genetic Code. Figure 2 The main molecular processes involved in the assemblage of free amino acids into proteins according to the rules of the genetic code. Three sets of molecules participate in the translation of the mRNAs: the encoded amino acids (20 different kinds), the aminoacyl-tRNA synthetases (aaRSs) (at least one for each encoded amino acid, not all shown here) and the tRNAs (at least 22 but often more than 30 different kinds, not all shown here). The example shows the events leading to the translation of a CUU codon of an mRNA. First, a leuRS binds a leucine and subsequently catalyses its activation through the hydrolysis of a molecule of ATP, resulting in an aminoacyl-adenylate (leu ~ AMP) and ppi [reaction (1)]. As soon as the leuRS recognizes a cognate tRNA (in this case, a tRNA with an anticodon GAG), the aminoacylation reaction occurs and the tRNA is thus charged with the amino acid [reaction (2)]. The AMP molecule is then released. After the [EF-Tu + GTP] complex has bound the amino acid at the 3' end of the tRNA (not shown here), the latter can participate in translation, occurring on a ribosome. In accordance with the wobble rules (Table 1), it can successfully bind the CUU codon, thereby enabling the formation of the peptide bond between the incoming leucine and the nascent protein [reaction (3)]. The previous tRNA (with the anticodon AAG) then leaves the ribosome. Note that the codons read from 5' to 3' (thus from right to left in the figure) and the anticodons from 3' to 5'. The arrow above the mRNA shows the starting place of translation, while the cutting of the strand into codons is indicated by small dots. For clarity, details on structures and events are omitted at the level of the ribosome.

codon at the point of being translated. At this stage, most of the non-cognate tRNAs are rejected. However, in addition to cognate tRNAs, some near-cognate tRNAs (whose anticodons can form partial base-pairs) may also cross this first stage, associated with the hydrolysis of the GTP. This hydrolysis results in a [▶conformational](#) change of EF-Tu, which subsequently leaves the 3' end of the tRNA. The amino acid is then free to move to the [▶peptidyl-transferase center](#) of the ribosome where peptide bond formation can take

place. Before that event however, almost all remaining near-cognate tRNAs are rejected (proofreading) (5).

Caveat

The standard rules of the genetic code can be altered during translation events that have been referred to as “recoding” (1). These alterations are specific to individual mRNAs, and involve three possible kinds of event, [▶programmed frame shifting](#), [▶translational bypassing](#) and [▶redefinition of codons](#). Albeit not

common, it is thought that any of these events may occur in all living organisms (1). They serve notably as control mechanisms of gene expression.

Order in the Genetic Code Table

Degeneracy

Since most of the amino acids are coded by more than one codon, the genetic code is degenerate (Fig. 3). The degeneracy is connected with the base in third position of the codons, which often has only a small influence on the nature of the encoded amino acid. A particular codon generally belongs to one of two major categories, the four-fold and the two-fold degenerate families. In two cases, the latter family is further split into individual codons, each with a specific assignment. The presence of a degeneracy symmetry over the entire table, appearing when the succession [U, C, G, A] (or its inverse) is used to write the table, shows that the splitting into four-fold and two-fold degenerate families is conditioned by the nature of the bases in the first and second positions of the codons (Fig. 3).

The origin of the degeneracy stems from the existence of wobble rules in the base-pairing of the third position of the anticodons that can be specific to the type of family (four-fold or two-fold degenerate), and which are controlled by the ribosome (Table 1). Thus, in the case of mitochondrial and chloroplastic systems where wobble rules are the most extensive, a tRNA with U in the third position of the anticodon can translate any of the four codons of the corresponding four-fold degenerate family, while a tRNA with U (G) in the third position of the anticodon can translate any of the two codons with G or A (U or C) in the third position of the anticodon of the corresponding two-fold degenerate family. As a result, only 22 different tRNAs are necessary in some systems (in which also the AGR family is reassigned to stop; see next subsection) in order to read all the codons for amino acids. The wobble rules are, however, generally less extensive and in most systems more than 30 kinds of tRNAs are used for the translation of all the codons. Moreover, the tRNAs often have modified bases in the third position of the anticodon (e.g. I, Q), each with particular matching properties (Table 1).

An important consequence of this wobble phenomenon, enabling a limited number of different tRNAs to cope with the whole set of codons, is an improvement in the efficiency of translation.

Moreover, some amino acids are present in more than one codon family. Thus, leu, ser and arg are coded by a total of six codons in two codon families in the canonical genetic code.

Reassignments

Although the general organization of the genetic code is conserved over all living organisms, many variants

1st pos. (5')	2nd pos.				3rd pos. (3')
	U _(Y,W)	C _(Y,S)	G _(R,S)	A _(R,W)	
U _(W)	phe	ser	cys	tyr	U } ^(Y)
	phe	ser	cys	tyr	C } ^(R)
	leu	ser	trp	stop	G } ^(R)
	leu	ser	stop	stop	A }
C _(S)	leu	pro	arg	his	U } ^(Y)
	leu	pro	arg	his	C } ^(R)
	leu	pro	arg	gln	G } ^(R)
	leu	pro	arg	gln	A }
G _(S)	val	ala	gly	asp	U } ^(Y)
	val	ala	gly	asp	C } ^(R)
	val	ala	gly	glu	G } ^(R)
	val	ala	gly	glu	A }
A _(W)	ile	thr	ser	asn	U } ^(Y)
	ile	thr	ser	asn	C } ^(R)
	met	thr	arg	lys	G } ^(R)
	ile	thr	arg	lys	A }

Genetic Code. Figure 3 Degeneracy in the canonical genetic code table. Two families of degeneracy are mainly present in the table: the four-fold degenerate families (gray rectangles) and the two-fold degenerate families (squares). In two cases (AUR and UGR), the latter type of degeneracy is further split into individual codons. The succession [U, C, G, A] used to write the table highlights a degeneracy symmetry over entire the table (dashed line), highlighting the fact that simple rules based on the nature of the bases in the first and second positions of the codon account for the distribution of the two main families in the table:

Four-fold degenerate families		Two-fold degenerate families	
1st pos.	2nd pos.	1st pos.	2nd pos.
(W)	(Y,S)	(W)	(Y,W), (R,S), (R,W)
(S)	(Y,W), (Y,S), (R,S)	(S)	(R,W)

where the following categories are used to differentiate the bases: S (strong matching) = G or C; W (weak matching) = A or U; R (purine) = A or G; Y (pyrimidine) = C or U

Genetic Code. Table 1 Wobble rules in the genetic code

3rd base anticodon (5')		3rd base codon (3')			
		Standard matching(s)	Limitation known so far	Possible extension of matchings	Limitation known so far
Standard bases	A	U		C,G,U(A)	ACN and CGN families in Mycoplasma ssp, yeast mt., nematode mt.
	C	G			
	G	C,U			
	U	A,G		A,C,G,U	All 4-fold degenerate families in mt. and ch.
Modified bases	I	A,C,U	4-fold degenerate families (except GGN)		
	Cm	A(G)	UUR family in E. coli		
	f ⁵ C	A,G	AUR family in nematode, bovine, squid mt. and Drosophila mt.		
	L	A	AUA in eubacteria and plant mt.		
	Q	C,U	2-fold degenerate families in eubacteria and eucaryotes		
	m ⁷ G	A,C,G,U	AGN family in echinoderm mt. and squid mt.		
	xo ⁵ U	A,G,U	UCN, GUN, GCN and ACN families in eubacteria		
	xm ⁵ Um, Um, xm ⁵ U	A,G	2-fold degenerate families in mt., eubacteria and eucaryotes		
	xm ⁵ s ² U	A(G)	2-fold degenerate families in eubacteria and eucaryotes		
	G in anticodon (UψG)	A,C,U	AAU, AAC, AAA in echinoderm mt.		

Main source: ref (6). N = U, C, G or A; R = A or G; Y = C or U. Mt. = mitochondria; ch. = chloroplasts

have been discovered in different taxa. In addition, the organelles (mitochondria and chloroplasts) have their own translation systems with specific variants of the canonical genetic code (Fig. 4).

An examination of all the variants discovered so far indicates that some codons are more prone to reassignment than others (6). The stop codons constitute a typical example; the two-fold degenerate family UAR is frequently reassigned to gln in the nuclear system, while UGA almost always codes for trp in mitochondria. This particular versatility may be explained by the rarity of these codons, which implies that their reassignments might not cause catastrophic disturbance. Another interesting change concerns the AUA codon, which often codes for met in the mitochondria. As a result, the degeneracy symmetry pointed out in Fig. 3 is entirely valid for the

mitochondrial systems that also have UGA coding for trp. These changes imply that the wobble rules can be altered within the codon families concerned.

Furthermore, it has been observed that all codons that have been reassigned in the nuclear systems have also been subject to reassignment in the mitochondria, while no codon has been reassigned in the nuclear systems only (6).

Some codon families seem to be excluded from any reassignment procedure. This is especially the case for all codons with G in first position, corresponding to the amino acids val, ala, gly, asp and glu, as compared with the codons with U, C or A in this same position, for which at least four changes have been reported (Fig. 4). Moreover, all these amino acids are only present in this part of the table in the canonical genetic code. This particularity shows that the genetic code has a core.

1st pos. (5')	2nd pos. hydrophobicity A G C U				3rd pos. (3')	prebiotic synthesis experiments
	U	C	G	A		
U	phe	ser	cys	tyr	U	< 1%
	phe	ser	cys	tyr	C	
	leu	ser	trp	stop ⁹	G	
	leu	ser ^{4*}	stop ⁵	stop ¹⁰	A	
C	leu ^{1*}	pro	arg ^{6*}	his	U	< 1%
	leu ^{1*}	pro	arg ^{6*}	his	C	
	leu ^{1,2}	pro	arg ⁶	gln	G	
	leu ^{1*}	pro	arg ^{6*}	gln	A	
G	val	ala	gly	asp	U	> 97%
	val	ala	gly	asp	C	
	val	ala	gly	glu	G	
	val	ala	gly	glu	A	
A	ile	thr	ser	asn	U	< 1%
	ile	thr	ser	asn	C	
	met	thr	arg ^{7*}	lys	G	
	ile ³	thr	arg ⁸	lys ^{11*}	A	

Genetic Code. Figure 4 Reassignments and parameters of order in the canonical genetic code table. The numbers refer to reassignments that have been reported so far [main source: ref (6)]: ¹ thr, ² ser, ³ met and unassigned, ⁴ stop, ⁵ trp and cys, ⁶ unassigned, ⁷ ser, gly, stop and unassigned, ⁸ ser, gly, stop and unassigned (⁷ and ⁸ changes are independent), ⁹ leu, ala and gln, ¹⁰ tyr, gln and glu, ¹¹ asn. Unassigned means that the codon(s) disappeared in the entire genome. Changes that have been reported in mitochondria only are indicated by an asterisk (*). The gray area within the table points out all codon families with G in first position, for which any reassignment has been reported so far. The corresponding amino acids generally constitute more than 97% of the amount of all amino acids produced in experiments thought to reproduce the conditions of the early Earth (the results however differ slightly depending on the hypothesized conditions). The hydrophobicity parameter is shown on the top of the table. It is connected with the 2nd anticodonic position (overdrawn above the codonic one), the succession [A, G, C, U] ranking the bases in decreasing order on the hydrophobicity scale. The hydrophobicity correlation between the base at this position and the amino acids result in most of the hydrophobic amino acids being on the left of the table, while most of the hydrophilic ones are on the right.

The Core of the Genetic Code

As revealed by statistical analyses of coding sequences, all codons of the genetic code do not occur with equal probability. Some codons occur very rarely while others

are frequent. Many factors can affect their distribution, but a general trend connected with the organization of the genetic code, which is independent of the organism or the genes under consideration can be highlighted.

The most general pattern occurring in coding sequences is GNN, meaning that the codons with G in first position are overrepresented in comparison with those with U, C or A at this same position (N stands for any nucleotide).

These codons code for the amino acids val, ala, gly, asp and glu, which display the simplest side chains of all the amino acids of the genetic code. This can explain their relative abundance in the proteins of which they constitute the scaffold.

One can, furthermore, connect this abundance to the observed stability of their assignment (see previous section). Changing the signification of any of the corresponding codons may cause malfunctions in some proteins, resulting in disturbances in biochemical processes (it has, however, been experimentally shown that an organism such as *E. coli* can survive treatment of this kind).

Interestingly enough, this set of amino acids generally constitute more than 97% of the total quantity of all amino acids found in the so-called prebiotic synthesis experiments (Fig. 4). It thus seems likely that the primitive code was limited to the use of these amino acids, which were only assigned to the codons with G in first position for physico-chemical reason. Later, more codons and more amino acids were added, eventually constituting the genetic code as we know it in present-day organisms.

A Hydrophobicity Parameter in the Table

Correlation studies on physico-chemical properties of the constituents of the genetic code have revealed that the **hydrophobicity** of the base in the second position of the anticodon is positively correlated with that of the encoded amino acid. It has also been established that the succession [A, G, C, U] ranks the bases from the most hydrophobic (A) to the most hydrophilic (U). Thus, Fig. 4 naturally arranges the amino acids into the genetic code table with regard to hydrophobicity; the most hydrophobic amino acids are mainly on the left of the table while the hydrophilic ones are on the right.

This order in the genetic code table is noteworthy since the hydrophobicity of the amino acids has a major influence on the folding of the proteins of which they are the constituents. Thus, an analysis of coding sequences based on this correlation can provide information on the structure of the corresponding encoded proteins (7).

Clinical Relevance

Since the aaRSs are responsible for the reliability of the coding rules, genetic diseases affecting these proteins can create important disorders in the organism.

Autoantibodies to five aaRRs have been described, and each is associated with a syndrome of inflammatory myopathy with interstitial lung disease and arthritis (8). At the level of the ribosome, antibiotics such as paromomycin and streptomycin can significantly affect the fidelity of translation. Paromomycin stabilizes the tRNAs irrespective of whether the codon-anticodon pair is cognate or near-cognate and hence increases amino acid misincorporation (5).

► **Nucleotide Biosynthesis**

► **Translational Control in Eukaryotes**

References

1. Baranov PV, Gesteland RF, Atkins JF (2002) Recoding: translational bifurcation in gene expression. *Gene* 286:187–201
2. Ibba M, Söll D (2002) Genetic code: introducing pyrrolysine. *Curr Biol* 12:R464–R466
3. Beuning PJ, Musier-Forsyth K (1999) Transfer RNA recognition by aminoacyl-tRNA synthetases. *Biopolymers (Nucleic Acids Sciences)* 52:1–28
4. Ribas de Pouplana L, Schimmel P (2001) Aminoacyl-tRNA synthetases: potential markers of genetic code development. *Trends Biochem Sci* 26:591–596
5. Rodnina MV, Wintermeier W (2001) Ribosome fidelity: tRNA discrimination, proofreading and induced fit. *Trends Biochem Sci* 26:124–130
6. Knight RD, Freeland SJ, Landweber LF (2001) Rewriting the keyboard: evolvability of the genetic code. *Nat Rev Gen* 2:49–58
7. Chiusano ML, Alvarez-Valin F, Di Giulio M et al (2000) Second codon positions of genes and the secondary structure of proteins. Relationships and implications for the origin of the genetic code. *Gene* 261:63–69
8. Hirakata M, Suwa A, Nagai S et al (1999) Anti-KS: identification of autoantibodies to asparaginyl-transfer RNA synthetase associated with interstitial lung disease. *J Immunol* 162:2315–2320

Genetic Counselling

Definition

The following definition of genetic counselling was adopted by the American Society of Human Genetics: “Genetic counselling is a communication process which deals with the human problems associated with the occurrence or the risk of an occurrence of a genetic disorder in the family. This process involves an attempt by one or more appropriately trained persons to help the individual or family to (1) comprehend the medical facts including the diagnosis probable course of the disorder and the available management; (2) appreciate the way heredity contributes to the disorder and the risk of recurrence in specified relatives; (3) understand the

alternatives for dealing with the risk of occurrence; (4) choose the course of action which seems to them appropriate in view of their risk their family goals and their ethical and religious standards to act in accordance with that decision; and (5) to make the best possible adjustment to the disorder in an affected family member and/or the risk of recurrence of that disorder.”

► **Fragile X Syndrome**

► **Peutz-Jeghers Syndrome**

Genetic Distance

Definition

Genetic distance describes the linear distance between genes on a genetic map, measured in Morgan (=100 centi-Morgan); a genetic distance of 1 centi-Morgan between two loci entails that they are expected to be involved, on average, in one recombination per 100 meioses.

► **Genetic Epidemiology**

Genetic Epidemiology

MICHAEL KRAWCZAK

Institute for Medical Informatics und Statistics,
Christian-Albrechts-University Kiel, Kiel, Germany
krawczak@medinfo.uni-kiel.de

Synonyms

Linkage analysis; positional cloning; gene mapping

Definition

Genetic epidemiology is the scientific discipline concerned with the causes and distribution of human diseases in groups of relatives and with the inherited predisposition to disease in populations. In many ways, genetic epidemiology is an interdisciplinary science that has more aspects of population and evolutionary genetics to it than of classical epidemiology. Only recently has the discipline changed its focus to target diseases that are characterized by the small-to-moderate relative risks that are usually the objective of conventional epidemiological research. So far however, genetic epidemiology has been most successful in elucidating the genetic basis of so-called “monogenic” disorders, including cystic fibrosis, Huntington disease and Duchenne muscular dystrophy (Table 1). The

Genetic Epidemiology. Table 1 A list of human disease genes characterized through genetic epidemiological research employing a positional cloning approach

Year	Gene	Disease	OMIM
1986	CYBB	chronic granulomatous disease	306400
1986	DMD	Duchenne muscular dystrophy	310200
1986	RB1	retinoblastoma	180200
1989	CFTR	cystic fibrosis	219700
1990	CHM	choroideremia	303100
1990	NF1	neurofibromatosis type 1	162200
1990	SRY	sex reversal	306100
1990	WT1	Wilms tumour	194070
1991	APC	familial adenomatous polyposis	175100
1991	FMR1	fragile X syndrome	309550
1991	KAL1	Kallmann syndrome	308700
1991	PAX6	aniridia	106210
1992	DMPK	myotonic dystrophy	160900
1992	NDP	Norrie disease	310600
1992	OCRL	Lowe syndrome	309000
1993	ABCD1	adrenoleukodystrophy	300100
1993	ATP7A	Menkes syndrome	309400
1993	ATP7B	Wilson disease	277900
1993	BTK	X-linked agammaglobulinemia	300300
1993	FMR2	fragile X syndrome	309548
1993	GK	hyperglycerolemia	307030
1993	HD	Huntington disease	143100
1993	NF2	neurofibromatosis type 2	101000
1993	PAFAH1B1	Miller-Dieker syndrome	247200
1993	SCA1	spinocerebellar ataxia type 1	164400
1993	TSC2	tuberous sclerosis	191092
1993	VHL	von Hippel Lindau syndrome	193300
1994	BRCA1	hereditary breast and ovarian cancer	113705
1994	DAX1	X-linked adrenal hypoplasia congenita	300200
1994	DRPLA	dentatorubral-pallidoluysian atrophy	125370
1994	EMD	Emery-Dreifuss muscular dystrophy	310300
1994	FGD1	Aarskog-Scott syndrome	305400
1994	FGFR3	achondroplasia	100800
1994	MJD	Machado-Joseph disease	109150
1994	PKD1	polycystic kidney disease type 1	173900
1994	SLC26A2	diastrophic dysplasia	222600

Genetic Epidemiology. Table 1 A list of human disease genes characterized through genetic epidemiological research employing a positional cloning approach (Continued)

Year	Gene	Disease	OMIM
1994	WAS	Wiskott-Aldrich syndrome	301000
1994	XK	McLeod syndrome	314850
1995	ARSE	X-linked recessive chondrodysplasia punctata	302950
1995	ATM	ataxia telangiectasia	208900
1995	BLM	Bloom syndrome	210900
1995	BRCA2	hereditary breast cancer	114480
1995	CAPN3	limb-girdle muscular dystrophy type 2A	253600
1995	CLN3	Batten disease	204200
1995	EXT1	multiple exostoses	133700
1995	OA1	ocular albinism	300500
1995	PHEX	hypophosphatemia rickets	307800
1995	POU3F4	X-linked mixed deafness	304400
1995	PSEN1	early onset familial Alzheimer disease	104311
1995	PSEN2	early onset familial Alzheimer disease	600759
1995	SGCB	limb-girdle muscular dystrophy type 2E	604286
1995	SGCG	limb-girdle muscular dystrophy type 2C	253700
1995	SMN1	spinal muscular atrophy type 1	253300
1996	CHS1	Chediak-Higashi syndrome	214500
1996	CSTB	progressive myoclonus epilepsy	254800
1996	ED1	ectodermal dysplasia type 1	305100
1996	EXT2	multiple exostoses type 2	133701
1996	FANCA	Fanconi anaemia	227650
1996	FRDA	Friedreich ataxia	229300
1996	GPC3	Simpson-Golabi-Behmel syndrome type 1	312870
1996	HFE	haemochromatosis	235200
1996	HPS	Hermansky-Pudlak syndrome	203300
1996	KCNQ1	long QT syndrome type 1	192500
1996	MTM1	X-linked myotubular myopathy type 1	310400
1996	PITX2	Rieger syndrome	180500
1996	PKD2	polycystic kidney disease type 2	173910
1996	PTCH	Gorlin syndrome	109400
1996	RECQL2	Werner syndrome	277700
1996	RPGR	X-linked retinitis pigmentosa	300389
1996	SGCD	limb-girdle muscular dystrophy type 2F	601287
1996	TAZ	Barth syndrome	302060
1996	TCF1	maturity-onset diabetes of the young type 3	600496

Genetic Epidemiology. Table 1 A list of human disease genes characterized through genetic epidemiological research employing a positional cloning approach (Continued)

Year	Gene	Disease	OMIM
1996	TCOF1	Treacher-Collins syndrome	154500
1997	ABCA4	Stargardt disease type 1	248200
1997	AIRE	autoimmune polyglandular syndrome type 1	240300
1997	DIAPH1	non-syndromic autosomal dominant deafness type 1	124900
1997	DYT1	early-onset torsion dystonia	128100
1997	JAG1	Alagille syndrome	118450
1997	MEFV	familial Mediterranean fever	249100
1997	MEN1	multiple endocrine neoplasia type 1	131100
1997	MID1	Opitz syndrome	300000
1997	MYOC	juvenile primary open angle glaucoma	137750
1997	NPC1	Niemann-Pick disease	257220
1997	RS1	retinoschisis	312700
1997	SCA7	spinocerebellar ataxia type 7	164500
1997	SLC26A4	Pendred syndrome	274600
1997	TBX5	Holt-Oram syndrome	142900
1997	TSC1	tuberous sclerosis	191100
1997	UBE3A	Angelman syndrome	105830
1997	ZIC3	situs inversus	306955
1998	CACNA1F	X-linked congenital night blindness type 2	300071
1998	CLN5	neuronal ceroid lipofuscinosis type 5	256731
1998	CTNS	cystinosis	219800
1998	DCX	X-linked lissencephaly	300067
1998	DFNA5	non-syndromic autosomal dominant deafness type 5	600994
1998	DYSF	limb-girdle muscular dystrophy type 2B	253601
1998	EPM2A	progressive myoclonus epilepsy (Lafora)	254780
1998	FCMD	Fukuyama type congenital muscular dystrophy	253800
1998	KCNQ2	benign familial neonatal convulsions type 1	121200
1998	MYO15A	non-syndromic sensorineural recessive deafness type 3	600316
1998	NBS1	Nijmegen breakage syndrome	251260
1998	NPHS1	nephrotic syndrome type 1	256300
1998	PABPN1	oculopharyngeal muscular dystrophy	164300
1998	PARK2	juvenile Parkinson disease	600116
1998	RP2	X-linked retinitis pigmentosa	312600
1998	SH2D1A	X-linked lymphoproliferative disease	308240
1998	STK11	Peutz-Jeghers syndrome	175200
1998	USH2A	Usher syndrome type 2A	276901

Genetic Epidemiology. Table 1 A list of human disease genes characterized through genetic epidemiological research employing a positional cloning approach (Continued)

Year	Gene	Disease	OMIM
1998	VMD2	Best disease	153700
1998	WFS1	Wolfram syndrome	222300
1999	ABCA1	Tangier disease	205400
1999	ATP2A2	Darier disease	124200
1999	CCM1	cerebral cavernous malformations type 1	116860
1999	CLN8	progressive epilepsy with mental retardation	600143
1999	EFEMP1	malattia leventinese	126600
1999	LMNA	Emery-Dreifuss muscular dystrophy	181350
1999	PRG4	camptodactyly-arthropathy-coxa vara-pericarditis syndrome	208250
1999	SLC17A5	Salla disease	604369
1999	SLC19A2	thiamine-responsive megaloblastic anaemia	249270
1999	SLC25A13	adult-onset citrullinemia type 2	603471
1999	SPG4	autosomal dominant hereditary spastic paraplegia type 4	182601
1999	WISP3	progressive pseudorheumatoid dysplasia	208230
2000	AIPL1	Leber congenital amaurosis type 4	604393
2000	ATP2C1	Hailey-Hailey disease	169600
2000	FGF23	autosomal dominant rickets	193100
2000	MCOLN1	mucopolipidosis type 4	252650
2000	MKKS	McKusick-Kaufman syndrome	236700
2000	MTMR2	Charcot-Marie-Tooth disease type 4B	601382
2000	MYH9	May-Hegglin anomaly	155100
2000	NPHS2	idiopathic steroid-resistant nephrosis	600995
2000	PRKAR1A	Carney complex	160980
2000	PVRL1	cleft lip/palate-ectodermal dysplasia syndrome	225000
2000	RELN	autosomal recessive lissencephaly	257320
2000	SPINK5	Netherton syndrome	256500
2000	USH1C	Usher syndrome 1C	276904
2001	BBS2	Bardet-Biedl syndrome	209900
2001	BBS4	Bardet-Biedl syndrome	209900
2001	BSND	Bartter syndrome	602522
2001	CDH23	Usher syndrome 1D	601067
2001	ELAC2	familial prostate cancer	176807
2001	FANCD2	Fanconi anaemia	227646
2001	PRPC8	retinitis pigmentosa 13	600059
2001	SOST	sclerosteosis	269500
2001	USH3A	Usher syndrome 3	276902

Genetic Epidemiology. Table 1 A list of human disease genes characterized through genetic epidemiological research employing a positional cloning approach (Continued)

Year	Gene	Disease	OMIM
2002	BBS1	Bardet-Biedl syndrome	209900
2002	GDAP1	Charcot-Marie-Tooth disease type 4A	214400
2002	RNASEL	hereditary prostate cancer type 1	601518
2002	SMARCA1	Schimke immuno-osseous dysplasia	242900
2002	TMC1	non-syndromic autosomal dominant deafness type 36	606705
2002	TRPM6	hypomagnesemia with secondary hypocalcemia	602014

OMIM: Entry number in "Mendelian Inheritance in Man Online" at ► <http://www.ncbi.nlm.nih.gov/omim/>

approach taken to detect the genetic variants responsible for these diseases is generally known as "positional cloning". Instead of relying upon prior knowledge about the disease-causing biochemical defect(s), positional cloning utilizes the segregation pattern of genetic markers (e.g. SNPs, microsatellites, RFLPs) in affected families to localize the genes involved in a given phenotypic trait ("linkage analysis"). The more often the trait and a particular marker allele are co-inherited by family members, the stronger the evidence that a gene in the vicinity of the marker influences the trait, i.e. that marker and disease gene are linked.

Characteristics

Gene Mapping and Meiotic ► Recombination

Formally, linkage analysis involves the assessment of the recombination fraction θ between two genetic loci like, in the context of genetic epidemiology, a marker and an unknown disease gene. Parameter θ equals the rate at which children receive from a given parent either the grand-maternal allele at one locus and the grand-paternal allele at the other or *vice versa*. Assuming Mendelian inheritance, $\theta = 1/2$ for a pair of genes located on different chromosomes. For two loci residing on the same chromosome, meiotic recombination is only possible *via* "► crossing-over", signifying the breakage and re-union of homologous, non-sister chromatids during the metaphase of meiotic division I (Fig. 1). Indeed, it can be shown mathematically that θ equals exactly half the probability of at least one crossing-over occurring between the two loci in question, provided that some critical assumptions about the randomness of crossing-overs are correct.

In any case, one corollary of the above is that θ represents an increasing function of the physical distance between two loci and therefore provides a key parameter for gene mapping. Unfortunately, θ is not an additive measure of distance since it can never exceed $1/2$. In order to facilitate gene mapping, θ

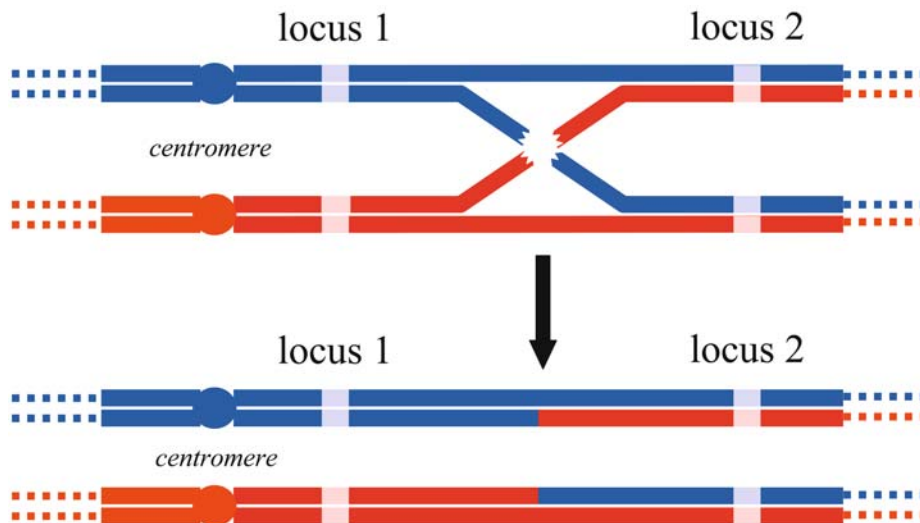
therefore has to be transformed into a linear "► genetic distance" $d(\theta)$ by some ► mapping function (the most widely used being $d(\theta) = 1/2 \ln(1-2\theta)$, proposed by British geneticist J.B.S. Haldane in 1919). Genetic distance is measured in units of Morgan (M) in order to honour T.H. Morgan, the American Nobel prize-winning biologist who first discovered the role of chromosomes in heredity. One centi-Morgan (cM) roughly corresponds to one expected recombination per 100 meioses.

Parametric Linkage Analysis

Linkage analysis has two major objectives, namely (i) to clarify with some statistical confidence whether $\theta = 1/2$ or $\theta < 1/2$, and (ii) to estimate θ in the latter case. Both goals are easily achieved in laboratory animals where controlled breeding can be performed such that, after a few generations, recombinants and non-recombinants can simply be counted. In humans as well as in animals with longer generation times however, linkage analysis has to fall back upon family data. How such data can be used to draw statistical inferences about linkage depends upon their complexity, i.e. on how much prior knowledge is available about the genetic, environmental and stochastic nature of the phenotypes of interest.

For the simple genotype-phenotype relationships encountered with most monogenic disorders, family data can be analysed by explicitly modelling the co-inheritance of the disease and marker in a family, based upon the underlying genotype frequencies and ► penetrances ("parametric linkage analysis"). In such cases, the likelihood L of a given recombination fraction θ_0 between disease gene and marker is a computable function of the phenotypic data D observed in the family. This leads to the definition of $z(\theta_0) = \log_{10} \{L(\theta = \theta_0|D)/L(\theta = 1/2|\Delta)\}$.

Quantity z , termed the "► lod score" and introduced by N.E. Morton in 1955, is used as a sequential statistic to



Genetic Epidemiology. Figure 1 Process of crossing-over during germ cell development. The two nearly duplicated chromosomes align during the late metaphase of meiotic division I, where an overlap and breakage of their constituent non-sister chromatids may occur (red: maternal chromosome, blue: paternal chromosome). Re-annealing and re-synthesis by the cellular repair mechanisms leaves two chromatids with genetic material flanking the site of crossing-over that is not of the same parental origin (i.e. the resulting chromosomes would represent recombinants with respect to the two loci shown).

test whether $\theta < \frac{1}{2}$, and to quantify the evidence in favour or against θ_0 . When $z(\theta_0) > 3$, then linkage between disease gene and marker locus is regarded as being proven and θ is estimated by that recombination fraction that yields the highest lod score. This procedure is exemplified in Fig. 2 for a large family affected by an autosomal dominant disorder. The results of a linkage analysis are usually presented in the form of lod score tables or graphs (Fig. 2b) where *ceteris paribus*, studies of independent (i.e. unrelated) families can be aggregated by summation of the family-wise lod scores. If a disease gene is to be integrated into a pre-existing map of linked markers, then this is most efficiently performed by parametric multi-locus linkage analysis, which has been shown to be up to twice as accurate as the pair-wise approach, measured in terms of the variance of the ensuing recombination fraction estimates.

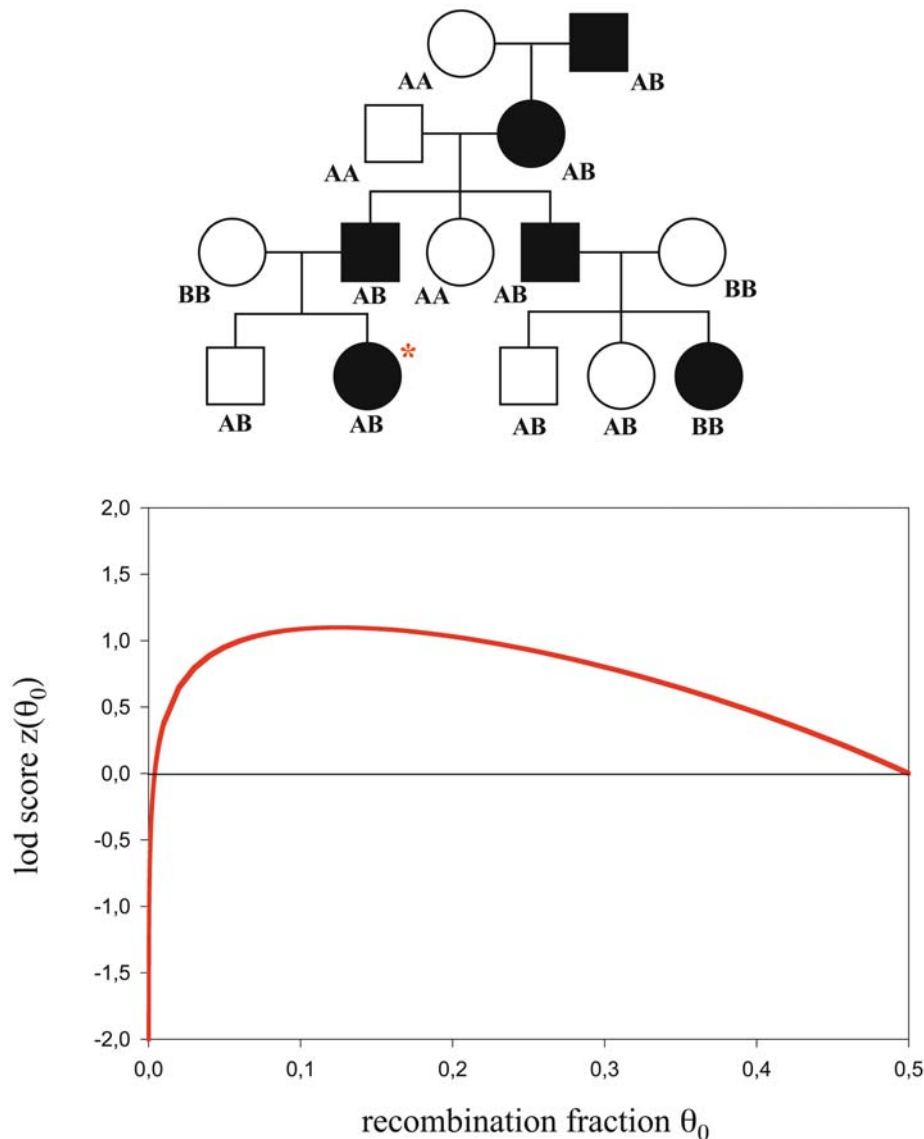
Non-Parametric Linkage Analysis of Complex Diseases

Increasingly, the major challenge to genetic epidemiology is being posed by so-called “complex” diseases, which comprise conditions such as diabetes, heart disease, cancer and psychiatric illness. When compared to monogenic disorders, this category of disease is characterized by

- a substantially higher population frequency,
- the involvement of multiple genes, most probably interacting with one another,

- relatively minor effects exerted by individual variants and
- an important modifying role of environmental factors.

Since no prior information is usually available as to how genetic variation at a given locus modifies the risk for a complex disease (i.e. the genetic model of the disease is unknown), gene mapping for complex diseases has to adopt robust, albeit less powerful, “non-parametric” or “model-free” linkage analysis such as, for example, the study of pairs of relatives. The idea underlying this approach, which is both simple and intuitively appealing, goes back to a 1935 paper by British medical geneticist L.S. Penrose. The number of sib-pairs, out of a total of n independent pairs, who share k parental alleles of an autosomal locus identical by descent (“ibd”) follows a multinomial distribution with parameters z_k , $k = 0, 1$, or 2 (Fig. 3). Under the null hypothesis of no etiological connection between marker and disease, the inheritance of a marker can be assumed to follow Mendelian rules and to be independent of the disease status of the siblings. This implies that, irrespective of whether the sibs are concordant or discordant, $z_0 = \frac{1}{4}$, $z_1 = \frac{1}{2}$ and $z_2 = \frac{1}{4}$. Any test for a deviation from these proportions represents a test for linkage between the marker and a putative disease gene. In its original form, the affected sib-pair test required that the ibd status at each marker be determined unequivocally for all sib-pairs.

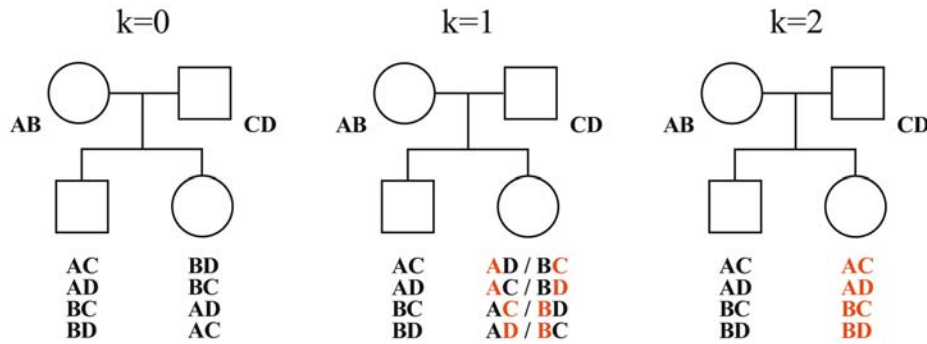


Genetic Epidemiology. Figure 2 Linkage analysis of a family affected by an autosomal dominant disease. (a) shows the pedigree with patients marked by black symbols. Genotypes observed for a biallelic marker locus with alleles “A” and “B” are displayed alongside each individual. (b) is a graphical display of the lod scores as calculated from the family data, assuming full penetrance and lack of *de novo* mutation. The lod score at $\theta_0=0.00$ equals minus infinity owing to a recombination that has occurred during the paternal meiosis leading to an affected girl in the most recent generation (marked by *).

However, a number of derivatives of the test have since been developed which incorporate posterior distributions on k (as inferred from other relatives), utilize relatives other than sibs or are based upon identity-by-state (“ibs”) rather than ibd allele sharing. In general, these methods are less powerful than the original test, but can extract mapping information that would otherwise not be used.

Quantitative Phenotypes

Non-parametric methods have also been devised for the genetic analysis of quantitative phenotypes which, in many instances, may be more powerful than the consideration of dichotomized disease outcomes (“affected” vs “non-affected”) derived from them. The idea underlying the concept of quantitative trait mapping is nevertheless the same as for qualitative characters in



Genetic Epidemiology. Figure 3 Level k of autosomal identical-by-descent (ibd) allele sharing between two sibs. For each value of k (i.e. $k=0, 1$, or 2), shared alleles are marked in red for the second sib.

Genetic Epidemiology. Table 2 Haplotype frequencies of two biallelic loci

Marker 2	Marker 1		Total
	allele 1A	allele 1B	
allele 2A	a	b	a+b
allele 2B	c	d	c+d
Total	a+c	b+d	1

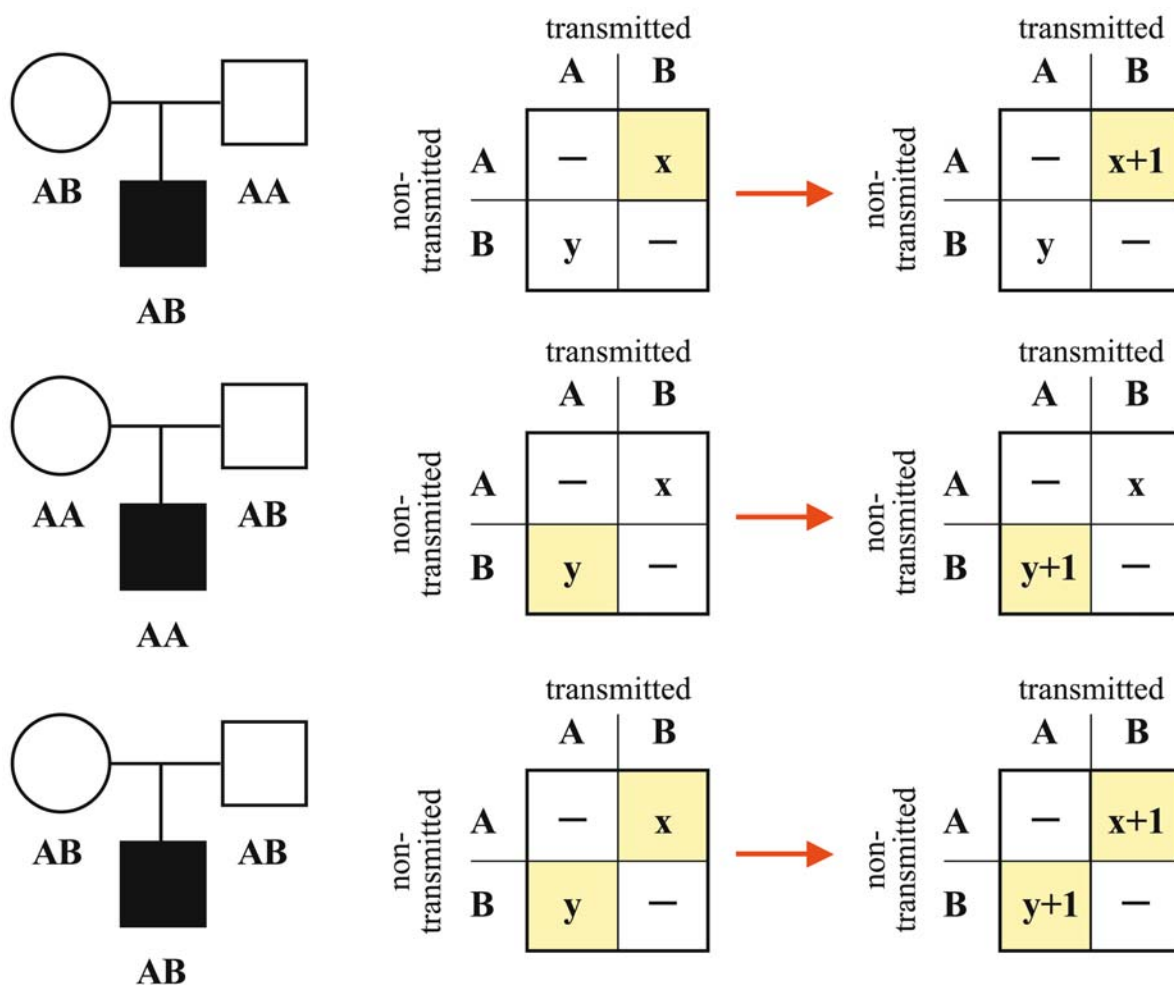
Linkage disequilibrium (LD) is usually measured by $D=ad-bc$. However, since D depends upon the marginal sums (i.e. the allele frequencies), LD is often quantified by $D'=D/D_{\max}$ instead. Here, D_{\max} denotes the maximum absolute value of D that is possible for the same allele frequencies. If $D>0$, then $D_{\max}=\min\{(a+c)(c+d), (a+b)(b+d)\}$; if $D<0$, then $D_{\max}=\min\{(a+b)(a+c), (b+d)(c+d)\}$.

that the observed level of marker ibd sharing between relatives is compared to their phenotypic similarity. If a marker is linked to a gene that influences the trait, then sibs with similar phenotypes, for example, will tend to share more than $1/2$ of their marker alleles ibd whereas dissimilar sibs will not. A formal test for this effect was proposed by American statisticians J.K. Haseman and R.C. Elston in 1972. For each sib-pair, the squared difference Y of their phenotypic values is calculated, and the number π of their ibd marker alleles determined (or estimated). A linear regression analysis of Y on π then reveals (i) whether the two variables are correlated and (ii) whether a significant relationship, if found, is a biologically plausible indication of linkage. The Haseman-Elston approach has since been expanded and refined, for example by considering the mean-corrected product of the sib-specific phenotypes instead of Y , so as to increase informativity about linkage. An alternative method aims at decomposing the variance of the phenotype into components that are due to genes which are either linked to the marker, or not ("variance component analysis"). For a marker

linked to a gene of strong effect, the phenotypic covariance among relatives should be positively related to their degree of marker ibd sharing, and this relationship translates into larger estimates of the corresponding variance components.

► Linkage Disequilibrium

Intervals of 1 Mb are generally regarded as the limit of mapping resolution that can be achieved using (family-based) linkage analysis, and more precise mapping of disease genes can only be expected from population-based association studies, exploiting linkage disequilibrium (LD). The most general definition of LD is the condition that the alleles of linked loci do not occur statistically independent on the chromosomes observed in a population. In the simplest case of two biallelic loci, haplotype frequencies can be arranged in a two-by-two table with cell probabilities a, b, c and d (Table 2), and LD is meaningfully quantified by the cross product $D = ad - bc$. When a new allele first arises in a population by either mutation or migration, it occurs as a single copy that resides on a certain haplotype background, together with certain alleles of other loci. Only in later generations will the allele become more frequent through either selection or genetic drift or both. In any case, chromosomes carrying the new allele will recombine with chromosomes carrying other haplotypes so that the strong original LD will erode with time. This loss of LD will be slower for closely linked loci and, under some simplifying assumptions about mutation rates, migration and mating patterns, D can indeed be shown to decrease by a factor of $1 - \theta$ in each generation. Therefore, strong LD is an indication of close linkage and assessment of LD between a marker and putative disease gene can be regarded as linkage analysis in a super-pedigree tying all analysed individuals together. In principle, any kind of genetic marker can be employed in disease association studies provided that (i) the marker mutation rate is low and (ii) the density of markers chosen for analysis is high enough to ensure



Genetic Epidemiology. Figure 4 Transmission disequilibrium test (TDT) of disease-association for biallelic marker genes. The sampling units of the TDT are nuclear families comprising both parents and an affected child ("trios"). Only transmissions from heterozygous parents to their children are evaluated (cells x and y in the table shown). The TDT statistic equals McNemar's $(x-y)^2/(x+y)$, which follows a χ^2 distribution with 1 degree of freedom under the null hypothesis of no association.

sufficiently strong LD with disease gene(s). Empirical data and theoretical considerations suggest that a sensible marker density should be no lower than approximately 1 in 50,000 nucleotides. Ideally, association markers should be chosen from within genes that represent biologically plausible candidates for an involvement in the disease of interest. The chance of detecting association would then be increased further if marker alleles were themselves of functional significance by altering, for example, the protein product or a regulatory sequence.

Family-based Association Studies

The simplest form of a population-based association study is that invoking a case-control design. As in classical epidemiology, relative risks or odds ratio of particular marker genotypes or haplotypes can be

estimated from the respective frequencies in patients and unrelated healthy controls. However, concerns have arisen over the potentially confounding effects of ethnic, social or geographical population stratification that would generate systematic differences between the genetic characteristics of the two samples, unrelated to disease. To solve this problem, family-based association designs have been proposed of which the transmission disequilibrium test (TDT) is the most widely used. A TDT is basically a McNemar test for preferential transmission of particular marker alleles from heterozygous parents to their affected offspring (Fig. 4). Any deviation of the transmission to non-transmission ratio from the expected 1:1 is indicative of both linkage between marker and disease gene in the presence of LD and of LD in the presence of linkage.

Since chromosomes of close relatives act as internal controls in the TDT and similar tests, it has become almost paradigmatic for the genetic epidemiology of complex disease that family-based association studies are superior to case-control designs. However, family-based designs also have disadvantages that might not always be fully outweighed by their apparent robustness. For example, gene-environment interactions cannot be analysed in family-based studies since no genuine controls are available for comparison to patients. Furthermore, parental genotypes are required for the TDT and these may be difficult to obtain for late-onset disorders. The use of other family members as surrogates to try to reconstruct parental genotypes with some certainty has been suggested in such instances. However, such methods are usually costly, inaccurate and inefficient. On the other hand, possible confounding of data by population stratification can be avoided in case-control studies through careful matching by ethnic and geographic origin. Furthermore, if a sufficiently large set of genetic markers is available that is not itself tested for association, then these markers can be used to estimate the level of population stratification and to correct the employed test statistic accordingly. Finally, although population stratification may represent a theoretical possibility, empirical evidence for its practical importance as a confounding factor in genetic epidemiology is still lacking.

Clinical Relevance

The reasons for the apparent lack of success that plagues genetic studies of complex disease are manifold and the most critical issue is probably the reduction in power caused by genetic heterogeneity. Genetic heterogeneity can occur at two levels, either within genes ("allelic heterogeneity") or between genes ("locus heterogeneity"). In order to increase the power of genetic studies of complex diseases, the major goal in their planning and performance is thus to control for genetic heterogeneity at all levels. First and foremost, this requires a careful choice of the population under study. Ideally for a disease association to be detectable, all copies of the predisposing allele should be *ibd* in patients. The best populations to analyse for LD are therefore those that have been small and isolated for most of their history or that have undergone recent expansion from a small number of founders. Not surprisingly, genetic epidemiology has been particularly successful in Finland and some inward breeding communities in the USA (e.g. Amish, Hutterites, Ashkenazim). In addition to population genetic issues, an appropriate definition of phenotypes, a breakdown by sub-phenotypes and the use of sensible covariates to define etiologically homogeneous sub-populations can help to reduce genetic heterogeneity further. Finally, genetic epidemiology

is a constantly evolving scientific discipline so that improved power may also arise from the development and application of new analytical tools that take genetic and etiological heterogeneity into account (e.g. multi-locus statistics, time series analysis).

Complex human diseases are typically common and have a substantial economic impact upon national health systems. The resulting public interest renders disorders such as cancer, heart disease and diabetes particularly attractive for genetic research and a large number of studies into these diseases are often being performed in parallel. On the other hand, for the reasons mentioned above, the prior probability of successfully mapping and characterizing genes for complex diseases is comparatively low and usually unknown. Even with high significance levels imposed, most positive gene mapping results may therefore be wrong. This implies that genetic epidemiological research will almost inevitably continue to be driven towards the generation of false positive results. Claims as to the elucidation of a disease predisposition should therefore always be received with some caution and judged as preliminary until confirmed by controlled replication, meta-analysis or independent laboratory experiments.

References

1. Elston RC (1998) Linkage and association. *Genet Epidemiol* 15:565–576
2. Schork NJ, Cardon LR, Xu X (1998) The future of genetic epidemiology. *Trends Genet* 14:266–272
3. Risch NJ (2000) Searching for genetic determinants in the new millennium. *Nature* 405:847–856
4. Terwilliger JD, Göring HHH (2000) Gene mapping in the 20th and 21st centuries: statistical methods, data analysis, and experimental design. *Hum Biol* 72:63–132
5. Sham P (2001) Shifting paradigms in gene-mapping methodology for complex traits. *Pharmacogenomics* 2:195–202

Genetic Hearing Disorder

Definition

Genetic hearing disorders mainly comprise of non-X linked genetic hearing loss, which is only observed in homozygous individuals having inherited the mutated gene from each parent.

► [Microvilli](#)

Genetic Heterogeneity

► [Heterogeneity/Heterogenous](#)

Genetic Immunization

Definition

Genetic immunization is a technique to induce specific immune responses by injecting antigen-encoding expression plasmid DNA.

► [DNA-based Vaccination](#)

Genetic Interactions

Definition

Genetic interactions describe interactions between two or more mutations that result in a phenotype.

► [Cell Polarity](#)

Genetic Map

Definition

Genetic map (also known as a linkage map) is a map of a genome, which shows the relative positions (order and distances) of the genes and/or markers on the chromosomes. The map is based on pairwise coinheritance (linkage) of markers. Genetic maps are generally composites of data from many experiments.

► [Chromosome 21, Disorders](#)

► [YAC and PAC Maps](#)

Genetic Modification

Definition

Genetic modification describes the introduction of a new nucleic acid into a cell, organism or micro-organism. In the context of clinical gene transfer, the term is used in conjunction with the transfer of a nucleic acid. This can be achieved by either introducing an expression construct or by modification of cellular nucleic acid e.g. by revision of a point mutation. Nucleic acids used (e.g. for triple helix formation or RNA with ribozyme function that is not part of a transgene) are not meant to lead to genetic modification as far as the term is used in the context of clinical gene transfer.

► [Clinical Gene Transfer](#)

Genetic Polymorphism

Definition

Genetic polymorphism is the presence of multiple inherited forms of a gene with at least an allele frequency of 1% within the population.

► [Pharmacogenomics](#)

Genetic Predisposition to Multiple Sclerosis

ALASTAIR COMPSTON

University of Cambridge Neurology Unit,
Cambridge, UK

alastair.compston@medschl.cam.ac.uk

Definition

► [Multiple sclerosis](#) is a typical complex trait. There is an increased familial recurrence risk and evidence from ► [linkage](#) and ► [association](#) studies for functional ► [polymorphisms](#) that increase disease susceptibility but without Mendelian patterns of inheritance. There is epidemiological evidence for the role of environmental factors in determining the distribution of the disease.

Characteristics

Familial Multiple Sclerosis

Multiple sclerosis has a familial recurrence rate of approximately 15%. Overall, the reduction in risk changes from 3% (relative risk 9) in first-degree relatives to 1% (relative risk 3.4) and 1% (relative risk 2.9) in second and third degree relatives, respectively, compared with a population lifetime rate of 0.3%. At one extreme, the risk of multiple sclerosis is 35% for the monozygotic twin partner of an affected proband, or the children of parents who are both affected (► [conjugal pairs](#)), compared to 0.3% for individuals related by adoption to a proband. Familial clustering is therefore genetically determined.

Cellular and Molecular Recognition The Analysis of Complex Traits

Two methods—linkage and association—underpin the analysis of complex traits. Linkage has low statistical power but operates within families across relatively large genetic distances. Association has high statistical power but is only informative within the boundaries of

►linkage disequilibrium present in the population under study. In founder populations, polymorphisms that increase susceptibility to disease are necessarily located within a large group of linked genes. This block is subject to recombination during subsequent meioses and is gradually whittled down until there is no residual linkage disequilibrium. It seems that genetic factors determine susceptibility and, to some extent, also shape the clinical course. The choice of markers has been driven either by *a priori* guesses on the nature of susceptibility (►candidate genes) or systematic screening of the genome. Candidates selected because they map within linked regions combine both strategies. Markers are analyzed individually (single point analysis) or corrected for information available from their neighbours (multipoint analysis). The cumulative probability that a given marker or region of interest is linked or associated with multiple sclerosis can be formally tested by ►meta-analysis of available studies. The lesson learned thus far is that no one gene makes a major contribution to susceptibility although collectively they determine a relative risk (for siblings) of around 20.

Candidate Genes in Multiple Sclerosis

Much effort has gone into the assessment of candidate susceptibility genes chosen on the basis of prevailing ideas concerning the pathogenesis of multiple sclerosis. Population studies comparing unrelated cases and controls show an association between the class II ►major histocompatibility complex alleles DR15 and DQ6 and their underlying genotypes (DRB1*1501, DRB5*0101 and DQA1*0102, DQB2*0602). This is seen in almost all populations (Caucasian, Oriental, Arab, Hispanic, Finnish, Russian and Jewish) although the strength of the association differs. Even those ethnic groups in which the frequency of multiple sclerosis is low, or the phenotype distinct from that usually observed in northern Europe, are now acknowledged to be primarily DR15 associated with one exception. In Sardinians, the association is with DR4 (DRB1*0405-DQA*0301-DQB1*0302). In some other Mediterranean populations (Canaries and Turkey), the association is with DR2 (DRB1*1501, DQA1*0102, DQB1*0602) and DR4 (DRB1*04, DQA1*03, DQB1*0302). Most investigators assume that—based on the genetics and obvious candidature through its role in restricting the immune response—DR (or DQ) is itself the susceptibility gene encoded at 6p21.

Outside the major histocompatibility complex, many candidates have been screened. The case made on the basis of ideas concerning the pathogenesis of multiple sclerosis is much strengthened by prior knowledge that the candidate gene also maps to a region already implicated by linkage studies (positional candidates). The range of candidates now studied includes adhesion

molecules, immune receptors, accessory molecules, cytokines, chemokines and their receptors or antagonists, structural genes of oligodendrocytes or myelin, and molecules regulating cell death and survival. In a small and regionally restricted population of Finns, multiple sclerosis is associated and linked to the gene for myelin basic protein, encoded on chromosome 18. The effect can be traced to a subset of families with common ancestry and does not hold up in the larger cohort. It is the nature of screening so many potential effects that a proportion will appear to be associated or linked but by chance. Equally, it would be difficult to show unambiguously that one or more genes exerting a small biological effect is making a contribution to susceptibility in a single study. That said, some plausible associations or linkages have been provisionally reported, although no one of these has yet stood up to repeated replication. These mainly involve factors that appear to increase susceptibility to multiple sclerosis, but there is provisional evidence for primary effects on resistance and effects on severity or clinical features of the disease.

DR15 is associated with younger age at diagnosis and female gender but does not distinguish features relating to disease course, outcome, specific clinical features or paraclinical investigations. This suggests that DR15 exerts its effect on susceptibility rather than modifying the course of multiple sclerosis. Loci apparently associated with disease protection are FAS-670, IL-12p40, FcR and MCP-3. Genes that may influence the course or phenotype of multiple sclerosis include CTLA4, IL-1Ra/IL-1B, IL-2, CCR5, oestrogen receptor, CNTF and Apo-E and mutations of mitochondrial DNA.

Linkage Genome Screens

The dividend from attempting to fast-track the solution to susceptibility in multiple sclerosis by the candidate gene approach has been small but the problem is also not solved by the nine whole genome linkage analyses using variable numbers of families from the United States, Canada, United Kingdom, Finland, Sardinia, Italy, Turkey, Scandinavia and Australia. These screens have involved between 21–225 families each typed for 257–443 microsatellite markers chosen to provide an average spacing of around 10 centiMorgans. Although several new genomic regions of interest were revealed, many are false positives. These whole genome screens have been used to explore regions of interest in more detail hoping to consolidate their status based on mapping but without picking out positional candidates. Linkage on chromosome 17q is supported by additional positional screens from Denmark, Canada, and Finland. There is collateral support for the involvement of chromosomes 5p, 7p and 12q based on direct evidence and synteny with genes determining susceptibility to experimental forms of demyelination.

Meta-analysis has been deployed in the expectation that this will reduce the evidence for false positive peaks and strengthen the candidature of those which are genuine providing the best guide to shared regions of interest as the map is serially up-dated. This was last completed in 2005.

Whole Genome Association Screening

Until recently, whole genome linkage disequilibrium mapping was considered impractical and dependent on chance co-localisation of susceptibility genes and markers applied randomly and at low density. This situation changed with the increased availability of widely distributed microsatellite markers and is set to increase further with the identification and mapping of [▶single nucleotide polymorphisms](#). A first pass at screening the genome for association was completed in 2003 based on a 0.5 cM map of microsatellite markers and using DNA pools derived from cases with multiple sclerosis and unrelated controls. Individual results provided provisional evidence for associations based on linkage disequilibrium outside the major histocompatibility complex on 6p21. The number of micro-satellite markers used necessarily made this a low-density screen, especially since the number of informative markers was less than the full set of 6000 used in this Genetic Analysis of Multiple sclerosis in EuropeanS (GAMES). With considerable variation depending on the stochastic nature of linkage disequilibrium on individual chromosomes in European populations, it may only have covered 10% of the genome in detail and another 20% in part, leaving much yet to be explored. Perhaps its main value lies in the exclusion of many microsatellite markers lying in blocks of linkage disequilibrium of varying size rather than in the provisional positive associations. New screens based on single nucleotide markers present on individual chips, and at a much higher density are now in progress.

Future Strategies for Identifying Susceptibility Genes

Once regions of interest are mapped, the next aim is to move from whole genome screening to the identification of functional polymorphisms which condition one component or another of the disease process and determine variations in the clinical course and features. How to reach that position is less clear and several parallel strategies have been suggested. One is to add incrementally to the number of available families until thresholds for linkage are reached for the identification of secure loci using statistical criteria for genome wide significances. An alternative is to accept that the combination of linkage and association now available is sufficient to concentrate the search for positional candidates within regions of interest. Each

provisional site already offers several interesting possibilities, although the number of genes encoding components of the nervous, immune and signalling systems is such as to make practically any region suggestive with respect to sensible candidates. Rapid progress is being made in characterising the whole genome for the size, distribution and diversity of blocks containing a restricted number of haplotypes. If the preliminary evidence holds up, it will be possible to tag the common variants within each block in populations (such as Europeans) retaining significant linkage disequilibrium and screen individuals for the susceptibility haplotypes with relative economy.

Clinical Relevance

[▶Concordance](#) within families can be used to gauge the influence of genetic factors in determining the clinical phenotype of multiple sclerosis. Time to reach the later stages of disability does not differ between familial and sporadic cases. Conjugal pairs show no evidence for clinical concordance, clustering at year of onset or distortion of the expected pattern of age at onset in the second affected spouse. The most recent assessment of concordance in co-affected siblings and parent-child pairs supports a role for genetic factors in determining age at onset and progression either from onset or after a phase of relapsing remitting disease, but not the initial presentation or disability. Concordant parent-child pairs show no distortion in the random distribution of male-female pairings and neither sex nor line of inheritance influence disability, age at onset or course. In this situation, disability is highest in the male offspring of affected fathers, who more commonly follow a primary progressive course.

The risk of [▶autoimmunity](#) is increased in the relatives of probands with multiple sclerosis. Three surveys, together involving around 4000 relatives of 1000 probands, have shown recurrence of multiple sclerosis in 15% with another autoimmune disease (Graves' disease, rheumatoid arthritis and diabetes) in about 5% of pedigrees. Several other disorders have been considered more frequent than expected in patients with multiple sclerosis. None of these is entirely secure but there may be co-morbidity between neurofibromatosis 1 and primary progressive multiple sclerosis.

A minority of patients who meet clinical criteria for the diagnosis of multiple sclerosis and in whom there are associated magnetic resonance imaging abnormalities and cerebrospinal fluid oligoclonal bands have an illness in which there is disproportionate involvement of the anterior visual pathway. These are commonly women with male relatives already known to be affected by [▶Leber's hereditary optic neuropathy](#) and they have pathological mutations of mitochondrial

DNA. The clinical features of demyelinating disease seen in Orientals and Africans are distinct and provide another example of clinical heterogeneity. In Japan, multiple sclerosis shows either a Western phenotype, in which a number of sites are involved, or an optico-spinal pattern in which the clinical picture is dominated by involvement of visual and spinal cord pathways with a specifically different genetic background (HLA-DP*1501 rather than DRB1*1501 seen with the Western phenotype). However, recent reports of multiple sclerosis in Japanese highlight the previously under-reported extent of the so-called Western phenotype. Demyelinating disease is considered extremely rare in Africans, but a number of cases are described and the phenotype is typically a severe illness dominated by one or more episodes usually affecting the anterior visual pathway and spinal cord—again combining the anatomical features of ►Devic's disease with the clinical course of moderately severe relapsing remitting multiple sclerosis. ►Phenocopies may confuse the analysis of complex traits where diagnosis depends on pattern recognition of symptoms, signs and laboratory investigations in the absence of a test for the disease. Reassuringly, one large cohort screened for other diseases was shown not to be contaminated by cases of ►CADASIL, ►spinocerebellar degeneration, or ►adrenoleukodystrophy.

Conclusions

Six main categories of susceptibility genes can be predicted: genes which determine susceptibility to the process of inflammation across a range of disorders – the *autoimmune* genes; those which determine the specificity of that process for the development of multiple sclerosis – the *ubiquitous* genes; those which are relevant for the pathogenesis in isolated populations – the *domestic* genes; those which determine particular phenotypes – the *pleiotropic* genes; those which determine variations in the clinical course – the *modifying* genes and those which cluster to provide specifically different (heterogeneous) contributions to the pathogenesis – the *epistatic* genes. A major part of future studies will be to resolve the question of disease heterogeneity in multiple sclerosis. When eventually in place, the potential of this genetic knowledge for improved understanding of the pathogenesis of multiple sclerosis and designing novel treatments is considerable. Resolving the issues of complexity and heterogeneity in multiple sclerosis and other complex traits has practical dividends. Without knowledge linking aetiology to pathogenesis and phenotype, putative new treatments will continue to be screened in cohorts who may or may not have an appropriate pathological substrate for that particular intervention.

References

1. Ebers GC, Bulman DE, Sadovnick AD et al (1986) A population based study of multiple sclerosis in twins. *N Engl J Med* 315:1638–1642
2. Gabriel SB, Schaffner SF, Nguyen H et al (2002) The structure of haplotype blocks in the human genome. *Science* 296:2225–2229
3. Harding AE, Sweeney MG, Brockington M et al (1992) Occurrence of a multiple sclerosis-like illness in women who have a Leber's hereditary optic neuropathy mitochondrial DNA mutation. *Brain* 115:989–989
4. Risch NJ (2000) Searching for genetic determinants in the new millennium. *Nature* 405:847–856
5. Sawcer S, Maranian M, Setakis E et al (2002) A whole genome screen for linkage disequilibrium in multiple sclerosis confirms disease associations with regions previously linked to susceptibility. *Brain* 125:1337–1347
6. The Transatlantic Multiple Sclerosis Genetics Cooperative (2001) A meta-analysis of genome screens in multiple sclerosis. *Mult Scler* 7:3–11

Genetic Redundancy

Definition

Genetic redundancy refers to the presence of genes in multiple forms in the DNA of eukaryotes. Two or more genes are capable of executing the same tasks, thus eliminating one gene's function. It does not alter the development of function.

- *Drosophila* Model of Cardiac Disease
- Muscle Development

Genetic Screen

Definition

Genetic screening is analysing a group of individuals to identify those who are at high risk of having or passing-on a specific genetic disorder. In experimental genetics, the term describes an approach to identifying genes and gene functions by random gene knock out and identifying genes causing a specific phenotype; or vice versa, identifying phenotypes caused by specific gene knock out.

- Genetic Screening in Populations
- Mutagenesis Approaches in the Zebrafish

Genetic Screening in Populations

VILMUNDUR GUDNASON

Icelandic Heart Association Heart Preventive Clinic
and Research Institute, Kopavogur, Iceland
v.gudnason@hjarta.is

Definition

“Genetic screening may be defined as any kind of test performed for the systematic early detection or exclusion of a genetic disease, the genetic predisposition or resistance to a disease or to determine whether a person carries a gene variant which may produce disease in offspring. Screening may be concerned with the general population or with specific sub-populations defined on some basis other than their health” (from the European Society of Human Genetics, ►<http://www.eshg.org/>).

Characteristics

Population screening has been applied in the newly born for a long time for medical conditions known to be preventable by medical intervention, such as phenylketonuria. In more recent times considerably more knowledge on underlying genetic defects for various diseases has become available. What conditions to screen for, medically, biochemically or genetically has been a matter of discussion for decades. Most common diseases have complex multifactorial and probably ►polygenic underlying causes. For that reason, genetic screening for complex diseases in populations is not really feasible for the time being. However, a number of genetic variations have been associated with the risk of developing disease and may be applied as specific tests in the evaluation of risk in groups of patients identified by a given disease. Despite that, only ►monogenic diseases or disorders fit the criteria set for appropriate population screening including genetic screening.

Wilson and Jungner (1) put forth ten criteria, which should be considered before population wide screening should be set up. These criteria are that:

1. the condition being screened for should be an important health problem.
2. the natural history should be well understood
3. there should be a detectable early stage
4. treatment at an early stage should be of more benefit than at a later stage
5. there should be a suitable test for identifying people at the early stage
6. the test should be acceptable
7. intervals for repeating the test should be determined

8. there should be adequate health service provision for the extra clinical workload resulting from the screening
9. the risk of screening, both physical and psychological should be less than the benefits
10. the costs should be balanced against the benefits

There are not many diseases or disorders that fulfil these criteria completely and far from all monogenic diseases do so. ►Familial hypercholesterolemia (FH) is one condition, which does fit exactly. FH is a monogenic disorder that is due to a defect in the low density lipoprotein (*LDL*) receptor gene with a lifelong elevation of blood cholesterol (2). It has a prevalence of 1 in 500 in most populations and a high risk of premature coronary artery disease and can be described as an important health problem. The natural history of FH is well understood and the disorder can be detected early in life. Treatment is available and there is strong evidence that early treatment is beneficial. The diagnosis of FH, whether performed by clinical or genetic testing does not require more complex intervention than venipuncture. Once the diagnosis is made, there is no need for further diagnosis. Considerable experience in screening for FH has accumulated in many populations and FH can be looked at as a paradigm for screening for other monogenic disorders or diseases.

Cellular and Molecular Regulation

Familial hypercholesterolemia is caused by a mutation in the ►*LDL receptor gene*. The ►LDL receptor takes up cholesterol from the bloodstream by binding LDL cholesterol particles, mainly in the liver and thus reduces blood levels of cholesterol. Individuals who lack LDL receptor or have a reduced number of functional receptors demonstrate considerably increased levels of blood cholesterol. FH is an autosomal dominant disorder with a gene dosage effect. This means that heterozygotes for a mutation in the *LDL receptor gene* frequently have double the amount of blood cholesterol compared to individuals from the general population and homozygotes or ►compound heterozygotes have several fold increase in blood cholesterol.

A variety of different mutations in the *LDL receptor gene* have been found to cause FH. These can be accessed on the FH web site www.ucl.ac.uk/fh/. The sheer number of mutations makes the identification of FH complicated. Most mutations are confined to a single or a few families, but others are more widespread. However, there are a number of mutations that have been found to be population specific allowing for population based screening for a given set of mutations. There are two main approaches for identifying previously unidentified individuals with FH; first, examination of first-degree relatives for a given index case and secondly, genealogical tracing in defined

populations to a common ancestor for a known mutation in the *LDL receptor gene*.

First Degree Relative Screening

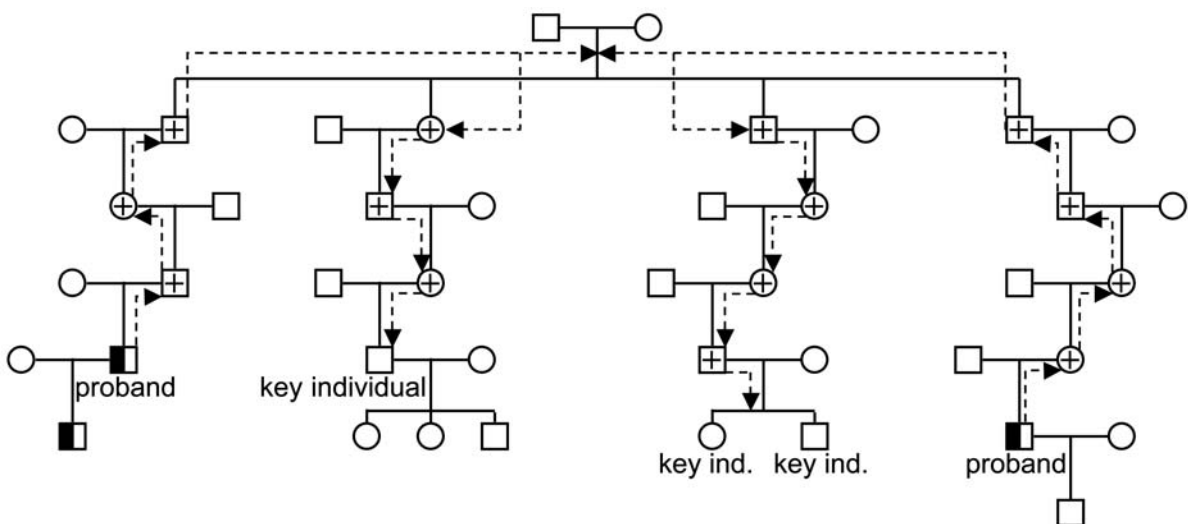
The conventional approach to identify new FH patients is to ask FH index cases for informed consent to contact their first-degree relatives. This is the method recommended by the Med Ped project (Make Early Diagnosis – Prevent Early Deaths in Medical Pedigrees) (<http://www.medped.org/>). The main advantage of such an approach is the high probability of diagnosing family members carrying the LDL-receptor mutation when contacting close relatives. The probability is 50% for first-degree relatives of the index case (parent, sibling and child) and then declines by about half for each generation. In the Netherlands, experience from FH screening using similar first degree relative approaches, revealed 2039 individuals identified as **heterozygous** FH of 5442 relatives tested (3). The main disadvantages are that the doctor needs to rely on consent from the patient to contact his/her relatives and that pedigrees based on information from the patients are seldom complete for more than one or two generations back. Therefore the search for new patients can reach a blind end.

Genealogy Tracing for Identification of Affected Individuals

The genealogy tracing approach is applicable where there is a common mutation in a specific population in an isolated area where genealogical information is available either from church records or other similar

records or genealogical information is electronically available such as in the example from Iceland described below (4).

A common mutation in the *LDL receptor gene* (14T +2C) has been identified and found to be responsible for up to 60% of FH in Iceland. For the genetic screening only probands with this mutation were included. These probands were genealogically traced to common ancestors by The Icelandic Genetic Council's family tracing office. The family tracing was performed through a partly computerized database derived from censuses (first carried out in Iceland 1703), church records and birth and marriage certificates. Once a common ancestor had been identified, a list of all descendants was produced. The oldest individual alive in each family lineage was identified as key individual and contacted for cholesterol measurements and for genetic testing (Fig. 1) after obtaining informed consent. If positive for the common mutation, his or her offspring were recruited for testing. Relatives of key individuals negative for the common mutation were not recruited. Fourteen probands positive for the common mutation were genealogically traced to four family clusters, one cluster with four probands, one with three probands and two with two probands each. The ancestors for the clusters were born in the late 18th century and early 19th century and were traced back for 3 and 4 generations. Three of the probands could not be linked to any other proband. The tracing revealed 2201 live individuals in the four family clusters and of these, 364 (17%) key individuals were identified (Fig. 1). Three



Genetic Screening in Populations. Figure 1 The pedigree shows how two individuals with FH (probands) are traced to a common ancestor (upwards tracing arrows). The oldest individual alive in each family lineage was identified as key individual and contacted for genetic testing (downward tracing arrows). Offspring of the key individuals positive for the mutation were recruited for testing. If positive, their offspring were also called in and so on. Relatives of key individuals negative for the mutation were not recruited. + means a deceased individual and black filling an affected individual.

hundred and six key individuals (84%) responded. Thirty five (11%) of the 306 key individuals who responded were positive for the common mutation or nearly one in every 9 key individuals tested. This yield is a fifty-six-fold enrichment from the 1 in 500 yield of screening the general population. No homozygotes were detected. Of the 35 positive key individuals, seven had not been diagnosed before.

This demonstrates that screening extended families is a feasible approach for achieving the goal of finding individuals with FH previously unknown to have this treatable condition. This approach may well be practical in other populations where genealogical information is available.

Clinical Relevance

Familial hypercholesterolemia is a condition with considerably elevated risk of developing coronary artery disease, which may eventually lead to a heart attack. A report from The World Health Organization (WHO) shows the mean age of onset for coronary heart disease in untreated individuals to be 45–48 years in males and 55–58 in females (5). This enormously raised risk of a heart attack that is potentially preventable calls for an active search for and identification of affected individuals at an early age and an aggressive treatment of all known risk factors for coronary heart disease. The WHO conference in Paris 1997 (5) urged an early diagnosis and treatment of individuals with FH. The main challenge is to prevent premature atherosclerosis in individuals with FH.

In recent years a new class of drugs for the treatment of hypercholesterolemia has become available. These are HMG CoA reductase inhibitors called statins, which directly affect intracellular production of cholesterol and hence lead to an increase in the number of LDL receptor molecules on the cell surface of the liver. This in turn leads to enhanced uptake of cholesterol rich particles from the circulation with a corresponding reduction in the level of blood cholesterol. It has been demonstrated that cholesterol-lowering drugs are effective in reducing coronary stenosis assessed by coronary angiography in patients with FH (6) and there is evidence for improved survival of patients with FH in recent years especially after the introduction of statin therapy (7).

The prognosis for primary prevention of coronary artery disease in heterozygous FH patients is excellent and for that reason it is a major important challenge to identify undiagnosed or inadequately treated individuals.

The above example of genetic screening fulfils all the criteria for screening in populations set by Wilson and Jungner in 1968 (1). There are a number of monogenic diseases that may benefit from the experience obtained from the effort of identifying new FH patients by systematic search.

References

1. Wilson J, Jungner YG (1968) Principles and practice of mass screening for disease. Public Health Papers 34, Report 65 (iv) WHO, Geneva
2. Goldstein JL, Hobbs HH, Brown MS (1995) Familial Hypercholesterolemia. In: Scriver CT, Beaudet AL, Sly WS (eds) et al The Metabolic and Molecular Bases of Inherited Diseases, McGraw-Hill, New York, pp 1981–2030
3. Eckenhausen MA, Defesche JC, Sijbrands EJ et al (2001) Review of first 5 years of screening for familial hypercholesterolaemia in the Netherlands. *Lancet* 357:165–168
4. Thorsson B, Sigurdsson G, Gudnason V (2003) Systematic Family Screening for Familial Hypercholesterolemia in Iceland. *Arterioscler Thromb Vasc Biol* 23:335–333
5. Report of a WHO Consultation (1997) Familial Hypercholesterolaemia (FH). WHO Human Genetics Programme, Geneva
6. Kane JP, Malloy MJ, Ports TA et al (1990) Regression of coronary atherosclerosis during treatment of familial hypercholesterolemia with combined drug regimens. *JAMA* 264:3007–3012
7. Scientific Steering Committee on behalf of the Simon Broome Register Group (1999) Mortality in treated heterozygous familial hypercholesterolaemia: implications for clinical management. *Atherosclerosis* 142:105–112

Genetic X-Linked Disease

Definition

A genetic x-linked disease is determined by mutation of a gene located on the X chromosome.

► **Microvilli**

Genetically Engineered Animals

► **Transgenic and Knock-out Animals**

Genome

Definition

The term genome refers to the complete set of genetic information contained in an organism or a cell, which includes both the chromosomes within the nucleus and in mitochondria.

- ▶ Biochemical Engineering of Glycoproteins
- ▶ Chromosome 21 Disorders
- ▶ COPD and Asthma Genetics
- ▶ Functional Genomics, the Systematic Analysis of Gene Function of all Genes and Gene Products in Parallel
- ▶ Protein Microarrays as Tool for Protein Profiling and Identification of Novel Protein-Protein Interactions

Genome Analysis in Plants

- ▶ Plant Genomics

Genome Engineering

- ▶ Cre/loxP Strategies

Genome Functionalization by Arrayed cDNA Transduction

- ▶ Automated High Throughput Functional Characterization of Human Proteins

Genome Instability

Definition

Genome instability describes processes in cells that accumulate mutations with high frequency. These mutations include point mutations, insertions, deletions and translocations.

- ▶ Chromosomal Instability Syndromes
- ▶ DNA-Repair Mechanisms

Genome Scan

Definition

Genome scan refers to a genetic research method in which the entire DNA of an organism is searched systematically for locations on the chromosomes that are inherited in the same pattern as a specific trait. This method is usually applied to collections of families that show multifactorial inheritance of specific traits, such as type 1 diabetes.

- ▶ Diabetes Mellitus, Genetics
- ▶ Manic Depression

Genome Screen

Definition

Genome screen describes the testing of a population group to identify a subset of individuals at high risk for having or transmitting a specific genetic disorder, by using several hundred markers selected from the whole genome to identify chromosomal regions that are co-inherited (linked) with a specific disease.

- ▶ Atopy Genetics
- ▶ COPD and Asthma Genetics
- ▶ Genetic Screening in Populations

Genome Walking

Definition

Genome walking is a local physical mapping technique for obtaining unknown DNA regions on either side of chromosomal regions of known nucleotide sequences.

- ▶ YAC and PAC Maps

Genome-Wide Analysis

Definition

A genome-wide analysis is the systematic investigation of all regions of the genome to determine those polymorphisms more often associated with a disease.

- ▶ Common Diseases, Genetics

Genomic Analysis of Single Disseminated Cancer Cells

CHRISTOPH A. KLEIN

Institute of Immunology, Ludwig Maximilian University, Munich, Germany
christoph.klein@med.uni-muenchen.de

Definition

The term “disseminated cancer cells” refers to cells that originate from a malignant primary tumour and are found at ectopic sites as single cells or small cell clusters. More than 80% of human malignant tumours stem from epithelial tissues (such as mammary glands, lung or gastro-intestinal organs). The major cause of death for patients that suffer from these types of cancer (carcinomas) is metastasis, i.e. emergence of tumour colonies that are found in organs distant from the primary site of tumour growth. Today many cancer patients are diagnosed before metastatic disease is detected by clinical imaging techniques (such as X-ray or CT scans) and will be submitted to surgery of their primary tumour. A substantial percentage of these patients, however, will eventually succumb to metastasis months, years or even decades after initial diagnosis of their primary cancer, indicating that tumour cells left the primary tumour before the surgeon removed it, and that some of these cells have the potential to found a metastasis. Therefore, great effort was undertaken to detect the precursor cells of later arising metastasis at the time of surgery in order to develop means for prevention of later arising metastasis. For cancers of epithelial origin, detection of disseminated cancer cells is often based on histogenetic markers that enable discrimination from the surrounding cells, i.e. identification of epithelial cells in purely mesenchymal organs. Clinically accessible mesenchymal organs are blood, bone marrow or lymph nodes. Thus the applied markers (e.g. epithelial cytokeratins) do not detect tumour cells directly but epithelial cells that are usually not found in mesenchymal organs of donors without epithelial malignancy. Clinical studies revealed that the finding of cytokeratin positive cells (with most studies performed on bone marrow or lymph node) at the time of surgery puts carcinoma patients at high risk for the development of metastasis later on. Depending on tumour type and clinical stage, approximately every third carcinoma patient without manifest metastases harbours disseminated cancer cells in bone marrow or macroscopically tumour-free lymph nodes. However, absolute and relative tumour cell

frequencies are extremely low, being about one or two cancer cells per two million bone marrow cells. Thus, disseminated cancer cells belong to the rarest cells in the human body and their molecular-genomic characterization requires special techniques.

Characteristics

(Whole) genome analysis of single disseminated cancer cells

The first studies that characterized cytokeratin-positive cells intended to confirm their malignant nature. It was shown by interphase fluorescence in-situ hybridisation (►FISH) that some cytokeratin-positive cells harbour chromosomal abnormalities. Then, protein expression of a variety of cancer-associated molecules was tested on cytokeratin-positive cells. Although sporadic insights into the biology of disseminated cancer cells could be obtained, double staining (or labelling in the case of FISH) was very cumbersome because of the extreme rarity of the investigated cells.

A single diploid cell contains about 6 pg genomic DNA. Multiple molecular-genetic analyses of such minute amounts therefore require amplification. To this end several methods have been developed with the most frequently used being the ►DOP-PCR (degenerate oligonucleotide-primed ►PCR) and the PEP (primer extension, preamplification) method or derivatives thereof. These methods use mixtures of degenerate or random primers to amplify the whole genome, which leads to the problem that it is impossible to control for equal binding of the primers to complex sequences for unbiased amplification. However unbiased amplification is mandatory for the application of whole genome screening techniques such as comparative genomic hybridization that measures numerical aberrations in tumour cell genomes. To circumvent this problem an adaptor-linker approach was developed. Here, the genome is cut into small fragments (about 100–2000 bp) by a frequently cutting restriction enzyme. Then, adaptors are ligated to both ends of the fragments and the fragmented genome is subsequently amplified using a single primer that binds to the adaptor and thereby amplifies all fragments alike.

The approach is increasingly applied to disseminated cancer cells of various types of tumours and the results have changed the prevailing view on metastatic progression. Of the many interesting findings perhaps the most important are that dissemination occurs often earlier than previously thought. The cancer cells often display less or different genomic aberration than their matched primary tumours. Thus metastases and primary tumours seem to develop independently to a large degree.

Gene expression analysis of single disseminated cancer cells

With the completion of the human genome project and the introduction of technologies such as DNA microarrays and laser microdissection, many fields in biology and medicine await the application of comprehensive gene expression analyses of specific cell types isolated from defined tissues. For the amplification of single cell mRNA the first protocols were introduced in the late eighties and early nineties of the last century. So far the protocols are based on either of two principal approaches, linear amplification by T7 RNA polymerase or PCR-based amplification.

As a general rule, PCR-based methods are easier to handle and less time consuming, while there are concerns about the quantitative reliability of measurements obtained after exponential amplification. The linear amplification achieved by T7 RNA polymerase, also referred to as the Eberwine protocol, has the advantage that a potentially occurring failure to amplify a given transcript will not be exponentially transmitted. Here, mRNA is transcribed by a primer containing the promoter of the T7 RNA polymerase. After **►cDNA** synthesis, in-vitro transcription is performed and the procedure is repeated once or twice. The effect of the few cycles is thought to change only marginally – if at all – the original template ratios. On the other hand, several groups have observed that the relative abundance of transcripts is also preserved by PCR-based methods – provided that the correct conditions are applied.

Our preferred method belongs to the approaches using PCR. The protocol uses a single primer that binds to two binding sites artificially introduced to all mRNA sequences. First, a poly-C flanking region is incorporated during cDNA synthesis and after reverse transcription a poly G-tail is added. Four aspects seem to be particularly important. Firstly, single cell mRNA is bound to a solid phase enabling the change of buffers and thereby always optimal conditions for each enzymatic reaction. Secondly, random primers for cDNA synthesis reduce the length of primary transcript and allow for subsequent amplification within the optimal range for PCR. Thirdly, a poly-G tail provides a much better primer binding site than a poly A or poly T tail. Fourthly, introducing a poly-C flank on one side of the template and a poly-G tail on the other makes all sequences equally G/C-rich at their primer-binding site. Adequate conditions for a single poly-C PCR primer, i.e. high annealing temperature and the addition of denaturing agents such as formamide enable highly specific and unbiased amplification of such sequences. With this amplification method in hand, gene expression profiling of single cells has become possible and first interesting results have been obtained. For example, we found that the minor histocompatibility

antigen HA-1 is aberrantly expressed on single disseminated cancer cells. This finding makes it reasonable to apply allogeneic bone marrow transplantation as immunotherapy in a HA-1 mismatch situation. We recently adopted the protocol for high-density oligonucleotide microarrays. Thus, screening of all expressed human genes may reveal new target structures on single disseminated cancer cells, the precursor cells of lethal metastasis.

Clinical Relevance

Clinically manifest metastatic disease can rarely be cured. One likely reason is that the cancer cells have genomically and phenotypically progressed so far that they are highly resistant against current ways to induce apoptosis by any type of treatment, another that the tumour burden is just too large for complete tumour cell eradication at tolerable drug doses. Therefore, systemic therapies are added to loco-regional treatment (e.g. surgery or irradiation therapy) before metastasis becomes manifest. Such therapies target the relatively few tumour cells that spread throughout the body, have therefore been called “adjuvant” and are currently in the centre of clinical efforts. The underlying rationale is to destroy the tumour seed timely when the tumour load is low and the cells still vulnerable. However, adjuvant chemotherapies, which are currently the best characterised and most effective, have not fulfilled the hopes so far. Although several therapy regimens improve significantly the overall- and the disease-free survival of the patients, the absolute benefit is rather low being in the range of few percent and improving survival time for the individual patient by few months. It is increasingly recognized that one reason for the failure of adjuvant therapies is the almost complete lack of knowledge about the target cells – the disseminated cancer cells. Disseminated cancer cells are genomically and phenotypically often very different from their matched primary tumours. Therefore, therapies that are based on mechanisms active in primary tumours do not necessarily exert an effect on disseminated cancer cells. Rather, direct analysis of single disseminated cancer cells promises to uncover novel molecular targets for effective adjuvant therapies.

References

1. Braun S et al (2000) Cytokeratin-positive cells in the bone marrow and survival of patients with stage I, II, or III breast cancer. *N Engl J Med* 342:525–533
2. Cole BF, Gelber RD, Gelber S et al (2001) Polychemotherapy for early breast cancer: an overview of the randomised clinical trials with quality-adjusted survival analysis. *Lancet* 358:277–286
3. Iscove NN et al (2002) Representation is faithfully preserved in global cDNA amplified exponentially from sub-picogram quantities of mRNA. *Nat Biotechnol* 20:940–943

4. Klein CA et al (1999) Comparative genomic hybridization, loss of heterozygosity, and DNA sequence analysis of single cells. *Proc Natl Acad Sci USA* 96:4494–4499
5. Klein CA et al (2002) Combined transcriptome and genome analysis of single micrometastatic cells. *Nat Biotechnol* 20:387–392
6. Klein CA (2003) The systemic progression of human cancer: A focus on the individual disseminated cancer cell – the unit of selection. *Adv Cancer Res* 89:35–67
7. Telenius H et al (1992) Degenerate oligonucleotide-primed PCR: general amplification of target DNA by a single degenerate primer. *Genomics* 13:718–725
8. Zhang L et al (1992) Whole genome amplification from a single cell: implications for genetic analysis. *Proc Natl Acad Sci USA* 89:5847–5851

Genomic Clone

Definition

Genomic clone denotes a fragment of cloned DNA originating from the genome of the organism of interest rather than from a reverse-transcript of an RNA.

►YAC and PAC Maps

Genomic Control

Definition

Genomic control describes a method to control for population stratification in association studies. The degree of genotype-phenotype association for a large number of neutral polymorphisms is measured and used to correct associations with candidate causal polymorphisms.

►COPD and Asthma Genetics

Genomic Imprinting

TAKUYA IMAMURA, ANDRAS PALDI
Ecole Pratique des Hautes Etudes, Evry, France
paldi@genethon.fr

Definition

Genomic imprinting, also called parental or gametic imprinting, is a process of epigenetic marking that

allows the cells to differentiate the alleles of paternal and maternal origin of a gene without changing their DNA sequence. The phenomenon of imprinting was explicitly recognized at the beginning of the 1980s on the basis of two types of observations. First, pronuclear transplantation studies on mouse zygotes demonstrated that monoparental conceptuses were not viable, suggesting that biparental contribution is necessary for mammalian development (1, 2). Second, systematic genetic studies of mice with chromosomal translocation showed that some chromosomal regions must be inherited from both parents for normal development (3). These pioneer studies led to the construction of a low-resolution chromosomal imprinting map of the mouse genome. It was postulated that the requirement for both parental genomes to be present in the same zygote was a consequence of differential epigenetic marks on a fraction of the paternal and maternal alleles. The parental origin-specific imprints on the two alleles of the same gene lead to their differential expression. Typically, one parental allele is silenced and only the other remains functional. It is precisely this property that has been used to identify imprinted genes. A gene is considered imprinted if it is expressed monoallelically and its allelic expression depends on its parental origin. However, the real situation is more complicated. The analyses of the known imprinted genes have demonstrated that most of them are expressed biallelically in some tissues, at least at some developmental stages. As a consequence, we face a paradoxical situation. It is quite difficult to prove that a gene is not imprinted, unless its allelic expression is examined in all tissues at all stages of the life. The ambiguity of the definition might be the reason for the difficulty estimating the number of imprinted genes in the genome. On the basis of the relatively low frequency of mutations with parental origin-dependent phenotype, the number of imprinted genes was initially estimated as not more than 300. However, a systematic transcriptome analysis suggested that there might be more than 2,000 genes with differential parental origin dependent expression in the mouse.

Characteristics

Characteristics of Imprinted Genes

In order to understand the phenomenon of imprinting, current research is focused on the following major questions: 1) How are the parental alleles of a gene marked without changing the DNA sequence? The imprint has to be sufficiently stable to be inherited through mitosis, but reversible during the meiosis. 2) When do the genes acquire their parental-specific imprints? 3) What is the biological significance of this phenomenon?

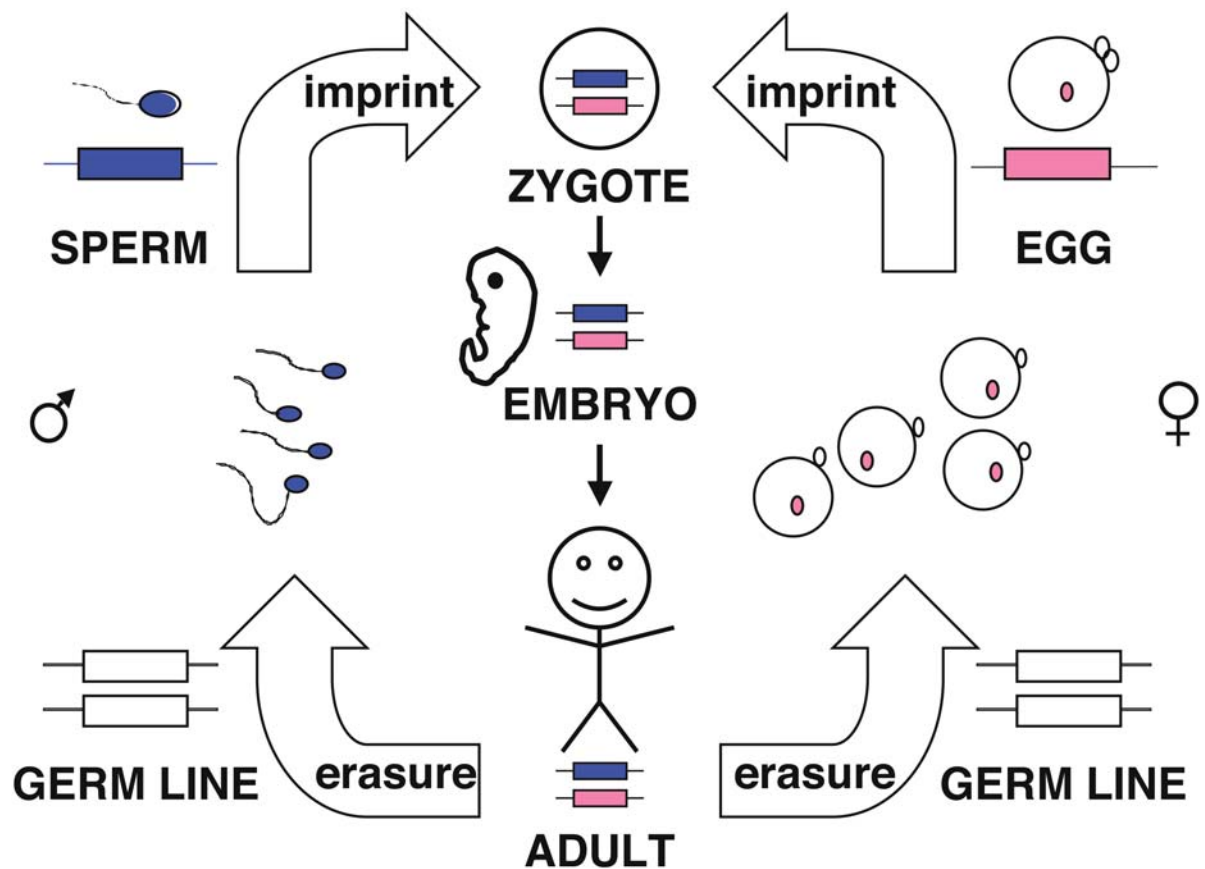
The first two imprinted genes, *Igf2r* and *Igf2*, were identified in 1991 on the basis of the parental origin-dependent phenotypes of heterozygous mutants. More than 70 genes have been identified in the mouse and human genome so far. (We quote here two comprehensive databases of imprinted genes that can be found on the web: ►<http://www.mgu.har.mrc.ac.uk/research/imprinting/imprinting2.html> and ►<http://cancer.otago.ac.nz/IGC/Web/home.html>). Detailed analyses of several of these genes made it possible to determine general characteristics of imprinted genes in addition to their monoallelic expression:

1. Imprinted genes are frequently associated with regulatory regions that carry differential ►**DNA methylation** on the two parental alleles. Several imprinted genes were identified on the basis of systematic search for ►**differentially methylated regions** (DMRs) in the genome. In general, DNA methylation upstream of genes, especially in the promoter region, is associated with attenuation of the expression.
2. It is impossible to classify the imprinted genes on the basis of the products they encode. Peptide hormones, growth factors, transcription factors, metabolic enzymes, cell surface receptors and many other proteins, but also several non-coding RNAs can be found among the products of imprinted genes.
3. Another characteristic feature of imprinted genes discovered so far is their non-random distribution in the genome. They are frequently clustered in well-defined genomic regions of up to several hundred kilobases. These clusters usually contain imprinted genes with either maternal or paternal monoallelic expression, but also genes that have been found to be biallelically expressed in all tissues analyzed so far.
4. Interestingly, many antisense transcripts are also detected in the imprinted genomic regions.
5. An important feature of the clusters is their ►**asynchronous replication**, which occurs in common as far as we know, suggesting that this phenomenon is one of the useful criteria that help determine imprinted genes (4). The paternal and maternal copies of the whole clusters replicate differentially during the mitotic cell cycle, including imprinted genes, intergenic sequences and the genes that are expressed biallelically. This characteristic is independent of the expression state of the genes in the cluster and is detected in all cell types. They all suggest that cluster level of imprinting regulation might be assigned in the context of chromatin structure as described below.
6. Imprinted chromosomal regions display strikingly different recombination frequencies during male and female meiosis. However, many non-imprinted

regions also show recombination with different frequencies during male and female meiosis.

Molecular Mechanisms

As indicated above, DNA methylation is a part of the mechanisms that differentiate the parental alleles of imprinted genes. Methylation of cytosines in CG dinucleotides (CpG methylation) is a well-known ►**epigenetic modification** of the DNA that regulates chromatin structures such as ►**heterochromatin** in concert with covalent ►**histone** modifications. The DMRs are usually relatively short, CG-rich DNA segments located at a distance from genes, but sometimes located in the promoter region or in the coding or intronic sequences. The best characterized DMRs in the human and mouse genome include those located in the regions of the *Igf2r*, *H19*, *Igf2*, *Snrpn*, *U2af1-rs1*, *Gnas* and *Gtl2* genes. The functional importance of these elements for the establishment and maintenance of the methylation imprint has been demonstrated by extensive targeted mutagenesis studies. Their deletions frequently perturb the function of the whole imprinted domain. These elements are usually called imprinting centers for their central role in the imprinting of a whole region. They most probably act as a structural organizer affecting gene expression over the whole imprinting cluster. This action is presumably mediated by recruiting various proteins, for instance the CTCF protein, that play a role in making up highly-ordered chromatin structure. Studies of chromatin structure around imprinted genes revealed differences in nucleosome positioning, histone acetylation or nuclease sensitivity between the parental alleles of imprinted genes. In general, methylated sequences are associated with hypoacetylated histones, whereas the unmethylated sequences are associated with hyperacetylated histones. The differences observed between the two alleles of an imprinted gene in the same tissue are similar to those typically observed between active and inactive copies of the same gene in different tissues. Naturally occurring, or experimentally induced mutations that alter the enzymatic mechanisms responsible for epigenetic modifications such as DNA methylation and histone modifications frequently disturb the normal imprinting process and modify the allelic expression of imprinted genes. In addition to the epigenetic modifications, the observation of several ►**antisense RNA** transcriptions in imprinted regions suggests that non-coding RNAs might be also involved in the maintenance of the characteristic chromatin structure and monoallelic expression of imprinted genes. A role of non-coding RNAs in this process has been suggested by analogy with the function of the Xist RNA in the inactivation of one of the X-chromosomes in females, although the role remains unknown.



Genomic Imprinting. Figure 1 The schematic representation of the cycle of acquisition/erasure of genomic imprinting in the germ line. Note that, in the germ cell lineage, the epigenetic marks should be erased and then re-established according to the sex of the individual.

Establishment of the Imprint

Each individual inherits a paternal and a maternal copy of every gene in the genome. However, both alleles are transmitted to the offspring either as paternal or as maternal copies depending on the individual's sex. Therefore, the parental imprint of a gene or a gene cluster has to be erased in the germ line of the individual and re-established according to its sex in mature gametes (Fig. 1). In order to follow this process, changes in CpG methylation pattern of DMRs were extensively studied at various imprinted loci. In general, the differences in CpG methylation pattern are erased from the parental alleles in early **▶primordial germ cell** (PGC) differentiation. The methylation pattern typical for the paternal or maternal alleles is established in meiotic cells. At the moment of fertilization, many DMRs are already differentially methylated and conserve their allele-specific methylation profiles at all subsequent stages of development while the bulk of the genome undergoes important methylation changes. However, some DMRs acquire their methylation profiles gradually during the somatic

cell division, suggesting that **▶CpG methylation** is not the only molecular mechanism that plays a role in marking the two parental alleles.

Other characteristics of the imprinted genomic regions follow different kinetics during development. For example, asynchronous replication of the two parental copies is maintained during the proliferation of PGC, when all methylation differences are already erased, suggesting that differences between the parental copies are still there even in the absence of methylation.

Clinical Relevance

The biological significance of the functional nonequivalence of the parental genomes is not yet known. Many hypotheses were proposed to explain why imprinting has evolved in mammals. The most popular hypothesis is the so-called parental conflict model. According to this model, imprinting has evolved in mammals because of the conflicting evolutionary interests of the paternal and maternal genomes over the allocation of parental resources. This hypothesis is based on the assumption that fetal imprinted genes

regulate resource transfer from the mother to the fetus. Therefore, parents are able to modulate the use of resources by transmitting epigenetically modified versions of imprinted genes to their offspring. Since the fetuses develop within the maternal uterus, the paternal investment in the offspring is obviously much lower than the maternal investment. This asymmetry leads to an asymmetry of the imprints on the resource usage-regulating genes.

Another possible explanation is that maintaining the differential chromosomal structure of the imprinted regions could be important for the coordinated replication of the genome and the correct segregation of the chromosomes during mitosis (5). The monoallelic expression of imprinted genes might be a byproduct of this process. Indeed, some experimental observations indicate that the parental copies of imprinted regions interact with each other during the somatic cell cycle in a way that is reminiscent of some trans-sensing phenomena observed in *Drosophila* or plants.

Whatever the biological function of parental imprinting, perturbations of the process lead to severe hereditary disorders that develop because of mutations in the active allele of imprinted genes, the normal but silenced allele being unable to compensate for the mutated copy (6). Biallelic expression of usually monoallelically expressed imprinted genes has also been implicated in various cancers. For example, Prader-Willi syndrome patients often display hypotonia, hyperphagia, obesity, hypogonadism and developmental delay. Angelman syndrome patients frequently show ataxia, tremulousness, sleep disorders, seizures and hyperactivity. Both syndromes may also show mental retardation and map to the imprinted gene cluster in human chromosome 15q11-13. Beckwith-Wiedemann syndrome maps to 11p15 and is characterized by general overgrowth with symptoms such as hemihypertrophy, macroglossia and visceromegaly.

►G-Proteins and G-Protein Mutations in Human Diseases

►Microdeletion Syndromes

►Prader Willi and Angelman Syndromes

References

- McGrath J, Solter D (1984) Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell* 37:179–183
- Surani MA, Barton SC, Norris ML (1984) Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature* 308:548–550
- Cattanach BM, Kirk M (1985) Differential activity of maternally and paternally derived chromosome regions in mice. *Nature* 315:496–498
- Kitsberg D, Selig S, Brandeis M et al (1993) Allele-specific replication timing of imprinted gene regions. *Nature* 364:459–463

- Paldi A (2003) Genomic imprinting: Could the chromatin structure be the driving force? *Curr Top Dev Biol* 53: 115–137
- Lalande M (1997) Parental imprinting and human disease. *Annu Rev Genet* 30:173–195

Genomic Information and Cancer

TRAVIS DUNCKLEY, KEITH D. COON,
DIETRICH A. STEPHAN
The Translational Genomics Research Institute,
Phoenix, AZ, USA
dstephan@tgen.org

G

Synonyms

Genomics

Definition

Genomics represents the systematic study of the entire genetic complement (DNA and RNA) of an individual or population of individuals. Since uncontrolled growth of cancerous cells results from inherited or somatically acquired mutations scattered throughout the complement of our chromosomes and often affects mRNA expression, the holistic tools of genomics are particularly relevant to understanding the molecular mechanisms underlying this set of diseases. As such, genomics has played a major role in advancing tumor subclassification, diagnosis and individualized treatment of patients.

Characteristics

Cancer results from the accumulation of genetic damage superimposed upon inherited predisposition. This damage manifests itself in the form of DNA mutations, chromosomal aberrations or epigenetic alterations in the chromatin structure. Elucidating the specific genetic events involved in the pathogenesis of different cancer types will be critical for the development of effective diagnostics and treatments. Initially, genetic studies in cancer focused primarily on heritable rare and highly penetrant alleles of cancer predisposition genes. Examples include the ►tumor suppressors *Rb1* and *p53*. However, heritable predisposition alleles account for a small percentage of cancer causing events. Multiple combinations of weak genetic variants may have a much larger impact on the development of cancer in the general population, the majority of whom do not have inherited alleles of the known highly penetrant cancer predisposing genes. Thus, cancer can be viewed as a multigenic disease. The challenge of

cancer genomics is to identify the multiple genetic variants that are involved in the development of cancer and to determine their effects on the molecular pathways of the premalignant cell type.

Identifying the genetic determinants of such a complex, multigenic disease as cancer necessitates the development and use of high throughput methods of genomic analysis. Toward that end, high throughput genomic DNA scanning technologies (sequencing, LOH, CGH and FISH) as well as microarray-based technologies have been critical. Microarray technology involves the fixation of DNA molecules to a slide or wafer. These DNA molecules can be placed on the microarray slides at very high densities, allowing for high throughput genome-wide analyses. Various microarray technologies exist, including ►[single nucleotide polymorphism](#) (SNP) arrays for linkage and LOH studies, ►[comparative genomic hybridization](#) (CGH) arrays, DNA sequencing arrays and perhaps most widely recognized, gene expression microarrays.

Single Nucleotide Polymorphism Genotyping for Linkage and LOH Studies

Traditionally, linkage analyses and ►[loss of heterozygosity](#) (LOH) have been done on a genome-wide scale using ►[microsatellite](#) markers at a density of ~10 cM (Mb) intervals. This is a tedious methodology. The resolution is appropriate for linkage studies but inadequate for LOH in the majority of cases. New information from the human genome project (HGP) has resulted in a high resolution SNP map of the human genome, as well as new technologies for rapidly genotyping these SNPs. Single nucleotide polymorphisms represent DNA base pair variations within a population of individuals. They can either have no effect on gene expression or may have subtle effects that, when combined, may lead to disease phenotypes. An SNP occurs, on average, once in every 1300 base pairs of the human genome and they account for the majority of genetic variability between individuals within a population. Although SNPs are biallelic and thus less informative, their density of ~every 30 kb (on the new Affymetrix 100k SNP array) allows larger haplotype block content to be inferred. Thus SNPs are likely to be equally informative over multiple adjacent SNPs as well as having the ability to identify smaller hemizygous deletions. SNP array technology has great potential for discovering multigenic contributions to cancer development. For example, by comparing the SNP profiles from a group of individuals that have a certain cancer type to the SNP profiles of unaffected individuals, one can identify a set of SNPs that are uniquely associated with that form of cancer (1). Since the sequence information for the SNPs is known, one can rapidly move towards identifying the relevant genes. Importantly, this technology is generally

applicable to the study of any disease with a genetic basis.

DNA Sequencing

DNA sequencing represents the ultimate in high-throughput cancer analysis. Once we fulfill the mandate of the HGP for rapid, whole-genome sequencing at minimal cost, we will revolutionize the ability to understand and diagnose cancer. Early attempts include massive parallel DNA arrays for hybridization-based sequencing. These arrays, designed by Perlegen Sciences, Inc. (CA, USA) consist of overlapping oligonucleotide probes that span the entire genome. This makes possible direct and rapid sequencing of the entire genome. This has clear advantages, particularly when an unidentified disease gene has been mapped previously using standard positional cloning strategies.

Comparative Genomic Hybridization

Comparative genomic hybridization (CGH) allows one to visualize gross chromosomal losses or gains by a modification of traditional karyotyping. The entire genomic complement of a normal individual is labeled and compared to a genomic sampling from a tumor that has been labeled with a different fluorophore. While revolutionary, the typical resolution of traditional CGH is ~10 Mb. Array CGH provides the advantage of higher resolution compared to traditional CGH methods (up to 0.5 Mb) and is generally useful for identifying large deletions, insertions, amplification events or overall changes in ploidy. Importantly, array CGH can be used as an adjunct to expression microarrays (see below) since the molecular events detected by CGH will have effects on gene expression levels.

Expression Microarrays

Gene expression microarrays are used to rapidly assess the gene expression profile of different cell types. For genetic mutations to affect cell proliferation, and hence the development of cancer, they must alter the function of at least some of the signaling pathways inside the affected cell. These effects can be seen as altered mRNA expression profiles (even members of phosphorylation cascades can be dysregulated) in malignant cell types and can be identified using expression microarrays.

There are two main variations of expression arrays currently in use, cDNA and oligonucleotide microarrays. Oligonucleotide arrays provide the benefits of greater specificity since the probes used are of shorter sequence (~25–70 nucleotides) than those used for cDNA arrays (200–2000 nucleotides). cDNA arrays have greater sensitivity but cannot, for example, discriminate between splice variants. For each type of array, sequences of DNA that are homologous to different genes of interest are attached to a glass slide at different locations. Each probe (or probe set for

oligonucleotide arrays) on a slide corresponds to a single gene and each slide can hold thousands of probes. One then hybridizes labeled cRNA from the cell type of interest to the microarray, rapidly generating an mRNA expression profile for thousands of genes.

Integration of LCM into ► Expression Profiling

One important limitation of expression profiling has been that the tissues used often contain multiple cell types in addition to the diseased or cancerous, cells. Additionally, cancers are almost always mosaic with respect to acquired somatic changes. Thus, they are heterogeneous with respect to the clinical and histopathological trait under study. This adds unwanted expression signatures to the global expression profile obtained and may generate misleading results. ► **Laser capture microdissection (LCM)** is increasingly being used to overcome this limitation. LCM uses an infrared laser to select only cell types of interest from a thin section of tissue sample. In the study of cancer, this allows the analysis of a nearly homogeneous population of malignant cells, generating a cleaner expression profile that is more indicative of the cancerous state. Because the volume of cells harvested using this technique is low, RNA amplification techniques must be used to generate enough RNA for expression profiling.

Clinical Relevance

Clearly, identification of the underlying DNA or RNA defects leading to cancer development and progression will lead to a greater understanding of tumorigenesis and will translate directly into drug design. SNP analyses provide the potential for early and reasonably noninvasive cancer diagnosis by detecting heritable cancer specific mutations in peripheral tissues, such as blood. As a result, treatments may be commenced earlier than was previously possible (2). Additionally, SNP analysis could become a routine screening procedure to identify individuals with SNP haplotypes that place them at risk for developing specific forms of cancer. This information could be used to direct at risk individuals to appropriate prevention strategies. As the technology matures, diagnosis and screening using this methodology may become economically feasible for general practice. In addition to providing a diagnosis method, SNP analyses will be useful for identifying causative mutations and affected molecular pathways in various cancers. Following validation of these pathways, new targets for therapy will emerge.

Gene expression profiling has numerous important clinical applications. First, expression profiling already has been used to predict the prognosis of disease course in ► **breast cancer** (3). In breast cancer susceptibility screening, genetic testing to identify individuals carrying known predisposition alleles has been used to indicate when more extreme prevention strategies,

such as surgical resection of at risk tissue, are needed. However, this approach suffers from the limitation that not all individuals carrying susceptibility genes will develop cancer. Further subclassifying a person's cancer risk based on additional genetic determinants will restrict such surgical prevention strategies to only a subset of individuals with the highest likelihood of developing particularly aggressive forms of disease. The above example with breast cancer illustrates the expectation that, as more tumors are profiled and subcategorized, expression profiling could become a general tool for predicting the course of cancer progression and for guiding prevention strategies in unaffected individuals. In the future, subcategorizing tumors based on their expression profiles may aid in patient-specific therapies that are designed to be most effective in the clinic on a real-time basis for the treatment of particular forms of cancer.

Microarrays are also important for identifying dysregulated genes and signaling pathways that are involved in tumor development and progression (4). Identification of these genes and pathways will have important implications for the development of novel anticancer therapies since they provide novel targets for treatment. As expression profiling in the field of cancer biology moves forward, a critical goal will be to translate the vast amounts of biological data into meaningful clinical advances. This will require large collaborative efforts pooling the combined knowledge and expertise of different institutions to accomplish successfully all of the required goals from tumor sample acquisition, to genomic analysis, to target identification and validation, to drug design and discovery.

Summary

Analysis of SNPs and gene expression profiling are valuable methods for identifying the genetic determinants of cancer. However, to realize the value of these techniques fully, it will be critical to translate the findings into practical applications that can benefit individuals who are suffering from cancer and those who are at risk of developing particular forms of cancer. Knowledge of an individual's innate susceptibility to various cancer types can be used to guide the course of prevention strategies that focus, for example, on lifestyle changes such as diet and exercise. In addition, SNP and expression profiles can be used both to diagnose cancers accurately and to subcategorize tumor types based on the severity of the malignant phenotype. This information could then be used to target the most aggressive therapies to patients with the most severe or invasive forms of disease. Ultimately, the information gleaned from these powerful genomics techniques will be used to identify novel targets for therapeutic intervention with the eventual endpoint of preventing tumor growth and metastasis.

References

1. Hoque MO, Lee CC, Cairns P et al (2003) Genome-wide genetic characterization of bladder cancer: a comparison of high-density single-nucleotide polymorphism arrays and PCR-based microsatellite analysis. *Cancer Res* 63:2216–2222
2. Sidransky D (2002) Emerging molecular markers of cancer. *Nat Rev Cancer* 2:210–219
3. van 't Veer LJ, Dai H, van de Vijver MJ et al (2002) Expression profiling predicts outcome in breast cancer. *Breast Cancer Res* 5:57–58
4. Mac Donald TJ, Brown KM, La Fleur B et al (2001) Expression profiling of medulloblastoma: PDGFRA and the RAS/MAPK pathway as therapeutic targets for metastatic disease. *Nat Genet* 29:143–152

Genomic Instability

Definition

Genomic instability describes a phenotypic feature of the cell, in which the genetic material mutates at a faster rate than normal as a consequence of a deficiency in proteins that function in ►DNA repair, cell cycle checkpoints, chromosome structure maintenance, and chromosome segregation, etc.

- Bloom Syndrome
- DNA Helicases
- DNA Repair Mechanisms

Genomics

Definition

The mapping, sequencing, and analysis of an organism's genome.

- Protein Databases

Genomics

- Genomic Information and Cancer
- Functional Genomics, the Systematic Analysis of the Function of All Genes and Gene Products in Parallel

Genotoxin

Definition

A genotoxin is a chemical, or another agent, which damages cellular DNA resulting in mutations and/or cancer.

- Chromosomal Instability Syndromes

Genotype

Definition

Genotype refers to the genetic constitution of an organism or cell; be it the alleles at a given locus, or those of several loci. For an autosome, the genotype for a specific chromosomal location would be 2 alleles.

- COPD and Asthma Genetics
- Diabetes Mellitus, Genetics
- Familial Dilated Cardiomyopathy
- Large-Scale ENU Mutagenesis in Mice
- Schizophrenia Genetics

Genotype-Driven Approach

Definition

Genotype-driven approach describes a plan of action based on the hypothesis that a specific gene is responsible for a specific function. To this end, the specific gene is mutated and the resulting phenotype (appearance) of the organism provides information about the function of the gene.

- Mouse Genomics

Genotype-Phenotype Correlations

Definition

Genotype-phenotype correlations describe the relationship between genotype (polymorphisms, sequence, variants, and mutations) and phenotype (their clinical expression).

- Heritable Skin Disorders

Genotyping

Definition

Genotyping is the determination of the specific allelic composition of a genome, a gene or a set of genes.

- ▶ [Cell Polarity](#)
- ▶ [SNP Detection and Mass Spectrometry](#)

Geranylgeranyl Pyrophosphate

Definition

Geranylgeranyl is a 20 carbon unit made up of four isoprene (dimethyl allyl) units, and in the form of the pyrophosphate, is a precursor molecule in cholesterol biosynthesis.

- ▶ [Protein Prenylation](#)
- ▶ [Tangier Disease](#)

Germ Cells

Definition

Germ cells are pre-meiotic or post-meiotic sperm cells and egg cells.

- ▶ [Mutagenesis Approaches in the Zebrafish](#)

Germinal Vesicle

Definition

Germinal vesicle is the meiotic prophase nucleus of an amphibian oocyte.

- ▶ [Xenopus as a Model Organism for Functional Genomics](#)

Germline (Gonadal) Mosaicism

Definition

The term mosaicism in general refers to an organism that is composed of two or more cell lines, which originate from only one zygote and differ in their genotype or chromosomal constitution. The different

cell lines have been formed by mutations. The period of development at which this mutation is formed determines the cell type of mosaicism: somatic or germline cells. The germline of one individual consists of two or more populations of cells, due to mutation(s) in one or more clonally expanded cell(s) in the population of germline cells. Germline mosaicism may vary between only a few cells or about 50% cells, and is the reason why genetically unaffected parents may have children with more than one X-linked or dominant genetic disorder.

- ▶ [Heritable Skin Disorders](#)
- ▶ [Neurofibromatosis Type 1 \(NF1\), Genetics](#)

Germline Mutation

Definition

Germline mutation denotes a mutation that affects the complete organism including the germ cells, and thus is passed on to the progeny of the affected individual.

- ▶ [Microarrays in Pancreatic Cancer](#)

Germline Transmission

Definition

Germline transmission refers to a process where the ES derived cells of a chimera contribute to the reproductive cells of a mammal (germ cells) and are genetically passed to its offspring.

- ▶ [Large-Scale Homologous Recombination Approaches in Mice](#)

GFACT Expression screening

- ▶ [Genome Functionalization by Arrayed cDNA Transduction \(GFACT\) Expression screening](#)

GFAP

Definition

The glial fibrillary acidic protein (GFAP) is an intermediate filament protein that is characteristic for astrocytes, but it is also expressed in certain

populations of ►[nestin](#)-expressing neural stem cells. It has thus become an ambiguous marker.

►[Neural Stem Cells](#)

GFFKR Domain

Definition

GFFKR domain refers to a conserved amino acid motif of the cytoplasmic tail of α integrin chains. It serves to stabilize the α and β integrin subunits in close spatial contact. This keeps the extracellular, ligand binding part of the integrin in a folded inactive form. Lysins, amino acids proximal to the GFFKR motif, are essential for interaction of the integrin α chain with RAPL. Binding of RAPL contributes to spatial separation of the cytoplasmic integrin chains, allowing unfolding of the extracellular part into a ligand binding integrin.

►[Focal Complexes/Focal Contacts](#)

GFP

►[Green Fluorescent Protein](#)

GGA Proteins

Definition

GGA proteins (Golgi-associated, γ -adaptin homologous, ARF-interacting proteins) constitute a conserved multidomain protein family involved in traffic between the Golgi complex and endosomes. They are recruited to membranes by GTP-bound ►[ARF](#). They can interact with trafficking motifs present on certain cargo, e.g. the mannose–6–phosphate receptor, and also interact with ►[clathrin](#), making them functionally analogous to ►[adaptor complexes](#).

►[Vesicular Traffic](#)

Giga-Seal

Definition

Giga-seal denotes the tight connection between the tip of a patch clamp pipette and the cell membrane during

the patch clamp analysis. The giga-seal is characterized by a large electrical resistance that reaches values in the giga-ohm range.

►[Patch Clamping](#)

Glanzmann's Thrombasthenia

Definition

Glanzmann's Thrombasthenia is an autosomal recessive disorder, characterized by the absence of dysfunction of the ►[GPIIb/IIIa complex](#), resulting in defective platelet aggregation.

►[Hereditary Hemostatic Defects and Recombinant Proteins for Treatment](#)

GLI

Definition

GLI comprise a family of zinc finger transcription factors involved in both developmental regulation and human diseases. Zinc-finger transcription factors of the GLI family play critical roles in the mediation and interpretation of Hedgehog signals. The *Drosophila* homologue is Cubitus interruptus (Ci).

►[Hedgehog Signalling](#)

►[Wnt/Beta-Catenin Signaling](#)

Glial Cells

Definition

Glial Cells are the non-neuronal cells of the nervous system. Glial cells do not carry nerve impulses (action potentials) but do have essential supportive functions, including physical support, provision with nutrients and trophic factors (astrocytes), insulation of axons (oligodendrocytes and Schwann cells), and phagocytic functions (astrocytes and microglia). During development, radial glial cells provide a scaffold for neuronal migration, and they function as neuronal progenitors.

►[Glial Cells and Myelination](#)

Glial Cells and Myelination

HAUKE WERNER, KLAUS-ARMIN NAVE

Department of Neurogenetics, Max Planck Institute of Experimental Medicine, Goettingen, Germany

hauke@em.mpg.de

nave@em.mpg.de

Definition

The large majority of non-neuronal cells in the nervous system are ►glial cells. In general, glial cells serve supportive functions for neurons, are not electrically excitable but may respond to neurotransmission. This overview focuses on two types of glial cells that myelinate axonal processes, ►oligodendrocytes in the central nervous system (CNS), and ►Schwann cells in the peripheral nervous system (PNS). ►Myelin enables ►axons to conduct action potentials much more rapidly, by insulating the axonal membrane, decreasing its capacitance and restricting ion fluxes to the ►nodes of Ranvier. Myelin is largely made during early postnatal life. Thus, developmental disorders of myelination or loss of myelin are severe neurological diseases, either inherited (►leukodystrophies, neuropathies) or acquired. One clinically important myelin disorder is multiple sclerosis, an inflammatory ►demyelination of the CNS.

Characteristics

The best understood function of oligodendrocytes and Schwann cells is to enwrap axonal processes with an insulating myelin sheath. Myelination occurs largely during early postnatal life and can be divided into several steps: (i) establishing contact between glial cell and axon (oligodendrocytes engulf multiple axons), (ii) spiral enwrapping of axonal segments with up to 50 layers of membrane, (iii) compaction of myelin by tight association of the intracellular and extracellular membrane surfaces, (iv) formation of functional nodes of Ranvier and paranodal structures. Although some of the molecules involved in myelination have been identified, the cellular mechanisms are not well understood. In humans, myelination begins around birth, with a peak in the first five years of life and is mostly completed by 11 years of age. Some active myelination has been observed until the 5th decade, possibly related to neuronal plasticity. During the peak of myelination, i.e. within a few days, oligodendrocytes produce a large amount of membrane material that exceeds their own weight several fold.

The highly periodic structure of myelin is best visualized in cross section under the electron microscope (Fig. 1). Proteins at the condensed cytoplasmic

membrane surface form the electron-dense major dense line (MDL), those at the condensed extracellular surface form the intraperiod line (IPL). The membrane itself is electron-lucent. The ultrastructure of CNS and PNS myelin is remarkably similar. Once myelinated, many axons become dependent on oligodendrocyte support. Thus, when oligodendrocytes or myelin degenerate in the course of a demyelinating disease, some axons will degenerate as well. The mechanism of this axon-glia interaction is not known.

Myelinated axons of the human PNS exhibit nerve conduction velocities (NCV) between 5 and 100 m/s, increasing with the axon diameter. In non-myelinated axons, NCV measures 0.5–5 m/s. To exhibit the same NCV, a non-myelinated axon would have to be of much larger diameter. Thus, myelin has evolved in vertebrates as means to achieve high NCV and reduced space requirements.

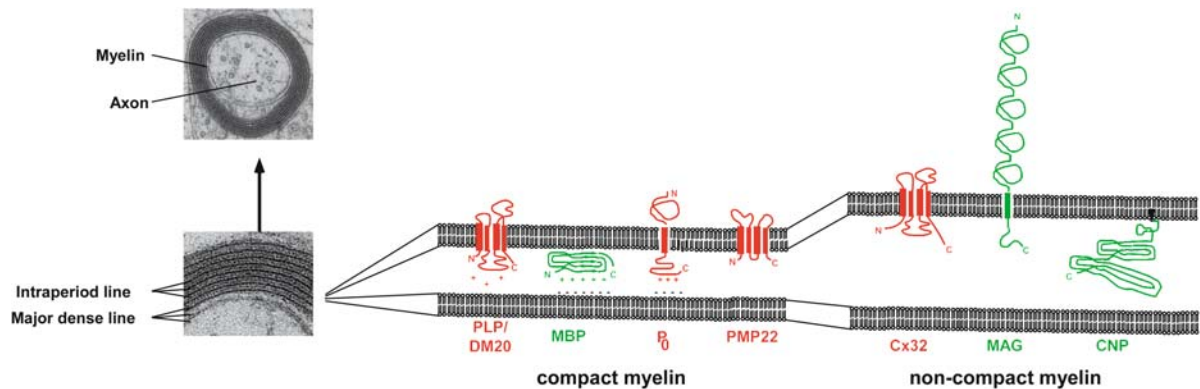
Molecular Interactions

Only axons with a diameter >1 µm are myelinated, smaller axons may be engulfed (i.e. in the PNS by non-myelin forming Schwann cells) but are not enwrapped. Axon diameter and myelin sheath thickness show a constant relationship (g-ratio). In the PNS, the axonal growth-factor neuregulin-I (Nrg1) plays a critical role in signaling size information to receptor tyrosine kinases erbB2/erbB3 on the Schwann cell and in regulating myelin membrane growth.

Molecularly, compact myelin is comprised of lipids and myelin proteins that differ in composition between CNS and PNS. Targeted gene inactivation in ('knock-out') mice has been used to explore the function of myelin-specific proteins systematically. Many of them are abundant integral or membrane-associated proteins with functions in membrane adhesion. Their CNS- or PNS-specific expression correlates well with the clinical picture of leukodystrophies and neuropathies respectively. Regions of non-compact myelin exist at the innermost myelin membrane, adjacent to the nodes of Ranvier and as Schmidt-Lanterman incisures in the internodal region. They provide radial connections between the periaxonal space and the glial cell soma, form a tight seal at the node of Ranvier (paranodal loops) and help organize the distribution of ion channels by interacting with proteins of the axonal membrane. Some structural proteins of non-compact myelin are quite abundant. For updated information on these genes and diseases refer to the McKusick entries of the Online Mendelian Inheritance in Man (►OMIM) database (►<http://www3.ncbi.nlm.nih.gov/Omim/>).

Myelin Lipids

Myelin has a lipid content of 70–80% (compared to less than 50% in other ►biological membranes) and lipids contribute to its insulating function. How myelin lipids



Glial Cells and Myelination. Figure 1 Cross section of a myelinated axon at the electron microscopic level (upper left), the ultrastructure of compact myelin (lower left), and schematic depictions of structural proteins in myelin. The condensed cytoplasmic membrane surfaces form the electron-dense *major dense line*, extracellular membrane adhesion forms the *intraperiod line*. The membrane itself is electron-lucent. Membrane proteins associated with human myelin diseases are depicted in red. PLP/DM20, proteolipid protein; MBP, myelin basic protein; P0, protein zero; PMP22, peripheral myelin protein of 22kD; Cx32, connexin of 32 kD; MAG, myelin-associated glycoprotein; CNP, cyclic nucleotide phosphodiesterase.

are enriched in the cell membrane is not understood. Myelin is particularly rich in cholesterol. Galactosyl-cerebroside (GalC) and its sulfated form (sulfatide) are nearly myelin-specific. Absence of GalC and sulfatide in mutant mice lacking UDP-galactose:ceramide galactosyl-transferase leads to progressive demyelination and early death. Thus, both lipids are essential for normal myelination, although glucosylcerebroside (an alternative product) may compensate for some functions of GalC. Patients with the Smith-Lemli-Opitz syndrome (McKusick #270400), a genetic disorder of cholesterol biogenesis, also have myelin abnormalities. The critical requirement for cholesterol in myelin assembly has been shown with conditional mouse mutants deficient in squalene synthase, a critical and specific enzyme of cholesterol synthesis.

Proteolipid Protein (PLP, McKusick *300401)

In the CNS, the most abundant protein of compact myelin is a hydrophobic integral membrane protein (proteolipid) with four transmembrane domains and its smaller splice isoform (DM20). PLP/DM20 may form homo-oligomers and associate with $\alpha(v)$ -integrin, but the function of these interactions is speculative. The tight association of PLP/DM20 with cholesterol may be required for membrane **raft** formation and normal membrane trafficking in oligodendrocytes. The ultrastructure of CNS myelin lacking PLP/DM20 in **knockout mice** suggests that the extracellular portion of PLP acts as a strut, organizing the extracellular apposition of myelin layers at the IPL, but myelination is possible in the absence of PLP. For comparison, point mutations in this gene (or PLP gene duplications) cause severe **dysmyelination** in Pelizaeus-Merzbacher disease (PMD) and in rodent PMD models. This is due

to ER retention of the mutant protein, unfolded protein response and oligodendroglial apoptosis

Myelin Protein Zero (MPZ, P0, McKusick *159440)

Myelin protein zero is the most abundant protein of compact PNS myelin, expressed exclusively by Schwann cells. With a single transmembrane domain and an extracellular Ig-like domain, P0 is a member of the Ig-superfamily of cell adhesion proteins. The crystal structure of the Ig-like domain, when combined with the analysis of P0-deficient mice, indicates that homo-tetrameric P0 engages in homophilic interactions with the opposing membrane layer, mediating membrane adhesion and formation of the IPL. Additionally, positive charges in the cytoplasmic domain contribute to the establishment of the MDL by direct interaction with negatively charged head groups of membrane phospholipids. Mutations of the human P0 gene cause a peripheral neuropathy (CMT1B).

Myelin Basic Protein (MBP, McKusick *159430)

Myelin basic protein refers to a group of related cellular proteins associated with both CNS and PNS myelin. The MBP gene encodes at least 5 splice isoforms, ranging from 14 to 21 kD in size. Positively charged amino acids interact with negatively charged head groups of membrane phospholipids causing MBP to mediate and stabilize myelin compaction at the MDL. The MBP gene is partially deleted in the natural mouse mutant *shiverer*, which presents with a severe demyelinated phenotype. The overall lack of myelin assembly is not yet fully explained. *Shiverer* mice provided the first opportunity to analyze the consequences of a missing myelin protein before transgenic knockout techniques became available. An involvement of the

MBP gene in a human leukodystrophy has not yet been demonstrated.

Peripheral Myelin Protein 22K (PMP22, McKusick *601097)

Peripheral myelin protein 22K is a glycosylated integral membrane protein of PNS myelin. By topology and hydrophobicity PMP22 is related to the proteolipids of CNS myelin. PMP22 interacts with myelin protein P0 in the myelin membrane and may stabilize the myelin architecture. Experiments carried out *in vitro* suggest that PMP22 also regulates Schwann cell proliferation and apoptosis, but its *in vivo* function is poorly understood. The PMP22 gene has captured interest because a gene duplication in humans underlies the most frequent peripheral neuropathy (CMT1A).

Myelin-associated Glycoprotein (MAG, McKusick *159460)

Myelin-associated glycoprotein is a member of the Ig-superfamily of cell adhesion proteins, with a single trans-membrane domain and 5 extracellular Ig-like domains. Its localization at the innermost (adaxonal) membrane of both CNS and PNS myelin suggested that MAG is engaged in adhesion and signaling events between glial cell and axon. The diameter of PNS axons is reduced in mice lacking MAG, suggesting that MAG-mediated myelin-to-axon signaling regulates the phosphorylation status of the axonal cytoskeleton. Axonal binding partners of MAG include the NoGo receptor (NoGoR / reticulon 4 receptor, McKusick *605566), and sialic acid residues of sialo-glycoproteins or sialo-glycolipids (gangliosides).

Connexin-32 (Cx32, McKusick *304040)

Connexin-32 is a member of the connexin family of gap junction proteins, permeable to molecules <1kD. The protein is expressed in many cell types, including oligodendrocytes and Schwann cells, where it is localized to Schmidt-Lanterman incisures and paranodal loops. The exact function of Cx32 in PNS myelin is not known, but its involvement in the non-classical gap junctions that form a connection between different myelin lamellae of the same Schwann cell is most likely. Radial transport may be required for second messenger molecules that are generated at the innermost myelin membrane. Mutations in the human Cx32 gene cause a peripheral neuropathy (CMT1X).

Claudin11/Oligodendrocyte-specific Protein (OSP, McKusick *601326)

Claudin11/oligodendrocyte-specific protein is a member of the claudin family of [tight junction](#)-specific integral membrane proteins with four transmembrane domains. Knockout mice revealed that OSP is an essential constituent of the radial component in CNS

myelin that stabilizes myelin through intramembranous tight junctions. The finding that OSP interacts with tetraspanin-3/OAP-1 suggests that other tight junction proteins have yet to be identified in myelin.

Nodal and Paranodal Specializations

For [saltatory nerve conduction](#), axonal voltage gated sodium (Na^+) channels must be clustered at the node of Ranvier, separated from fast potassium (K^+) channels that assemble beneath the myelin sheath at the juxtaparanode. K^+ channels are associated with Caspr2, a member of the neurexin family of [adhesion molecules](#), probably *via* a PDZ domain protein adapter. Caspr2 in turn is associated with TAG-1, a GPI-anchored cell adhesion molecule of the Ig-like superfamily on the glial adaxonal membrane. Knockout experiments in mice demonstrated that axonal Caspr2 is required to maintain K^+ channel clusters at the juxtaparanode. More devastatingly, in the absence of TAG-1, axonal Caspr2 and K^+ channels are unclustered at the juxtaparanode. Na^+ channel distribution is unaffected by deletion of juxtaparanodal proteins.

The assembly sites of axonal Na^+ and K^+ channels are divided by the paranode, a region devoid of ion channels. The paranodal axon is tightly attached to the glial paranodal loops by a septate-like junctional structure that seals against ion flux and separates Na^+ from K^+ channels. The paranodal axon is molecularly defined by a complex of the Ig-like GPI-anchored cell adhesion molecule F3/contactin associated with the neurexin Caspr1. This complex interacts with neurofascin155 on the paranodal loop. Disruption of individual components in knockout mice impedes the septate-like junction.

References

1. Arroyo EJ, Scherer SS (2000) On the molecular architecture of myelinated fibers. *Histochem Cell Biol* 113:1–18
2. Werner H, Jung M, Klugmann M et al (1998) Mouse models of myelin diseases. *Brain Pathol* 8:771–793
3. Salzer JL (2003) Polarized domains of myelinated axons. *Neuron* 40:297–318
4. Schwab ME (2004) Nogo and axon regeneration. *Curr Opin Neurobiol* 14:118–214
5. Colognato H, French-Constant C (2004) Mechanisms of glial development. *Curr Opin Neurobiol* 14:37–44

Global Genome Repair

Definition

Global genome repair designates the branch of nucleotide excision repair that occurs in all regions of

DNA, with the exception of actively transcribed genes, and removes damage that could block DNA replication.

► [Nucleotide Excision Repair](#)

Glomerular Filtrate

Definition

Urine formation in the kidney begins when the fluid portion of the blood leaves the glomerulus and enters the glomerular capsule as glomerular filtrate. Glomerular filtrate consists of water and small size components of blood, separated from blood cells. The glomerular filtrate flows into the tubules, where further water is extracted from the filtrate, and minerals and other body chemicals are absorbed from or secreted into the filtrate.

► [Diabetes Insipidus, a Water Homeostasis Disease](#)

Glomerulonephritis

Definition

Glomerulonephritis (GN) refers to inflammation of the capillary loops of the glomeruli.

► [Morbus Wegener](#)

► [SLE Pathogenesis Genetic Dissection](#)

Glomerulus

Definition

Glomerulus is the network of blood capillaries in the cup-like end (Bowman's capsule) of the nephron. It is where waste products are filtered from the blood into the kidney tubule.

► [Kidney](#)

Glucocorticoid/Mineralocorticoid Receptors

Definition

Glucocorticoid receptor (GR) and mineralocorticoid receptors (MR) are members of a nuclear receptor

superfamily that mediate an organism's response to glucocorticoids or mineralocorticoids, respectively, by changing the transcription rates of glucocorticoid- or mineralocorticoid-responsive genes.

► [Steroid Hormone Receptor Defects, Molecular Basis](#)

Glucocorticoid/Mineralocorticoid Resistance

Definition

Glucocorticoid/mineralocorticoid resistance are pathologic conditions that demonstrate several manifestations caused by partial insensitivity of tissues to glucocorticoid or mineralocorticoid hormones. These are frequently due to inactivating mutations in the glucocorticoid or mineralocorticoid receptors.

► [Steroid Hormone Receptor Defects, Molecular Basis](#)

Glucocorticoids

Definition

Glucocorticoids are steroid hormones that are synthesised from cholesterol by cytochrome P450 dependent steroidhydroxylase, mainly in the adrenal cortex.

► [Mendelian Forms of Human Hypertension and Mechanisms of Disease](#)

Glutamate

Definition

L-Glutamate is an excitatory amino acid neurotransmitter. It influences almost all neurons in the brain. Glutamatergic neurotransmission has been associated functionally with a number of physiological and pathophysiological processes related to neuronal plasticity and memory.

► [Addiction, Molecular Biology](#)

Glutathione Peroxidase

Definition

GPx defines a family of homologous proteins that is characterized by a catalytic triad composed of a (seleno) cysteine, a glutamine and a tryptophan. These enzymes reduce H_2O_2 and other

►Free Radicals

Glycoconjugate

Definition

Glycoconjugate refers to a compound that is composed of an oligosaccharide which is linked to a protein or lipid.

►Biochemical Engineering of Glycoproteins

Glycan

Definition

Glycan is a general term for a polymer of mono-saccharide units joined by glycosidic bonds. It may or may not have other components.

►Biochemical Engineering of Glycoproteins

►Glycosylation of Proteins

Glycoform

Definition

Glycoform describes various forms of a particular species of glycoproteins that differ in the structures and/or types of glycans.

►Glycosylation of Proteins

Glycated Protein

Definition

Glycated protein designates a protein containing carbohydrate that was added by a nonenzymatic, chemical modification, usually through a Schiff-base reaction with the amino group of the side chain of lysine, and subsequent Amadori rearrangement, to give a stable conjugate.

►Glycosylation of Proteins

Glycogen Synthase Kinase-3

Definition

Glycogen synthase kinase-3 (GSK3) is a constitutively active kinase that undergoes inhibition by hormones and growth factors. One of its functions is to inhibit Wnt/ β -catenin signaling by phosphorylation of β -catenin thereby targeting it for degradation. Also called zeste white 3/shaggy in *Drosophila*.

►Wnt/Beta-Catenin Signaling Pathway

Glycine

Definition

Glycine is an amino acid that is derived from dietary sources, but is also generated endogenously from glyoxylate. Glycine serves as an important inhibitory neurotransmitter, predominantly in the spinal cord, brain stem and retina.

►Peroxisomal Disorders

Glycohemoglobin

Definition

Glycohemoglobin stand for glycosylated hemoglobin. The ratio of glycohemoglobin and total hemoglobin is indicative of a person's average blood glucose level over the last months.

►Affinity Chromatography and *In Vitro* Binding (Beads)

Glycolysis

Definition

Glycolysis is the metabolic pathway that occurs in the cytoplasm of cells, and by which glucose is broken down to pyruvic acid.

► [Limb Girdle Muscular Dystrophies](#)

Glycoprotein

Definition

Glycoprotein defines a protein with one or more carbohydrate moieties that are covalently bound to it.

► [Affinity Chromatography and In Vitro Binding \(Beads\)](#)

► [Biochemical Engineering of Glycoproteins](#)

► [Glycosylation of Proteins](#)

► [Protein Databases](#)

Glycoproteomics

Definition

Glycoproteomics refers to the science of defining the structures of glycoproteins, and the sites of attachment and structure of glycans to proteins.

► [Glycosylation of Proteins](#)

Glycosaminoglycan

Definition

Glycosaminoglycan refers to polysaccharide side-chains of proteoglycans or free complex polysaccharides that are composed of linear disaccharide repeating units, each composed of a hexosamine and a hexose or a hexuronic acid (heparin, heparan sulfate, chondroitin sulfate, dermatan sulfate, and hyaluronan).

► [Glycosylation of Proteins](#)

Glycosidic Linkage

Definition

Glycosidic linkage describes the linkage of a monosaccharide to another residue via the anomeric hydroxyl group. The linkage generally results from the reaction of a hemiacetal with an alcohol (e.g. a hydroxyl group on another monosaccharide or amino acid) to form an acetal.

► [Glycosylation of Proteins](#)

Glycosylase

Definition

Glycosylase is an enzyme that catalyzes the cleavage of an N-C1' glycosylic bond, which links a DNA base to the deoxyribosephosphate backbone of DNA.

► [Base Excision Repair](#)

Glycosylation of Proteins

RICHARD D. CUMMINGS

University of Oklahoma Health Services Center,
Department of Biochemistry and Molecular Biology,
Oklahoma City, Oklahoma, USA
richard-cummings@ouhsc.edu

Definition

► [Glycoproteins](#) are proteins that contain one or more covalently attached carbohydrates, including ► [monosaccharides](#) and ► [oligosaccharides](#). The attached carbohydrates are also termed ► [glycans](#). In typical glycoproteins the glycans can contribute up to ~20% of the total weight. ► [Proteoglycans](#) are a special class of glycoproteins that contain at least one large-sized (typically >5 kD), acidic polysaccharide (► [glycosaminoglycan](#)) attached to protein. ► [Mucins](#) are another special class of glycoproteins with a repeating peptide motif that usually contains multiple Ser and Thr residues to which relatively small-sized (usually <3 kD) glycans are attached. The mucin glycans can contribute more than 25% of the total weight. The ► [storage polysaccharides](#) starch and glycogen, which contain glucose polysaccharide linked to a core protein, also represent a special class of glycoprotein. In these cases the glycan portion

contributes more than 90% of the total weight. Another type of glycoprotein is the glycosylphosphatidylinositol-anchored or ►**GPI-anchored glycoprotein**. These contain a C-terminal amino acid that is linked to ethanolamine, which is linked to the glycans of the GPI anchor. GPI-anchored glycoproteins, which may also contain covalently attached N- and/or O-glycans at other residues within the polypeptide, are usually anchored to plasma membranes of cells by insertion of the acyl chains of the GPI moiety into the membrane outer leaflet. Thus, glycoproteins are found in many different sizes, ranging from several thousand Daltons to millions of Daltons.

The presence of carbohydrate on protein is a type of ►**post-translational modification**, which along with phosphorylation constitutes one of the most common types of such modifications. The majority of the nearly 30,000 proteins expressed in human cells are glycoproteins (1). The carbohydrates of glycoproteins are added enzymatically to specific sites on a protein. This contrasts with what is seen for ►**glycated proteins** in which carbohydrate addition occurs through the chemical or non-enzymatic addition of a free monosaccharide, usually glucose or galactose, to amino groups in proteins through formation of a Schiff base and rearrangement to a stable oxoamine adduct known as an Amadori product. This process is termed glycation and is often seen in patients with diabetes and children with galactosemia. Animals, plants, fungi, Protocista, archaea and bacteria synthesize glycoproteins and many animal viruses contain glycoproteins. In animals, most of the proteins that are on cell surfaces and those that are secreted by cells are glycosylated. Glycoproteins in membranes occur as integral or intrinsic membrane glycoproteins and may contain one or more transmembrane domains. GPI-anchored glycoproteins are also considered to be integral membrane proteins. Glycoproteins can also occur as extrinsic membrane glycoproteins that associate with the membrane through other mechanisms. Many proteins within the cytosol of eukaryotes are also glycoproteins. In animal cells two of the major classes of intracellular glycoproteins are the storage polysaccharides and those that contain O-linked GlcNAc or O-GlcNAc (termed O-GlcNAcylated) modifications. The attached glycans in different glycoprotein species often exhibit tremendous diversity in size and structure. Within a single glycoprotein species these structural differences are often denoted by the term microheterogeneity and the varied forms of a single glycoprotein species are termed ►**glycoforms**. Different glycoproteins from the same cell may be glycosylated very differently, depending on the primary, secondary, tertiary and/or quaternary structure, association with other proteins and subcellular localization. However, glycosylation differences are greater among glycoproteins from different cell types within an organism; the

differences are often even greater in glycoproteins between different organisms. Identifying the structures of the glycans made by different organisms is known as the field of glycomics and identifying sites of glycan attachment in glycoproteins is known as the field of ►**glycoproteomics**.

Characteristics

Glycans on glycoproteins can vary in the types of sugars that are attached, the ►**anomeric** configuration and structure of the attached sugars, the numbers of attachments and the sites of attachments (2). The linkage of one monosaccharide to another is typically *via* a ►**glycosidic linkage**, characterized by the acetal structure. The part of the glycan linked to protein is termed the reducing end and the opposite, unattached end(s) of the glycan is termed the non-reducing end or terminal region. Although there are many types of sugar-protein linkages, which are also glycosides, most sugar-protein linkages are of two types, N-glycosides, in which the amide of Asn (and in some organisms Arg) forms the linkage group (-C-N-C-) and O-glycosides, in which the hydroxy group of hydroxyamino acids, such as Ser, Thr, Tyr, hydroxylysine (Hyl) and hydroxyproline (Hyp), forms the linkage group (-C-O-C-). C-glycosides are an exception to this generalization, here a -C-N-C- bond links the sugar to an amino acid, as seen in Man-C-Trp. Acetal or glycosidic linkages between sugars are stable to alkali. By contrast, the linkage of sugar to protein *via* Asn, Ser or Thr residues is labile to relatively mild alkali. The sugar-protein linkages *via* Tyr, Hyl and Hyp are resistant to alkali. The glycosidic linkages within a glycan and all glycosidic linkages of sugars to proteins can be hydrolyzed by treatment with strong acid. An exception to this generalization is the C-glycoside linkage, which cannot be hydrolyzed by treatment with either alkali or acid.

Table 1 lists some of the common sugar-protein linkages found in glycoproteins from animals, plants, fungi, archaea and bacteria. Many dozens of linkages are now known, but mammalian cells appear to generate about a dozen or so. In most cases the monosaccharide residues shown in Table 1 are extended by the addition of other sugar residues. For examples see the composite animal cell glycoproteins shown in Fig. 1. The typical mammalian glycoprotein may contain N- and/or O-glycans of the types shown. One exception to the generalization that glycans are extended by addition of other sugars is found in eukaryotic cytosolic proteins that contain O-linked GlcNAc (GlcNAc β -O-Ser/Thr), where this residue is not further modified. It is also one of the few examples where the sugar addition is reversible, i.e. the GlcNAc residue is removed and added back multiple times on a mature protein. This probably serves an important

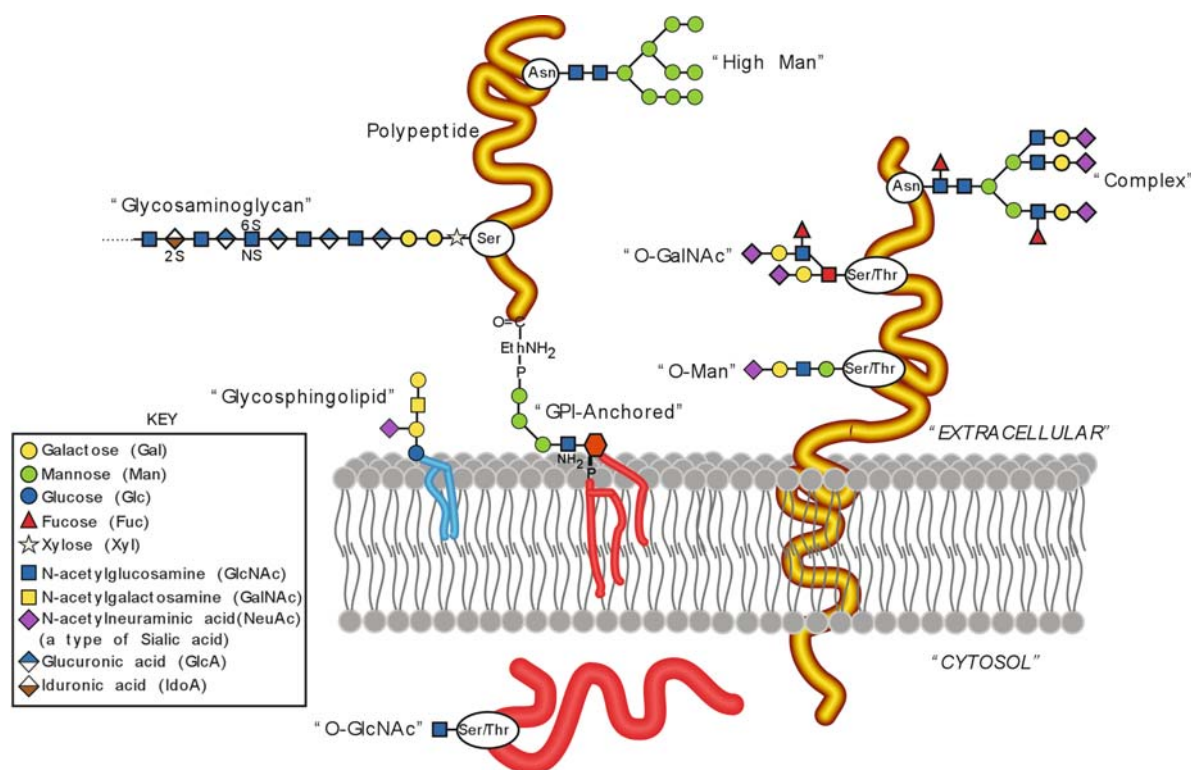
Glycosylation of Proteins. Table 1 Examples of Sugar-Protein Linkages in Glycoproteins from Different Organisms

Kingdom	Sugar-Protein Linkage
Animalia	GlcNAc β -N-Asn* GalNAc α -O-Ser/Thr GlcNAc β -O-Ser/Thr Man α -O-Ser/Thr Xyl β -O-Ser Fuc α -O-Ser/Thr Glc β -O-Ser/Thr Glc α -O-Tyr (in glycogen) Gal β -Hyl Man-C-Trp
Plantae	GlcNAc β -N-Asn Gal α -O-Hyp Gal α -O-Ser Ara β -O-Hyp Glc β -Arg (in starch)
Fungi	GlcNAc β -N-Asn Man α -O-Ser/Thr
Protocista (algae, sea-weeds and protozoa)	GlcNAc β -N-Asn Ara β -O-Hyp GlcNAc α -O-PO ₃ ⁻ -Ser GlcNAc α -O-Ser/Thr GlcNAc α -Hyp
Archaea	Glc β -N-Asn GalNAc β -N-Asn Rha-N-Asn Gal-O-Thr
Bacteria (Eubacteria)	Gal β -O-Tyr Gal β -O-Ser/Thr GalNAc β -O-Ser/Thr Glc α -O-Ser

* Abbreviations: GlcNAc, N-acetylglucosamine; Asn, Asparagine; GalNAc, N-acetylgalactosamine; Ser, Serine; Thr, Threonine; Man, Mannose; Xyl, Xylose; Fuc, Fucose; Glc, Glucose; Tyr, Tyrosine; Gal, Galactose; Hyl, Hydroxylysine; Trp, Tryptophan; Hyp, Hydroxyproline; Ara, Arabinose; Arg, Arginine; Rha, Rhamnose; Pro, Proline; Cys, Cysteine; Gly, Glycine; Xaa, any amino acid, except as indicated

regulatory function for cytosolic glycoproteins, akin to the action of protein phosphorylation, which is also reversible. (The other example of reversible

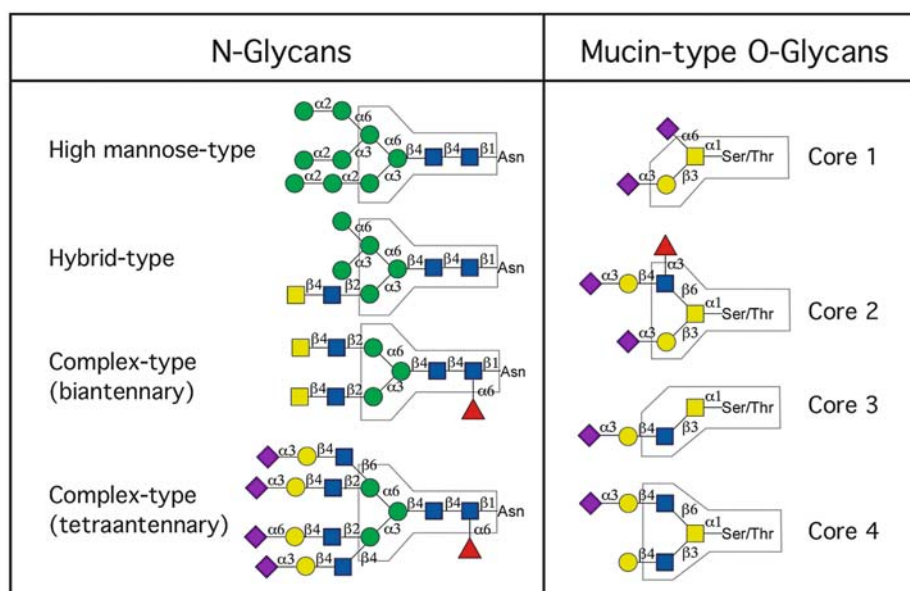
glycosylation is found in glucosylation of N-glycans as part of the [▶quality control](#) system for glycoprotein folding, as discussed below.)



Glycosylation of Proteins. Figure 1 Examples of different types of protein glycosylation. Shown are examples of composite membrane glycoproteins in animals that may contain one or more O- or N-glycans and a GPI-anchor or a transmembrane domain. Glycoproteins in the cytosol may also contain O-GlcNAc residues. The key for the symbols and abbreviations of monosaccharides is indicated and used in other figures.

N-glycans in higher animals typically contain 7–20 monosaccharide residues, whereas O-glycans typically have 2–10 residues. However, in yeast, many N-glycans contain mannan, a polysaccharide of mannose, which can contain hundreds of mannose residues. In proteoglycans the attached glycosaminoglycans can be hundreds of residues in length. Some examples of common types of N- and O-glycans in mammals are shown in Fig. 2. Mammalian glycoproteins are largely composed of the ten sugars or building blocks shown, which are the **▶hexoses** galactose (Gal), glucose (Glc) and mannose (Man), the **▶deoxyhexose** fucose (Fuc), the **▶hexosamines** N-acetylglucosamine (GlcNAc) and N-acetylgalactosamine (GalNAc), the **▶uronic acids** glucuronic acid (GlcA) and iduronic acid (IdoA), the pentose xylose (Xyl) and the 9-carbon carboxylated amino sugar **▶sialic acid** (Sia) and its multiple derivatives. In humans Sia occurs primarily as N-acetylneuraminic acid (NeuAc). Other organisms use many of these same monosaccharide residues, but also use novel ones not found in animals. For example, both Gal and GlcNAc are commonly found in glycoproteins from all the known kingdoms. Rhamnose (Rha) and arabinose (Ara) are found in plant, but not animal, glycoproteins, whereas sialic acids are found in animal,

but not plant, glycoproteins. Some bacteria synthesize sialic acid, however it has not yet been identified on bacterial glycoproteins. Some glycoproteins from nematodes contain tyvelose (3,6-dideoxy-D-arabinohexose - Tyv), which is not found in vertebrate glycoproteins. Some of these monosaccharide residues may themselves be further modified before or after incorporation into the glycan moiety of the glycoprotein to provide even more diversity of structure, as seen for NeuAc, which may be O- and/or N-acetylated at various positions, GlcA and IdoA, which can be N-sulfated and O-sulfated at various positions, and GlcNAc and Gal, which may be O-sulfated. Thus, the number of possible glycan structures is astronomically large and the upper limit of the number of structures is not known for any organism. It is important to note that polypeptides are typified by a linear structure in which two amino acids (L-amino acids in animals) are linked by a peptide bond. By contrast, glycans in glycoproteins are usually branched structures, where the monosaccharides are linked to each in multiple ways, such as different anomeric configuration (α versus β) to various positions on a residue (position C-2, C-3, C-6, etc.). In addition, the sugars in glycoproteins can be in either pyranose (6-membered ring) or furanose



Glycosylation of Proteins. Figure 2 Examples of different types of N- and O-glycans. Animal cell N-glycans shown on the left side have a common pentasaccharide core (shown in the boxed structure), which is composed of a trimannosyl sequence linked to a chitobiosyl disaccharide, which is in turn linked to an Asn residue through N-glycosidic linkage. The N-glycans are generally classified as high mannose-, hybrid- or complex-type as shown. The complex-type N-glycans may have multiple branches or antennae, described as mono-, bi-, tri-, or tetra-antennary, etc. Animal cell mucin-type O-glycans shown on the right side have a common structure of GalNAc linked to either Ser or Thr. This GalNAc residue may be modified in various ways to generate a variety of core structures (shown in the boxed structures).

(5-membered ring) structures and the residues may be either or D- or L-enantiomers (mirror images). Considering all these possibilities, it is easy to see that two identical amino acids linked together in a protein give a single dipeptide structure, whereas two identical hexoses may be linked together to give 64 possible isomeric disaccharide structures. If two different hexoses are linked together it is possible to obtain 128 different isomers (3).

The N-glycans, also called Asn- or **N-linked oligosaccharides**, in animals, plants, fungi and protista contain the common trimannosyl core structure that is linked *via* a chitobiosyl core (-GlcNAc-GlcNAc-) to Asn, as highlighted in Fig. 2. There are various types of N-glycans in animal cells, distinguished by the outer or terminal sugar structure, as seen for **high mannose-type**, hybrid-type and **complex-type** sequences. The O-glycans in mucins of animal cells, which are also called Ser/Thr- or O-linked oligosaccharides, are characterized by the linkage to Ser/Thr residues *via* GalNAc, as shown in Fig. 2. The GalNAc residue may be modified in different ways by linkage to other sugars to give various core structures. Altogether there are at least 8 different core structures in mucin-type O-glycans of animals. Some of the more common ones are highlighted in Fig. 2 as cores 1–4.

The attachment of N-glycans to Asn residues of secreted and membrane-bound glycoproteins occurs within a **consensus sequence** or sequon –Asn-X-Ser/Thr- (or Cys) (Table 2), although not all Asn residues within the sequon of such glycoproteins are always used. Asn residues outside this sequon are not N-glycosylated. In addition, cytoplasmic proteins with the N-glycosylation sequon are not N-glycosylated, because the pathway of N-glycosylation occurs within the lumen of the **endoplasmic reticulum** (ER), as discussed below. Although the N-glycosylation sequon is the most well known glycosylation sequon, a few other sugar-amino acid linkages also occur in definable sequences of proteins, as seen for addition of O-Glc, O-Fuc, C-Man and O-Gal (collagen) (Table 2). For most other attachments of sugars to proteins however, there are no clearly predictable consensus sequences, although the probability of attachment of some sugars to protein, such as GalNAc or GlcNAc to Ser/Thr residues in mucins and in cytosolic glycoproteins, appears to be enhanced by clusters of Ser and/or Thr residues and nearby amino acids (e.g. Pro) (Table 2). Some mathematical algorithms have been developed based on this information to predict sites of addition of GalNAc to Ser/Thr residues in animal mucins.

Glycosylation of Proteins. Table 2 Some Protein Consensus Sequences for Sugar Addition

Sugar-Protein Linkage	Consensus Sequence*
GlcNAc β -N-Asn	- Asn -Xaa-Ser/Thr- (where Xaa \neq Pro) [#] - Asn -Xaa-Cys- (where Xaa \neq Pro) <i>rare</i> (animals, plants, fungi, Protoctista)
Fuca α -O-Ser/Thr	-Cys-Xaa-Xaa-Gly-Gly- Ser/Thr -Cys- (animals)
Glc β -O-Ser/Thr	-Cys-Xaa- Ser -Xaa-Pro-Cys- (animals)
Gal β -Hyl	-Gly-Xaa- Hyl -Gly- (animal collagen)
Man-C-Trp	- Trp -Xaa-Xaa-Trp- (animals)
Xyl β -O-Ser	- Ser -Gly- (or Ala) (indefinite) (animals)
GalNAc α -O-Ser/Thr	clustered Ser/Thr near Pro (indefinite) (animals)

*Amino acids to which sugars are linked are in bold

[#]Abbreviations: See Table 1

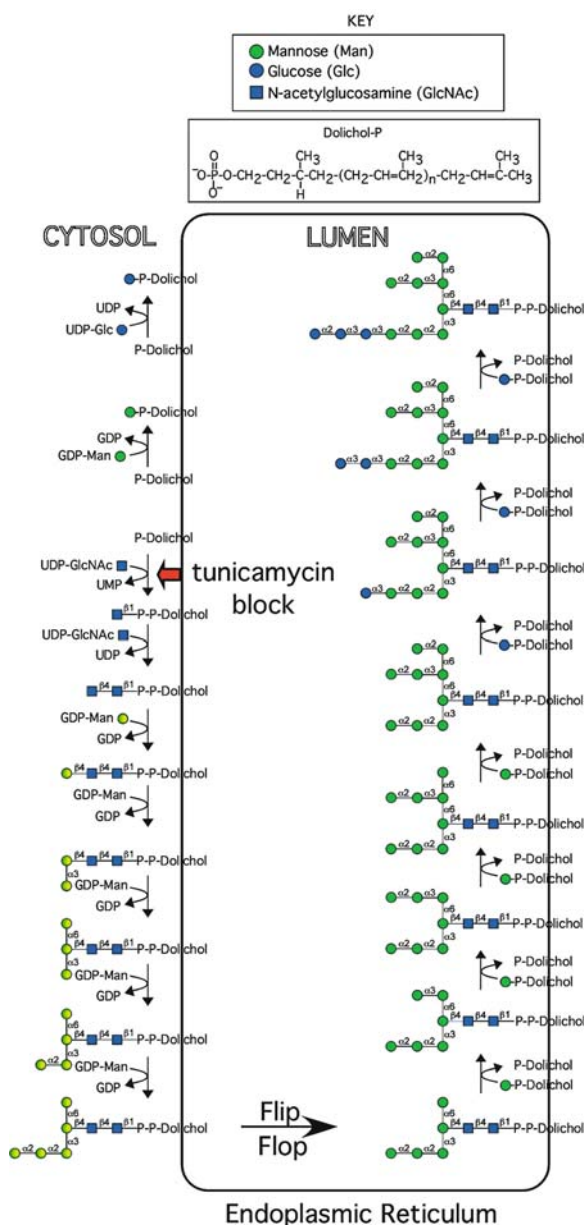
Glycoprotein Biosynthesis

N-Glycan Biosynthesis

In animal cells glycoprotein biosynthesis occurs in several cellular compartments. The primary sites for N-glycan biosynthesis are the ER and **Golgi apparatus**. N-glycans are generated by a unique pathway, involving the addition of a preformed glycan to Asn residues in the consensus sequence –Asn-X-Ser/Thr- in nascent or forming polypeptides during their translation (**co-translational modification**) through the translocon Sec61p, the pore-forming protein in the ER associated with ribosomes on the cytoplasmic face of the ER. The preformed glycan added to newly synthesized glycoproteins occurs as a lipid-linked donor **dolichol**-pyrophospho-oligosaccharide (dol-P-P-oligosaccharide), which in vertebrates contains 14 monosaccharide residues in the formula Glc₃Man₉GlcNAc₂-dolichol (Fig. 3). This large oligosaccharide can be eventually converted, as described below, to the high-mannose-, hybrid- and complex-type N-glycans discussed above. The enzyme that transfers this preformed glycan is called the **oligosaccharyltransferase** (OST) and occurs in all eukaryotes as a complex, hetero-oligomeric enzyme associated with the ER membrane. Dolichol (dol) is a polyisoprenoid alcohol (prenol) whose general formula can be seen from the structure of dolichol-phosphate (P-dol) (Fig. 3). Dolichol contains 75–95 carbons and is one of the largest, and most unusual lipids found in animals. It is synthesized from the same initial precursors and using the same early enzymatic steps that are used to generate sterols, such as cholesterol.

The synthesis of the dol-P-P-oligosaccharide also occurs in the ER and is initiated by the addition of N-acetylglucosamine from the sugar nucleotide donor

UDPGlcNAc to generate GlcNAc-P-P-dol (4). This step in the biosynthesis of N-glycans is blocked by the naturally occurring inhibitor **tunicamycin**, a transition state analog of UDP-GlcNAc, which was originally identified in the fungus-like soil bacterium *Streptomyces lysosuperificus*. (The name tunicamycin derives from its discovery as an antiviral agent that blocked viral coat (tunica) formation.) Treatment of animal cells with tunicamycin blocks N-glycosylation of proteins by blocking formation of the precursor dol-P-P-oligosaccharide) and results in cell death, due to the inability to synthesize and correctly fold glycoproteins, as discussed below. Interestingly, the synthesis of GlcNAc-P-P-dol and several other steps beyond this occurs on the cytoplasmic side of the ER. Following this first step to synthesize GlcNAc-P-P-dol, additional GlcNAc and mannose (Man) residues are added stepwise from their respective sugar nucleotide donors (Fig. 3). After the formation of Man₅GlcNAc₂-P-P-dol, which faces the cytoplasm, the Man₅GlcNAc₂-P-P-dol is “flipped” across the membrane of the ER by an unknown mechanism so that it now faces the lumen or inner region of the organelle. This Man₅GlcNAc₂-P-P-dol is then further elongated in the lumen of the ER by donation of Man residues from the dolichol intermediate dol-P-Man (Fig. 3). In vertebrate cells, following completion of the mannose addition to give Man₉GlcNAc₂-P-P-dol, 3 glucose residues are added by the intermediate donor dol-P-Glc, to generate the final product Glc₃Man₉GlcNAc₂-dol. Each step from the formation of GlcNAc-P-P-dol to the formation of Glc₃Man₉GlcNAc₂-P-P-dol is catalyzed by a distinct enzyme. Mutations in any steps in the pathway usually result in an inability to add other sugars, thus truncated dol-P-P-oligosaccharides are generated, which are often

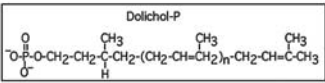


Glycosylation of Proteins. Figure 3 Biosynthesis of dolichol-P-P-oligosaccharide in higher animals. The pathway shown occurs in the cytosolic and luminal regions of the endoplasmic reticulum (ER). A separate enzyme catalyzes each step in the pathway shown. The mechanism by which the intermediate dolichol-P-P-oligosaccharide is reoriented in the ER membrane by “flip-flop” is not yet understood.

not efficiently utilized by the OST. The blockage in the addition of Dol-P-man, because of an inability to synthesize it, was first identified in a cultured mammalian cell line and resulted in the accumulation of $\text{Man}_5\text{GlcNAc}_2\text{-P-P-dol}$, leading to the addition of

$\text{Glc}_3\text{Man}_5\text{GlcNAc}_2$ instead of $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ to proteins. Such mutant cell lines, which are viable *in vitro*, helped to elucidate one of the key intermediate steps in this complex pathway of N-glycan biosynthesis. The addition of $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ to proteins occurs on the nascent polypeptides and multiple $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ residues can be added to Asn residues within the N-glycosylation sequon as they emerge translationally into the ER through the translocon Sec61p (Fig. 4). But protein folding can also begin with polypeptide intermediates and such folding may interfere with glycan addition to Asn residues. Thus, while Asn residues in some N-glycosylation sequons are quantitatively and efficiently N-glycosylated, Asn residues within other sequons may be only partly or inefficiently N-glycosylated. This partial glycosylation at some Asn residues may not be accidental however, and may be under metabolic control and help to regulate glycoprotein function. Following the addition of $\text{Glc}_3\text{Man}_9\text{GlcNAc}_2$ to protein, the glucose residues are removed sequentially from nascent polypeptides and completely translated glycoproteins by two different α -glucosidases (I and II) in the ER to generate $\text{Man}_9\text{GlcNAc}_2\text{-Asn}$ (Fig. 4). Glucosidases are examples of enzymes termed glycosidases, which are able to cleave glycosidic linkages. The actions of the α -glucosidases are the first steps in glycoprotein biosynthesis termed processing, where specific sugars are removed from a newly synthesized glycoprotein in an orderly fashion. These α -glucosidases are inhibited by some sugar analogs, such as australine, which inhibits α -glucosidase I and castanospermine and 1-deoxynorijirimycin, which inhibit both α -glucosidases I and II.

Following removal of the three Glc residues, a single glucose residue can be added back to $\text{Man}_9\text{GlcNAc}_2\text{-Asn}$ to generate $\text{Glc}_1\text{Man}_9\text{GlcNAc}_2\text{-Asn}$ (Fig. 4). This re-glycosylation is catalyzed by the enzyme UDPGlc: glycoprotein glucosyltransferase (UGGT), a protein found in the ER (5). UGGT adds Glc from the sugar nucleotide donor UDPGlc, and generates a UDP byproduct. The action of UGGT is part of a quality control process in which glycoprotein folding in the ER is regulated. Improperly folded proteins either fail to leave the ER and are degraded there or leave to be degraded by the **proteasome** machinery located in the cytoplasm. Mature proteins have a specific shape that results from folding the polypeptide backbone and chemical cross-linking between Cys residues to form disulfide bonds. While protein folding is spontaneous, it is relatively slow and inefficient. Protein folding during biosynthesis is rapid and this is usually achieved through the assistance of **molecular chaperones**. Chaperones are proteins that assist other proteins in acquiring their mature and active forms. This is often associated with proper protein folding and prevention of



generally defined as non-immune proteins that recognize and bind to specific glycan structures without catalyzing a chemical modification. Calnexin/calreticulin bind in a reversible manner and their binding is probably associated with binding of other chaperones that recognize specific peptide features of the protein. Once released from calnexin/calreticulin, the Glc₁Man₉GlcNAc₂-Asn is subject to action of α -glucosidase II, resulting in reformation of Man₉GlcNAc₂-Asn. If a glycoprotein is still not properly folded, the UGGT adds back Glc to regenerate Glc₁Man₉GlcNAc₂-Asn. UGGT participates in quality control of N-glycan biosynthesis and protein folding through this cycle of

glucosylation/deglucosylation, which is repeated until a glycoprotein assumes a conformation that blocks its interaction with UGGT. Blocking the action of α -glucosidase with castanospermine or other glucosidase inhibitors can result in glycoprotein accumulation in the ER, due to inefficient protein folding. This single Glc residue serves as a type of ER-retention signal, preventing glycoprotein exit from the ER.

Following the formation of $\text{Man}_9\text{GlcNAc}_2\text{-Asn}$ on a folded glycoprotein, the ER α -mannosidase removes one of the mannose residues of $\text{Man}_9\text{GlcNAc}_2\text{-Asn}$ to form $\text{Man}_8\text{GlcNAc}_2\text{-Asn}$ (Fig. 4). Interestingly, some cells contain an endomannosidase that can remove Glc-Man disaccharide from $\text{Glc}_1\text{Man}_9\text{GlcNAc}_2\text{-Asn}$ in an α -glucosidase II-independent pathway to form an alternative $\text{Man}_8\text{GlcNAc}_2\text{-Asn}$. The endomannosidase can also act on $\text{Glc}_{1-3}\text{Man}_9\text{GlcNAc}_2\text{-Asn}$ derivatives. Following formation of $\text{Man}_8\text{GlcNAc}_2\text{-Asn}$ on ER glycoproteins, they usually exit to the Golgi apparatus by vesicular transport involving COP-coated vesicles. The Golgi apparatus is recognized as a multi-compartment organelle, with cis, medial and trans compartments and a terminal compartment called the transGolgi network (TGN). The Golgi apparatus is usually positioned in the cell so that the cis-Golgi is nearest or proximal to the ER and the trans-Golgi is away from or distal to the ER.

Upon reaching the cis-Golgi the $\text{Man}_8\text{GlcNAc}_2\text{-Asn}$ in glycoproteins is subjected to further processing by α -mannosidase I (Fig. 4). This enzyme removes 3 additional α -linked mannose residues from $\text{Man}_8\text{GlcNAc}_2\text{-Asn}$ to generate $\text{Man}_5\text{GlcNAc}_2\text{-Asn}$. Not all the high-mannose N-glycans, however, are susceptible to α -mannosidase I. The $\text{Man}_8\text{GlcNAc}_2\text{-Asn}$ in some glycoproteins are not very accessible to α -mannosidase I, leading to formation of mature glycoproteins having $\text{Man}_8\text{GlcNAc}_2\text{-Asn}$ or partly processed forms such as $\text{Man}_7\text{-}$, $\text{Man}_6\text{-}$, or $\text{Man}_5\text{GlcNAc}_2\text{-Asn}$. The action of α -mannosidase-I can be inhibited by several drugs, including the plant alkaloid kifunensine and the mannose derivative 1-deoxymannojirimycin.

Acid hydrolases that are destined to enter lysosomes are subject to the action of an alternative pathway in which they acquire phosphorylated mannose residues (Man-6-P) that are recognized by the mannose-6-phosphate (Man-6-P) receptors (Fig. 4). These receptors help to deliver lysosomal enzymes to endosomes, from which the **lysosomal acid hydrolases** can enter mature lysosomes and the Man-6-P receptors recycle to the Golgi for additional rounds of delivery (6). The high mannose-type N-glycans of lysosomal acid hydrolases in the early Golgi apparatus are recognized by the UDPGlcNAc:lysosomal enzyme phosphotransferase. Many of the lysosomal acid hydrolases have unique 3-dimensional structures that generate a surface patch, which is a basic region that includes

Lys residues. The phosphotransferase recognizes the signal patch and adds GlcNAc-1-P from the donor UDPGlcNAc to nearby Man residues on the high mannose-type N-glycans to generate the phosphodiester $\text{GlcNAc-1-P-6-Man}_{5-8}\text{GlcNAc}_2\text{-Asn}$. Following formation of $\text{GlcNAc-1-P-6-Man}_{5-8}\text{GlcNAc}_2\text{-Asn}$ on lysosomal enzymes, they are subjected to the action of the α -N-acetylglucosamine-1-P phosphodiesterase ("uncovering" enzyme or UCE), which removes the α -linked GlcNAc residue, resulting in formation of the phosphomonoester structure $\text{P-6-Man}_{5-8}\text{GlcNAc}_2\text{-Asn}$. Thus, lysosomal enzymes acquire one or more Man-6-phosphate- (Man-6-P) phosphomonoester residues on high mannose-type N-glycans. The presence of Man-6-P blocks action of α -mannosidases on the specifically phosphorylated Man residues. Thus, as described below, phosphorylated glycans cannot be converted to complex-type N-glycans, although they can be converted to Man-6-P-containing hybrid-type N-glycans.

It is important to note that sugar nucleotides, which are important donors for glycosyltransferases in the lumen of the ER and Golgi apparatus, are synthesized in the cytosol. The sugar nucleotides are imported into the lumen of the ER and Golgi apparatus by specific transporters. Most of these transporters function as antiporters; they move a sugar nucleotide into the lumen and the cognate nucleoside monophosphate into the cytosol for reutilization. Defects in transporter function are associated with human genetic diseases, as discussed below.

The recognition of lysosomal hydrolases by the phosphotransferase is dependent on the folded structure of lysosomal enzymes and unfolded proteins are not recognized by the phosphotransferase (6). As discussed below, mutations in the phosphotransferase can result in lack of addition of GlcNAc-1-P to the more than 50 lysosomal enzymes. Alternatively, the phosphotransferase can stochastically fail to add the GlcNAc-1-P to a fraction of the lysosomal acid hydrolases. Such non-phosphorylated glycans of lysosomal acid hydrolases are subject to further processing and modification in the Golgi apparatus and can acquire complex-type N-glycan structures that contain sialic acid and other terminal sugars.

The $\text{Man}_5\text{GlcNAc}_2\text{-Asn}$ structures are potential acceptors for the enzyme N-acetylglucosaminyltransferase I (GNT-I), which adds GlcNAc from the donor UDPGlcNAc to form $\text{GlcNAcMan}_5\text{GlcNAc}_2\text{-Asn}$ (Fig. 4). This is a hybrid-type N-glycan, that contains non-reducing terminal Man residues and other terminal sugars, such as GlcNAc. In vertebrate cells the product $\text{GlcNAcMan}_5\text{GlcNAc}_2\text{-Asn}$ is usually acted upon by α -mannosidase II, which specifically recognizes $\text{GlcNAcMan}_5\text{GlcNAc}_2\text{-Asn}$ and removes two mannose residues to form $\text{GlcNAcMan}_3\text{GlcNAc}_2\text{-Asn}$. The

concerted action of α -mannosidases and GNT-I, which appears to occur largely in the cis-Golgi region, generates the trimannosyl structure common to all complex-type N-glycans (7). This GlcNAcMan₃-GlcNAc₂-Asn is usually acted upon by N-acetylglucosaminyltransferase II, which adds GlcNAc from UDPGlcNAc to form the product GlcNAc₂Man₃-GlcNAc₂-Asn. This is an example of a biantennary complex-type N-glycan, which is characterized by the lack of non-reducing terminal Man residues and the presence of other terminal sugars. The biantennary nature refers to the presence of two branches of the complex-type N-glycan. But within the cis-Golgi additional N-acetylglucosaminyltransferases (GNT-III through VI) can add additional GlcNAc residues to the Man residues to form bisected N-glycans, or multi-antennary N-glycans, such as tri-, tetra-, penta- and hexa-antennary structures. Within the more distal trans-Golgi apparatus, the GlcNAc₂Man₃GlcNAc₂-Asn is subjected to modification by galactosyltransferases, causing addition of Gal residues to GlcNAc residues from the donor UDPGal to form Gal β 4GlcNAc-R sequences. This disaccharide terminal sequence is termed N-acetylglucosamine (LN). In vertebrates the pituitary glycoprotein hormones, such as lactating hormone and follicle stimulating hormone, acquire biantennary N-glycans but are subject to addition of GalNAc residues from the donor UDPGalNAc to form the sequence GalNAc β 4GlcNAc-R; this terminal disaccharide is termed lactosamine-di-N-acetyl (LacdiNAc or LDN). This formation of LDN sequences on pituitary glycoprotein hormones requires the action of a specific N-acetylgalactosaminyltransferase that appears to recognize primary sequences within the hormones, and does not generally act on other glycoproteins within the pituitary (8). But less specific N-acetylgalactosaminyltransferases are also expressed in other cells to generate the LDN structure on non-pituitary glycoprotein hormones. The LDN termini of pituitary glycoprotein hormones are sulfated at the C-3 position of the terminal GalNAc residues by a PAPS:GalNAc 3-O-sulfotransferase to form S-3-GalNAc moieties. The resultant formation of (S-3-GalNAc)₂ GlcNAc₂Man₃GlcNAc₂-Asn on pituitary glycoprotein hormones promotes their recognition and clearance from the blood circulation by a liver receptor for S-3-GalNAc.

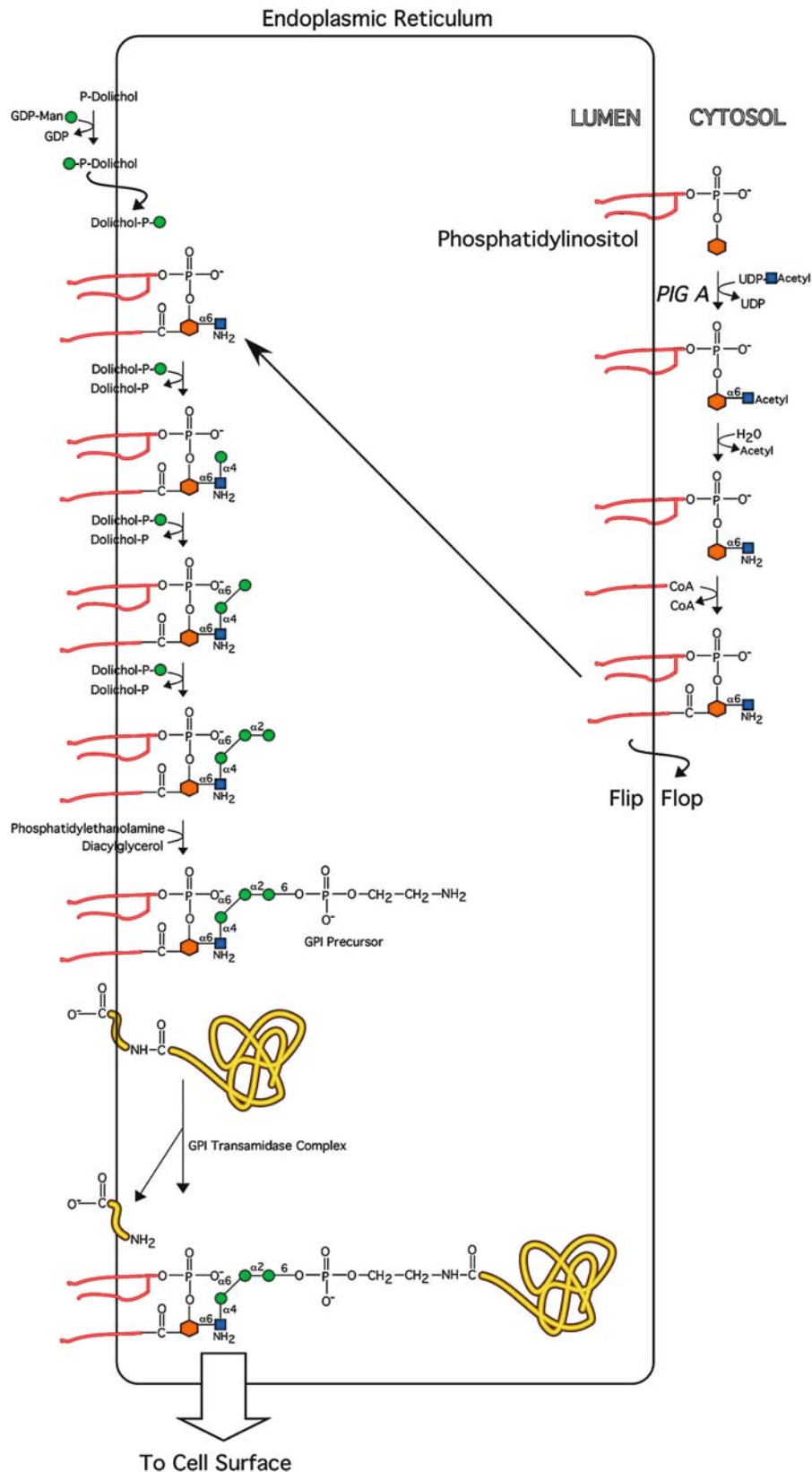
In most glycoproteins following the addition of Gal residues to GlcNAc residues, the complex-type N-glycans can acquire other modifications in the trans-Golgi and TGN. These include addition of sialic acid from CMPNeuAc, fucose from GDPFuc, and other residues. Each addition or modification is catalyzed by a separate enzyme. A tremendous variety of modifications are possible depending on a wide variety of factors, such as expression of the modifying enzymes

and the availability of the glycoprotein glycans to the enzymes. Following these modifications of the N-glycans in the Golgi apparatus, secretory glycoproteins are released into the extracellular space by secretory vesicles, while membrane-bound glycoproteins may be targeted to the plasma membrane or lysosomes.

GPI-Anchor Biosynthesis

Many glycoproteins in eukaryotes contain a novel C-terminal modification of a glycosylphosphatidylinositol lipid anchor, the GPI anchor (9). The addition of the GPI-anchor to proteins occurs in the ER and involves recognition of a C-terminal domain of newly synthesized protein (Fig. 5). A preformed lipid-linked precursor is generated from phosphatidylinositol by initial reactions in the cytosolic face of the ER. An intermediate containing glucosamine is then reoriented (“flip-flop”) by translocation across the ER membrane to allow further elongation of the precursor by addition of Man residues from dol-P-Man. Ethanamine phosphate is added to Man residues by donation from phosphatidylethanolamine to generate a GPI precursor. In all organisms this GPI precursor is characterized by having glucosamine linked to inositol and the trimannosyl sequence linked to ethanolamine in a “core” structure. But this core structure may be differentially modified in a tremendous variety of ways, depending on the organism, by addition of other sugars, e.g. Man and Gal residues, additional ethanolamine residues, addition of phosphate, fatty acylation of the sugars and further acylation of inositol.

The GPI precursor is the substrate for a transamidation reaction by the GPI transamidase complex. This enzyme complex recognizes the C-terminal lipophilic portion of some proteins with a GPI anchor sequence, causing the cleavage of the polypeptide bond and transfer of the GPI precursor to the new C-terminal amino acid. The GPI transamidase forms a carbonyl intermediate with the substrate protein. The signal sequence for GPI anchor addition is a C-terminal region with an amino acid to which the anchor is eventually attached that is termed the ω site. The amino acids that are two residues to the carboxyl side of ω residue (the $\omega + 2$ site) have small side chains, whereas the residues at the $\omega + 1$ site can have large side chains. In all cases the $\omega + 2$ site is followed by a stretch of 5–10 hydrophilic amino acids and then 15–20 hydrophobic residues at or very near the carboxyl or C-terminus of the protein. Following the addition of the GPI anchor, the GPI-anchored glycoproteins move to the plasma membrane. These glycoproteins usually have other sugar residues attached to other amino acids, such as N-glycans, that may or may not be processed within the ER and Golgi apparatus. It is interesting that the formation of N-glycans and GPI-anchored glycoproteins uses a common intermediate, i.e. dol-P-Man.



Glycosylation of Proteins. Figure 5 Biosynthesis of GPI-anchored glycoproteins. The pathway shown is for human GPI anchor biosynthesis from phosphatidylinositol, which occurs in the cytosolic and luminal regions of the ER. Following generating of the GPI anchor precursor, the GPI anchor is added *en bloc* to the C-terminal region of an ER protein by the transamidase complex, resulting in the cleavage and release of a C-terminal peptide.

Some individuals have a mutation in the gene (termed *PIG A*) encoding the first enzyme of the pathway for GPI anchor biosynthesis that normally adds GlcNAc from UDPGlcNAc to phosphatidylinositol. Thus, these individuals are defective in generating the mature GPI anchor precursor and are deficient in generating GPI-anchored glycoproteins. Such individuals are often clinically recognized as having paroxysmal nocturnal hemoglobinuria (PNH), a form of hemolytic anemia. GPI-anchored glycoproteins also occur in many protozoans, and have been especially well characterized in African trypanosomes, where the GPI-anchored glycoprotein is recognized as a highly antigenic variant surface glycoprotein (VSG).

Mucin-Type O-glycan Biosynthesis

Many glycoproteins within the Golgi apparatus are modified to contain GalNAc α 1-Ser/Thr residues, typically found in animal mucins, by the action of a family of UDPGalNAc:polypeptide α -N-acetylgalactosaminyltransferases (ppGalNAcTs). While mucins may contain hundreds of such linkages, some glycoproteins, such as the transferrin receptor, contain only a single O-glycan. Yet, all such linkages are categorized as mucin-type. The ppGalNAcTs recognize Ser and Thr residues in glycoproteins and add GalNAc in O-glycosidic linkage from the donor UDPGalNAc to these amino acid side chains to form GalNAc α 1-Ser/Thr, which is also called the Tn antigen (10). Well over a dozen different ppGalNAcTs are known and many of these are expressed simultaneously within cells. Some of these enzymes may have unique, but partly overlapping, recognition of Ser/Thr residues within the polypeptide sequence. Interestingly, many ppGalNAcTs are dual function enzymes containing a catalytic domain that transfers GalNAc and a lectin domain (ricin- or R-type) that binds to GalNAc residues. Thus, addition of GalNAc to some Ser/Thr sites may promote further modification by attracting more ppGalNAcTs. Such concerted actions of ppGalNAcTs in the Golgi apparatus may promote the relatively efficient modifications of hundreds of Ser/Thr residues within some very large mucin polypeptides, some of which have over 10,000 amino acids.

Following the formation of GalNAc α 1-Ser/Thr residues, glycoproteins are subjected to the action of a β -3-galactosyltransferase, also called the T-synthase, to form the disaccharide Gal β 3GalNAc α 1-Ser/Thr, which is called the Thomsen-Friedenrich, TF or simply T antigen, using the donor UDPGal. The T antigen disaccharide is also the simplest core 1 O-glycan structure (Fig. 2). However, occasionally the GalNAc α 1-Ser/Thr residues may be sialylated to generate the disaccharide NeuA α 6GalNAc α 1-Ser/Thr (sialyl Tn antigen), which cannot be further modified. Upon formation of Gal β 3GalNAc α 1-Ser/Thr, the core 1 structure may be modified by an

N-acetylglucosaminyltransferase (the core 2 GlcNAcT) to generate the trisaccharide Gal β 3(GlcNAc β 6)GalNAc α 1-Ser/Thr (core 2 O-glycan) from the donor UDPGlcNAc. This core 2 O-glycan can be subsequently modified by addition of other sugars, such as galactose, fucose and N-acetylneuraminic acid, and/or sulfate residues on selected sugars to generate a wide variety of O-glycan structures.

Biosynthesis of Other O-Glycans

The biosynthesis of non-mucin type O-glycans is incredibly varied depending on the cellular compartment. O-GlcNAc residues are added to proteins in the cytoplasm, as discussed below. Glycosaminoglycan addition is initiated by UDPXyl:core protein β -D-xylosyltransferases I and II, which transfer Xyl from UDPXyl to specific Ser residues in proteoglycan core proteins in the ER. The Xyl residue is subsequently modified by addition of Gal and GlcA residues by galactosyltransferases and glucuronyltransferases respectively, to form the core linkage tetrasaccharide of glycosaminoglycans, GlcA β 1-3Gal β 1-3Gal β 1-4Xyl β 1-Ser, which occurs in all proteoglycans. This synthesis of the glycosaminoglycan core region may be completed in the ER, while the subsequent elongation of the glycans and sulfation and epimerization, which is an orchestrated and incredibly complex series of reactions, appear to occur primarily in the Golgi apparatus. The elongation of glycosaminoglycans on proteoglycans can be partly averted by feeding cells β -xylosides, that act as acceptors for addition of Gal, thus effectively decreasing elongation of glycosaminoglycan within the proteoglycan acceptors. Remarkably, β -xylosides appear capable of penetrating the ER and possibly the Golgi apparatus of cells. This inhibition by competition can result in synthesis of free glycosaminoglycans on the β -xyloside and reduced addition of glycosaminoglycan to proteoglycans. O-Mannosylation of proteins in yeast is initiated in the ER by transfer of Man from the donor dol-P-Man using a specific O-mannosyltransferase. Further elongation to generate mannose-containing polysaccharides in yeast occurs by Man donation from GDPMan in the Golgi apparatus by additional mannosyltransferases. An equivalent enzyme in animals, termed POMT1, may initiate O-Man formation on selective Ser/Thr residues in glycoproteins in the ER using dol-P-Man as the donor, while further elongation and addition of other sugars may occur in the Golgi apparatus. O-fucosylation and O-glucosylation of EGF-like domains on glycoproteins are catalyzed by specific enzymes that transfer Fuc or Glc from GDPFuc or UDPGlc, respectively, in the Golgi apparatus. Collagen is glycosylated in the ER following hydroxylation of Lys residues to generate hydroxylysine (Hyl). The addition of Gal to Hyl is catalyzed by a collagen-specific enzyme UDPGal:

procollagen-5-hydroxy-L-lysine D-galactosyltransferase, which adds Gal to Hyl residues on procollagen in the ER during procollagen biosynthesis and concomitantly with Hyl formation on nascent polypeptides catalyzed by lysyl hydroxylase activity.

Glycosylation in the Cytosol

Many cytosolic proteins in animals (and probably plants) contain one or more residues of β -linked GlcNAc in O-glycosidic linkage to Ser/Thr residues (11). These O-GlcNAcylated proteins (O-GlcNAc-containing glycoproteins) are generated by the action of the UDPGlcNAc:polypeptide O-acetylglucosaminyltransferase (O-GlcNAc transferase), which transfers GlcNAc from UDPGlcNAc to selected Ser/Thr residues of cytosolic proteins. Some of the more prominent O-GlcNAcylated glycoproteins include RNA polymerase II, c-myc and the estrogen receptor. O-GlcNAcylation is one of the only types of glycosylation that is reversible. The O-GlcNAc may be selectively removed by the action of an O-GlcNAc specific acetylglucosaminidase (O-GlcNAcase) in the cytosol. This alternating addition and removal of O-GlcNAc by these two enzymes is akin to reversible phosphorylation and dephosphorylation of cytosolic proteins. O-GlcNAcylation may serve to regulate many metabolic pathways and is required for animal and plant cell growth.

The storage polysaccharide glycogen, which is a glycoprotein in animals, is generated on the core protein glycogenin within the cytosol of animals, by its autocatalytic “self-glucosylation” of a Tyr residue at position 194 using UDPGlc as a donor. The Glc-O-Tyr is then elongated by addition of other Glc residues (up to ~ 10) from UDPGlc by glycogenin activity. The Glc-containing oligosaccharide on glycogenin is then elongated by glycogen synthase. A similar type of activity may occur on the starch protein amylogenin.

Plant Glycoproteins

Many plant glycoproteins contain N-glycans, which are also synthesized *via* the dolichol pathway in the ER. They can also be subsequently modified by processing reactions and addition of other sugars to generate high mannose-, hybrid- and complex-type N-glycans. Many plant wall proteins are typically glycoproteins rich in the amino acids hydroxyproline (**►hydroxyproline-rich glycoprotein**, HRGP), proline (proline-rich protein, PRP), and glycine (glycine-rich protein, GRP). The O-glycans in HRGPs may account for up to 95% of the glycoprotein weight and the glycans can range in size from a single attached Ara residue to large **►arabino-galactans** containing nearly 100 residues of Ara and Gal. Many of these glycoproteins form rods (HRGP, PRP) or β -pleated sheets (GRP). Extensin is one of the best-studied HRGPs. HRGP expression is increased by

plant wounding and pathogen attack. The pistil and pollen tube extracellular matrix are enriched in these highly glycosylated proteins.

Bacterial Glycoproteins

Glycoproteins are also found in prokaryotes and in archaeobacteria, although the general structures of the attached glycans and sugar residues are very different from those found in animals and plants. Among the best studied prokaryotic glycoproteins are the cell surface or S-layer glycoproteins (12). Such S-layer glycoproteins can assemble into ordered lattice-like structures on the cell surface. Each S-layer glycoprotein may contain more than one attached glycan, which can be linked *via* Asn or other amino acid residues (Table 1). In many cases the bacterial S-layer glycan chains are linear or branched homo- or hetero-saccharides having 20–50 identical repeating units. By contrast, archaeal S-layer glycoproteins have shorter glycans, generally lacking repeating units. Although the exact mechanisms of S-layer glycoprotein biosynthesis are not yet defined, it appears that most sugar residue addition occurs in the outer membrane following protein translocation.

Many Factors Regulate Protein Glycosylation

As discussed above, two of the major factors regulating protein glycosylation are the sequence motifs within the primary structure of glycoproteins and the site of biosynthesis. But many other factors also contribute to regulation of protein glycosylation. These include the expression of glycosyltransferases, expression of glycosidases, secondary, tertiary and/or quaternary structures of proteins, availability of donor substrates, e.g. sugar nucleotides and dolichol, cations, e.g. magnesium and manganese, temperature and membrane lipid composition and structure. Many of these factors, especially expression of glycosyltransferases, vary tremendously between cell types. Dozens of different glycosyltransferase genes encoding enzymes that act on glycoproteins exist in the genomes of most multicellular organisms. Together, these many factors help to explain the huge differences in glycosylation observed between different cells and tissues.

Glycoproteins Have Many Biological Functions

Because glycoproteins are so common in all cells, it is not surprising that the glycan moieties have many different functions. Although many of the specific functions of glycoproteins are being defined, it is likely that the complete picture of glycoprotein functions will take many years to complete. Some of the known functions of glycoproteins and their attached glycans include cell-cell adhesion, cell-matrix interactions, glycoprotein targeting to organelles and cell signaling. For example, glycoproteins regulate many different

types of cell adhesion, including sperm-egg adhesion, leukocyte-platelet-endothelial cells adhesion, recognition and phagocytosis and neuronal cell-matrix adhesion. Some of the non-specific functions of glycoprotein glycans include protein folding and assembly, protein protection and stability against proteases, control of the circulatory half-life of glycoproteins, regulation of protein conformation and thermal stability and control of enzyme kinetics. Many glycoprotein glycan functions are generated by glycan recognition through carbohydrate-binding proteins or lectins. Lectins are made by all organisms, including animal, plants, bacteria and viruses.

Human Disorders Associated with Defective Protein Glycosylation

There are many human disorders associated with an altered ability to add carbohydrate residues to glycoproteins. One of the first defined examples of this was [►I-cell disease](#), where patients were found to have a recessive genetic mutation of the gene encoding the phosphotransferase activity that helps to generate Man-6-P residues on lysosomal acid hydrolases. Consequently, their cells are unable to synthesize lysosomal acid hydrolases with Man-6-P residues efficiently (13). Most of the non-phosphorylated lysosomal acid hydrolases from these patients become processed within the Golgi apparatus, acquire sialic acid and other sugar residues on hybrid- and complex-type N-glycans and are secreted into body fluids. The patients accumulate undegraded macromolecules in lysosomes due to lack of acid hydrolases and these accumulations are recognized microscopically as inclusion or I-cells, hence the name I-cell disease. Another historically important defect in glycoprotein glycosylation associated with human disease is PNH. Patients with PNH have reduced ability to generate the GPI anchor, due to mutation in the X-linked *PIG A* gene. Hemolytic anemia results due to deficiencies in the normally GPI-anchored glycoproteins termed decay accelerating factor (DAF or CD55) and membrane inhibitor of reactive lysis (MIRL or CD59), which function to decrease autolysis of erythrocytes by activated complement.

Many of the genetic defects in the ability to glycosylate proteins are now recognized within the broad category of [►congenital disorders in glycosylation](#) (CDGs) (14). The CDGs are highly varied depending on the glycan structures made and are recognized as different types, such as Type 1a, 1b, 1c, 1d, 1e, and IIa. Each type of CDG results from mutations in one of the many genes encoding proteins involved in N-glycosylation *via* the dolichol pathway or subsequent processing and glycosylation reactions, or in genes regulating organelle trafficking and biosynthesis. CDGs are often diagnosed by examining the N-glycosylation pattern of serum glycoproteins, such as transferrin, where altered

N-glycosylation can affect mobility upon isoelectric focusing chromatography. CDG patients, depending on the altered gene, exhibit a variety of changes in physiognomy and suffer from neurological, liver and/or intestinal problems. Children with CDG, depending on the type of genetic mutation, exhibit impairments in cognitive ability, speech and balance and motor skills. Other disorders where altered protein glycosylation is observed include several forms of congenital muscular dystrophy, such as Fukuyama congenital muscular dystrophy, [►limb-girdle muscular dystrophy](#), muscle-eye-brain disease and Walker-Warburg syndrome (15). Many of these diseases are associated with mutations in genes encoding glycosyltransferases that add GlcNAc or Man residues to generate O-linked Man-containing glycans on α -dystroglycan, which is a membrane-associated glycoprotein that helps to link neuronal cells and their cytosolic signaling machinery to extracellular matrix molecules such as laminin. Patients with progeroid-type Ehlers-Danlos (E-D) syndrome have defects in a galactosyltransferase that is required to synthesize the common linkage region of glycosaminoglycans. Another disorder where altered protein glycosylation is observed is leukocyte adhesion deficiency type II (LAD II), where patients lack the ability to add fucose to glycoproteins. This deficiency in fucosylation results in a lack of leukocyte adhesion to selectins, a group of carbohydrate-binding proteins that recognize fucose-containing O-glycans and serve to regulate leukocyte trafficking from the bloodstream. Some LAD II patients have mutations in the gene encoding the Golgi transporter for GDPFuc, thus preventing normal movement of GDPFuc from its site of synthesis in the cytosol into the Golgi apparatus for utilization by fucosyltransferases. Finally, defects in glycoprotein glycosylation are also seen in patients with some autoimmune diseases, such as may occur in congenital dyserythropoietic anemia type II, where a defect in N-glycosylation may occur due to deficiency of α -mannosidase II activity and in IgA nephropathy, where a subset of IgA molecules lack appropriate O-glycan structures within the hinge region. These are just a few of the many examples where protein glycosylation is essential to biological processes.

[►Biochemical Engineering of Glycoproteins](#)

[►Protein Databases](#)

[►Recombinant Protein Production in Mammalian Cell Culture](#)

References

1. Apweiler R, Hermjakob H, Sharon N (1999) On the frequency of protein glycosylation, as deduced from analysis of the SWISS-PROT database. *Biochim Biophys Acta* 1473:4–8
2. Spiro RG (2002) Protein glycosylation: nature, distribution, enzymatic formation, and disease implications of glycopeptide bonds. *Glycobiology* 12:43R–56R

3. Laine RA (1994) A calculation of all possible oligosaccharide isomers both branched and linear yields 1.05×10^{12} structures for a reducing hexasaccharide: the Isomer Barrier to development of single-method saccharide sequencing or synthesis systems. *Glycobiology* 4:759–767
4. Helenius J, Aebi M (2002) Transmembrane movement of dolichol linked carbohydrates during N-glycoprotein biosynthesis in the endoplasmic reticulum. *Semin Cell Dev Biol* 13:171–178
5. Parodi AJ (2000) Protein glucosylation and its role in protein folding. *Annu Rev Biochem* 69:69–93
6. Kornfeld S (1987) Chromatography: A review of clinical applications. *FASEB J* 1:462–468
7. Schachter H (2000) The joys of HexNAc. The synthesis and function of N- and O-glycan branches. *Glycoconj J* 17:465–483
8. Baenziger JU, Green ED (1988) Pituitary glycoprotein hormone oligosaccharides: structure, synthesis and function of the asparagine-linked oligosaccharides on lutropin, follitropin and thyrotropin. *Biochim Biophys Acta* 947:287–306
9. McConville MJ, Menon AK (2000) Recent developments in the cell biology and biochemistry of glycosylphosphatidylinositol lipids (review). *Mol Membr Biol* 17:1–16
10. Brockhausen I (1999) Pathways of O-glycan biosynthesis in cancer cells. *Biochim Biophys Acta* 1473:67–95
11. Wells L, Vosseller K, Hart GW (2001) Glycosylation of nucleocytoplasmic proteins: signal transduction and O-GlcNAc. *Science* 291:2376–2378
12. Schaffer C, Messner P (2004) Surface-layer glycoproteins: an example for the diversity of bacterial glycosylation with promising impacts on nanobiotechnology. *Glycobiology* 14:31R–42R
13. Raas-Rothschild A, Cormier-Daire V, Bao M et al (2000) Molecular basis of variant pseudo-hurler polydystrophy (mucopolidosis IIIC). *J Clin Invest* 105:673–681
14. Marquardt T, Denecke J (2003) Congenital disorders of glycosylation: review of their molecular bases, clinical presentations and specific therapies. *Eur J Pediatr* 162:359–379
15. Endo T, Toda T (2003) Glycosylation in congenital muscular dystrophies. *Biol Pharm Bull* 26:1641–1647

Glycosylphosphatidylinositol (GPI) Anchors

Definition

In the lumen of the endoplasmic reticulum, the GPI anchor is covalently attached to the C terminus of proteins destined for the plasma membrane, and the transmembrane segment of the protein is cleaved off. As proteins are only attached to the exofacial leaflet of the plasma membrane by the GPI anchor, they can be

released in soluble form from the cell surface by the action of specific phospholipases.

- ▶ [Epithelial Cells](#)
- ▶ [Glycosylation of Proteins](#)

Glycosyltransferase

Definition

Glycosyltransferase is a member of a large family of enzymes expressed in the endoplasmic reticulum and Golgi apparatus, which catalyze the transfer of a monosaccharide unit from a sugar-nucleotide donor, typically to the non-reducing terminus of an oligosaccharide chain in glycoproteins and glycolipids.

- ▶ [Glycosylation of Proteins](#)
- ▶ [Limb Girdle Muscular Dystrophies](#)
- ▶ [Methylation of Proteins](#)

Glyoxylate

Definition

Glyoxylate is a toxic compound, generated *in vivo*, which needs to be eliminated by conversion into glycine via the peroxisomal enzyme alanine glyoxylate aminotransferase.

- ▶ [Peroxisomal Disorders](#)

Glypidation

Definition

Glypidation describes the attachment of a glycosylphosphatidylinositol-(GPI)-anchor to certain integral membrane proteins. The anchor is composed of the lipid phosphatidylinositol to which a carbohydrate and an ethanol and phosphate moiety is linked. The GPI-anchor is attached post-translationally in the lumen of the endoplasmic reticulum, thereby replacing a transient transmembrane region of the modified protein. The ▶ [GPI-anchor](#) attaches proteins to the exoplasmic leaflet of membranes, possibly to certain subdomains such as caveolae and lipid-rafts.

- ▶ [Fatty Acid Acylation of Proteins](#)
- ▶ [Glycosylation of Proteins](#)

Golgi Apparatus (Golgi Complex)

Definition

Golgi apparatus (Golgi complex) refers to a cytoplasmic organelle in eukaryotes consisting of stacked, flattened membrane cisternae, surrounded by vesicles, which is involved in transport and post-translational modification (especially glycosylation) of proteins on their journey through the secretory pathway. The Golgi complex is also a central sorting station in the secretory pathway; on the trans- or exit-side of the Golgi complex, proteins get sorted into several distinct vesicle types for transport to different final destinations.

- ▶ Biochemical Engineering of Glycoproteins
- ▶ Exocytotic Pathway
- ▶ Glycosylation of Proteins
- ▶ Limb Girdle Muscular Dystrophies
- ▶ Rho, Rac, Cdc42
- ▶ Vesicular Traffic

Gomori Trichrome

Definition

Gomori trichrome is a mixture of chemical compounds that stain mitochondria red.

- ▶ Mitochondrial Myopathies

Gonadal Mosaicism

- ▶ Germline (Gonadal) Mosaicism

Gonadotropin Deficiency

Definition

Gonadotropin deficiency describes the absence, decreased production or dysfunction of anterior pituitary hormone (LH and/or FSH), which results in a decreased or lack of testosterone in males and estrogen

in females. This causes delayed or no pubertal development and infertility.

- ▶ Hypothalamic and Pituitary Diseases Genetics

Gonadotropins

Definition

Gonadotropins are pituitary hormones that influence the functions of the ovary: ▶ follicle stimulating hormone (FSH) and ▶ luteinizing hormone (LH).

- ▶ SRY – Sex Reversal
- ▶ Hypothalamic and Pituitary Disease, Genetics

Gordon's Syndrome

Definition

Gordon's syndrome is also known as type 2 Pseudo-hypoaldosteronism (PHA2).

- ▶ Mendelian Forms of Human Hypertension and Mechanisms of Disease
- ▶ Type 2 Pseudohypoaldosteronism

Gorlin's Syndrome

Definition

Gorlin's syndrome (also known as naevoid basal cell carcinoma syndrome (NBCCS)) is a rare autosomal dominant cancer disorder characterised primarily by a predisposition to several tumours, most commonly basal cell carcinoma (BCC). In addition to cancer susceptibility, this syndrome is also associated with a range of defects resulting from abnormal embryonic development. Gorlin's syndrome results from mutation of the *patched* gene which functions in the hedgehog signalling pathway. The developmental defects are believed to result from ▶ haploinsufficiency, with subsequent mutation of the remaining allele resulting in tumour formation.

- ▶ Hedgehog Signalling

GPCRs

► G-Protein Coupled Receptors

G-Phase

► Cell Cycle – Overview

GPI (-Anchored) Protein

Definition

A GPI-anchored protein is a protein that is anchored in the membrane by glycosylated derivatives of phosphatidylinositol (GPI). The carboxyl group of the C-terminal amino acid is connected through an amide link to phosphoethanolamine, which is attached to a core tetrasaccharide composed of three mannose sugars and a single glucosamine sugar. The tetrasaccharide is in turn attached to phosphatidylinositol embedded in the membrane. Proteins with this type of lipid anchor are only found on the extracellular face (outer leaflet) of the membrane, and can be released in soluble form on the cell's surface by the action of specific phospholipases.

- Biological Membranes
- Epithelial Cells
- Glycosylation of Proteins

GPIIb/IIIa Complex

Definition

The GPIIb/IIIa complex is a platelet membrane glycoprotein complex mediating platelet aggregation and adhesion to endothelial cells. The complex is an integrin that recognises the arginine-glycine-aspartic acid (rgd) sequence present on several adhesive proteins. The GPIIb/IIIa complex functions as a receptor for fibrinogen, von Willebrand Factor (vWGF), fibronectin, vitronectin, and thrombospondin. Deficiency of GPIIb/IIIa causes ► [Glanzmann's Thrombasthenia](#).

G-Protein Coupled Proteolytic Site

A G-protein coupled proteolytic site is a peptide sequence found in a number of G-protein coupled receptors that acts as the target for specific proteolytic cleavage, which releases the extracellular portion of the receptor from the rest of the molecule.

- Autosomal Dominant (Inherited Disorder)
- G-Proteins and G-Protein Mutations in Human Diseases
- Polycystic Kidney Disease, Autosomal Dominant

G-Protein Coupled Receptors

Definition

G-protein coupled receptors (GPCRs) comprise of the largest family of cell surface receptors, which communicate their signal through G-proteins. A common structural feature of GPCRs is the presence of seven hydrophobic transmembrane helices. Several hundred subtypes exist, which bind a huge variety of ligands, such as hormones and neurotransmitters. About 50% of the currently used drugs are directed at GPCR's.

- Cardiac Signaling: Cellular, Molecular and Clinical Aspects
- Cytokine Receptors
- G-Proteins and G-Protein Mutations in Human Diseases
- Growth Factors
- Methylation of Proteins
- Photoreceptors
- Seven-Transmembrane Receptors
- Wnt/Beta-Catenin Signaling Pathway

G-Proteins

Definition

- Diabetes Insipidus, a Water Homeostasis Disease
- G-Proteins and G-Protein Mutations in Human Diseases
- Molecular Motors

G-Proteins and G-Protein Mutations in Human Diseases

THOMAS GUDERMANN

Phillips-University Marburg, Marburg, Germany
gudermann@staff.uni-marburg.de

Synonyms

Heterotrimeric guanine nucleotide-binding proteins

Definition

A plethora of extracellular signaling molecules like hormones, neurotransmitters, autacoids and growth factors convey information between cells of a living organism. The majority of these extracellular signaling molecules transmit their signal by interacting with a three-protein transmembrane signal transduction system composed of receptors, G-proteins and cellular effectors. All mammalian cells are endowed with a complement of G-protein-coupled receptors and several types of heterotrimeric G-proteins. Over the last few years systematic biochemical, cell biological and structural studies have laid a solid foundation for our understanding of G-protein dependent signal transduction processes. The characterization of genetically engineered mice carrying mutations in different G-protein genes as well as the clinical phenotype of patients affected by mutated G-proteins have greatly furthered our understanding of the biological functions of heterotrimeric G-proteins as central processors of information.

Characteristics

Basic Structures, Mechanisms and Classifications

The vast majority of extracellular signals interact with transmembrane receptors which couple to heterotrimeric guanine nucleotide-binding proteins (G-proteins) acting as transducers and signal amplifiers (1). Activated G-proteins then modulate the activity of cellular effectors. A comprehensive analysis of the human and mouse genomes defined a repertoire of G-protein-coupled receptors (GPCRs) for endogenous ligands comprising close to 400 genes (2, 3). GPCRs form a large and functionally diverse superfamily, participate in a variety of physiological processes and are prime targets for pharmaceutical drugs. They are integral membrane proteins, characterized by 7 α -helical transmembrane domains arranged in an anti-clockwise bundle (as viewed from the extracellular side) and connected by alternating extracellular and intracellular loops of variable lengths. The crystal structure of the prototypical GPCR **►rhodopsin** in the

inactive ground state has been resolved, by and large confirming the overall architecture deduced from analogy-based biomodelling and various mutagenesis approaches (4). As yet, it is only rudimentarily understood how ligand-induced conformational change of a GPCR is translated into G-protein activation.

G-Protein Composition and Structural Aspects

G-proteins are heterotrimers composed of α , β , and γ subunits. The α subunit is responsible for guanine nucleotide binding and GTP hydrolysis; the β and γ subunits are associated in a tenaciously linked $\beta\gamma$ complex and can be regarded as one functional unit. To date 16 distinct genes for α , 5 for β and 12 for γ subunits have been identified and characterized functionally (Table 1). Access to a deeper understanding of G-protein structure and function at the atomic level has been granted by solving crystal structures of G-protein α subunits in the GDP- and GTP-bound forms as well as in the transition state (5). In addition, atomic structures of G-protein $\beta\gamma$ dimers and of $\alpha\beta\gamma$ heterotrimers are also available. $G\alpha$ proteins are principally composed of a **►Ras-like GTPase** and an α -helical domain, both forming a deep cleft harboring the guanine nucleotide. $G\beta$ subunits fold into a highly symmetrical β propeller with an approximate 7-fold symmetry. $G\gamma$ binds to $G\beta$ in an extended conformation devoid of intrachain tertiary interactions. Those segments in the $G\alpha$ protein that undergo structural changes upon GTP hydrolysis are named switch I, II, and III regions. The most obvious changes occur in the switch II region. A mechanistic explanation for subunit dissociation and reassociation contingent upon the guanine nucleotide bound to the α subunit can be derived from the observation that the most extensive contact area between $G\alpha$ and $\beta\gamma$ comprises the protein surface around the switch II region of $G\alpha$. So far, structural information does not provide an obvious cue as to the specificity in the pairing of a particular α subunit with a defined $\beta\gamma$ dimer.

Because effector contact sites of $G\alpha$ have also been mapped to areas around the switch II region, $G\beta\gamma$ - and effector-interacting surfaces of $G\alpha$ overlap significantly. Thus, $G\alpha$ cannot interact with cellular effectors unless it dissociates from $G\beta\gamma$. Conversely, activation of $G\beta\gamma$ is a consequence of its release from $G\alpha$, which functions as a negative regulator of free $\beta\gamma$ subunits within the cell. The regions of $G\beta\gamma$ that interact with downstream effectors map to an N-terminal $G\beta$ fragment of approximately 100 amino acids. As yet, it is not understood how an activated receptor catalyses the dissociation of GDP from a G-protein heterotrimer. The most clearly defined $G\alpha$ contact sites with heptahelical receptors are located in the C-terminal region of the α subunit. However, compelling evidence

G-Proteins and G-Protein Mutations in Human Diseases. Table 1 G-protein α subunits

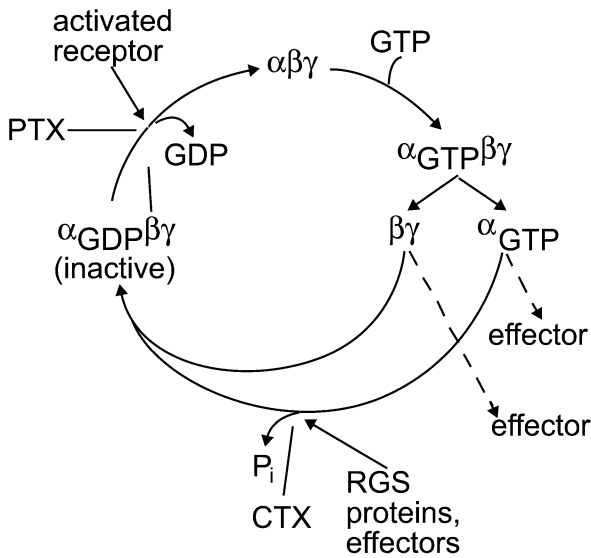
Name	Expression	Examples of effectors	Bacterial Toxins
α_s subfamily			
α_s	ubiquitous	AC \uparrow , VGCC, Src, RGS-PX1 (GAP, sorting nexin)	CTX
α_{olf}	olfactory epithelium, brain		CTX
α_i subfamily			
α_{t-r}	retinal rods	cGMP PDE \uparrow	CTX, PTX
α_{t-c}	retinal cones	cGMP PDE \uparrow	CTX, PTX
α_{gust}	taste cells	PDE ?, PLC- β ?	CTX, PTX
α_{i1}	mainly neuronal cells	AC \downarrow , Src, Rap1GAPs	PTX
α_{i2}	ubiquitous		PTX
α_{i3}	widely expressed		PTX
α_o	neuronal, neuroendocrine cells		PTX
α_z	neuronal cells, thrombocytes		
α_q subfamily			
α_q	ubiquitous	PLC- β \uparrow , LARG-RhoGEF	
α_{11}	widely expressed		
α_{14}	kidney, lung, spleen		
$\alpha_{15/16}$	hematopoietic cells		
α_{12} subfamily			
α_{12}	ubiquitous	RhoGEFs, E-cadherin	
α_{13}	ubiquitous		

AC, adenylyl cyclase; CTX, cholera toxin; GAP, GTPase-activating protein; GEF, guanine nucleotide exchange factor; PDE, phosphodiesterase; PLC, phospholipase C; PTX, pertussis toxin; RGS, regulator of G-protein signaling; VGCC, voltage-gated calcium channel

has also been presented for participation of the $G\alpha$ N-terminus as well as the $\beta\gamma$ dimer in receptor interaction. Considering the proximity of the $G\alpha$ N-terminus and $G\gamma$ C-terminus on a common face of the G-protein heterotrimer anchored to the plasma membrane *via* lipid modifications, cytoplasmic receptor domains have access to large surface areas of α , β and γ subunits. However, according to the crystal structures available, the distance between the posttranslationally modified $G\alpha$ N-terminus and $G\gamma$ C-terminus (estimated minimum of 40 Å) appears to be too large to interact with rhodopsin's cytoplasmic surface area (maximal distance between interacting loop sites approximately 35 Å) at the same time, thus implying a two-step sequential interaction mechanism. The structural basis underlying the selectivity of receptor-G-protein interaction still remains only partially defined.

G-Protein Cycle

Binding of an agonist to a heptahelical receptor entails the formation of a ternary complex consisting of agonist, receptor and heterotrimeric G-protein in its GDP-liganded form. The binding of $\beta\gamma$ subunits to $G\alpha$ stabilizes the flexible switch regions and hence the GDP-dependent inactive $G\alpha$ conformation. The activated receptor fulfills the role of a catalytically acting **▶guanine nucleotide exchange factor** that facilitates the release of GDP (Fig. 1). The short-lived guanine nucleotide-free G-protein heterotrimer stabilizes the receptor in its high-affinity conformation. Due to its high intracellular concentration, GTP is rapidly incorporated into the guanine nucleotide-binding pocket in $G\alpha$, resulting in a conformational change in the α subunit and a dissociation of GTP-bound $G\alpha$ and $G\beta\gamma$. Both reaction products are intracellular signaling



G-Proteins and G-Protein Mutations in Human Diseases. Figure 1 The G-protein cycle. The activated receptor functions as a guanine nucleotide exchange factor to release bound GDP. G-protein $\beta\gamma$ subunits stabilize the inactive, GDP-bound $G\alpha$ conformation. Pertussis toxin (PTX) modifies the C-terminus of some G-protein α subunits and uncouples these G-proteins from the receptor. Both GTP-bound α subunits and $\beta\gamma$ dimers are signaling proteins in their own right and interact with effector proteins. $G\alpha$ activation is terminated by hydrolysis of GTP. The endogenous GTPase activity is accelerated by effectors such as phospholipase C- β or RGS (for “regulator of G-protein signaling”) proteins. Cholera toxin (CTX) modifies a highly conserved arginine residue in some α subunits, thereby abolishing GTPase activity and rendering the G-protein α subunit constitutively active. For further details see text.

proteins in their own right, activating distinct and overlapping portfolios of cellular effectors. In contrast to the situation with monomeric GTPases of the Ras family, G-protein α subunits are endowed with an endogenous GTPase activity. A highly conserved arginine residue in the helical domain, Arg²⁰¹ in $G\alpha_s$, directly participates in GTP hydrolysis by stabilizing the negative charge on γ -phosphoryl oxygen atoms in the transition state. Ras proteins lack such a residue and are essentially inactive as GTPases. The conformation of GDP-bound $G\alpha$ allows for the reassociation with the $\beta\gamma$ dimer and the inactive $G\alpha_{GDP}\beta\gamma$ complex is now prone to another round of activation and deactivation. Two bacterial toxins interfere with the GTPase cycle by covalently modifying G-protein α subunits. ▶**Cholera toxin** (CTX) ADP-ribosylates the aforementioned conserved arginine residue in the GTPase domain of some G-protein α subunits (Table 1), thus blocking

GTP hydrolysis and rendering the α subunit constitutively active. ▶**Pertussis toxin** (PTX) adds the ADP-ribosyl moiety to a cysteine residue in the C-terminus of most G_i proteins, thereby effectively uncoupling modified G-proteins from the receptor. Both toxins have been extensively used to dissect G-protein-mediated signaling pathways.

In a physiological setting, the GTPase activity of α subunits is controlled by a diverse family of multifunctional signaling proteins, ▶**regulators of G-protein signaling** (RGS), which bind directly to activated $G\alpha$ subunits in order to accelerate the rate of GTP hydrolysis by several orders of magnitude. In addition, G-protein effectors such as phospholipase C- β can also potentiate GTPase activity of α subunits (Fig. 1). Apart from deactivation of G-protein α subunits and termination of downstream signals, RGS proteins may also be viewed as *bona fide* effectors setting in motion various intracellular signaling cascades by means of protein-protein interactions mediated by a host of defined structural signaling motifs. Furthermore, they may profoundly shape the dynamics and determine the efficacy of G-protein cycling. An approximately 120-amino acid region that directly interacts with GTP-bound α subunits, called the RGS domain, is the structural hallmark of more than 30 RGS family members. The majority of RGS proteins appear to target G_i and G_q family members, whereas some (p115-RhoGEF, PDZ-RhoGEF, LARG-RhoGEF) specifically interact either with $G\alpha_{12/13}$ proteins or with $G\alpha_s$ (RGS-PX1). The exact physiological functions of RGS proteins are still poorly understood.

Within the past few years alternative modes of signal input into G-protein cascades have been discovered which are independent of heptahelical receptors. Four proteins, called AGS1-4 (for ▶**activators of G-protein signaling**), have been identified that engage G-protein-dependent signaling pathways in the absence of a classical receptor. AGS1 is a distinct member of the large superfamily of Ras-related proteins and targets α subunits. AGS2 represents a $G\beta\gamma$ -binding component of the cytoplasmic motor protein dynein and has undefined roles in cellular signaling. The third protein, AGS3, possesses seven tetratricopeptide (TRP) repeats and four amino acid repeats termed GoLoco motifs (meaning $G\alpha_{i/o}$ -Loco interaction motif) also described as G-protein regulatory (GPR) motifs. The GoLoco motif, which can also be found in other signaling proteins like RGS12 and 14 functions as a selective $G\alpha$ binding partner. The GoLoco/ $G\alpha$ interaction releases $\beta\gamma$ subunits and at the same time inhibits guanine nucleotide exchange in the bound α subunit, thus “freezing” monomeric $G\alpha$ in the inactive, GDP-bound state. Recently, an additional AGS protein, AGS4, was identified which contains three GPR motifs and regulates the activation state of $G\alpha_i$. At present, the

cell physiological role of AGS proteins is still fairly obscure. Possibly, these proteins are involved in the regulation of basic cellular processes like the maintenance of cell polarity and cell division. By acting in concert with GPCRs they may provide for a signal amplification mechanism and at the same time allow signal transmission *via* G-proteins independent of heptahelical membrane receptors.

G-Protein Families and Effectors

Based on the primary amino acid sequence of their α subunits, G-proteins are subdivided into four distinct families: G_s , G_i , G_q , and G_{12} (Table 1) (6, 7). Concentrations of G_i proteins in the cell considerably exceed those of other families, and in brain G_o may amount to 1–2% of total membrane protein. Some G-protein α subunits are characterized by a very restricted expression pattern (Table 1), while others like G_s , G_{i2} , G_q , G_{12} , and G_{13} are ubiquitously expressed. G-proteins can also be classified on the basis of the cellular effectors to which they couple. G_s proteins classically stimulate adenylyl cyclase activity, while G_i proteins inhibit adenylyl cyclase *via* their α subunits, but activate inwardly rectifying potassium channels and inhibit P/Q-, N- and R-type voltage-gated calcium channels *via* $\beta\gamma$ subunits released upon GTP binding to $G\alpha_i$ proteins. G_q proteins activate phospholipase C- β isoforms, and G_{12} proteins couple to Rho guanine nucleotide exchange factors resulting in Rho activation and stress fiber formation (Table 1). Both GTP-loaded α subunits and $\beta\gamma$ dimers are *eo ipso* signaling proteins exerting their action through activation or inhibition of an ever expanding list of cellular effector proteins (for $G\alpha$ effectors see: Table 1). Effectors for G-protein $\beta\gamma$ subunits include inwardly rectifying potassium channels (Kir3.1–3.4), G-protein-coupled receptor kinases (GRKs), adenylyl cyclases (adenylyl cyclases II and IV), phospholipases C (PLC)- β_1 , β_2 and β_3 and phosphatidylinositol 3-kinases (PI3K) β and γ .

Considering that hundreds of GPCRs transduce signals by interacting with a limited number of G-proteins, the question of coupling specificity is worth considering. The concept of linear G-protein-mediated signal transduction pathways, i.e. one receptor coupling to one distinct G-protein activating one receptor, appears to be inadequate to describe physiology. G-protein-mediated signal transduction is a complex signaling network with diverging and converging transduction steps at each coupling interface. Deciphering the mechanism of signal specificity in living cells still remains a scientific challenge of paramount importance.

G-Protein Mutations as the Molecular Basis of Human Diseases

Because of their central role in controlling many physiological functions, naturally occurring activating

and inactivating mutations in GPCRs and G-proteins are responsible for an increasing number of human diseases. Functional variability resulting from polymorphisms may underlie interindividual differences in response to endogenous ligands as well as drugs. At present, the $G\alpha_s$ gene (*GNAS1*) is the only G-protein gene that has been unequivocally shown to be afflicted with activating or inactivating mutations that cause human diseases (8–10).

Activating Mutations in $G\alpha_i$

Mutations in the $G\alpha_{i2}$ gene were diagnosed in fixed sections of human ovarian sex chord stromal tumors and adrenal cortical tumors. In a few affected specimens, the highly conserved Arg¹⁷⁹ corresponding to the aforementioned Arg²⁰¹ in $G\alpha_s$ in the helical domain was found to be exchanged for a histidine (Arg¹⁷⁹H) giving rise to a $G\alpha_{i2}$ protein devoid of any GTPase-activity. Constitutively active $G\alpha_{i2}$ was subsequently referred to as the *gip2* oncogene. Most notably, this finding could not be confirmed by subsequent studies on fresh surgically resected tumors, and transgenic animals expressing the *gip2* oncogene in selected tissues have not been reported yet. Therefore, one has to conclude at this point that the oncogenic potential as well as the frequency of activating mutations of $G\alpha_{i2}$ appear to be rather low.

Activating Mutations in $G\alpha_s$

In many endocrine glands, cAMP stimulates proliferation, differentiation and hormone secretion. A first hint to the possible causative contribution of activating mutations in *GNAS1* arose from the identification of a subset of growth hormone (GH)-secreting pituitary tumors characterized by high intracellular cAMP concentrations and increased adenylyl cyclase activity. These adenomas accounting for approximately 40% of GH-secreting tumors were shown to harbor heterozygous missense mutations in *GNAS1* exons 8 or 9 giving rise to Arg²⁰¹Cys/His and Gln²²⁷Arg/Lys missense mutations, respectively. Both the highly conserved Arg and Gln residues are essential for GTP hydrolysis to occur in the α subunit with the requirement of Gln²²⁷ to orient and polarize the catalytic water in the transition state. Therefore, these missense mutations ablate the endogenous GTPase activity of $G\alpha_s$ and render the α subunit constitutively active, leading to uncontrolled, excessive cAMP production in somatotrophs. As cAMP stimulates proliferation and differentiation in these cells, the GTPase-deficient $G\alpha_s$ mutants have been designated *gsp* oncogenes.

Gsp mutations are also rarely observed in other pituitary tumors like corticotrophs, resulting in increased ACTH release (Table 2). Besides, approximately 10% of non-functional pituitary adenomas carry

G-Proteins and G-Protein Mutations in Human Diseases. Table 2 Diseases associated with *GNAS1* mutations

Type of mutation	Disease	Mode of inheritance
Gain of function	GH-secreting pituitary adenomas, thyroid adenomas and carcinomas, Leydig cell adenomas, pheochromocytoma, parathyroid adenoma, McCune-Albright syndrome, osseous fibrous dysplasia	somatic mutations
Loss of function	Albright hereditary osteodystrophy	germline mutations,
	PHP Ia	maternal transmission
	PPHP	paternal transmission
	progressive osseous heteroplasia (POH)	paternal transmission?
	PHP Ib ($G\alpha_s\text{-}\Delta I^{382}$)	maternal transmission
Loss and gain of function	PHP Ia and testotoxicosis ($G\alpha_s\text{-}A^{366}S$)	maternal transmission
<i>GNAS1</i> imprinting defect	PHP Ib	maternal transmission

GH, growth hormone; PHP, pseudohypoparathyroidism; PPHP, pseudopseudohypoparathyroidism

gsp mutations. Several studies have confirmed the presence of activating mutations in $G\alpha_s$ in up to 30% of toxic thyroid adenomas and in less than 10% of thyroid carcinomas. Sporadically, *gsp* mutations are found in parathyroid and adrenocortical tumors as well as in ►[pheochromocytomas](#).

The ►[McCune-Albright syndrome](#) (MAS) is classically defined by the clinical triad of *café-au-lait* hyperpigmented skin lesions, precocious puberty and polyostotic fibrous dysplasia of the bone. Apart from the gonads, other endocrine glands such as the pituitary, adrenal cortex and thyroid that are sensitive to trophic cAMP-dependent stimuli were also found to be hyperfunctional in MAS. Nodular and diffuse goiters as well as benign thyroid nodules are associated with MAS. The sporadic occurrence of thyroid cancer (papillary and clear cell thyroid carcinoma) in MAS patients suggests that additional mutational or epigenetic events in addition to gain-of-function $G\alpha_s$ mutations are mandatory for thyroid carcinogenesis in these patients. In 1986 the dermatologist Happle suggested MAS to be caused by a dominant somatic mutation as an early postzygotic event resulting the mosaic pattern of clinical stigmata. Mutations in *GNAS1* have been confirmed in affected endocrine tissues and in hyperpigmented skin lesions of all MAS patients. Interestingly, missense mutations were detected at only one position, i.e. Arg²⁰¹His/Cys. The overall clinical picture of an individual MAS patient is determined by the distribution of cells bearing the somatic *gsp* mutation. It is tempting to speculate that germline *gsp* mutations are incompatible with life. In contrast to the latter concept, one patient with severe

developmental, skeletal and endocrine abnormalities was described to harbor a germline Arg²⁰¹Leu mutation in *GNAS1*.

$G\alpha_s$ mutations have also been found in all cases of fibrous dysplasia (FD) of the bone. The majority of FD patients has only a single bone defect, a small group suffers from multiple bone lesions or has other features of MAS. Missense mutations in *GNAS1* identical to those in MAS patients, i.e. Arg²⁰¹His/Cys, were diagnosed in nearly all forms of FD. A possible explanation for the clinical phenotype relates to the general concept that elevated intracellular cAMP levels in osteogenic precursors entail increased proliferation and decreased differentiation of these cells resulting in benign fibrous bone lesions. Activating $G\alpha_s$ mutations have also been reported to occur in isolated intramuscular myxomas and those that present in conjunction with FD (Mazabraud syndrome)

Loss-of-Function Mutations in $G\alpha_s$

More than 60 years ago, Fuller Albright and his colleagues described several patients presenting with short stature, obesity, skeletal abnormalities, mental retardation and often subcutaneous ossification. This syndrome, now collectively called ►[Albright's hereditary osteodystrophy](#) (AHO) frequently concurs with resistance to parathyroid hormone (PTH) and other hormones like GH-releasing hormone, thyrotropin and gonadotropins acting *via* G_s -coupled receptors. AHO in conjunction with this kind of hormone resistance gives rise to the complex syndrome of ►[pseudohypoparathyroidism](#) (PHP) type Ia (Table 2). Patients with pseudopseudohypoparathyroidism (PPHP) show the

typical features of AHO, yet do not suffer from any kind of hormone resistance. On the contrary, PHP Ib patients present with symptoms of isolated PTH resistance, but lack the typical AHO phenotype. A mild form of thyrotropin resistance has recently been observed in PHP Ib patients raising the possibility that other endocrine systems may also be affected. Subsequent systematic studies were able to allocate the molecular defect in these three main forms of PHP to *GNAS1*.

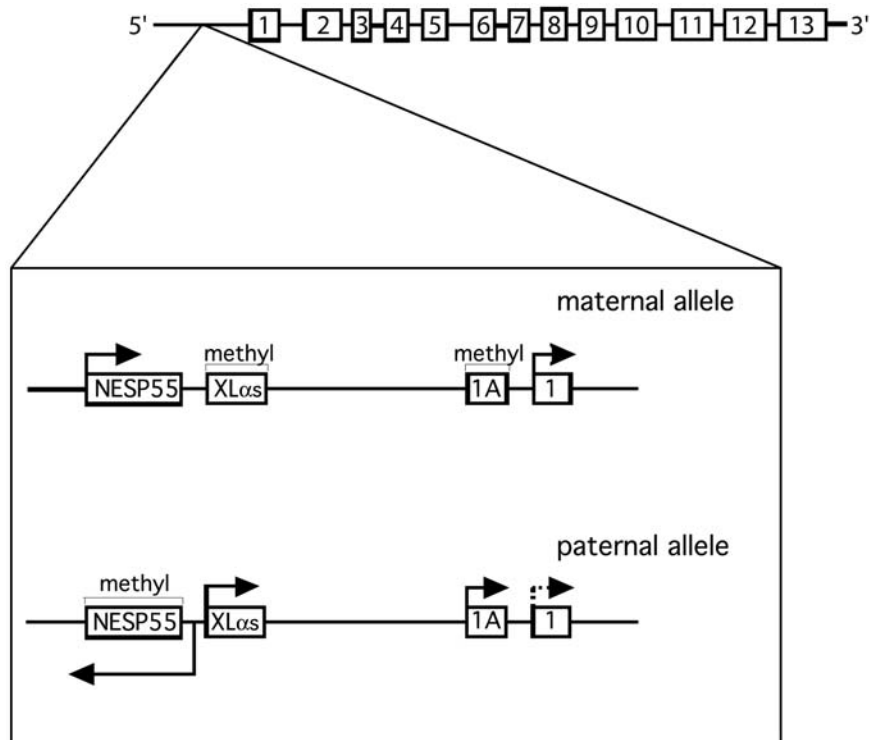
Heterozygous inactivating mutations affecting one of the $G\alpha_s$ -specific exons are the molecular cause of PHP Ia, PPHP and of progressive osseous heteroplasia (POH). POH patients suffer from severe heterotopic ossification involving skeletal muscle and deep connective tissue. They frequently lack hormone resistance and typical AHO features. Many of the *GNAS1* mutations are deletions or insertions that give rise to frameshifts and premature stop codons, nonsense mutations or splice junction mutations. In addition, a number of missense mutations adversely affect protein stability. The latter scenario is exemplified by a missense mutation, A³⁶⁶S, in the critical guanine nucleotide binding motif of the GTPase domain, leading to an accelerated release of GDP from the α subunit and marked instability of the guanine nucleotide-free protein at the body core temperature of 37 °C. At lower ambient temperatures, for instance in the testis, protein stability is not impaired and the accelerated nucleotide exchange manifests as constitutive $G\alpha_s$ activity. Therefore, the clinical phenotype of PHP Ia and excessive testicular testosterone production (testotoxicosis) arise from the intriguing A³⁶⁶S mutation (Table 2). In most tissues, an approximately 50% reduction of functional $G\alpha_s$ activity significantly reduces cAMP formation in the case of inactivating $G\alpha_s$ mutations. However, the cAMP levels that can still be generated are sufficient to maintain physiological functions. Thus, there is no evidence for haploinsufficiency to explain the hormone resistance observed in patients.

Retrospective analyses of PHP patients revealed that the clinical phenotype was strongly influenced by the parent transmitting the mutated allele. Any inactivating $G\alpha_s$ mutation leads to AHO irrespective of the parent transmitting the defective gene. Hormone resistance characteristic of PHP Ia occurs only if the genetic defect is inherited from a mother suffering from either PHP Ia or PPHP. Conversely, there is mounting evidence that POH is inherited from the father. This conspicuous parent-of-origin-specific inheritance pattern suggests that *GNAS1* is imprinted.

The human gene for $G\alpha_s$ is a single-copy gene located at 20q13.2-13.3. $G\alpha_s$ is encoded by exons 1–13 (Fig. 2). Alternative splicing of exon 3 produces two

long and two short forms of $G\alpha_s$. There is little evidence to suggest that these splice variants have distinct signaling properties. During the past few years it has become obvious that *GNAS1* not only codes for $G\alpha_s$, but also for several other transcripts by using 4 alternative promoters and first exons which splice onto a common exon 2 (Fig. 2). The most upstream alternative promoter gives rise to transcripts coding for the chromogranin-like neuroendocrine secretory protein 55 (NESP55) whose entire coding sequence resides in the upstream exon, thus leaving $G\alpha_s$ exons 2–13 within the 3' untranslated region of the NESP55 transcript. The NESP55 promoter is methylated on the paternal allele, so that the NESP55 gene is exclusively transcribed maternally. The XLa_s transcript encodes a protein with an extended N-terminus when compared to $G\alpha_s$ and is transcribed from the paternal allele only (Fig. 2). The C-terminal 348 amino acid residues are identical to $G\alpha_s$. XLa_s is highly expressed in the pituitary, is targeted to the plasma membrane, interacts with $\beta\gamma$ subunits and can be activated by non-hydrolysable GTP analogs. However, there is no evidence that XLa_s is regulated by GPCRs. An additional transcript derived from the sense strand of the paternal allele uses exon 1A (exon A/B) as the first exon and also splices onto exons 2–13 (Fig. 2). However, exon 1A generates transcripts that are presumably untranslated. Upstream of the XLa_s exon, a promoter for antisense transcripts traversing the NESP55 exon has been identified. These NESP55 antisense transcripts are only expressed from the paternal allele and may contribute to the imprinting of NESP55 by silencing the NESP55 promoter on the paternal allele.

Around 100 autosomal genes are subject to [genomic imprinting](#). One genomic region controlled by this epigenetic phenomenon is located in the distal portion of chromosome 2 and encompasses *GNAS1*. All imprinted genes have one or more regions in which the cytosines within CpG dinucleotide stretches are methylated on one parental allele only. Very often these methylated regions coincide with gene promoters. As described above and illustrated in Fig. 2, the promoter regions of *GNAS1* display a complex imprinting pattern. To complicate the scenario even further, the promoter region giving rise to $G\alpha_s$ transcripts does not exhibit allele-selective methylation and in most tissues expression occurs biallelically. This situation notwithstanding, paternal $G\alpha_s$ expression is silenced in a few tissues by a mechanism that is presently unknown. In proximal renal tubular cells, adipocytes, pituitary gland, thyroid and gonads, $G\alpha_s$ expression is largely driven by the maternal allele. In PHP Ia patients, renal proximal tubule cells are resistant to PTH action, because $G\alpha_s$ expression is restricted



G-Proteins and G-Protein Mutations in Human Diseases. Figure 2 Organization and imprinting of the *GNAS1* locus. *GNAS1* is characterized by 4 alternative first exons which splice onto exon 2. Methylation patterns (*methyl*) and transcriptional activation (*arrows*) of the maternal and paternal allele are indicated. The *hatched arrow* for exon 1 of the paternal allele indicates that it does not contribute to Gα_s expression in all cells. An antisense mRNA is transcribed across the NESP55 exon on the paternal allele.

to the maternal allele that carries an inactivating mutation. Yet PHP Ia patients are not prone to hypercalciuria, suggesting that the anticalciuric PTH action in the thick ascending limb is fully operative because of biallelic Gα_s expression in this part of the nephron. Thus, tissue- and cell-specific imprinting represents the molecular mechanism underlying the clinical features of PHP Ia, while haploinsufficiency alone may lead to AHO.

A first glance at the mechanism of tissue-specific Gα_s imprinting was granted by studies on patients with PHP Ib. The vast majority of these patients who present with renal PTH resistance, sometimes accompanied by partial TSH resistance, exhibit a loss of methylation at the *GNAS1* exon 1A, while lacking mutations in the exons coding for Gα_s. This loss of the maternal allele-specific methylation pattern, linked to an upstream 3 kb deletion, makes the maternal allele look like the paternal one, resulting in silencing of maternal Gα_s expression in renal proximal tubules. One possible explanation is based on the hypothesis that the non-methylated exon 1A region allows for the binding of a

tissue-specific repressor protein that hampers Gα_s expression. Alternatively, the deletion may disrupt a cis-acting imprinting control element necessary for the methylation imprint at exon 1A (exon A/B). As the described deletions disrupt another gene, *STX16* coding for Syntaxin 16, one may speculate that the *STX16* region comprises such an imprinting control element. The relevance of these epigenetic changes, i.e. the loss of maternal-specific methylation of *GNAS1*, for the clinical PHP Ib phenotype is emphasized by a patient with paternal uniparental disomy of chromosome 20q. In this situation both long arms of chromosome 20q are of paternal origin resulting in PTH resistance, but not in AHO.

A unique heterozygous 3 bp deletion causing loss of Ile³⁸² in the C-terminus of Gα_s was detected in 3 affected boys with PHP Ib. When heterologously expressed, the mutant Gα_s was found to be unable to couple to the PTH receptor, while interaction with the G_s-coupled thyrotropin and luteinizing hormone receptors was unaffected. These results explain PTH-specific hormone resistance in the affected patients.

The absence of any phenotype in the mother and maternal grandfather carrying the same mutation is commensurate with our current understanding of paternal imprinting of the *GNAS1* gene.

The $G\beta_3$ -C825T Polymorphism in Multigenic Disorders

A single base substitution (C825T) in the $G\beta_3$ subunit leading to a truncated protein has originally been reported in association with primary hypertension. More recently, genetic associations with a number of other disorders such as obesity and insulin resistance have been suggested. So far, the underlying mechanism by which the $G\beta_3$ variant causes the different phenotypes remains elusive.

Conclusions

In the past few years significant progress has been made towards a truly molecular understanding of receptor/G-protein-mediated signal transduction. One crystal structure of a heptahelical receptor, rhodopsin, and several of G-proteins provide a solid foundation for future work on the mechanisms of receptor and G-protein activation. An important goal will be to determine the structural differences between the inactive and active receptor conformations as well as the structure of receptors in complex with heterotrimeric G-proteins. Studies on engineered gene-deficient mice as well as the thorough *in vivo* and *in vitro* characterization of naturally occurring G-protein mutations detected in patients have taught us invaluable lessons on the physiology of these cardinal signaling proteins. Studying clinical and molecular aspects of the different forms of PHP has highlighted the complex regulation of $G\alpha_s$ expression and provided remarkable insights into the basic mechanisms of genomic imprinting. Our understanding of receptor and G-protein-mediated signaling processes has shifted from studying linear signaling cascades towards the consideration of complex signaling networks which will require novel collaborative research initiatives to integrate bits and pieces of knowledge into a coherent instructive model.

►Cardiac Signaling: Cellular, Molecular and Clinical Aspects

References

1. Gudermann T, Kalkbrenner F, Schultz G (1996) Diversity and selectivity of receptor-G-protein interaction. *Annu Rev Pharmacol Toxicol* 36:429–459
2. Schöneberg T, Schulz A, Gudermann T (2002) The structural basis of G-protein-coupled receptor function and dysfunction in human diseases. *Rev Physiol Biochem Pharmacol* 144:143–227
3. Pierce KL, Premont RT, Lefkowitz RJ (2002) Seven-transmembrane receptors. *Nat Rev Mol Cell Biol* 3:639–650
4. Palczewski K, Kumasaka T, Hori T et al (2000) Crystal structure of rhodopsin: A G-protein-coupled receptor. *Science* 289:739–745
5. Wall MA, Posner BA, Sprang SR (1998) Structural basis of activity and subunit recognition in G-protein heterotrimers. *Structure* 6:1169–1183
6. Offermanns S (2003) G-proteins as transducers in transmembrane signalling. *Progr Biophys Mol Biol* 83:101–130
7. Cabrera-Vera TM, Vanhauwe J, Thomas TO et al (2003) Insights into G-protein structure, function, and regulation. *Endocr Rev* 24:765–781
8. Spiegel AM, Weinstein LS (2004) Inherited diseases involving G-proteins and G-protein-coupled receptors. *Annu Rev Med* 55:27–39
9. Weinstein LS, Liu J, Sakamoto A et al (2004) *GNAS*: normal and abnormal functions. *Endocrinology* 145:5459–5464
10. Bastepe M, Juppner H (2005) *GNAS* locus and pseudohypoparathyroidism. *Horm Res* 63:65–74

GPx

►Glutathione Peroxidase

G-Quartet DNA

Definition

G-quartet DNA (also know as G4 DNA) defines a four-stranded DNA structure formed by nucleic acid rich guanine/cytosine regions. This structure is highly stabilized by a planar array of four hydrogen-bonded guanine bases.

►DNA Helicases

Graafian Follicles

Definition

Graafian follicles designate cellular components of the ovary, each consisting of a germ cell (oocyte), surrounded by somatic cells (follicle cells), and a large

fluid-filled cavity from which the unfertilized egg emerges.

► [Mammalian Fertilization](#)

Grade of Malignancy

Definition

Grade of malignancy designates the histomorphological assessment of the malignant behavior of a tumor, as estimated by cytological criteria such as nuclear pleomorphism and number of mitoses, and histological criteria such as the formation of differentiated structures. Usually, three grades (G1, well differentiated; G2, moderately differentiated; and G3, poorly differentiated; with increasing aggressiveness in this order) are distinguished.

► [Breast Cancer](#)

Granulation Tissue

Definition

Granulation tissue defines a new connective tissue that is formed during the wound repair process and temporarily replaces the lost dermal part of the skin. The name derives from the granular appearance of numerous new capillaries.

► [Wound Healing](#)

Granuloma

Definition

Granuloma represents a chronic inflammatory lesion initiated by various infectious and non-infectious agents. Granuloma consists of either small, nodular aggregations of mononuclear inflammatory cells or of aggregations of different cells, usually modified macrophages surrounded by lymphocytes and multinucleated giant cells. Sometimes granuloma may also contain eosinophils and B cells, and are surrounded by fibrotic tissue.

► [Morbus Wegener](#)

Granulomatosis

Definition

Granulomatosis refers to a multisystem disease that is characterized by an inflammation of the blood vessels (► [vasculitis](#)) involving the upper and lower respiratory tracts and variable degrees of systemic, small vessel vasculitis, which is generally considered to represent a hypersensitivity reaction to an unknown antigen.

► [Recombinant Protein Expression in Bacteria](#)

GRAS

Definition

Generally Recognised As Safe: the US Congress established this concept and regulatory policy in 1958 as part of its food safety legislation. Judged by qualified experts, it means that ingredients or hosts are safe when used in food or food production to accomplish their technical or nutritional purposes.

► [Recombinant Protein Expression in Yeast](#)

Grb2

Definition

Grb2 stands for Growth-factor-receptor-bound protein 2. It is an adaptor protein containing src homology domains, one of which binds to and translocates the guanine nucleotide exchange factors ► [SOS](#). It is involved in activation of Ras, but can also play a role in other signaling pathways in mammalian cells.

► [Ras Signalling Pathway](#)

► [Signal Transduction: Integrin-Mediated Pathways](#)

► [Tyrosine Kinase](#)

Green Fluorescent Protein

Definition

GFP stands for Green Fluorescent Protein. It is a natural, 27 kDa fluorescent protein, originally produced by the marine jellyfish *Aequorea victoria*, and fluoresces or

glows green visible light when excited by UV light (395 nm). GFP is commonly used in the laboratory for labeling, detecting and tracking proteins and biological processes. It can be cloned without co-factors in most organisms, and is used as a reporter molecule in light microscopic imaging in co-expression assays to visualize cellular structures and molecules. The protein has been mutated to generate blue, cyan, yellow, photo-activatable, and monomeric variants.

► *C. Elegans* as a Model Organism for Functional Genomics

► Electron Tomography

► FCS

► FRAP

► Functional Assays

► High-Throughput Approaches to the Analysis of Gene Function in Mammalian Cells

► Immunochemical Methods, Localization

► Large-Scale Homologous Recombination Approaches in Mice

► Medaka as a Model Organism for Functional Genomics

► Transgenic and Knockout Animals

Greig's Cephalopolysyndactyly

Definition

Greig's cephalopolysyndactyly refers to a syndrome that affects embryonic development of the limbs, skull and face. Major features include polysyndactyly of the hands and feet, broad thumbs, macrocephaly and a high prominent forehead. This disorder is due to mutation of the gene encoding the ►GLI3 zinc finger transcription factor which is involved in mediation of the ►Hedgehog Signal Transduction Pathway.

► Hedgehog Signalling

Growth Factor Receptors

Definition

Growth factor receptors are cell surface molecules that bind growth factors, and initiate an intracellular signal that affects gene transcription in cells.

► Growth Factors

► Receptor Serine/Threonine Kinase

► Tyrosine Kinases

Growth Factors

ULF HEDIN¹, JOY ROY²

¹Department of Surgical Sciences, Karolinska Hospital, and ²Department of Surgery, St Gorans Hospital, Stockholm, Sweden
ulf.hedin@kirurgi.ki.se

Definition

Polypeptide growth factors are proteins with a fundamental role during embryogenesis and regeneration of tissues. In contrast to some hormones, which regulate growth of entire organisms, growth factors are essential for replication of individual cells and to the maintenance of normal cell function. Some growth factors stimulate cell division in numerous different cell types, while others are specific for particular cells. Growth factors mediate their effects on cells by binding to specific surface receptors. Binding of growth factors to their corresponding receptors activates signaling systems inside cells, which regulate transcription of genes involved in cellular processes such as differentiation, proliferation, migration, protein synthesis and metabolism. In adult organisms, certain growth factors are essential for the regeneration of cells in the bone marrow and for tissue repair, whereas other factors regulate the development and growth of tissues in the embryo (see ►Limb Development). In human disease, growth factors contribute to the abnormal regulation of cell proliferation found in cancer but also regulate cellular processes in cardiovascular and inflammatory diseases. The progressive accumulation of new information regarding the function of growth factors in human health and disease is providing new alternatives for future treatment strategies.

Characteristics

In 1986, Rita Levi-Montalcini and Stanley Cohen were awarded the Nobel Prize in Physiology or Medicine for their discoveries of the first identified growth factors, nerve growth factor (►NGF) and epidermal growth factor (►EGF). NGF regulates the development and survival of neurons whereas EGF stimulates cell proliferation in a number of different cell types. Ever since, numerous other growth factors have been identified and their functions in human development, homeostasis and disease explored. These discoveries have to a large extent been promoted by cancer research, especially investigations into how viruses cause tumors. Some viruses that infect cells cause permanent mutations in the host cell DNA, which lead to the expression of genes that disturb the internal control of the cell's ability to proliferate. These so

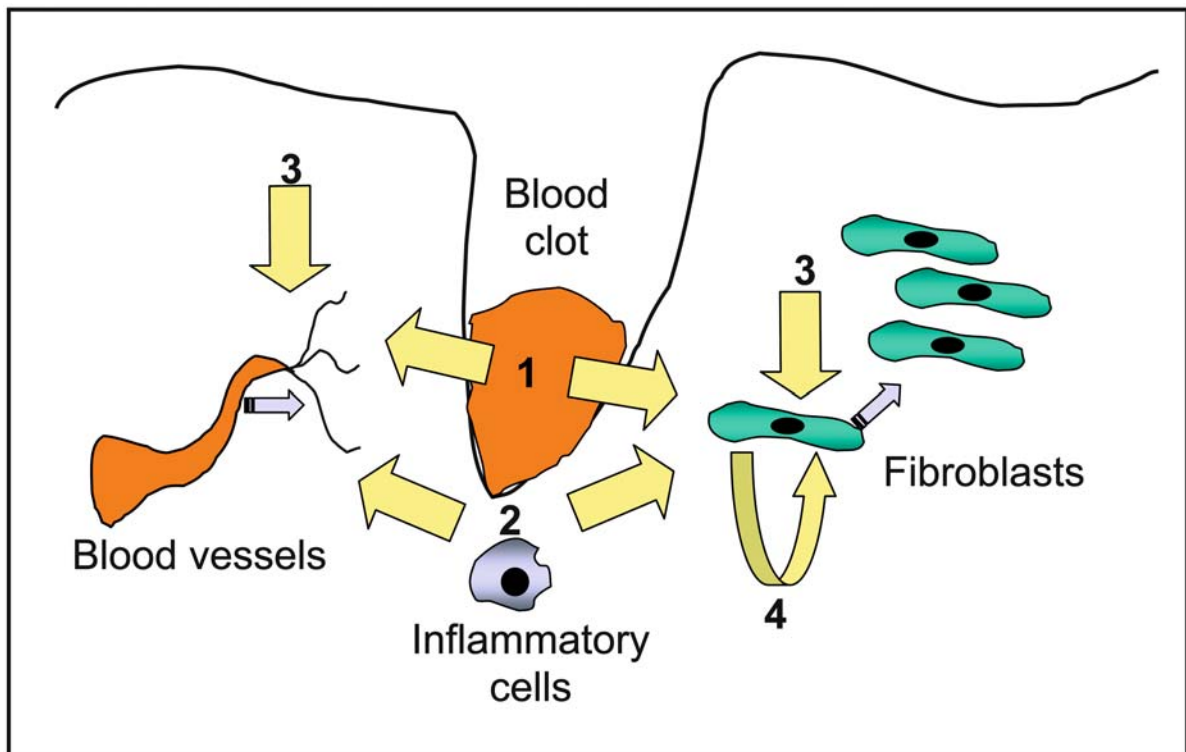
called ►**oncogenes** can promote abnormal cell proliferation, a hallmark of cancer cells, by changing the function of any protein involved in the normal control of cell replication. Thus oncogenes can lead to abnormal expression of a particular growth factor, altered expression of growth factor receptors, increased growth factor receptor activity or any other disruption of the intracellular machinery that regulates cell division. These insights have led to the discovery of individual growth factors, growth factor receptors and intracellular signaling pathways that transmit growth stimuli to the cell nucleus. These naturally expressed growth regulatory proteins are called proto-oncogenes and historically, they have been named after the virus that gave rise to the growth disturbing mutation. For example, the proto-oncogene that encodes for platelet-derived growth factor (►**PDGF**) was named *sis* after the virus simian sarcoma virus, which induces proliferation in infected cells through over-expression of a PDGF-like protein.

The polypeptide growth factors include a wide variety of signaling molecules that can be categorized into several groups or families. They can be produced and secreted by cells in order to act locally on neighboring cells (►**paracrine** function) or actually even on the same cells that produced them (►**autocrine** function). Some growth factors are also produced in organs but exert their action on target cells after being transported in the blood to a distance from the source (►**endocrine** function).

One of the first growth factors characterized, EGF, is found in salivary glands in the gastrointestinal tract and promotes proliferation of a large variety of cells, epithelial cells and mesenchymal cells included. ►**Erythropoietin** is a growth factor produced in the kidney, which stimulates proliferation of immature red blood cells in the bone marrow. Some growth factors are stored extracellularly in tissues or in cells and can be released to stimulate cells in the immediate vicinity. For example, fibroblast growth factor (►**FGF**), a member of a large family of growth factors, has the capacity to be stored in tissues by binding to sugar residues on proteins in the extracellular matrix, the so-called proteoglycans, and can be released after injury to participate in tissue repair by stimulating cell proliferation. Several members of the FGF family stimulate proliferation of endothelial cells and participate in the formation of new vessels, angiogenesis. FGF is also an important growth factor in the developing embryo and mutations in receptors for these growth factors have been associated with several different bone disorders, for example achondroplasia (dwarfism). There are numerous other growth factors that are involved in angiogenesis. Vascular endothelial cell growth factor (►**VEGF**) can be produced in tissues with a deficient blood supply, for example after myocardial infarction,

and selectively stimulate proliferation of endothelial cells to recruit new vessels into the anoxic tissue. VEGF can also be released from cancer cells and stimulate growth of vessels from the surrounding tissues, which ensures a blood supply in the growing tumor. PDGF is stored in blood platelets and can act on cells after platelet degranulation in association with tissue injury, bleeding and blood clotting. Together with other growth factors such as EGF, which is also stored in platelets, PDGF acts locally on cells in the injured tissue and promotes cell proliferation in the healing process (Fig. 1). Insulin-like growth factor-1 (►**IGF I**) chemically resembles the hormone insulin and is mainly produced in the liver under the control of growth hormone. During development, this growth factor participates in the regulation of skeletal growth and maturation after birth, but it is also involved in tissue repair and may be important for abnormal proliferation of cancer cells. The transforming growth factor family (►**TGF**) is a large family of different growth factors that were initially characterized by their ability to transform normal cells into tumor cells in culture. They have profound effects on cell metabolism and cellular synthesis of extracellular matrix proteins, and in some cell types they rather prevent than stimulate cell proliferation. ►**Cytokines** are a unique family of growth factors, which primarily act on cells in the immune system and stimulate proliferation of lymphoid cells. Cytokines such as the interleukins (►**IL**), a large family with more than 20 members, regulate proliferation on a variety of lymphocytes but also affect differentiation and growth of cells in the bone marrow.

The identification and characterization of growth factors have yielded potential therapeutic tools for the management of a large variety of human diseases. In patients with chronic kidney failure, the metabolic dysfunction of the organ will lead to anemia due to insufficient production of erythropoietin. Today, patients suffering from this condition receive injections of recombinant erythropoietin that stimulate production of red blood cells in the bone marrow. A number of other conditions where the bone marrow does not produce enough blood cells can also be corrected through the addition of specific growth factors. For example, deficient production of white blood cells in the bone marrow, a side effect of cancer treatment with chemotherapy, can now be corrected through injections of growth factors that specifically stimulate proliferation of leukocytes. In some disorders, especially in cancer, pharmacological approaches are taken to develop drugs that prevent the effects of growth factors, for example drugs that interfere with the binding between specific growth factors and their corresponding receptors on the surface of cells.



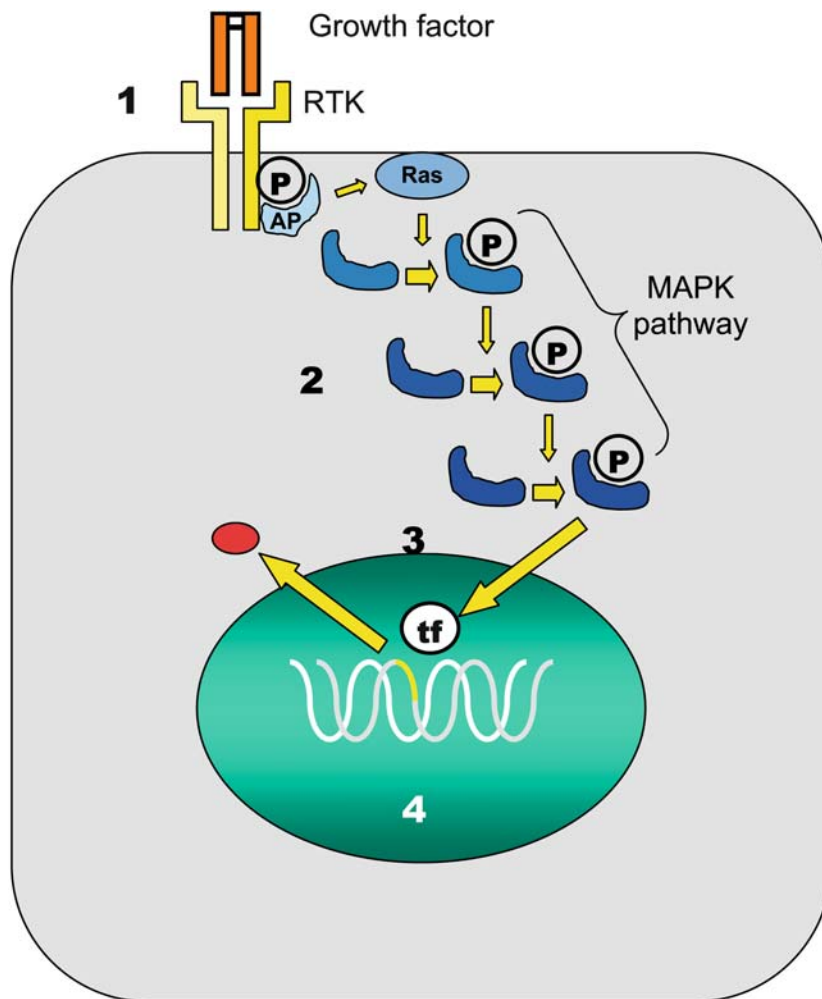
Growth Factors. Figure 1 Growth factors regulate healing responses. After tissue injury, bleeding and formation of a blood clot, growth factors can be released from degranulating platelets (1) and stimulate fibroblast proliferation or growth of new blood vessels in the injured tissue. In addition, inflammatory cells recruited from the blood stream can produce growth factors (2); they may be released from storage pools in the tissue (3) or they may be released from proliferating cells (4).

Regulatory Mechanisms

Growth factors are secreted into the extracellular environment of cells where they may come in contact with specific receptors on the same cell that secreted them, on nearby cells or on distant cells after transportation through the blood stream. Polypeptide growth factors usually bind with high affinity to their corresponding receptors. The receptors are transmembrane proteins with an extracellular domain that has a specific binding site for the growth factor. Upon binding of the ligand, a signal is sent to the intracellular domain, which initiates an intracellular signal transmitted from the receptor to the cell nucleus through specific signaling pathways. Once the signal has reached the nucleus, a specific response is elicited in the regulation of gene transcription and the cell will express specific genes, for example genes involved in the control of proliferation.

Growth factor receptors can be classified into three different groups, receptor tyrosine kinases (▶RTK), ▶tyrosine-kinase-associated receptors, and G-protein coupled receptors (▶GPCR). Most growth factor receptors are RTKs. These growth factor receptors have a

single transmembrane domain, an extracellular domain that binds to the ligand and a cytosolic domain, which harbors an enzymatic activity that catalyzes addition of phosphate residues to the amino acid tyrosine, ▶tyrosine kinase activity. Many growth factors, such as PDGF, are dimers, two separate polypeptides linked to each other, which bind to two separate receptor subunits at the cell surface. This dimerization of the receptor subunits facilitates tyrosine kinase activity in the cytosolic domains whereby tyrosine residues are phosphorylated, a phenomenon termed ▶autophosphorylation. Other monomeric growth factors, such as FGF, may instead form pairs by binding to sugar structures at the cell surface and thereby achieve simultaneous binding of two corresponding cell surface receptors and receptor dimerization. Autophosphorylation of tyrosine residues in the cytosolic domains of RTKs allows binding and initiation of catalytic activity in intracellular proteins that transmit signals to the nucleus through different pathways. These signaling pathways generally function by a series of kinases that step by step are phosphorylated and activated one after the other so that the signal is propagated to the final activation of gene expression in the nucleus (Fig. 2).



Growth Factors. Figure 2 Growth factors regulate gene transcription in cells. 1) Binding of a growth factor to a corresponding receptor (receptor tyrosine kinase; RTK) on the cell surface induces dimerization of two receptor subunits, which elicits phosphorylation (P) of tyrosine residues on the cytosolic domains of the receptor. 2) Specific intracellular adaptor proteins (AP) bind to the receptor and facilitate activation of catalytic proteins, such as Ras, which thereafter activate intracellular signaling through the mitogen activated protein kinase (MAPK) pathway. 3) In this pathway, phosphorylation of amino acid residues activates a series of kinases and the last activated protein in the pathway enters the cell nucleus (4) where it induces activation of transcription factors (tf), which promote expression of specific genes, for example genes coding for proteins necessary for cell proliferation.

Cytokines bind to tyrosine-kinase-associated receptors that are structurally similar to RTKs but lack intrinsic tyrosine kinase activity in their cytosolic domains. Instead, these growth factor receptors are associated with molecules that have tyrosine kinase activity. Receptor binding of cytokines also facilitates dimerization of receptor subunits, which then mediate phosphorylation of associated tyrosine kinases. Activation of these tyrosine kinases leads to phosphorylation of

tyrosine residues on the receptor, which provides binding sites for downstream signaling molecules and initiation of signaling cascades. The receptors of the GPCR family are traditionally regarded as mediators of signaling for substances that regulate specific physiological responses in differentiated cells, for example contraction and relaxation in muscle cells. Lately, it has been understood that these receptors may also transmit growth signals in less differentiated and proliferative

cells after stimulation by factors not normally perceived as growth factors, for example angiotensin II which stimulates contraction of smooth muscle cells in the vessel wall but also initiates proliferation in de-differentiated smooth muscle cells in diseased vessels. GPCRs have an extracellular domain that binds the ligand, seven transmembrane segments and an intracellular domain that associates with a guanine nucleotide-binding protein (hence the name G-protein). Ligand binding, for example of thrombin or angiotensin, to its specific GPCR extracellular domain, leads to a conformational change in the cytosolic domain of the receptor. The altered receptor structure facilitates binding of G-proteins to the intracellular domain, which in turn activates an enzyme attached to the plasma membrane. This enzyme catalyzes a reaction leading to the release of a second messenger, which can then activate intracellular signaling molecules that reach the nucleus and affects gene transcription.

Cascades of intracellular signaling molecules constitute links between growth factor binding to a growth factor receptor and expression of genes in the nucleus that control cell proliferation. The ►**mitogen activated protein kinase** (MAPK) pathways are the most well studied and understood intracellular signaling pathways, which are activated after binding of growth factors to their corresponding cell surface receptors. For example, binding of PDGF to the PDGF receptor leads directly to receptor dimerization and autophosphorylation of tyrosine residues on the cytosolic domains of the receptor. Tyrosine autophosphorylation provides binding sites for an adaptor protein called Grb2. Grb2 in turn binds to another adaptor protein SoS (son of sevenless) that can activate the small GTP binding proto-oncogene protein Ras. Ras then phosphorylates a Map-kinase-kinase-kinase, which in turn phosphorylates a MAP-kinase-kinase and then finally, the last signaling molecule of the pathway, a MAP-kinase (MAPK) is activated. After phosphorylation, MAPK translocates into the nucleus and activates a set of ►**transcription factors**, which promote expression of growth-related genes (Fig. 2).

The ►**cell cycle** is divided into four phases, G0, G1, S, G2 and M. Non-malignant eukaryotic cells are normally resting in the G0 phase. In cells that harbor the capacity to proliferate, such as fibroblasts, growth factors stimulate the cells to leave the G0 phase and enter the G1 phase. When cells have been stimulated with growth factors for a specific time period in the G1 phase, further progression in the cell-cycle and cell division will inevitably follow, even if the growth factor is removed. This so called ►**restriction point** is followed by the S-phase where DNA replication takes place, the G2 phase when factors necessary for the physical division of the cell are produced and finally the M phase when mitosis occurs. Passage through the cell-cycle is controlled by periodic

changes in the expression and activity of families of cell-cycle regulatory proteins termed ►**cyclins** and cyclin dependent kinases (►**cdk**). The expression of some cyclins is largely dependent on growth factors and activation of growth factor receptors. After growth factor stimulation, cyclins are synthesized, which form complexes with cdks. These complexes catalyze massive phosphorylation of the retinoblastoma protein, ►**Rb**. In resting cells, Rb is bound to the transcription factor E2F but upon hyperphosphorylation, E2F is released and can activate transcription of a number of genes necessary for S-phase initiation and further cell-cycle progression.

References

1. Oxford Reference Online. Oxford University Press. BIBSAM, 2004. ►<http://www.oxfordreference.com>
2. Bast et al (2000) Cancer Medicine, 5th edn. BC Decker Inc, Hamilton
3. Lodish et al (2003) Molecular cell biology, 5th edn. WH Freeman and Co, New York
4. Alberts et al (1994) Molecular Biology of the Cell, 3rd edn. Garland Publishing, New York and London
5. Heldin and Purton (1996) Signal transduction, 1st edn. Chapman & Hall, London

Growth Hormone Deficiency

Definition

Growth hormone deficiency is caused by absence, decreased production or dysfunction of the anterior pituitary hormone, which results in dwarfism or short stature and possibly some metabolic abnormalities (such as hypoglycemia).

►**Hypothalamic and Pituitary Diseases Genetics**

Growth Plate

Definition

The growth plate is a cartilaginous structure at the end of bones that generates the entire longitudinal growth through proliferation and differentiation of chondrocytes, and the conversion of cartilage into bone.

►**Bone Disease and Skeletal Disorders, Genetics**

Gamma-Secretase (Complex)

► γ -Secretase

GSK3

► Glycogen Synthase Kinase-3

GST

Definition

GST stands for Glutathione-S-transferase, a 26.9 kDa protein-fragment that is used to tag recombinant proteins.

► Recombinant Protein Expression in Bacteria

GST Pull-Down Experiment

Definition

In this kind of assay, a recombinant affinity tag ► [fusion protein](#) is used as bait to capture ('pull down') binding partners out of a cell lysate. The cell lysate is applied to the immobilised bait protein, or, alternatively, bait protein and lysate are mixed in solution and complexes are captured by affinity chromatography afterwards. Glutathion-S-transferase (GST) is often used as the affinity tag.

► [Affinity Chromatography and *In Vitro* Binding \(Beads\)](#)

GTP

Definition

GTP stands for Guanosine 5'-triphosphate. It is produced by phosphorylation of GDP (guanosine 5'-diphosphate).

► [Rho, Rac, Cdc42](#)

GTPase-Activating Proteins

Definition

GTPase-activating proteins (GAPs) comprise of proteins that bind to a GTP-binding protein and inactivate it by stimulating its GTPase activity so that it hydrolyzes its bound GTP to GDP.

► [Rho, Rac, Cdc42](#)

GTPases

Definition

GTPases comprise of a large group of intracellular signaling proteins characterized by an active state when GTP-bound, and an inactive state when GDP-bound. Their enzymatic activity hydrolyzes GTP.

► [Cardiac Signaling: Cellular, Molecular and Clinical Aspects](#)

► [Diabetes Insipidus, a Water Homeostasis Disease](#)

► [Rho, Rac, Cdc42](#)

Guanine Nucleotide Dissociation Inhibitors

Definition

Guanine nucleotide dissociation inhibitors (GDIs) are protein factors that inhibit the dissociation of guanine nucleotides (GDP/GTP) from GTPbinding proteins.

► [Rho, Rac, Cdc42](#)

Guanine Nucleotide Exchange Factors

Definition

Guanine nucleotide exchange factors (GEFs) are proteins that catalyze the release of guanine nucleotides (mostly GDP) from monomeric or heterotrimeric GTPases, thereby allowing them to bind GTP in its place. In the latter case, heptahelical receptors serve as GEFs.

► [G-Proteins](#)

► [Rho, Rac, Cdc42](#)

► [Tight Junctions](#)

Gut Epithelium

MICHÈLE KEDINGER, JEAN-NOËL FREUND
INSERM U682, Strasbourg, France
michele.kedinger@inserm.u-strasbg.fr

Definition

The intestinal epithelium lines the gut lumen, dividing the outside from the interior of the body. It is implicated in most digestive and absorptive functions of the gut, but also in the immunological and non-immunological protection against nutritional, bacterial, viral and parasite aggressions. Its importance in terms of evolution is demonstrated by its presence early in the animal kingdom long before the spinal cord or neuronal cells. The intestinal epithelium forms invaginations into the stroma, the crypts, and outwardly projecting villi in the small intestine or flat cuffs in the colon. Taking into account the length of the gut (5–6 m in humans), these structures, and additional microscopic foldings at the apical membrane of epithelial cells, the microvilli, provide an overall absorptive surface of about 300 m². The gut epithelium exhibits a morphological and functional regionalization along the proximo-distal axis, mainly obvious between the small intestine and colon.

Characteristics

1. The gut epithelium develops in interaction with multiple partners.
The intestinal epithelium derives from the mid and posterior parts of the definitive endoderm (midgut and hindgut), which itself originates from the ingression of primitive streak cells at gastrulation. Specification of the presumptive intestinal endoderm is driven by interactions with lateral mesoderm. Morphogenesis of the gut generates the single layer columnar epithelium laid on and separated from lamina propria myofibroblasts by a ►**basement membrane**, which is a specialization of the ►**extracellular matrix**. Intestinal epithelial cells also interact with lymphocytes (either intra-epithelial lymphocytes or organized into sub-jacent nodules), and with neuronal cell processes. On the luminal side, nutrients and the intestinal microflora are able to signal and modulate the epithelial cell behavior.
2. The gut epithelium is highly polarized.
Intestinal epithelial cells exhibit an asymmetrical morphology shaped by a highly organized network of microtubules, intermediate filaments and actin filaments connected to the plasma membrane. Apical-lateral ►**junctional complexes** (tight and

adherent junctions) link the cells together, regulate the barrier function, and delineate two compartments of the plasma membrane, the apical and the basolateral membranes. These complexes are composed of at least 40 different proteins, among which are transmembrane proteins such as occludin, claudins and cadherins, and cytoplasmic proteins, among which are some which can shuttle to the nucleus like the transcription factor ZONAB associated to ZO-1 and β -catenin which connects E-cadherin to actin filaments. The apical membrane, called the brush border, is formed by microvilli sustained by an actin core associated to a number of cytoskeletal proteins among which is villin, which regulates the plasticity of the brush border. The functional proteins (digestive enzymes, outside-in transporters) are anchored into the brush border membrane. The basolateral membrane carries inside-out transporters and ion channels involved in the processing of nutritional metabolites, but also receptors to basement membrane molecules. These receptors, including ►**integrins** and ►**dystroglycan**, ensure the link with the basement membrane and also act synergistically with growth factor receptors to initiate intracellular signaling and to participate in the regulation of cell functions.

3. The gut epithelium is composed of several functionally different cell types.
The main cell type in the small intestine (about 90%) is a columnar absorptive cell responsible for the digestive functions, the enterocytes. Additional cytotypes belong to the secretory cell lineages and comprise (i) mucus-producing goblet cells, which produce the protective mucus layer on top of the epithelium, (ii) entero-endocrine cells composed of at least 15 different cell subtypes classified on the basis of their hormonal content, which control important physiological functions such as glycemia, exocrine pancreatic secretion and growth/repair of the gut epithelium and (iii) Paneth cells, which are involved in the mucosal defense function and produce a variety of antibiotic and antimicrobial factors. The last cytotype is composed of M cells, which correspond to a specialization of absorptive cells facing lymphoid nodules and participate in the control of intestinal immunity and tolerance. Mucous cells represent the major cell type in the colon.
4. The gut epithelium is continually and rapidly renewed throughout life.
The adult intestinal epithelium is compartmentalized. Stem cells confined near the crypt bottom generate proliferative transit amplifying cells, which cycle every 12 h, and migrate vertically towards the crypt mouth. When leaving the crypts, epithelial cells abruptly stop proliferating and become differentiated. All the differentiated cell types – except the

Paneth cells, which migrate downwards to the crypt bottom – migrate to the villus tip in the small intestine or to the surface of the colon cuffs, and subsequently undergo apoptosis and exfoliation into the gut lumen. The overall life cycle of the gut epithelial cells is around 5 days in humans.

5. Multipotent stem cells are maintained in the adult gut epithelium.

The continuous renewal of the gut epithelium implies the presence of stem cells located in a specific niche near the crypt bottom surrounded by sub-epithelial myofibroblasts and specific extracellular molecules. Although intestinal stem cells have still not been isolated, several experimental approaches – transgenesis, mouse embryo aggregation chimeras, mutagenesis, regeneration after X-ray irradiation – have provided evidence that all the differentiated cell types of the gut epithelium derive from ►multipotent stem cells (1, 2). Stem cells undergo asymmetrical division to stochastically produce one new self-maintaining stem cell and one daughter cell that cycles and fuels a population of potential clonogenic stem cells. These cells are further displaced into the compartment of transit amplifying cells that eventually go into the differentiation process. Potential clonogenic stem cells, unlike transit amplifying cells, can replace true stem cells if this population is altered. The process towards differentiation uses two minor pathways consisting of long-lived progenitors of absorptive and of mucous cells respectively and one major pathway in which short-lived progenitors can produce both absorptive and mucous short-lived progenitors. Interestingly, glucagon-like peptide-2 produced by one of the enteroendocrine cell types – known to prevent intestinal damage and to facilitate repair – signals through enteric neurons to produce a mediator that stimulates the production of long-lived progenitors of the absorptive lineage (2).

6. Malignant epithelial tumors develop in the colon and rectum.

Colorectal cancer is a major disease in terms of incidence and malignancy. It results from imbalanced cell proliferation, differentiation, migration and apoptosis in colonic crypts. The vast majority of tumors are of sporadic origin while a small proportion is familial. The major familial form, ►Hereditary Non-Polyposis Colon Cancer (HNPCC), is characterized by microsatellite DNA instability resulting from germ line mutations in the MLH1 or MSH6 genes, which cause defects in the DNA mismatch repair system. The second familial form, ►Familial Adenomatous Polyposis (FAP), is characterized by chromosomal instability and is linked to germ line mutations in the tumor suppressor gene ►APC. Loss of function of

APC leads to inappropriate activation of the Wnt/ β -catenin signaling pathway (3) and to chromosomal instability linked to altered kinetochores. Rare cases of FAP without germ line mutation of APC are associated to biallelic germ line mutations in the base excision repair gene MYH. The sporadic form of colorectal cancer is mainly related to chromosomal instability and to gradual histological changes, the adenoma-carcinoma sequence, associated to the accumulation of somatic alterations in a number of tumor suppressor genes (APC, p53) and oncogenes (K-ras, Bcl2). Malignant tumors occur almost exclusively in the colon and rectum, while atypical tumors develop in the small intestine in the context of chronic inflammatory bowel disease. The primary incidence of tumors in the colon correlates with the colon specific expression of the anti-apoptotic gene Bcl2 at the stem cell positions in the crypt base.

Regulatory Mechanisms

1. Intestinal epithelial identity and homeostasis: involvement of the ►Cdx genes.

►Homeobox genes belong to a large family of transcription factors acting at multiple levels during embryonic development. *Cdx1* and *Cdx2* are two paralogue homeobox genes, which are expressed in the presumptive gut endoderm and in posterior organs in the embryos, and specifically in the gut epithelium throughout adult life. *Cdx2* displays the homeotic function devoted to defining the intestinal identity (4). Indeed, ectopic expression of *Cdx2* in the stomach epithelium – normally devoid of *Cdx* gene expression – converts gastric mucosa to an intestinal phenotype, whereas loss of expression in the gut endoderm leads to a gastric transdifferentiation. *Cdx1* and *Cdx2* are also modulators of cell renewal *via* distinct and complementary functions; *Cdx1* stimulates cell proliferation, resistance to apoptosis and eventually cell differentiation, whereas *Cdx2* reduces cell proliferation and stimulates differentiation. The *Cdx1* and/or *Cdx2* proteins act directly on a panel of target genes and cellular functions including regulators of cell cycle and apoptosis (p21WAF, Bcl2), transcription factors (KLF4), proteins involved in cell interactions (L1-cadherin), in calcium metabolism (vitamin D receptor, calbindin-D9k) and in glucose metabolism (glucagon, glucose-6-phosphatase), digestive enzymes (sucrase, lactase) and mucus production (Muc2). Interestingly, the expression profiles of *Cdx1* and *Cdx2* are altered in colorectal cancers and pro-oncogenic pathways have opposite effects on the two homeobox genes. For instance *Cdx1* is upregulated by the Wnt/ β -catenin pathway,

whereas *Cdx2* is down-regulated by the ►PI3K/Akt pathway. These observations suggest that *Cdx1* and *Cdx2* dysfunctions contribute to cancer progression. In accordance with this, the alteration of the *Cdx2* status sensitizes the colon epithelium to carcinogenesis, linked to a lower capacity to switch on apoptosis. Thus, in addition to its homeotic function during embryonic development of the digestive tract, the *Cdx2* homeobox gene is a gut-specific tumor suppressor gene in the adult colon.

2. The Wnt/ β -catenin signaling pathway.

Cell growth and polarity are major attributes of the gut epithelium, and the Wnt/ β -catenin pathway plays a pivotal role in this balance. β -catenin contributes to cell polarity by cross-linking the membrane E-cadherin to the actin cytoskeleton. In differentiated epithelial cells, an excess of β -catenin molecules not coupled to E-cadherin are loaded on a complex comprising APC, Axin and the CKI and GSK3 β kinases that phosphorylate β -catenin and target it to the proteasome degradation system. During intestinal development as well as in the crypts, the Wnt/ β -catenin signaling pathway is activated by secreted morphogens of the Wnt family that bind to Frizzled receptors and activate several downstream pathways, one of which leads to the inhibition of GSK3 β activity. In this context, β -catenin escapes degradation and translocates into the nucleus to bind HMG-box transcription factors of the Tcf/Lef family. These factors play a major role in the maintenance and self-renewal of the stem cell stock, since Tcf4-deficient mice lack proliferative cells in the prospective intervillous regions; in contrast over-expression of Lef1 causes increased stem cell apoptosis. Major targets of the activated Wnt/ β -catenin pathway are the proto-oncogene *c-myc* and cyclin D₁, which subsequently down-regulates the cell cycle inhibitor p21WAF (5). Interestingly, a transcriptomic approach has identified a set of genes of the *c-myc* cascade in the stem cells/progenitors compartment, including regulators of *c-myc* gene transcription and protein stability and *c-myc* downstream targets (6). As mentioned above, a direct Wnt/ β -catenin tissue-specific target is *Cdx1*, and an indirect target is *Cdx2*, through another HMG-box factor, SOX9. Finally, the Wnt/ β -catenin pathway also controls crypt cell sorting *via* the regulation of combinations of ephrin receptors EphB2/EphB3 and their ligands EphrinB1/EphrinB2 (5). Thus, crypt formation, cell sorting and the mechanism of stem cell selection appear to depend on an adequate threshold of β -catenin-mediated signaling during normal intestinal homeostasis, whereas inappropriate activation of this pathway contributes to colorectal tumorigenesis. Cooperating with the WNT/ β -catenin pathway,

BMP (Bone Morphogenetic Protein) signaling and Hh (Hedgehog) signaling also control the proliferation/differentiation equilibrium of the gut epithelium (7, 8, 9).

3. Molecular determinants of gut stem cells and progenitors. Although gut stem cells and progenitors have been described near the crypt base, molecular markers remain elusive. Yet, a gene product, *Msi1*, has been reported in intestinal crypts, specifically in individual cells that display the theoretical location attributed to stem cells (1). *Msi1* is an RNA-binding protein associated with asymmetrical division of neural progenitors. *Msi1* controls the translation of several RNAs; in particular, it acts as a translational repressor of *numb*, an antagonist of Notch signal activation. The Notch pathway is involved in the maintenance of an undifferentiated state by a lateral inhibition mechanism by which a cell differentiating along a given pathway produces a signal that prevents neighboring cells from differentiation along the same pathway. Upon activation by their ligands, the intracellular domains of Notch receptors are released and translocated into the nucleus where they bind to CSL (CBF1/Suppressor of hairless/Lag1) DNA binding proteins. This results in the transcriptional activation of downstream targets encoding ►bHLH transcription factors of the Hes family, among which is *Hes1*. These are negative regulators of differentiation, by repressing other bHLH genes that promote differentiation. Indeed, a series of gene invalidations led to the conclusion that cell fate commitment in the intestine depends essentially on a genetic cascade controlled by bHLH transcription factors. Firstly, *Hes1*, which is expressed in crypt cells like *Math1* and *Ngn3*, maintains the precursor pool expansion and prevents premature endocrine and mucous cell differentiation. Secondly, *Hes1* antagonizes *Math1*, which is required for the commitment of the three secretory cell lineages of the gut epithelium (goblet, endocrine and Paneth cells). This suggests that the choice between absorptive and secretory cell lineages is balanced by *Hes1* and *Math1* (2). Thirdly, within the population of *Math1* expressing-cells, *Ngn3* specifies the endocrine progenitors (10), whereas *NeuroD*, which is expressed in the villus cells, is required for the differentiation of a subset of endocrine cells. Genetic modulations of the Notch pathway have confirmed its influence on the maintenance of undifferentiated, proliferative cells in the crypts and on intestinal cell lineage specification by acting on the balance between bHLH factors (11).

In conclusion, the renewal of the gut epithelium has long been recognized as a paradigm in cell biology as

regards well-ordered cell proliferation followed by differentiation from self-renewing stem cells. Recent investigations have provided new insights into the molecular mechanisms implicated in multiple aspects of this process, such as the determination of intestinal identity, the cell commitment into differentiated lineages. Emerging concepts propose integrative models involving the interplay of local and reciprocal stimulatory and inhibitory signals between epithelial cells and the underlying myofibroblasts, to fine-tune the homeostatic balance between stemness, commitment, proliferation and differentiation (7, 9). Furthermore, these results enlarge our knowledge of the molecular alterations at the basis of malignant transformation in ►colorectal cancers. However, the exact nature of the gut stem cells and their relationship and regulation by neighboring elements of the stem cell niche remain to be elucidated. Understanding the biology of the gut stem/progenitor cells is a challenge for the future that should open new avenues in the field of cancer therapy, intestine regeneration and cellular therapy of type-1 diabetes.

References

1. Booth C, Potten CS (2000) Gut instincts: thoughts on intestinal epithelial stem cells. *J Clin Invest* 105:1493–1499
2. Brittan M, Wright NA (2002) Gastrointestinal stem cells. *J Pathol* 197:492–509
3. Giles RH, van Es JH, Clevers H (2003) Caught up in a Wnt storm: Wnt signaling in cancer. *Biochim Biophys Acta* 1653:1–24
4. Freund J-N, Domon-Dell C, Kedinger M et al (1998) The Cdx1 and Cdx2 homeobox genes in the intestine. *Biochem Cell Biol* 76:957–969
5. Van de Wetering M, Sancho E, Verweij C et al (2002) The beta-Catenin/TCF-4 complex imposes a crypt progenitor phenotype on colorectal cancer cells. *Cell* 111:241–250
6. Stappenbeck TS, Mills JC, Gordon JI (2003) Molecular features of adult mouse small intestinal epithelial progenitors. *PNAS* 100:1004–1009
7. He XC, Zhang J, Tong WG et al (2004) BMP signaling inhibits intestinal stem cell self renewal through suppression of Wnt-beta-catenin signaling. *Nat Genet* 36:1117–1121
8. Madison BB, Braunstein K, Kuizon E et al (2005) Epithelial hedgehog signals pattern the intestinal crypt-villus axis. *Development* 132:279–289
9. Radtke F, Clevers H (2005) Self-renewal and cancer of the gut: two sides of a coin. *Science* 307:1904–1909
10. Jenny M, Uhl C, Roche C, Duluc I et al (2002) Neurogenin3 is differentially required for endocrine cell fate specification in the intestinal and gastric epithelium. *EMBO J* 21:6338–6347
11. Van Es JH, van Gijn ME, Riccio O et al (2005) Notch/gamma-secretase inhibition turns proliferative cells in intestinal crypts and adenomas into goblet cells. *Nature* 435:959–963

Hair

► Skin and Hair

Hair Bulge

Definition

Hair bulge is that part of the outer root sheet of the hair follicle that is located above the insertion of the arrector pili muscle. It corresponds to the lowest non-cycling part of the hair follicle and contains multipotent skin stem cells.

► Skin and Hair

Hair Matrix

Definition

Hair matrix consists of transit-amplifying, epithelial cells that are located at the base of the hair follicle, the hair bulb. The progeny of hair matrix cells withdraw from the cell cycle and move upwards in cellular cohorts while undergoing terminal differentiation.

► Skin and Hair

Hairpin End

Definition

Hairpin end refers to the end of a DNA double strand that is sealed in the form of a hairpin by the covalent linkage of the 5' strand to the 3' strand.

► DNA Recombination

Hairpin Ribozyme

Definition

A hairpin ribozyme can cleave oligoribonucleotides at specific sites, GUC.

► Catalytic RNA

Half-Life

Definition

Half-life defines the time required to decrease the concentration of a substance/drug to the half of its original value.

► Biochemical Engineering of Glycoproteins

Hamartoma

Definition

Hamartoma is a benign outgrowth of tissue composed of normally differentiated cells and matrix components typical for that site without signs of dysplasia.

► Peutz-Jeghers Syndrome

► Tuberous Sclerosis

Hammerhead Ribozyme

Definition

Hammerhead ribozymes can cleave oligoribonucleotides at specific sites, e.g. NUX, where N and X are A, G, C, U and A, C, U, respectively, with most efficient cleavage at GUC triplets.

► Catalytic RNA

Haploid

Definition

The term haploid describes a cell that has only a single copy of each chromosome (e.g. post-meiotic germ cells).

► [Mutagenesis Approaches in the Zebrafish](#)

Haploid Analysis

Definition

Haploid analysis refers to an analysis technique aimed at separating two paternal and maternal chromosome pairs into two different cell lines. Haploid analysis makes it possible to separately detect a mutation in each chromosome.

► [Spinal Muscular Atrophy](#)

Haploinsufficiency

Definition

Haploinsufficiency describes the situation in which an individual, who is heterozygous for a certain gene ► [mutation](#) or ► [hemizygous](#) at a particular ► [locus](#), is clinically affected because the protein produced by a single copy of an otherwise normal gene is not sufficient to assure normal function.

- [Bone and Cartilage](#)
- [Hereditary Diseases Genetic Basis](#)
- [Homeodomain Transcription Factors](#)
- [Mouse Genomics](#)
- [Protein/DNA Interaction](#)
- [Replication Fork](#)

Haplotype

Definition

The pattern of polymorphic alleles in a genomic region of a single chromosome. Alleles at multiple loci inherited together on the same chromosome.

- [Atopy, Genetics](#)
- [Chromosome 21, Disorders](#)
- [Diabetes Mellitus, Genetics](#)
- [Schizophrenia, Genetics](#)

Haplotype Block

Definition

Haplotype block is a section of the genome around 25 kilo bases showing minimal diversity amongst individuals of the same ethnic group.

► [Genetic Predisposition to Multiple Sclerosis](#)

Hapten

Definition

Hapten refers to a small molecule that reacts with a specific antibody but cannot induce the formation of antibodies unless bound to a carrier protein or other large antigenic molecule.

► [Protein Microarrays as Tool for Protein Profiling and Identification of Novel Protein-Protein Interactions](#)

Harker Section

Definition

Harker section refers to a special plane within the Patterson map, which contains unusually high peaks caused by vectors between atoms that are related by symmetry elements of the point group (e. g. between heavy atoms). A 6-fold screw axis in space group $P6_1$ or $P6_5$ has five Harker sections at $u, v, 1/6$; $u, v, 1/3$; $u, v, 1/2$; $u, v, 2/3$; $u, v, 5/6$.

► [MAD Phasing](#)

Hayflick Limit

Definition

Hayflick limit is defined as the specific number of proliferations that a primary human cell can perform in culture. After this limit has been reached, the so-called replicative or cellular senescence occurs.

- [Ageing](#)
- [Steroid Hormone Receptor Defects, Molecular Basis](#)

HCAb

- Heavy-Chain Antibody

HCC

- Hepatocellular Carcinoma

HCM

- Hypercalcemia of Malignancy

HCS

- High Content Screening

HD

- Chorea Huntington (HD)
- Huntington's Disease

HDAC

Definition

- Histone Deacetylase

HDAC Inhibitor(s)

Definition

HDAC inhibitor(s) refer(s) to a class of drugs that inhibit histone deacetylases, thereby increasing global histone acetylation.

- Polyglutamine Disease, the Emerging Role of Transcription Interference

HDL

- High-Density Lipoprotein

Health Literacy

Definition

Health literacy is the degree to which individuals are able to obtain, process, and understand basic health information and services that is needed to make appropriate decisions [www.healthypeople.gov].

- Ethical Issues in Medical Genetics

Heart

INGO MORANO

Max-Delbrück-Center for Molecular Medicine,
Berlin, Germany
imorano@mdc-berlin.de

Definition

The heart is a four-chamber pump, the size of a fist, mainly made of muscle tissue (the ►myocardium), located between the lungs behind the breastbone. It is separated into a right heart, which pumps blood through the lungs, and a left heart, which pumps blood through the peripheral organs, roughly 70 times per min during the whole of life.

Characteristics

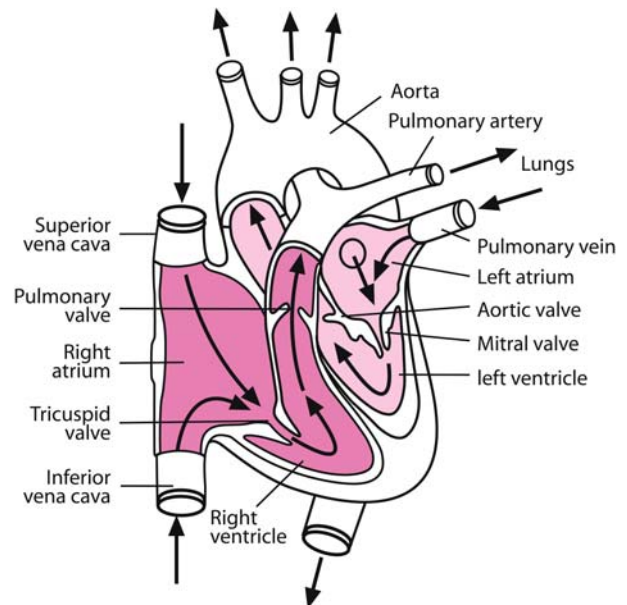
The heart is embedded into a fibrous pericardium, which secretes lubricating liquid. It consists of left and right atria and ventricles, which are separated by the interventricular septum. Desoxygenated blood from the superior and inferior vena cavae of the body enters through the right atrium into the right ventricle, which ejects it through the pulmonary artery into the lungs. Oxygenated blood coming from the lungs enters through the pulmonary vein into the left atrium and then into the left ventricle, which pumps it through the aorta into the body circulation. Atria, ventricles and the outflow tracts are separated by valves, called atrioventricular valves (tricuspid valve right, mitral valve left) and semilunar valves that separate the outflow tracts (pulmonary valve between right ventricle and pulmonary artery, aortic valve between left ventricle and aorta) (Fig. 1). The heart consists of the epicardium, that covers the outer surface and the endocardium, which covers the inner surfaces. The ventricular myocardium, the largest mass, lies between epi- and endo-cardia and consists of overlapping sheets of muscle bundles which follow spiral paths from the base to the apex of the heart.

► **Cardiomyocytes** of the working myocardium are cross-striated, longitudinally arranged cells (around 50 µm in length and 20 µm in width) with one or two nuclei and a well-developed ► **sarcoplasmic reticulum** (SR). The SR communicates with invaginations of the plasmalemma, the T-tubuli. Cardiomyocytes are inter-connected by electron-dense structures, the ► **intercalated discs**, consisting of specialized cell-cell junctions, which provide mechanical and electrical communication between cells (Fig. 2).

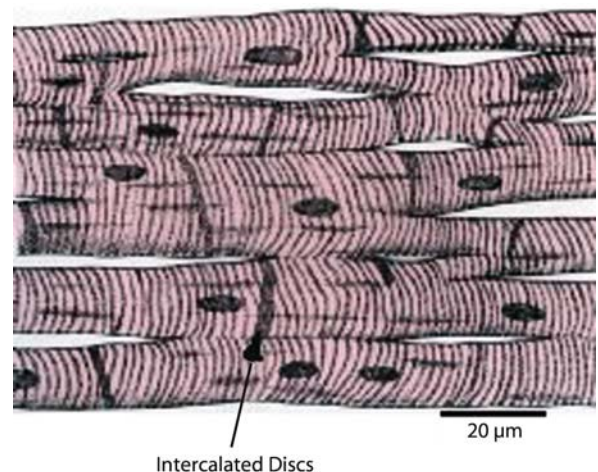
Cardiomyocytes represent just about 1/3 of the total cells of the heart and remain terminally differentiated. However, multipotent cardiac stem cells able to regenerate the myocardium during life may exist. This process may also be relevant for myocardial repair after heart infarction (1).

The right and left coronary arteries, which arise at the root of the aorta, run as epicardial blood vessels and provide the blood supply of the right and left atria and ventricles, respectively. The left main ► **coronary artery** divides into the left anterior descending and the circumflex branches. The venous effluent is collected in veins that parallel the epicardial coronary arteries. The main venous drainage of the left ventricle enters the coronary sinus, emptying into the right atrium. The main venous drainage of the right (and some of the left) ventricle reaches the right atrium by the anterior coronary veins. A small fraction drains through the thebesian veins directly into the cavities of the right and left ventricle.

Activation of the heart by depolarization and generation of ► **action potentials** normally starts in pacemaker



Heart. Figure 1 Scheme of the structure of the heart and blood flow. Desoxygenated blood (dark red) from the superior and inferior vena cava of the body enters through the right atrium into the right ventricle, which ejects it through the pulmonary artery into the lungs. Oxygenated blood (light red) coming from the lungs enters through the pulmonary vein into the left atrium and then into the left ventricle, which pumps it through the aorta into the body circulation.



Heart. Figure 2 Scheme of myocardial cells of the ventricle. The mono- or binucleated cross-striated cardiomyocytes are connected by intercalated discs.

cells of the sinoatrial (SA) node, specialized cells on the junction between the superior vena cava and the right

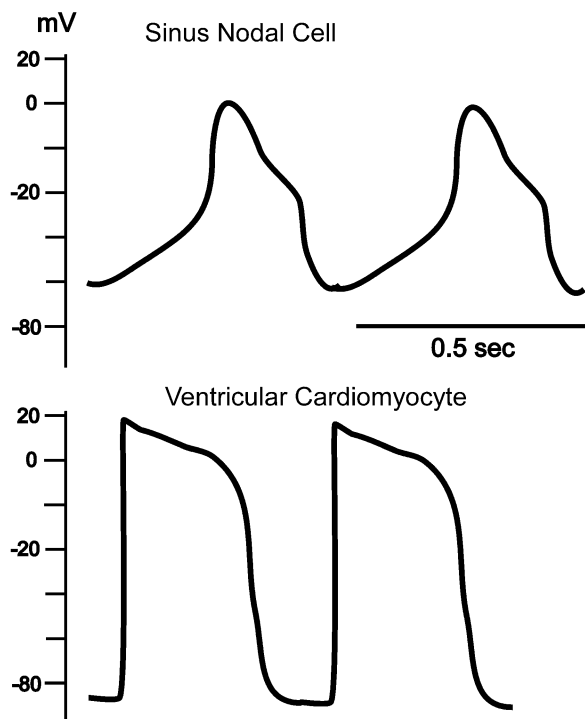
atrium. Action potentials cross the atrioventricular (AV) node and reach the left and right ventricles through AV bundles (or His bundles), which divide into left and right branches. Ramifications of the bundle branches continue as a subendocardial network of Purkinje fibers. Electrical impulses then propagate between cardiomyocytes by a special class of ion channels, each made up of 6 connexin molecules, found in the gap junctions (or nexus) of the intercalated discs. The heart therefore represents a functional syncytium.

Regulatory Mechanisms

The Heartbeat

A heartbeat, or cardiac cycle, consists of a contraction (systole) and a relaxation (diastole) generating blood ejection (normal stroke volume around 70 ml) and blood filling (normal filling volume around 130 ml), respectively. Cardiac output is the product of beating frequency (normal 60–80× per min) and stroke volume, representing an important factor of arterial blood pressure determination and organ perfusion. The heart is innervated by both sympathetic nerve fibers (which originate from the stellate ganglia and cardiac plexus, innervated mainly from the 4th and 5th thoracic segments of the spinal cord) and parasympathetic vagal nerve fibers, (which originate in the dorsal efferent nuclei of the medulla oblongata). Whereas postganglionic sympathetic fibers innervate all heart regions (with a bilateral functional asymmetry), parasympathetic innervation supplies mainly the SA and AV nodes. Vagal stimulation decreased, sympathetic stimulation increased the beating frequency. Beating frequency of the completely denervated heart is considerably above normal, which demonstrates the predominance of parasympathetic tone during rest.

Compared to nerve fibers or skeletal muscle fibers, the action potential of a heart cell is very long, roughly 200–300 ms, almost the duration of a contraction-relaxation cycle. Autonomous activation by SA node cells relies on an unstable resting potential (Fig. 3). A continuous decline in the conductance of K^+ channels and an inward Na^+/Ca^{2+} current depolarize the membrane potential of SA node cells to the threshold potential, around -40 mV, where voltage-operated Na^+ and Ca^{2+} channels open to elicit an action potential which is then propagated through the whole heart (see above). Repolarization is associated with the inactivation of Na^+ and Ca^{2+} channels as well as K^+ outward currents. Action potentials of cardiomyocytes of the ventricular myocardium start with a rapid depolarization caused by gating of the voltage-operated Na^+ channels and a subsequent characteristic long-lasting plateau phase due to a Ca^{2+} current through voltage-operated L-type Ca^{2+} channels (Fig. 3). Inactivation of the Na^+ and Ca^{2+} channels as well as



Heart. Figure 3 Action potentials of sinus nodal cells (top) and ventricular cardiomyocyte (bottom).

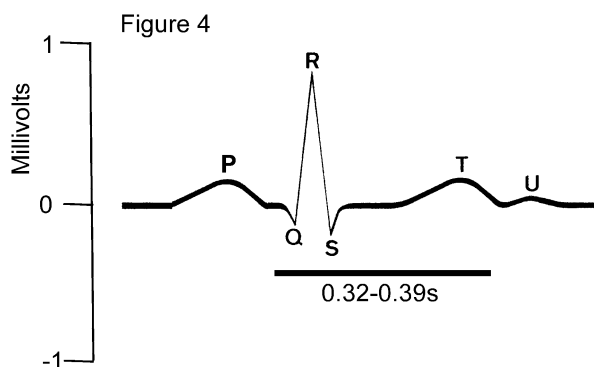
outward K^+ currents through various types of K^+ channels are involved in the repolarization mechanism. Acetylcholine (ACh), the parasympathetic transmitter, binds to muscarinic ACh receptors, 7 membrane-spanning receptors that are associated with small regulatory heterotrimeric inhibitory G-proteins. The activated inhibitory G_i -protein dissociates into its $G_{i\alpha}$ and $G_{i\beta\gamma}$ subunits. The latter directly increased the conductance of K^+ channels, thus stabilizing the resting potential of SA node cells. This causes a prolongation of the time-to-threshold potential, thus decreasing frequency of pacemaker action potentials and therefore heartbeats. G_{α} inhibits adenylyl cyclase (AC), thus reducing the intracellular cAMP levels, antagonizing the influence of sympathetic activation. Norepinephrine, the major sympathetic transmitter, binds to its β -receptor, a representative of the 7 membrane-spanning receptors associated with stimulating Gs-proteins. Upon binding of NE to its β -receptor and subsequent activation of Gs, AC is activated by the G_{α} subunit, thus increasing cAMP levels. The second messenger cAMP activates protein kinase A (PKA) which phosphorylates K^+ channels and Ca^{2+} channels, thus decreasing the outward K^+ current and increasing the inward Ca^{2+} current and thereby rendering the resting potential of pacemaker cells even more unstable. Time-to-threshold potential becomes lessened, pacemaker frequency and therefore beating frequency increase.

The ► Electrocardiogram (ECG)

Electrical changes in the beating myocardium can be recorded using electrodes placed on the skin on opposite sides of the heart. A normal ECG is composed of a P-wave, a QRS-complex and a T-wave, which can be recorded during atrial depolarization, ventricular depolarization and ventricular repolarization, respectively (Fig. 4). ECG recordings allow insights into the anatomical orientation of the heart, relative sizes of the chambers, disturbances of rhythm and conduction, extent, location and progress of ischemic damage, effects of altered electrolyte concentrations and the influences of certain drugs.

Excitation-Contraction Coupling of the Ventricular Cardiomyocyte

The small Ca^{2+} inward current during the action potential triggers the release of large amounts of Ca^{2+} from the sarcoplasmic reticulum (SR) through Ca^{2+} release channels (ryanodine receptors; RyR), which are inserted into the SR membrane (Ca^{2+} -induced Ca^{2+} release). Free cytosolic Ca^{2+} rises to about 600 nm, which roughly half-maximally activates the ►myofilaments of the cardiomyocytes, generating force (for ventricular pressure) and shortening (for blood ejection). Activation of the myofilaments of the heart by Ca^{2+} via binding to troponin C, conformational changes in the regulatory troponin-tropomyosin system and chemo-mechanical energy transformation by the ATP-driven interaction between actin filaments and myosin cross-bridges are very similar to those of skeletal muscle (c.f. chapter "Muscle Contraction"). Hence, force generation in the heart can be regulated by alteration of both the free systolic Ca^{2+} -concentration and the sensitivity of the myofilaments to Ca^{2+} . During diastole, systolic intracellular Ca^{2+} is reduced to resting levels mainly by an ATP-driven Ca^{2+} pump located in

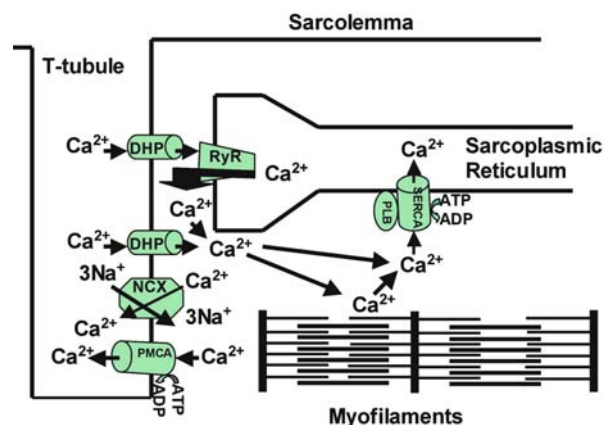


Heart. Figure 4 A normal electrocardiogram. P wave (atrial depolarization), QRS complex (ventricular depolarization prior contraction) and T wave (ventricular repolarization).

the SR membrane (SERCA) and a $\text{Na}^+/\text{Ca}^{2+}$ exchange (NCX) (2). Decreased free cytoplasmic Ca^{2+} inactivates the myofilaments; the heart relaxes (Fig. 5).

Modulation of Cardiac Contraction by Sympathetic Innervation

Since a normal systole caused just around half-maximal activation of the myofilaments, there is a large contractile reserve, which can be recruited by sympathetic stimulation. As mentioned above, β -adrenergic stimulation causes an increased beating frequency (positive chronotropy), which itself increases free cytoplasmic Ca^{2+} -activation. Furthermore, activated PKA upon β -adrenergic stimulation phosphorylates L-type Ca^{2+} channels and ryanodine receptors. These interventions cause an increased Ca^{2+} current through the L-type Ca^{2+} channels and Ca^{2+} release from the SR, respectively, which caused an increased contraction (positive inotropy). In addition, PKA phosphorylates the small SERCA inhibitory protein phospholamban, which thereby loses its inhibitory effect. Hence, SERCA activity rises, causing an accelerated Ca^{2+} sequestration into the SR and therefore accelerated relaxation (positive lusitropic effect). PKA-dependent phosphorylation of troponin I reduces Ca^{2+} sensitivity of the myofilaments, supporting the relaxation process. In addition, the connexin molecules of the gap junctions are phosphorylated by PKA which is believed to increase the ion conduction between the cardiomyocytes, thus accelerating the wave of electrical current through the heart (positive dromotropy) (4).



Heart. Figure 5 Scheme of excitation-contraction coupling in cardiomyocytes. DHP = dihydropyridine receptors or L-Type Ca^{2+} channels; RyR = ryanodine receptors or Ca^{2+} release channels; NCX = $\text{Na}^+/\text{Ca}^{2+}$ exchanger; PMCA = plasmalemmal Ca^{2+} ATPase; SERCA = sarcoplasmic-endoplasmic Ca^{2+} ATPase; PLB = phospholamban.

Cardiac Mechanics

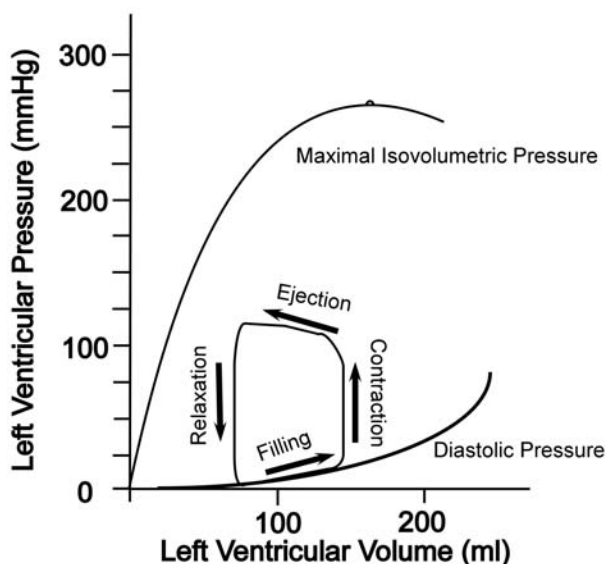
The structure of the cardiac **myofibrils** with its organization of contractile proteins into **sarcomeres** as well as the regulation of activation of the myofibrils are very similar to those of skeletal muscle (cf. Chapter on “Muscle Contraction”). There are actin and myosin filaments anchored into the Z-line and M-line, respectively as well as titin filaments that connect the Z-line with the M-line. A troponin-tropomyosin regulatory system located at the actin filament confers Ca^{2+} regulation of actin-myosin interaction. However, structural and functional features of the sarcomeric proteins differ between cardiac and skeletal muscles due to the expression of different genes and/or alternative splicing of genes. Thus, mechanical features of the two muscle types vary and there are specific regulatory systems that are unique to cardiomyocytes. The resting cardiomyocyte is much stiffer than the skeletal muscle fiber, possibly due to the presence of titin isoforms with different elastic properties. In addition, the presence of the incompressible pericardium restricts extensibility of the whole heart *in situ*. The systolic pressure developed by the heart decreases as its diastolic volume is altered in either direction from an optimum value (law of Frank-Starling) (Fig. 6). In accordance with the sliding filament theory (c.f. Chapter “Muscle Contraction”), developed force rises with increasing sarcomere length to an optimum length of $2.2\ \mu\text{m}$ and declines beyond $2.3\ \mu\text{m}$ (2). In addition, length-dependent differences of Ca^{2+} release from the SR and changes in the Ca^{2+} sensitivity of myofilaments may also contribute to the Frank-Starling law of the heart.

Functional Adaptations of the Heart

Increased work demands on the heart cause a panel of acute and chronic functional adaptations. Thus, increased filling pressure could be acutely compensated by an increased stroke volume (cf. Frank-Starling law, cf. above). Furthermore, sympathetic stimulation increases beating frequency and the ability of the heart to generate tension, thus increasing stroke volume and cardiac output. Chronic overload induces hypertrophy of the myocardium, elicited by quantitative and qualitative changes of gene expression. Chronically enhanced sympathetic drive and increased free Ca^{2+} concentrations upon cardiac overload are translated into hypertrophic gene expression patterns, associated with the activation of numerous signaling pathways in the cardiomyocyte. In particular the calcineurin-NFAT pathway, stimulated by Ca^{2+} -calmodulin, activates the hypertrophic reprogramming of cardiac gene expression (6).

Clinical Relevance

Disorders of the heart can be acquired (e.g. chronic hypertension, viral or bacterial infections, coronary



Heart. Figure 6 Scheme of pressure-volume relationship of the heart. Left ventricular filling increases ventricular pressure during diastole. Upon stimulation, systole starts and the heart contracts, thus increasing left ventricular pressure until the aortic valve opens. Blood is ejected into the aorta. During diastole, aortic valves close and the heart relaxes.

artery disease, valve diseases etc) or inherited. Mutations affecting genes coding for myofibrillar, plasmalemmal, cytoplasmic or Ca^{2+} handling proteins are associated with different types of cardiomyopathies that cause congestive heart failure (3). Mutations that disrupt sarcolemmal-myofibrillar functions are more involved in dilated cardiomyopathy, while mutations causing myofibrillar disruption may lead to hypertrophic cardiomyopathy (7). Mutations in ion channels cause arrhythmic disorders of the heart. Thus, loss of function of K^{+} channels has been implicated in heritable long-QT syndrome. Mutations of Na^{+} channels were associated with idiopathic ventricular fibrillation (8).

►Cardiac Signaling: Cellular, Molecular and Clinical Aspects

References

1. Anversa P, Nadal-Ginard B (2002) Myocyte renewal and ventricular remodelling. *Nature* 415:240–243
2. Bers DM (2002) Cardiac excitation-contraction coupling. *Nature* 415:198–206
3. Chien KR (2003) Genotype, phenotype: upstairs, downstairs in the family of cardiomyopathies. *J Clin Invest* 111:175–178
4. Katz AM (2001) Physiology of the heart. 3rd edn, Lippincott Williams and Wilkins, Philadelphia
5. Fabiato A, Fabiato F (1975) Dependence of the contractile activation of skinned cardiac cells on the sarcomere. *Nature* 256:54–56

6. Frey N, Olson EN (2003) Cardiac hypertrophy: the good, the bad, and the ugly. *Ann Rev Physiol* 65:45–79
7. Towbin JA, Bowles NE (2002) The failing heart. *Nature* 415:227–233
8. Marban E (2002) Cardiac channelopathies. *Nature* 415:213–218

Heat Capacity Change

Definition

Heat capacity change (DC_p) represents the second temperature derivative of the free energy change at constant pressure. It describes the temperature dependence of the enthalpy and entropy changes.

► [Thermodynamic Properties of DNA](#)

Heat Shock

Definition

Heat shock during the first cell cycle inhibits the cell division and leads to homozygous diploid embryos.

► [Mutagenesis Approaches in the Zebrafish](#)

Heat Shock Protein

Definition

Heat shock proteins (Hsp, stress proteins) belong to a family of proteins conserved through prokaryotic and eukaryotic cells. They are expressed in response to environmental stresses including, but not restricted to, elevated temperatures. They increase cellular tolerance to stress.

► [DNA-based Vaccination](#)

► [Protein Folding](#)

► [Proteomics in Cardiovascular Disease](#)

Heat Shock Response

Definition

Heat shock is induced by elevated temperature, as well as a variety of other cell stresses such as exposure to heavy

metals, hypoxia and lack of glucose. Characteristic of the heat shock response is the induction of heat shock protein (hsp) expression. Heat shock proteins can prevent cell death. Many of the hsp family members are molecular ► [chaperones](#), which are involved in protein folding and transport and the assembly of multiprotein complexes. During cell stress they protect proteins by being involved in protein repair and/or degradation of destroyed proteins through the ubiquitin-proteasome pathway.

► [HSF](#)

► [Protein Folding](#)

► [Translational Control in Eukaryotes](#)

► [Ubiquitin-Proteasome Pathway](#)

Heavy-Chain Antibody

Definition

Heavy-chain antibodies (HCAb) are functional antibodies that naturally occur in Camelidae species. The molecules are composed of a homodimer of heavy chains and are lacking in light chains.

► [Camel as a Model for Functional Genomics](#)

HECT-Domain

Definition

The HECT-domain (short for homologous to the E6–AP Carboxyl terminus) is a characteristic catalytic domain for a subclass of E3 ubiquitin ligases, which can form a covalent bond with ubiquitin. The domain is defined by analogy to the viral factor E6–AP (homology to E6–AP Cterminus).

► [Ubiquitination](#)

Hedgehog

Definition

Hedgehog (HH) is a term for a family of signaling molecules (morphogens) that control development of the limb, the neural tube, the brain (sonic hedgehog) and differentiation of cartilage (indian hedgehog) due to activation of transcription.

- Bone Disease and Skeletal Disorders, Genetics
- Hedgehog signalling
- Wnt/Beta-Catenin Signaling Pathway

Hedgehog Signalling

CAROL WICKING, TIMOTHY M. EVANS

The Institute for Molecular Bioscience, The University of Queensland, St Lucia, Queensland, Australia
c.wicking@imb.uq.edu.au

Definition

The hedgehog signalling pathway is central to the embryonic development of a vast array of vertebrate organ structures (reviewed in 1). Aberrant functioning of this pathway results not only in a range of dysmorphology syndromes, but also contributes to a number of common cancers in both their familial and sporadic forms (reviewed in 2, 3, 4). This highlights the importance of hedgehog signalling in embryonic development as well as in further growth and differentiation in the adult.

Hedgehog signalling is highly conserved from *Drosophila* through to humans, with much of what we know about this pathway gleaned from classic studies in the fly. As is the case with many processes, hedgehog signalling is complicated in the vertebrate by the presence of families of homologous molecules, which in *Drosophila* are represented by a single protein. In some cases the role of these family members can be inferred to some extent from their expression patterns, as is the case with the three vertebrate hedgehog molecules, Sonic hedgehog (Shh), Indian hedgehog (Ihh) and Desert hedgehog (Dhh). In other cases such as the three Gli molecules, which mediate the transcriptional output of hedgehog signalling, their role is regulated not only by site of expression but also by post-translational modification in a context-dependent manner.

Characteristics

Hedgehog Signalling and Developmental Disease

Mutation of key members of the hedgehog signalling pathway has been implicated in a range of human disease states (reviewed in 2, 3). The array of phenotypes highlights the importance of hedgehog signalling in regulating a large number of diverse processes in the developing embryo. *SHH* itself has been shown to be mutated in a subset of familial ►holoprosencephaly (HPE), a midline defect of the forebrain and face. The gene encoding the hedgehog receptor patched (*PTCH*) is mutated in ►Gorlin's syndrome, which is characterised by tumour susceptibility

as well as associated developmental defects. The gene encoding the downstream transcription factor *GLI3* has also been implicated in developmental anomalies, with mutations identified in ►Greig's cephalopolysyndactyly (GCPS) and a number of other syndromes primarily characterised by limb defects.

Hedgehog Signalling and Tumorigenesis

In terms of tumorigenesis, up-regulation of hedgehog signalling has been associated with a number of common cancers (reviewed in 3, 4). The *PTCH* gene acts as a ►tumour suppressor gene and has been implicated as a key factor in the development of a range of tumour types including the common skin tumour ►basal cell carcinoma (BCC) and ►medulloblastoma, a paediatric brain tumour. While to date no mutations have been detected in human ►rhabdomyosarcoma, heterozygous *PTCH* knockout mice develop these muscle-related tumours, suggesting a role for hedgehog signalling in their genesis. Mutations in the suppressor-of-fused (*SUFU*) gene, which encodes a downstream negative effector of hedgehog signalling, have also been identified in medulloblastoma.

In keeping with an up-regulation of hedgehog signalling leading to cancer formation, activating oncogenic mutations in BCCs and brain tumours have been identified in genes encoding positive regulators of the hedgehog pathway, including smoothened (SMO), also a component of the hedgehog receptor complex. In addition, several mouse models of BCC result from activation of the pathway by over-expression of either *Gli1*, *Gli2*, *Shh* or an activated *Smo* allele in the skin. Until recently, dysregulation of hedgehog signalling was associated primarily with cancers that occur with increased incidence in Gorlin's syndrome, such as skin and brain tumours. However, several reports now implicate up-regulation of this pathway in small cell lung carcinoma (SCLC), pancreatic adenocarcinoma, prostate cancer and digestive tract tumours (reviewed in 4), thus suggesting a more widespread role for hedgehog signalling in tumour formation.

It is clear that a range of tumours result from constitutive activation of hedgehog signalling, and for a long time it was believed that this was due to the nature of the downstream target genes regulated by this pathway. While this is likely to be a contributing factor, the demonstration that hedgehog signalling is directly involved in regulation of a number of cyclin genes suggests that some of the tumorigenic effects of this pathway may result from a more direct effect on cell-cycle progression.

Molecular Interactions

Reception and Transduction of the Hedgehog Signal

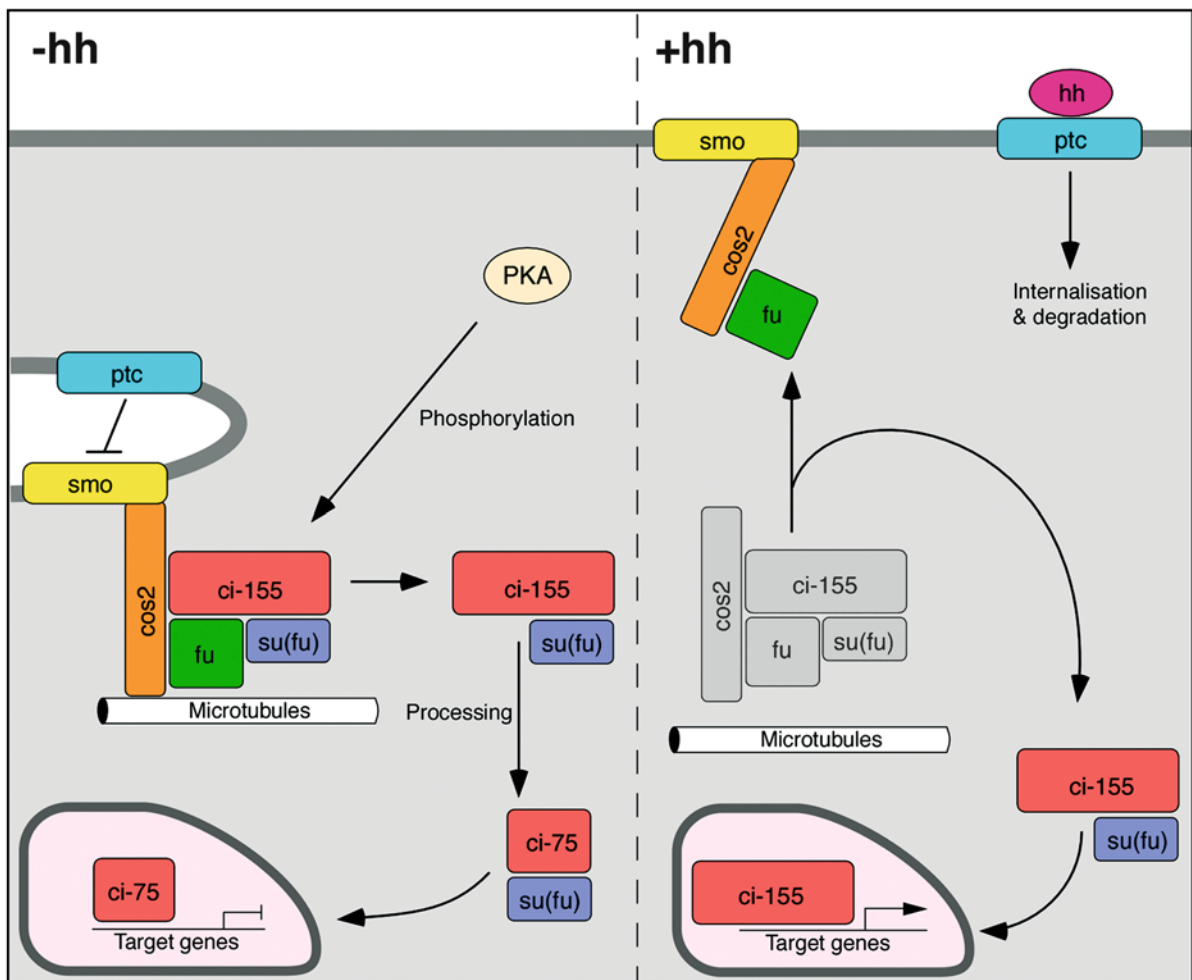
The prevailing model of hedgehog signalling predicts that reception of the hedgehog ligand at the cell surface

is mediated by a multimeric complex involving two transmembrane molecules, patched and smoothened (Fig. 1). In the absence of a hedgehog ligand, patched, a twelve-pass transmembrane protein, is thought to inhibit the activity of smoothened, a protein with homology to a G-protein coupled receptor. On binding of hedgehog to patched this inhibition is relieved and smoothened transduces the hedgehog signal into the receiving cell.

A number of other molecules have more recently been implicated at this stage of hedgehog signalling but to date their role is less clearly defined. These include dispatched, gas1 and hedgehog interacting protein (Hip).

Transcriptional Mediation

Within the cell the hedgehog signal is mediated primarily by the zinc finger ►transcription factor cubitus interruptus (ci) in *Drosophila*. In vertebrates there are three members of the Gli family, which together have evolved to perform the function of ci. In flies the single hedgehog molecule regulates whether the bipartite factor ci acts as an activator or a repressor through proteolytic cleavage of the full-length protein, which is capable of acting as an activator, into an N-terminal repressor form. A major question in understanding vertebrate hedgehog signalling is how the role of ci has been partitioned between the three Gli molecules in a context-dependent manner. This has



Hedgehog Signalling. Figure 1 Model for hedgehog (hh) signal transduction in *Drosophila*. In the absence of hh (-hh) patched inhibits the activity of smoothened. The cytoplasmic hedgehog signalling complex (HSC) which includes costal2 (cos2), fused (fu), suppressor of fused (su(fu)) and ci is bound to microtubules. PKA phosphorylates ci-155, which triggers processing and cleavage by the proteasome to generate ci-75 which enters the nucleus and represses target gene expression. Upon hh stimulation (+hh), hh binds ptc at the plasma membrane and together they are internalised and degraded. The HSC complex is translocated to the plasma membrane and components are phosphorylated. Cleavage of ci is inhibited and the full-length form (ci-155) enters the nucleus to activate target gene transcription.

been the subject of intense debate for some years and only now is starting to be unravelled (reviewed in 5). It now appears that Gli1 and 2 act primarily in gene activation, while Gli3 is predominantly a transcriptional repressor. Gli1 contains only an activator domain and is not proteolytically cleaved. Interestingly Gli1 is also a universal transcriptional target of hedgehog signalling. Gli2 contains both activator and repressor domains and can undergo cleavage under certain conditions, but its major role *in vivo* is thought to be as an activator. In contrast Gli3 contains activator and repressor domains and undergoes proteolytic cleavage in an analogous manner to ci to generate a transcriptional repressor. While the major role of Gli3 is to repress target genes, recent studies have shown that, at least in the spinal cord, the full-length form is also capable of activation (6).

A number of the known downstream target genes regulated by Gli molecules have been extrapolated from studies in *Drosophila*. However, with the advent of transcriptional profiling approaches such as microarray analysis, we now have the opportunity to identify novel downstream targets of hedgehog signalling on a genomic scale and thus potentially link the hedgehog pathway to a host of other signalling networks within the cell.

The Hedgehog Signalling Complex (HSC) – Linking the Cell Membrane to the Nucleus

One of the remaining questions in understanding the molecular basis of hedgehog signalling is the nature of the Smo signal and how this is transmitted to the ci/Gli molecules in the nucleus. This has at least in part been solved by a number of recent studies showing a direct interaction between Smo and the kinesin-like protein Costal-2 (Cos2) (reviewed in 7). Previous studies in *Drosophila* have shown that in the absence of hedgehog, ci is bound to microtubules in a complex involving cos2, the serine-threonine kinase fused (fu) and the novel protein suppressor of fused (su(fu)). In this state, phosphorylation of full-length ci (ci-55) by protein kinase A (PKA) triggers proteolytic cleavage to a 75 kD N-terminal repressor fragment (ci-75) which enters the nucleus and inhibits transcription of target genes. Stimulation by hedgehog results in translocation of this complex to the cell membrane and phosphorylation of component molecules. At the same time cleavage of ci is inhibited and under certain conditions, the full-length molecule enters the nucleus and activates transcription. As discussed above a similar process in vertebrates is thought to be mediated by the three Gli molecules. These studies provide a direct link between the hedgehog receptor complex and transcriptional outputs in the nucleus and take us a step forward in understanding the events involved in transmission of the hedgehog signal.

In addition to the key components of hedgehog signalling discussed above there exist a number of accessory proteins that have been implicated in hedgehog signalling. In addition, there will undoubtedly be a host of yet to be discovered molecules that interact with the core signalling components to modulate the final output of the hedgehog pathway.

Regulatory Mechanisms

The mechanisms regulating hedgehog signalling at each step from reception and transduction of the signal through to transcriptional mediation of downstream target genes are extremely complex and still poorly defined. At the level of signal reception and transduction there is mounting evidence to suggest that both intracellular sterol levels and trafficking events play key regulatory roles. A role for sterols in hedgehog signalling is supported by a number of independent observations. Firstly, Shh is autoproteolytically cleaved into an active 19 kD molecule which is modified by addition of a cholesterol moiety to its C-terminus and is also N-terminally palmitoylated. Second, mutations in Shh have been implicated in holoprosencephaly, which is also seen in a subset of individuals with ►Smith-Lemli-Opitz syndrome, (SLOS), which results from a defect in cholesterol biosynthesis. In addition, the offspring of animals fed with the steroidal alkaloid ►cyclopamine, which also directly inhibits Shh signalling, develop HPE. The final link between hedgehog signalling and cholesterol came with the discovery that patched contains a putative sterol-sensing domain (SSD). While the role of the SSD in patched is yet to be fully elucidated, a similar domain in a number of other proteins, mostly involved in cholesterol homeostasis, has been shown to be responsive to sterol levels. A role for trafficking events in regulating hedgehog signalling is supported by the finding that the Rab23 vesicular transport protein acts as a negative regulator of hedgehog signalling (8). Rab23 is a member of the Rab GTPase family of proteins, and was shown to be mutated in the open-brain mouse whose phenotype resembles a partial loss of patched function (8). While the exact function of Rab23 is yet to be determined, recent studies suggest that it is localised to both the plasma membrane and the endocytic pathway, and that it is likely to act downstream of patched and smoothed in hedgehog regulation (9).

Since the first discovery of the *Drosophila* hedgehog mutant over twenty years ago genetic, biochemical and cell biological approaches have combined to unravel the central role of the vertebrate hedgehog pathway in embryonic development and disease. This highlights the value of comparative analysis across species in the elucidation of developmental signalling pathways, although it should be noted that a number of hedgehog pathway components are thought to be mammalian

specific. With the advent of large scale functional genomics and proteomics type approaches, it is likely that the coming years will see an expansion of our knowledge of the intricacies of hedgehog regulation and of how links with other signalling networks combine to orchestrate complex developmental processes.

References

1. Ingham PW, McMahon AP (2001) Hedgehog signaling in animal development: paradigms and principles. *Genes Dev* 15:3059–3087
2. Mullor JL, Sanchez P, Ruiz i Altaba A (2002) Pathways and consequences: Hedgehog signaling in human disease. *Trends Cell Biol* 12:562–569
3. Wicking C, Smyth I, Bale A (1999) The hedgehog signalling pathway in tumorigenesis and development. *Oncogene* 18:7844–7851
4. Beachy PA, Karhadkar SS, Berman DM (2004) Tissue repair and stem cell renewal in carcinogenesis. *Nature* 432:324–331
5. Koebernick K, Pieler T (2002) Gli-type zinc finger proteins as bipotential transducers of Hedgehog signaling. *Differentiation* 70:69–76
6. Bai CB, Stephen D, Joyner AL (2004) All mouse ventral spinal cord patterning by hedgehog is gli dependent and involves an activator function of gli3. *Dev Cell* 6:103–115
7. Collins RT, Cohen SM (2004) The Secret Life of Smoothened. *Dev Cell* 5:823–824
8. Eggenschwiler JT, Espinoza E, Anderson KV (2001) Rab23 is an essential negative regulator of the mouse Sonic hedgehog signalling pathway. *Nature* 412:194–198
9. Evans TM, Ferguson C, Wainwright BJ et al (2003) Rab23, a negative regulator of hedgehog signaling, localizes to the plasma membrane and the endocytic pathway. *Traffic* 4:869–884

Helicases

Definition

Helicases are proteins that are involved in the unwinding of double-stranded DNA and RNA, and are of pivotal importance for replication, DNA repair, transcription and recombination. They are of special importance for repair of damage by oxidative stress.

- ▶ Ageing
- ▶ DNA Polymerases
- ▶ Nucleotide Excision Repair
- ▶ Steroid Hormone Receptor Defects, Molecular Basis

Helix

- ▶ Alpha-Helix

Helix Boundary Motif

Definition

At either end of the α -helical rod domain, forming the main structural feature of all intermediate filament proteins, is a recognizable and well conserved stretch of amino acid sequence known as a helix boundary motif. The helix initiation motif (HIM) is located at the amino terminal end of the rod domain, and the helix termination motif (HTM) is located at the carboxyl end is. The HTM is recognizable in all intermediate filament sequences from different organisms. The HIM is more specific to each type of intermediate filaments. These helix boundary motifs are involved in end-to-end overlaps of the subunits in the assembled filament, and mutation in these regions can severely compromise the function of intermediate filaments, leading to cell fragility.

- ▶ Intermediate Filaments

Helix Initiation

Definition

Helix initiation designates a process by which complementary DNA strands form an initiation complex.

- ▶ Thermodynamic Properties of DNA

Helix Initiation Peptide Helix Termination Peptide

Definition

Helix initiation and Helix termination peptides are highly conserved sequences at the beginning and end of the central keratin rod domain, crucial for heterodimer alignment and keratin intermediate filament assembly.

- ▶ Heritable Skin Disorders

Helix Notation

Definition

In visual pigments and ▶GPCRs the seven transmembrane helices are assigned by Roman numbers. Starting from the first N-terminal helix (H-I) and ending with the C-terminal helix (HVII). An additional amphipathic

helix, following H–VII, is denoted H–VIII. In bacterial rhodopsins capital letters are used, i.e. H-A to H-G).

► [Photoreceptors](#)

Helix Propagation

Definition

Helix propagation designates the extension of base pairing subsequent to the formation of the initiation complex.

► [Thermodynamic Properties of DNA](#)

Helix Termination Peptide

► [Helix Initiation Peptide; Helix Termination Peptide](#)

Helix-Loop-Helix

► [Basic Helix-Loop-Helix](#)

Helix-Turn-Helix

Definition

Helix-turn-helix is a structural binding motif that consists of two α helices oriented in a conserved angle bind to DNA. This motif is found in ► [homeotic proteins](#).

► [Transcription Factors and Regulation of Gene Expression](#)

Helix-Turn-Helix Transcription Factors

Definition

Helix-turn-helix (HTH) transcription factors comprise of a large family of bacterial and eukaryotic

transcription factors, which use the helix-turn-helix motif for sequence specific recognition in the major groove.

► [Helix-Turn-Helix Transcription Factors](#)

► [Protein/DNA Interaction](#)

► [Transcription Factors and Regulation of Gene Expression](#)

Helper Phage

Definition

Helper phages (e.g. M13K07) provide all the necessary gene products for particle formation when using ► [phagemid](#) vectors. They are mutated wild-type phage containing the whole genome, with a defective origin of replication or packaging signal, and hence, are inefficient in self-packaging. When infecting *E. coli* containing a phagemid, phage progeny is produced with only the phagemid being packaged efficiently. Therefore, the resulting phage virion contains a mixture of wild-type and fusion coat protein, and the genetic information of the fusion protein is encoded by the packaged phagemid.

► [Protein Interaction-Phage Display](#)

Hematopoiesis

Definition

Hematopoiesis is the maintenance, production and development of all hematopoietic (blood) cell types throughout the entire lifespan of an organism from hematopoietic stem cells in the bone marrow.

► [Bone and Cartilage](#)

► [Hemochromatosis](#)

► [Interleukins and Their Receptors, a Role in Regulating Hematopoietic Stem Cells](#)

► [Stem Cells - Overview](#)

► [Wnt/Beta-Catenin Signaling Pathway](#)

Hematopoietic Stem and Progenitor Cells

Definition

Hematopoietic stem and progenitor cells (HSCs/HPs) are multipotent hematopoietic cells that form the basis

of the hematopoietic hierarchy. These cells are capable of producing all differentiated mature hematopoietic cell types. Hematopoietic stem cells have the capacity to extensively self-renew and reconstitute the hematopoietic system of an HSC ablated individual, while progenitor cells have limited self-renewal capacity and limited differentiation ability.

► [Interleukins and Their Receptors, a Role in Regulating Hematopoietic Stem Cells](#)

► [Stem Cells - Overview](#)

Hemichannels

Definition

Hemichannels are connexin hexamers destined to dock with partners in neighbouring cells thus generating a ► [gap junction](#). They also function independently in the release by cells of ATP during metabolic stress.

► [Gap Junctions](#)

Hemidesmosomes

SANDY H.M. LITJENS, ARNOUD SONNENBERG

Division of Cell Biology, The Netherlands Cancer Institute, Amsterdam, The Netherlands

s.litjens@nki.nl

a.sonnenberg@nki.nl

Definition

Hemidesmosomes are multi-protein complexes that provide firm adhesion of basal epithelial cells to the underlying ► [basement membrane](#) in (pseudo-) stratified and some complex epithelia. Hemidesmosomes were initially defined as distinctive electron-dense structures, composed of an inner plaque connected to the intermediate filament system, an outer plaque, which is linked to the extracellular basement membrane and a sub-basal dense plate located beneath the outer plaque in the lamina lucida of the basement membrane (Fig. 1).

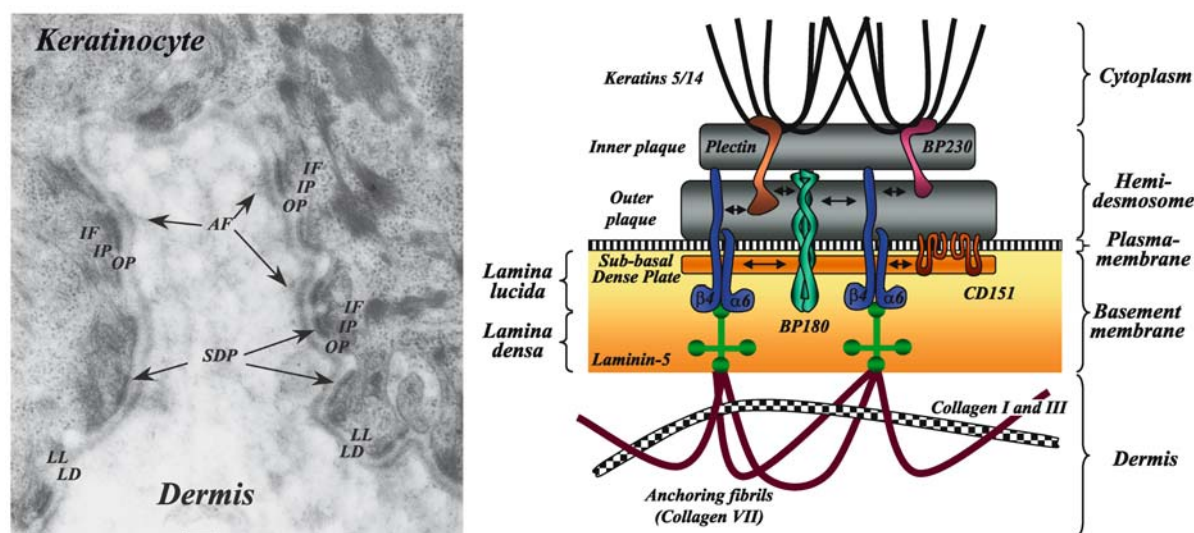
Characteristics

Anchoring of basal cells to the basement membrane is of fundamental importance for maintaining the integrity of epithelial tissues. For instance in the skin,

hemidesmosomes are crucial for the stable adhesion of the epidermis to the underlying dermis and mutations in components of these adhesion-complexes result in tissue fragility and blistering of the skin. In order to form a stable adhesion structure, six hemidesmosomal proteins are assembled into large clusters at the basal side of cells (1, 2) (Fig. 1). The ► [integrin](#) $\alpha 6 \beta 4$ is the central component of hemidesmosomes and is essential for their efficient assembly. Compared to other integrin β subunits, the β subunit of $\alpha 6 \beta 4$ has an unusually large cytoplasmic domain, containing two pairs of fibronectin type III (FNIII) repeats, which are separated by a connecting segment (Fig. 2). The integrin forms a link between its extracellular ligand ► [laminin-5](#) and the inner plaque proteins plectin and BP230 (Fig. 1). Plectin and BP230 are members of the ► [plakin](#) family of ► [cytoskeletal linker](#) proteins. In hemidesmosomes, plectin and BP230 serve as a bridge between the outer plaque and the ► [intermediate filament](#) system, to which the hemidesmosome is connected. The other two hemidesmosomal components are the collagenous transmembrane protein BP180, also referred to as type XVII collagen and the ► [tetraspanin](#) CD151, which like $\alpha 6 \beta 4$, reside in the outer plaque of the hemidesmosome (Fig. 1). BP180 occurs as a homotrimeric complex and is thought to contribute to the stable attachment by linking the hemidesmosome to the extracellular matrix. The role of CD151 in hemidesmosomes is largely unknown.

Molecular Interactions

Some simple epithelia (e.g. the epithelium of the gut) do not contain the complex ultra structure of the epidermal hemidesmosome. However, these epithelia do contain clusters of $\alpha 6 \beta 4$ bound to laminin-5, in which plectin and possibly CD151 are also localized. These complexes are called type II hemidesmosomes, as opposed to type I hemidesmosomes, which also contain BP180 and BP230. The presence of type II hemidesmosomes, which can be regarded as rudimentary hemidesmosomes, suggests that the interaction of the integrin $\alpha 6 \beta 4$ with plectin is the first step in hemidesmosome assembly (3). Indeed, single amino acid substitutions in the $\beta 4$ subunit that abrogate its interaction with plectin result in a complete loss of hemidesmosome formation *in vitro* and a reduced number of rudimentary hemidesmosomes *in vivo*. The main interaction between $\beta 4$ and plectin occurs *via* the first pair of FNIII repeats and the N-terminal part of the connecting segment of $\beta 4$ (PBS-1) and the actin binding domain (ABD) of plectin. However, secondary binding sites have been determined in the connecting segment (PBS-2) and the C-terminal tail of $\beta 4$ (PBS-3) by which this subunit interacts with the plakin domain of plectin (Fig. 2). These additional binding sites may



Hemidesmosomes. Figure 1 Structure of the hemidesmosome. (left) Electron micrograph of a skin sample containing hemidesmosomes. AF, anchoring fibrils; LL, lamina lucida; LD, lamina densa; SDP, sub-basal dense plate; OP, outer plaque; IP, inner plaque; IF, intermediate filaments. (right) Schematic representation of a hemidesmosome illustrating the molecular interactions involved in its assembly.

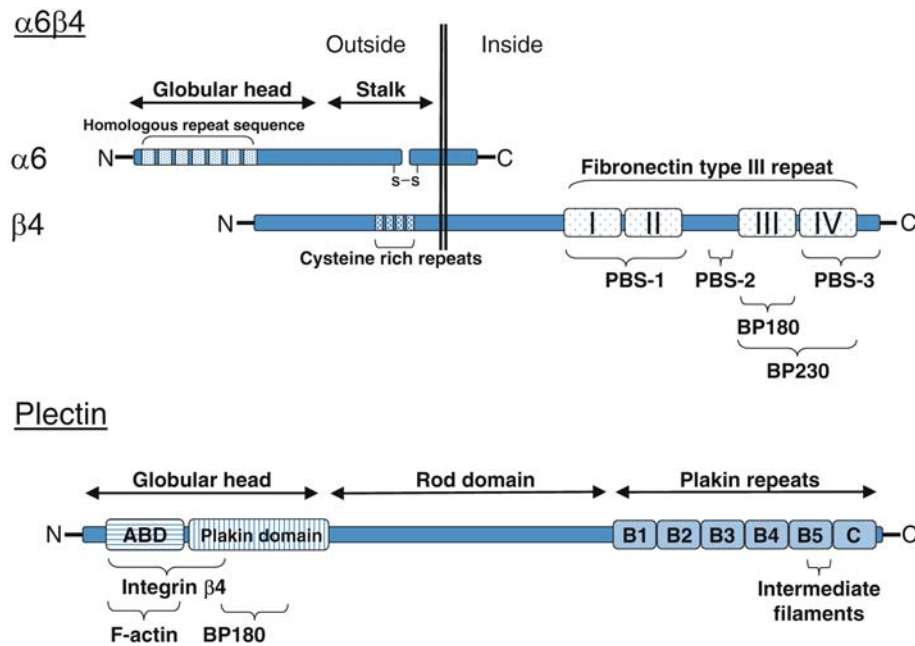
stabilize the original interaction between $\beta 4$ and the ABD of plectin. However, they are not capable of generating a primary interaction between these two proteins, e.g. when the binding site for the plectin-ABD is destroyed.

All the other hemidesmosomal components also bind to the $\alpha 6 \beta 4$ integrin (Fig. 1). The tetraspanin CD151 interacts with the $\alpha 6$ subunit of the integrin. Tetraspanins contain short cytoplasmic carboxy and amino terminal domains, a small cytoplasmic loop, and one small and one large extracellular loop (4). They typically contain conserved polar residues in their transmembrane domains and 4–6 conserved extracellular cysteine residues linked into 2–3 disulfide bonds in their large extracellular loop. Four of these cysteine residues are absolutely conserved, and two of these appear in a CCG (cysteine-cysteine-glycine) motif. These features distinguish tetraspanins from other proteins containing four transmembrane domains. In the large extracellular loop of CD151, the residues QRD^{194–196} are involved in binding to the integrin $\alpha 6$ subunit. The binding of $\alpha 6$ to CD151 and the binding of $\beta 4$ to plectin are early steps in hemidesmosome assembly. Once plectin is recruited by $\alpha 6 \beta 4$, BP180 can bind to both the α - and the β -subunit of the integrin as well as to plectin, thereby stabilizing the ternary complex (3). The extracellular domains of BP180 and $\alpha 6$ interact, whereas the cytoplasmic domain of BP180 binds to the third FNIII repeat of the $\beta 4$ cytoplasmic domain and the plakun domain of plectin (Fig. 2). Although BP230 can directly bind to the second pair of FNIII repeats of $\beta 4$ (Fig. 2), this interaction is not

sufficient for its recruitment into hemidesmosomes. The presence and interaction with BP180 is also necessary for the incorporation of BP230 into hemidesmosomes (3). Both plectin and BP230 contain an intermediate filament-binding domain in their C-terminus, and link the hemidesmosome to intermediate filaments (keratin 5/14 in the basal keratinocytes of the epidermis). Plectin has been shown to interact with various types of intermediate filaments, including several keratins, vimentin and desmin. BP230, however, only binds to keratin 5/14, and not to keratin 8/18 or vimentin. Finally, to link the hemidesmosomal complex to the basement membrane, both the integrin $\alpha 6 \beta 4$ and BP180 bind to ligands in the extracellular matrix. The main ligand for $\alpha 6 \beta 4$ in the skin is laminin-5, but other laminins (1, 2/4, 10/11) can also bind to this integrin. The ligand for BP180 is unknown, although there is data suggesting that BP180 might bind to laminin-5, albeit with low affinity.

Acquired and Genetic Skin Diseases Involving Hemidesmosomes

The degree and complexity of the interactions between the components of the hemidesmosome indicate that the integrity of the complex is crucial for its role in anchoring cells to the basement membrane. Indeed, down-regulation or deletion of any of the hemidesmosomal components can result in severe and sometimes fatal blistering diseases of the skin, both in human patients and in mouse models. Diseases affecting hemidesmosomes can be roughly classified into two



Hemidesmosomes. Figure 2 Domain structure of two major hemidesmosomal components, the integrin $\alpha 6 \beta 4$ and plectin. The binding sites for plectin (PBS-1 to 3), BP180 and BP230 on the $\beta 4$ cytoplasmic domain and the binding sites for $\beta 4$, F-actin, BP180 and intermediate filaments on plectin are indicated. ABD, actin binding domain; N, N-terminus; C, C-terminus.

groups, acquired and hereditary skin diseases. Acquired skin diseases are characterized by the presence of autoantibodies against one or more of the hemidesmosomal components (5). Autoantibodies against the $\beta 4$ subunit of the $\alpha 6 \beta 4$ integrin, plectin, BP180, BP230 and laminin-5 have all been identified in patients with different kinds of pemphigoid diseases (Table 1). The main characteristic of these diseases is blistering of the skin and sometimes of other epithelia, e.g. the oral mucosa. Hereditary skin diseases are caused by mutations in the genes coding for hemidesmosomal components, leading to a reduced expression, absence or dysfunction of the protein. Genetic diseases are also characterized by severe and often fatal blistering of the skin. Mutations have been identified in all the genes encoding hemidesmosomal components, except BP230 (*BPAG1*) (6) (Table 1). *BPAG1* encodes multiple neuronal isoforms. Consequently, *BPAG1* knockout mice not only suffer from blistering of their skin, but also develop a neurological disease characterized by severe neuro-degeneration with dystonia and ataxia. Similarly, if mutations in *BPAG1* occur in humans, they might lead to such severe neurological diseases that they are lethal in the embryonic stage, or before skin disorders can manifest themselves.

The loss of any of the hemidesmosomal components causes fragility of the skin and blistering upon application of mechanical stress. Nonetheless, the loss

of each of them produces a characteristic cleavage pattern of the epidermis, often reflecting the position and function of the protein involved. For example, defects in plectin cause rupture of cells at the level of the hemidesmosomes, leading to epidermolysis bullosa simplex (EBS). When $\alpha 6 \beta 4$ is absent, the split occurs either within the cell at the location of the hemidesmosome and/or through the extracellular basement membrane, producing a combined cellular (EBS) and junctional (JEB) phenotype. Additional complications that characterize defects of plectin and $\alpha 6 \beta 4$ are **muscular dystrophy** (MD) and pyloric atresia (PA), respectively. Loss of BP180 has been mainly associated with cleavage of the skin at the basement membrane zone as is the case when laminin-5 is absent. Notably, while loss of BP180 only causes a mild form of **epidermolysis bullosa** (simplex) (GABEB), the absence of laminin-5 results in a severe and fatal skin disorder (JEB). Absence of CD151 also leads to a relatively mild form of EB. However, patients without CD151 suffer from fatal renal failure. In contrast to the other available knock-out mice (see Table 1), *CD151* knock-out mice do not have the same phenotype as human patients; i.e. they do not have epidermal blistering or renal failure. These mice only show minor abnormalities in tissue homeostasis. Probably, compensatory mechanisms are responsible for this much milder phenotype. Cleavage above the hemidesmosome, leading to EBS are due to mutations in either

keratin 5 or 14, whereas that beneath the basement membrane in the sub-basal lamina lead to dystrophic epidermolysis bullosa (DEB). The latter defect is caused by mutations in type VII collagen, which forms the anchoring fibrils of the hemidesmosomal adhesion complex (Fig. 1). Thus, the clinical features associated with the loss of these hemidesmosomal components may vary from a relatively benign form of epidermolysis bullosa (in the case of loss of BP180) to very severe forms of junctional epidermolysis bullosa (when $\alpha 6\beta 4$ or laminin-5 is absent) (6) (Table 1).

Regulatory Mechanisms

The skin is a dynamic tissue that is constantly renewed and therefore the assembly and disassembly of hemidesmosomes needs to be tightly regulated. When the basal cells of the epidermis differentiate and move upwards, the expression of $\alpha 6\beta 4$ integrin is down regulated and hemidesmosomes are disassembled. Furthermore, during healing of cutaneous wounds, the stable hemidesmosomal adhesions must be weakened to allow basal keratinocytes to migrate into the

wound site and to re-establish the epidermis. Not only is regulation of hemidesmosome assembly and disassembly important for the proper functioning of normal cells. Also in cancer cells, hemidesmosomes are disassembled to enable invasion into the surrounding tissues (7). In many carcinomas, $\alpha 6\beta 4$ expression is increased; however, the integrin is not clustered in hemidesmosomes but evenly distributed over the membrane. Furthermore, abnormal expression of $\alpha 6\beta 4$ in carcinoma cells is correlated with an invasive phenotype, both in humans and in mouse models. How these processes are regulated is not well understood. It is clear that the phosphorylation state of the $\beta 4$ integrin is crucial for the nucleation of hemidesmosome assembly, and therefore also for their disassembly (1, 2, 3). Phosphorylation of the $\alpha 6$ subunit appears to be less crucial. In fact, there are two cytoplasmic variants of the $\alpha 6$ subunit, A and B, of which only $\alpha 6A$ is a target for phosphorylation by protein kinase C. In the adult skin, $\alpha 6A$, but not the $\alpha 6B$ subunit is expressed. However, replacement of $\alpha 6A$ by $\alpha 6B$ in mice had no effect on the ability of this integrin to induce the

Hemidesmosomes. Table 1 Involvement of hemidesmosomal components in acquired and genetic skin diseases. CP, cicatricial pemphigoid; DEB, dystrophic epidermolysis bullosa; EBA, epidermolysis bullosa acquisita; EBS, epidermolysis bullosa simplex; GABEB, generalized atrophic benign epidermolysis bullosa; GP, gestational pemphigoid; HN, hereditary nephritis; JEB, junctional epidermolysis bullosa; LAD, linear IgA dermatosis; LPP, lichen planus pemphigoid; MD, muscular dystrophy; PA, pyloric atresia; PEB, pretibial epidermolysis bullosa; PNP, paraneoplastic pemphigus

Morphological structure	Protein	Disease		Affected gene	Animal model
		Autoimmune	Inherited		
Intermediate filaments	Keratin 5 and 14	-	EBS	<i>KRT5</i> , <i>KRT14</i>	Yes
Hemidesmosomes	$\alpha 6$ integrin	-	PA-JEB	<i>ITGA6</i>	Yes
	$\beta 4$ integrin	CP	PA-JEB	<i>ITGB4</i>	Yes
	Plectin	PNP	EBS-MD	<i>PLEC1</i>	Yes
	BP230	BP PNP	-	<i>BPAG1</i>	Yes
	BP180	BP	GABEB	<i>BPAG2</i>	No
		CP			
		GP			
		LAD			
		LPP			
	CD151	-	HN-PEB	<i>CD151</i>	Yes
Basement membrane	Laminin-5	CP	JEB	<i>LAMA3</i> , <i>LAMB3</i> , <i>LAMC2</i>	Yes
Anchoring fibrils	Collagen type VII	EBA	DEB	<i>COL7A1</i>	Yes

formation of hemidesmosomes, suggesting that phosphorylation of the $\alpha 6$ subunit is not involved. Ligation of the integrin to laminin-5 also triggers specific phosphorylation and dephosphorylation events downstream of the receptor, which may have an impact on its ability to initiate hemidesmosome assembly (8). Furthermore, there is evidence suggesting that the composition of the substrate is of key importance in the assembly of hemidesmosomes. Both the $\alpha 3$ and the $\gamma 2$ chains of laminin-5 can be cleaved and, depending on whether laminin-5 in the basement membrane is in its cleaved or non-cleaved form, it either enhances cell motility or induces stable adhesion. A further factor found to play a role in the assembly of hemidesmosomes is the distribution and density of laminin-5. If laminin-5 is deposited into dense clusters, it efficiently induces hemidesmosome assembly. Conversely, if it is evenly distributed, the assembly is less efficient.

The kinases found to regulate hemidesmosome assembly through phosphorylation of the integrin $\beta 4$ subunit include members of the protein kinase C family, which are ser/thr kinases, and the tyrosine kinase Fyn. However, there is some contradiction in the results. For example, in some studies PKC α has been implicated in the assembly of hemidesmosomes, whereas in others PKC α was involved in the disassembly and PKC δ in the assembly of the hemidesmosomal complex. Furthermore, it has been reported that a wide range of [▶receptor tyrosine kinases](#), e.g. the EGF-receptor (erbB-1), erbB-2, Met (receptor for hepatocyte growth factor) and Ron (receptor for macrophage stimulating protein), can associate directly or indirectly with the $\alpha 6\beta 4$ integrin and influence hemidesmosome assembly and disassembly (8). Some of these associations may be cell type specific. For instance, in carcinoma cells $\alpha 6\beta 4$ was found to be associated with Met, whereas in many kinds of normal cells Met and $\beta 4$ are independently expressed. Thus, the association of $\alpha 6\beta 4$ with Met might be more general in transformed cells, for which it is essential to mediate their invasion into surrounding tissues. Nevertheless, extracellular factors, such as the matrix to which the cells are attached and growth factors present in the matrix, are important regulators of hemidesmosome formation and their breakdown. Clearly, further studies need to be carried out to unravel the precise role of signaling via $\alpha 6\beta 4$ in the assembly and disassembly of hemidesmosomes.

References

1. Borradori L, Sonnenberg A (1999) Structure and function of hemidesmosomes: more than simple adhesion complexes. *J Invest Dermatol* 112:411–418
2. Jones JC, Hopkinson SB, Goldfinger LE (1998) Structure and assembly of hemidesmosomes. *Bioessays* 20: 488–494
3. Koster J, Borradori L, Sonnenberg A (2004) Hemidesmosomes: molecular organization and their importance for cell adhesion and disease. *Handbook of Experimental Pharmacology*, vol Cell Adhesion, pp 243–280
4. Hemler ME (2003) Tetraspanin proteins mediate cellular penetration, invasion, and fusion events and define a novel type of membrane microdomain. *Annu Rev Cell Dev Biol* 19:397–422
5. Zillikens D (1999) Acquired skin disease of hemidesmosomes. *J Dermatol Sci* 20:134–154
6. Pulkkinen L, Uitto J (1999) Mutation analysis and molecular genetics of epidermolysis bullosa. *Matrix Biol* 18:29–42
7. Mercurio AM, Rabinovitz L, Shaw LM (2001) The $\alpha 6\beta 4$ integrin and epithelial cell migration. *Curr Opin Cell Biol* 13:541–545
8. Giancotti FG, Tarone G (2003) Positional control of cell fate through joint integrin/receptor protein kinase signaling. *Annu Rev Cell Dev Biol* 19:173–206

Hemizygous

Definition

Hemizygous describes an individual who has only one member of a chromosome pair or chromosome segment rather than the usual two. It refers in particular to X linked genes in males who under usual circumstances have only one X chromosome.

[▶Bone and Cartilage](#)

Hemochromatosis

ERNEST BEUTLER

The Scripps Research Institute, Department of Molecular and Experimental Medicine, La Jolla, CA, USA

beutler@scripps.edu

Synonyms

Bronzed diabetes; iron storage disease

Definition

Hemochromatosis is a metabolic disorder in which excessive amounts of iron accumulate in the body, leading to tissue damage. It is to be distinguished from hemosiderosis in which excess iron may be present in

tissues, but without tissue damage. Hemochromatosis occurs in hereditary (or primary) and in secondary forms. The most common primary form of the disease is designated hereditary hemochromatosis, once termed idiopathic hemochromatosis. It is the adult onset form of the disease, usually diagnosed after the age of forty. Although it can be a serious disorder characterized by cirrhosis, diabetes, and cardiac disease, in most cases there are no clinical manifestations. Juvenile hemochromatosis is an early onset form of the disease with a different genetic basis and usually severe clinical manifestations. Neonatal hemochromatosis occurs at birth, but it is not clear that the iron deposition that occurs in this poorly defined disorder plays a primary role in pathogenesis. It is probably a heterogeneous disorder, not actually a form of hemochromatosis, and will not be discussed further here.

Excess iron accumulation may be secondary to a variety of disorders of ▶[hematopoiesis](#). This type of hemochromatosis is usually designated secondary hemochromatosis. The most common cause is β -thalassemia major. It also occurs in ▶[congenital dyserythropoietic anemias](#), ▶[hereditary sideroblastic anemia](#), hemolytic anemia due to ▶[pyruvate kinase deficiency](#) and other erythroenzymopathies and in patients with myelodysplastic syndromes such as acquired sideroblastic anemia. It is particularly likely to occur in patients who have ▶[ineffective erythropoiesis](#) and who receive many blood transfusions.

Characteristics

Fully developed hemochromatosis is characterized by the accumulation of large amounts of iron in the form of ▶[ferritin](#) and hemosiderin in many organs. Iron storage is particularly prominent in the liver, where the parenchymal cells become heavily iron loaded, but the macrophages (▶[Kupffer cells](#)) are actually iron-poor. In the liver the reaction to excess iron is fibrosis, occasionally with progression to frank cirrhosis. Liver damage can be very severe and is a common cause of death in the disease. But tissue damage is not limited to the liver alone. Compromise of pancreatic function leads to diabetes mellitus. Deposition of iron in joints is thought to result in arthritis and in the heart in cardiomyopathies. Darkening of the skin is the consequence of both the presence of hemosiderin in the skin, and interference by iron with melanin metabolism.

The characteristics of milder forms of hemochromatosis are by no means so clear, however. It has been suggested that symptoms such as fatigue, impotence and heart disease may occur in such patients. However, controlled studies have shown no increase in such putative manifestations in persons with the biochemical phenotype (1). It has even been proposed that the

heterozygous state is associated with an increased incidence of cancer and/or heart disease, but evidence for this concept has been, at best, contradictory.

Cellular and Molecular Regulation

Iron enters most cells through what has been termed the ▶[transferrin](#) cycle. Transferrin is a plasma protein that binds iron with high affinity at neutral or alkaline pH levels, but readily releases it when the pH is lowered. Iron-containing transferrin binds to a ▶[transferrin receptor](#) on the plasma membrane. The transferrin receptor and transferrin are internalized and transported to the interior of the cell as a vacuole in which the pH is lowered to a point at which the iron dissociates from transferrin and can be transported through the vesicle wall into the cytoplasm. The iron-free transferrin (apotransferrin) is then carried back to the membrane, where it is ready to bind another molecule of transferrin. Although most iron is delivered to cells in this manner, iron enters macrophages in the form of phagocytosed erythrocytes.

Iron regulates the expression of several proteins involved in its metabolism, and the best understood mechanism is a post-transcriptional one (2). Either the 5' or the 3' untranslated portion of the mRNA contains a loop structure that is known as an iron responsive element (IRE). This structure can bind an iron regulatory protein (IRP). There are two types of IRPs, IRP-1 and IRP-2. IRP-1 has been found to be cytoplasmic aconitase. It has an iron-sulfur cluster, and the saturation of this cluster with iron affects binding to the IRE. IRP-2 is regulated by iron-induced ▶[ubiquitination](#) and degradation, and its availability as a ligand for the IRE is therefore a function of iron concentration. The binding of an IRP to a 5' IRE serves to inhibit translation; binding of an IRP to a 3' IRE stabilizes the mRNA.

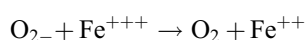
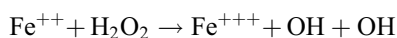
Although there is considerable understanding of intracellular trafficking and regulation of iron, the regulation of the total iron content is less well understood. Since there is very little loss of iron from the body, regulation of body iron content depends upon increasing or decreasing the amount of iron absorbed from the gastrointestinal tract. Much of what we know about this process is derived from mutations that cause hemochromatosis in humans and genes, targeted disruption of which causes excess iron storage in mice. The most powerful regulator of iron absorption seems to be a small peptide that has been designated hepcidin. It is up-regulated when iron stores are increased and down-regulated in iron deficiency. A lack of hepcidin, both in humans and in mice, causes rapid accumulation of iron (3). The receptor for hepcidin is the transport protein ferroportin. Another protein that influences body iron content is Hfe. It is the product of a gene

designated ►*HFE* that resides on chromosome 6 of man. While it is known that Hfe interacts with the transferrin receptor, it is not clear how this causes increased iron absorption from the intestine. However, in the absence of Hfe, increased body iron fails to stimulate production of hepcidin. This could indicate that Hfe is involved, directly or indirectly, in the regulation of hepcidin levels, and that its effect on iron homeostasis is the result of its regulation of hepcidin.

Clinical Relevance

General Considerations

Under normal circumstances most iron in the body is firmly bound to proteins, particularly transferrin in the plasma and ferritin within cells. However, as the iron burden increases it is likely that the amount of free iron also increases, and this is generally believed to result in tissue damage, particularly to the liver, heart, and endocrine organs through the Haber-Weiss reaction:



Hereditary Hemochromatosis

The most common mutation of the *HFE* gene is 187 G→C predicting a H63D substitution. This mutation is relatively mild with respect to its effect on indicators of iron homeostasis. Its distribution is worldwide, with the highest gene frequencies of approximately 0.16 found in European populations. A second mutation, 845 G→A resulting in the C282Y substitution is confined almost entirely to the European population and generally has a gene frequency of approximately 0.07 in northern Europe (4). This mutation has a stronger phenotypic effect than does the H63D mutation. Most homozygotes for the C282Y mutation and compound heterozygotes for the C282Y and H63D mutation manifest increased saturation of transferrin with iron and usually also have increased ferritin levels during adult life. Although the penetrance of these genotypes is quite high with respect to biochemical parameters, it is very low with respect to clinical manifestations. Indeed, in large controlled population studies it has been impossible to show an effect on either general health or on longevity. There is, however, an increased frequency of abnormal liver function tests in homozygotes for the C282Y mutation. Nonetheless, there is a small subset of patients who develop severe liver disease. The other factors that cause these patients to develop disease remain undefined, although environmental factors such as alcohol intake and hepatitis infections, and probably some genetic factors presumably play a role.

Mutations of transferrin receptor-2 are rare causes of the same clinical syndrome that is caused by *HFE* mutations (5).

Juvenile Hemochromatosis

Juvenile hemochromatosis is a much more severe form of iron storage disease, with an earlier onset than the type of hereditary hemochromatosis that is a result of the *HFE* mutations. Cardiomyopathies and endocrine disturbances are the most prominent clinical features, although cirrhosis of the liver also occurs. The disease locus has been mapped, in most patients, to 1q. The gene that causes the disease has been cloned and is designated *HJV*. In a few families in which the mutation does not map to 1q it has been identified as being a mutation in the hepcidin gene (6).

Iron Overload in Patients with Ferroportin Mutations

Excess iron storage in macrophages has been encountered in a number of families with mutations of ferroportin, an iron transport protein. Unlike the other inherited forms of iron storage disease, transmission is as an autosomal dominant disorder. Several different mutations have been observed, but a recurrent mutation appears to be the deletion of one of a series of 3 sequential valines at amino acid positions 160-162 (7). Mild anemia has been reported in some patients with this form of iron storage disease.

African Iron Overload

This disorder is much less well defined than are hereditary hemochromatosis and juvenile hemochromatosis. Among Africans in South Africa iron overload is extremely common. It has been associated with the ingestion of a large amount of a local beer that has a very high iron content because it is brewed in iron pots. However, familial aggregation of elevated ferritin levels has been documented, affecting even family members who do not ingest the beer. Moreover, African-Americans also have increased serum ferritin levels when compared with whites.

Secondary Hemochromatosis

Accumulation of iron in children with β -thalassemia major is the result of transfusions and of the increased iron absorption that is characteristic of patients with ineffective erythropoiesis. It is the most common cause of death in this disorder, the increased iron burden causing severe cardiomyopathies and cirrhosis of the liver. The same syndrome occurs when excessive iron accumulates in patients with other disorders characterized by anemia, transfusion, and ineffective erythropoiesis (see Definition).

Treatment

Treatment of iron overload is directed at mobilizing the stored iron and removing it from the body. This is best accomplished in patients who have intact erythropoiesis by multiple phlebotomies. As iron-containing blood is removed, the body mobilizes stored iron, which is then incorporated into newly formed erythrocytes in the marrow. This approach cannot be used in patients who have abnormal erythropoiesis, and in such patients the stored iron must be removed by the use of chelating agents, such as ►desferrioxamine.

References

1. Beutler E, Felitti VJ, Koziol JA et al (2002) Penetrance of the 845G→A (C282Y) HFE hereditary haemochromatosis mutation in the USA. *Lancet* 359:211–218
2. Rouault TA (2002) Post-transcriptional regulation of human iron metabolism by iron regulatory proteins. *Blood Cells Mol Dis* 29:309–314
3. Ganz T (2003) Heparin, a key regulator of iron metabolism and mediator of anemia of inflammation. *Blood*
4. Beutler E, Bothwell TH, Charlton RW et al (2001) Hereditary Hemochromatosis. In: Scriver CR, Beaudet AL, Sly WS (eds) *The Metabolic and Molecular Basis of Inherited Disease*. McGraw-Hill, New York, p 3127–3161
5. Camaschella C, Roetto A, Cali A et al (2000) The gene TFR2 is mutated in a new type of haemochromatosis mapping to 7q22. *Nat Genet* 25:14–15
6. Roetto A, Papanikolaou G, Politou M et al (2003) Mutant antimicrobial peptide hepcidin is associated with severe juvenile hemochromatosis. *Nat Genet* 33:21–22
7. Njajou OT, Vaessen N, Joosse M et al (2001) A mutation in SLC11A3 is associated with autosomal dominant hemochromatosis. *Nat Genet* 28:213–214

Hemophilia A

Definition

Hemophilia A is a hereditary X-linked bleeding disorder resulting from a congenital deficiency of plasma coagulation factor VIII (antihemophilic factor), the product of the F8 gene on chromosome Xq28.

►Hereditary Hemostatic Defects and Recombinant Proteins for Treatment

►Methylation of Proteins

Hemophilia B

Definition

Hemophilia B is an inherited hemostatic defect that is caused by low concentration of coagulation factor IX.

►Hereditary Hemostatic Defects and Recombinant Proteins for Treatment

Heparin Sulfate Proteoglycans

Definition

Heparin sulfate proteoglycans (HSPGs) belong to the family of proteoglycans. They are cell surface molecules that are part of the extracellular matrix. HSPGs consist of a protein core, such as syndecan and glypican, to which long unbranched chains of sulphated repeating disaccharides are attached. Dally is the *Drosophila* proteoglycan while in vertebrates Knypek and Syndecan-1 typify this family. HSPGs have been implicated in regulating the signaling activities and distribution of both Wnt and Hedgehog.

►Wnt/Beta-Catenin Signaling Pathway

►Sonic Hedgehog

Hepatitis B Virus

Definition

Hepatitis B virus belongs to the hepadna virus family. The virus infects and replicates in liver, where it can induce chronic disease. Hepatocellular carcinoma is the frequent consequence of chronic hepatitis B virus infection. The enveloped virus particles contain small circular DNA genomes with four major genes. At least one of its proteins, the X protein, has transforming features. Broad vaccination against Hepatitis B virus has significantly lowered the prevalence of Hepatitis B. This provides the first example of a highly efficient prophylaxis against virus associated tumour forms in the general population.

►Viral Oncogenesis

Hepatocellular Carcinoma

Hepatocellular carcinoma (HCC) refers to liver cancer as a frequent consequence of chronic hepatitis B virus infection.

►Wnt/Beta-Catenin Signaling Pathway

Hepatocyte Nuclear Factor 4a

Definition

Hepatocyte nuclear factor 4a is a member of the hormone receptor superfamily of transcription factors. Mutations in the HNF4A gene cause the MODY1 form of diabetes.

► [Diabetes Mellitus, Genetics](#)

HER1

Definition

HER1 is a transmembrane glycoprotein receptor of 185 kD (also called ErbB2/neu), which belongs to the epidermal growth factor receptor (EGFR) family of receptors, containing an intramolecular tyrosine kinase domain. HER receptors have been widely acknowledged as playing a crucial role in tumorigenesis and tumor progression.

► [Breast Cancer](#)

Hereditary Cancer

Definition

Hereditary cancer defines a cancer that is caused by the inheritance of a mutant allele, most commonly a mutation in a tumor suppressor gene.

► [Tumor Suppressor Genes](#)

Hereditary Coproporphyrria

Definition

Hereditary coproporphyrria is an acute hepatic porphyria that results from defects in the enzyme coproporphyrinogen oxidase. Coproporphyrria is inherited as autosomal dominant trait. Disease symptoms are abdominal pain, neuropathies, constipation, and skin changes.

► [Acute Intermittent Porphyrria](#)

Hereditary Disease, Genetic Basis

JÖRG SCHMIDTKE

Institute of Human Genetics, Hannover Medical School, Hannover, Germany

Schmidtke.Joerg@mh-hannover.de

Definition

Hereditary disease is due to mutations of the genetic material of germ cells. Such mutations can occur at the level of a chromosome (chromosomal mutations), the genome (genome mutation), or a gene (gene mutations). Mutations alter either the structure or the dosage of one or several genes, the functional units of inheritance. Most, but not all, hereditary diseases are individually rare (1), but taken together they comprise about 5% of the disease load in developed countries.

Characteristics

Chromosomal Disorders

The normal human chromosome complement of somatic cells consists of 46 chromosomes. Every individual inherits one half set of 23 chromosomes from each parent, and thus carries two functional homologues of each gene, with the exception of those genes that are located on the sex chromosomes, X and Y. Males inherit a 23,Y complement from their fathers and a 23,X complement from their mothers (male karyotype: 46,XY), while females receive 23,X complements from both parents (female karyotype 46,XX). The 22 chromosomes shared by male and female germ cells are called autosomes. Most, but not all genes of one of the two X chromosomes are functionally silenced (inactivated) in a random fashion in female somatic cells.

Chromosomes can be mutated either with respect to their number [numerical aberrations (“aneuploidies”)] or their structure (structural aberrations). Numerical aberrations are the result of misdivision of the chromosome complement during parental meiosis. Numerical aberrations that can be found in live-born individuals comprise trisomies 13, 18, and 21, as well as sex chromosome aneuploidies including 45,X (Turner syndrome), 47,XXX (Triple-X syndrome), XXY (► [Klinefelter syndrome](#)), and 47,XYY. Occasionally, numerical aberrations occur in a mosaic fashion, i.e. only a proportion of tissues or cells within a tissue carry the abnormal chromosome complement; such changes are due to postzygotic (somatic) mutations within the affected individual. Chromosomal ► [mosaicism](#) is usually associated with milder disease manifestation than non-mosaic cases. Autosomal and

X-chromosomal ►aneuploidies with added chromosomes in newborns are associated with increased maternal age. Numerical chromosomal aberrations are a frequent cause of spontaneous abortions (trisomy 16, 45,X-karyotype). The same is true for genome mutations (triploidy, tetraploidy), which are extremely rare among new-borns.

Structural ►chromosomal aberrations are due to rearrangement of chromosomal material. Inversions are the result of chromosomal breakage at two locations and re-insertion of the central fragment in the opposite orientation; translocations are the result of interchange of fragments between non-homologous chromosomes; loss of fragments of chromosomes results in ►deletions; extra fragments are called duplications (sometimes called partial trisomy). Apart from cases where one of the chromosomal breakpoints is intragenic, structural aberrations are associated with a pathological state only if they are unbalanced with regard to the amount of genetic material. Infertility with recurrent abortions is sometimes due to a balanced parental translocation that has segregated into an unbalanced state in the fetus. Male sterility is sometimes caused by deletions of Y-chromosomal material, involving genes required for normal spermatogenesis.

Single-Gene Disorders

About 4000 different disorders without visible chromosomal alterations have been shown to follow Mendelian rules (at least in some families) and are thus assumed to be caused by mutations of single genes. The great majority of single-gene disorders is encoded by nuclear genes. A small number of heritable diseases is due to mutations in the mitochondrial genome. The on-line catalogue of Mendelian inheritance in man (2) contains comprehensive clinical and genetic information, as well as useful links to other resources regarding these disorders. In mid-2003, disease-causing mutations were known for 1446 nuclear genes, and 1933 clinical phenotypes were associated with a mutation in a nuclear gene (3). However, because of locus and allelic heterogeneity, molecular diagnosis is usually not possible in all patients affected by a given disease.

At the gene level, structural alterations comprise nucleotide exchanges, deletions or insertions of single or multiple nucleotides, and more complex rearrangements. A mutation can be pathogenic if it results in a structurally and hence functionally altered protein product, or if it leads to altered, usually reduced dosage of gene products (►haploinsufficiency).

The inheritance pattern of genetic disease depends on the chromosomal location (autosomal or sex-chromosome linked) and on the phenotypic result of the mutation. If a disease fully manifests itself in the

►heterozygous carrier of the disease-causing mutation, the inheritance pattern of the disease is called dominant. If a disease manifests itself only in the homozygous or ►compound heterozygous carriers of the disease-causing mutation, the inheritance pattern of the disease is called recessive. If the phenotype of the heterozygote is intermediate between the homozygous affected and non-affected individuals, the pattern of inheritance is called codominant (or intermediary). If the disease-causing mutation is in an autosomal gene, individuals are affected independent of gender. If a disease causing mutation is on the X-chromosome and the mode of inheritance is recessive, only males will be affected. If a disease-causing mutation is on the Y chromosome, the mode of inheritance is dominant and the disease restricted to males. The only known disease that is Y-linked is male sub- or infertility, the latter being heritable only through assisted reproduction.

In practical clinical genetics, the terms dominant and recessive are used in a pragmatic fashion and are associated with both the mode of inheritance and the disease itself. The vast majority of dominant disorders are strictly speaking co-dominant in that homozygosity for the underlying mutations would result in a much severer phenotype including early lethality. Heterozygous mutation carriers for a recessive mutation often show mild phenotypic expression (heterozygote manifestation); in the special case of X-linked recessive disease, the phenomenon can be ascribed to deviation from randomness regarding X-chromosome inactivation.

It should also be noted that some disorders following a dominant pattern of inheritance in pedigrees, require ►homozygosity (or usually compound-heterozygosity) at the cellular level. Thus, most familial cancer syndromes follow an autosomal-dominant pattern, because a mutation in a tumor-suppressor gene is transmitted through generations. In order for the disorder to become manifest, a second mutation must occur in one or more somatic cells of the germ-line mutation carrier.

Contiguous Gene Syndromes

Disorders that are caused by a mutation affecting a contiguous set of genes, usually a deletion, are referred to as contiguous gene syndromes. These mutations are not normally visible using classical cytogenetic methods, and their detection requires a combination of cytogenetic and molecular methodology, ►FISH (fluorescence-in situ-hybridization). They usually arise *de novo* and act through haploinsufficiency.

Genetic Risk

In disorders following a Mendelian mode of inheritance, the expected genetic risk to members of affected

families can easily be calculated. In autosomal-recessive disorders, the disease risk is 1 in 4 for sibs of affected individuals; 2 out of 3 healthy sibs will be heterozygous carriers. In autosomal-dominant disorders, the risk is 1 in 2 for offspring of affected individuals. In X-linked recessive disorders, the disease risk is 1 in 2 for male offspring of heterozygous female carriers. All daughters of affected males and 1 in 2 daughters of female heterozygous carriers will be heterozygous carriers; sons of affected males will not be affected. It should be noted that absence of a family history in the case of a hereditary disease can either reflect recessivity or can be ascribed to a *de novo* mutation with a dominant effect.

However, a number of factors lead to deviations from calculated risk figures regarding disease manifestation.

- **germ line mosaicism**
Occasionally, a person who does not carry a disease-causing mutation in his or her somatic cells has multiple affected offspring. Given that repeated *de novo* mutations in the same gene are extremely unlikely, the phenomenon is ascribable to some or all germ-line cells of this person carrying the mutation.
- **late or sex-skewed disease manifestation**
A number of single-gene disorders become manifest in middle or late adult life, and the time of onset may vary greatly even within families. Thus a disease may be inherited from an ancestor who did not reach the age of disease manifestation him- or herself (phenotypic “skipping” of generations). Some mutations, such as those within genes predisposing to familial breast cancer, rarely manifest themselves in males but frequently in daughters of male carriers.
- **reduced genetic penetrance**
Genetic penetrance is the proportion of persons manifesting a disease among those that carry the responsible genotype. In single-gene disorders, full penetrance is indeed the exception rather than the rule, whereas chromosomal disorders and contiguous gene syndromes are usually fully penetrant. Which factors cause reduced penetrance is usually not known. Both other genetic (presence or absence of mutations in other genes contributing to the pathogenesis or of protective variants) and environmental factors (presence or absence of precipitating or protective agents) could be involved. Thus, the notion of penetrance mediates a transition to the concept of polygenic and multifactorial inheritance typical for “complex” diseases. Late and sex-skewed

disease manifestation can also be regarded as the result of age- or sex-dependent penetrance.

- **variable expressivity**
Variable expressivity is the term used to describe the situation when a disease, even within a family, differs greatly with respect to severity. The contribution of other genes or environmental factors, as well as allelic heterogeneity may be responsible for this common phenomenon.
- **locus heterogeneity**
This term is used when a disease may be caused by mutations in different genes. In the case of autosomal-recessive disorders, this could lead to none of the offspring of affected parents developing the disease.
- **genomic imprinting**
If the phenotype depends on the gender of the transmitting parent, genomic imprinting may be invoked. Here, ►**epigenetic** chemical change (methylation) of genomic DNA during the paternal or maternal gametogenesis leads to inactivation of specific parental alleles.
- **uniparental disomy**
Rarely, both members of a chromosome pair are of identical parental origin. This can be the result of the attempt of the zygote to repair a trisomy that was caused by non-disjunction during gametogenesis. Deleterious effects may arise in conjunction with genomic imprinting or recessivity.
- **anticipation**
Some diseases occur at progressively earlier age and/or with pronounced severity in successive generations. Unless due to ascertainment bias, the phenomenon can be ascribed to dynamic repeat expansion mutations.
- **mitochondrial inheritance**
A small number of diseases are caused by mutations in one of the 37 genes of the mitochondrial genome. Due to the fact that the mitochondrial material of the zygote is (almost) entirely derived from the egg, mitochondrial disorders are inherited in a strictly maternal fashion. Variable expressivity of mitochondrial disease can be ascribed to ►**heteroplasmy**.

References

1. ORPHANET: www.orpha.net
2. OMIM: www3.ncbi.nlm.nih.gov/Omim
3. Stenson, PD, Ball, EV, Mort, M et al (2003) Human Gene Mutation Database (HGMD): 2003 update. *Hum Mutat* 21:577–581

Hereditary Hemostatic Defects and Recombinant Proteins for Treatment

ELISABETH ERHARDTSEN

Novo Nordisk A/S, Bagsværd, Denmark
eer@novonordisk.com

Definition

Hereditary hemostatic defects caused by human gene mutations can be divided into the more commonly occurring coagulation factors defects (e.g. ►[hemophilia A](#) and [B](#)), the rare defects responsible for platelet adhesion and aggregation defects (e.g. ►[Glanzmann's thrombasthenia](#) and ►[Bernard-Soulier syndrome](#)) and defects within ►[fibrinolysis](#). This essay will cover the most frequently reported coagulation factor defects and the recombinant proteins for their treatment.

Hemophilia A and B, caused by deficiencies of coagulation factors VIII (FVIII) or IX (FIX), are the best described hemostatic defects of high prevalence, causing lifelong bleeding disorder in mild, moderate or severe forms. Von Willebrand factor deficiency leads to ►[von Willebrand's disease](#) with an estimated prevalence of 1% of the general population. A range of more rare inherited coagulation factor deficiencies are reported, such as deficiency of coagulation factors V, VII, X, XI, or XIII.

The hereditary coagulation factor deficiencies are either ►[X-linked](#) ►[recessively](#) or ►[autosomal](#) recessively transmitted gene defects.

Characteristics

X-linked Gene Mutations

Coagulation FVIII Deficiency (Hemophilia A)

Hemophilia A caused by the deficiency of coagulation FVIII has a prevalence of 1:5,000 males. The bleeding symptoms depend upon the FVIII concentration in the patient's blood, and may be characterized by spontaneous joint and muscle bleeds in severe hemophilia (FVIII<1%) and excessive bleeding after surgery or trauma in mild, moderate and severe hemophilia.

Coagulation FIX Deficiency (►Hemophilia B)

Hemophilia B is caused by coagulation FIX deficiency, with a prevalence of 1:30,000 males. The bleeding symptoms depend upon the FIX concentration in the patient's blood with a pattern similar to that affecting hemophilia A patients.

Table 1 shows the laboratory and clinical manifestations of hemophilias A and B.

Bleeding Pattern in Hemophilia

A characteristic of hemophilia is the frequent and painful joint bleedings. Persistent or recurrent joint bleeding may result in the development of chronic arthropathy and incapacity with fixed flexion and other deformities particularly of the knees, ankles and elbows. Associated muscle wasting can be profound and in the worst scenario the patient becomes confined to a wheelchair.

Muscle bleeds may be life- or limb-threatening due to compression of blood vessels and nerves. Before effective therapy became available, a cut in the finger could result in death. Today, any bleeding episode in a severe hemophilia patient is treated promptly.

Autosomal Recessive Gene Mutations

Von Willebrand's Disease (vWD)

vWD is caused by a deficiency or abnormality of ►[von Willebrand Factor](#) (vWF) which is a multimeric adhesive protein playing an important role partly in the initial hemostasis by promoting platelet adhesion at site of vascular injury and partly as the carrier of coagulation FVIII to which it is associated in plasma as a tight complex. A clear distinction between FVIII and vWF deficiencies was not made until 1957.

vWD has a prevalence of up to 1% in the general population; however, only about 1/10,000 inhabitants receive treatment for the disorder. More than 20 subtypes of vWD are described with a broad division into two types, i.e. associated with quantitative deficiency (types 1 and 3) and qualitative deficiency (type 2). Diagnosis of vWD, especially type 1 accounting for ~70% of all vWD cases, may be difficult because the laboratory phenotype is highly ►[heterogeneous](#) and factors such as e.g. blood group influence the vWF level. It has been established that type 1 vWD represents the ►[heterozygous](#) or carrier state for type 3 vWD.

Bleeding Pattern

Mucosal bleeding is predominant in vWD, whereas hematomas and muscle bleeds are seen only in type 3 vWD.

Coagulation Factor VII Deficiency

Factor VII (FVII) deficiency is a common autosomal recessive coagulation disorder with a prevalence of 1:500,000 in the general population; a severe bleeding phenotype is only seen in patients with FVII activities <2% of the normal level but the severity of symptoms is reported to be poorly correlated with plasma levels. The majority of FVII-deficient patients are either asymptomatic or the clinical phenotype is unknown and the mutations are often identified through pre-operative screening.

Hereditary Hemostatic Defects and Recombinant Proteins for Treatment. Table 1 Laboratory and clinical manifestations of hemophilia A and B

Factor VIII/IX Activity Level			
Normal	Severe	Moderate	Mild
(0.5 – 1.5 U/mL)	(<0.01 U/mL)	(0.01 – 0.05 U/mL)	(>0.05 U/mL)
% of all hemophilia A	70%	15%	15%
% of all hemophilia B	50%	30%	20%
Bleeding Manifestations			
Age of onset	≤1 year	1-2 years	2 years – adult
Neonatal symptoms	PCB: common ICH: occasionally	PCB: occasionally ICH: uncommon	None Rare
Muscle/joint hemorrhage	"Spontaneous"; requires no trauma	Requires minor trauma	Requires major trauma
CNS hemorrhage	High risk (2-5%)	Moderate risk	Rare
Post surgical hemorrhage (without prophylaxis)	Frank bleeding, severe	Wound bleeding common	Wound bleeding with factor <0.3 U/ml
Oral hemorrhage following trauma, tooth extraction	Common	Common	Often

PCB: postcircumcisional bleeding; ICH, intracranial hemorrhage

Bleeding Pattern

Epistaxis and mucocutaneous bleedings are the most frequently observed bleedings in FVII-deficiency.

Coagulation Factor XI Deficiency

Factor XI (FXI) deficiency is an injury-related autosomal bleeding disorder with a frequency of up to 8% of Ashkenazi Jews. It is characterized by a decrease in FXI activity usually accompanied by correspondingly low levels of FXI antigen; however, a clear relationship between FXI plasma level and bleeding tendency has not been demonstrated.

Bleeding Pattern

Patients who are mildly affected with FXI deficiency may have oral and postoperative bleeding, while severe FXI deficiency (<1%) is characterized by bleeding symptoms after trauma or surgery.

Inhibitory Antibodies

Spontaneous inhibitors may develop in patients lacking various coagulation proteins as a response to treatment with blood components containing the missing factor. Such inhibitors are most often directed against FVIII; however, any coagulation protein may be affected. So far, no clinical consequences of inhibitory antibodies have been reported in FVII-deficient patients, whereas inhibitors to FXI and von Willebrand patients have been reported to have clinical consequences.

Development of inhibitory antibodies to treatment occurs mainly in severe hemophilia, with a rate of 30–50% in hemophilia A and 1.5–3% in hemophilia B patients. The risk is associated with the type of mutation: mutations that severely truncate or prevent production of FVIII lead to a frequency of 35% of inhibitor development, while a missense mutation or small deletions lead to about 5% inhibitor development in hemophilia A patients. Hemophilia B patients with large gene deletions (1–3% of patients) or rearrangements have a risk of about 50% and are at particular risk of developing anaphylaxis when given therapy with the missing factor. Frameshift, premature stop, or splice-site mutations lead to a risk of about 20% of inhibitor development.

Cellular and Molecular Regulation Factor VIII

The mutation usually occurs in males in whom the mutant gene (X_H) is inherited from the carrier mother (X_H/X). However, in approximately 30% of cases the mutation occurs spontaneously. The gene measures 186 kb and contains 26 exons, with a large inversion and translocation of exons 1–22 as the most common genetic defect.

Further information on mutations, phenotype data etc. is available on the Internet ► <http://europium.csc.mrc.ac.uk>.

Factor IX

The FIX gene is located at band Xq27 on the X-chromosome, contains eight exons and measures 33.5 kb. More than 2100 mutations are assembled in a worldwide database of mutations for FIX gene causing hemophilia B that is accessible through the Internet ►<http://www.kcl.ac.uk/ip/petergreen/haemB-database.html>).

von Willebrand Factor

The vWF cDNA was cloned in 1985; the gene spans 178 kb on the short arm of chromosome 12 and contains 52 exons. Different mutations, missense, nonsense or frameshift, will result in loss of or unstable expression of vWF with missense mutations being the basis for most qualitative vWD variants.

Factor VII

The FVII gene is small (12 kb, nine exons) and found on chromosome 13 adjacent to the FX gene. A mutation database can be found on the Internet at ►<http://europium.csc.mrc.ac.uk> with data from 238 individuals described in the world literature.

Factor XI

The gene is found on chromosome 4 and is 23 kb in length. Two mutations in the gene encoding FXI are responsible for most cases of FXI deficiency, a nonsense mutation Glu117Stop and a missense mutation Phe283Leu.

Clinical Relevance

The progress made in understanding the mechanisms of coagulation defects over the past three decades has provided a basis for development of recombinant products in the treatment of patients suffering from the coagulation deficiencies with a great impact on their quality of life.

Recombinant Proteins for Treatment

The genes for FIX and FVIII were cloned in 1982 and 1984 resulting in molecular characterization of the defects. The development of recombinant DNA technologies soon led to the production of recombinant proteins for treatment of hemophilia. The alternatives to plasma-derived products have reduced the risk of transmission of blood-borne pathogens; some countries prioritize treatment with recombinant products to patients with newly diagnosed hemophilia.

Factor VIII

The first rFVIII preparations were introduced in the early 1990s with an excellent efficacy and high correlation between the dose given and the FVIII plasma level. Recently, a new rFVIII preparation has been licensed, which is a sucrose-formulated,

full-length rFVIII produced by baby hamster kidney cells into which the human FVIII gene has been introduced. Compared to its predecessor product, it incorporates a revised purification and formulation process that eliminates the addition of human albumin. Future strategies are aimed at producing FVIII molecules with an enhanced expression, stability, prolonged half-life and reduced antigenicity and immunogenicity.

Factor IX

Only one rFIX product is licensed for treatment of patients with hemophilia B. This rFIX product is expressed in Chinese hamster ovary cells cultured in a medium without any blood or plasma products. It has shown satisfactory results although the plasma level obtained is lower than that found after treatment with conventional plasma-derived FIX concentrates. A wide patient-related variability in the recovery rate is found which seems to impede the development of next generation rFIX products.

Treatment

The aim of the treatment is to raise the plasma level of FVIII or FIX to levels that cause arrest of bleeding and prevent rebleeding. Treatment is done on demand as bleeding occurs or as ►[prophylaxis](#). The aim of prophylaxis is to abolish spontaneous joint bleeding and the long-term consequences such as wheel-chair confinement; it is now the recommended strategy for children with severe hemophilia in many countries even though it may be demanding for the families and many questions remain to be answered, e.g. at what age should prophylaxis begin.

Patients with a moderate bleeding pattern are either treated before major physical activities or when a bleeding starts.

Immune Tolerance

Inhibitory antibodies in hemophilia patients are a frequent and significant complication and management is difficult. High-responding inhibitors to FVIII or FIX may be eradicated by immune tolerance regimens, i.e. regular infusions with either plasma factor concentrates or recombinant factor products. Several protocols have been established for immune tolerance regimens, mainly with the use of very high daily doses of factor VIII or FIX in combination with immunosuppressive therapy.

Recombinant Activated Coagulation Factor VII (rFVIIa)

Recombinant activated factor VII is a bypassing product licensed in 1996. It has been shown to induce hemostasis in hemophilia patients with inhibitors against FVIII or FIX independent of FVIII/FIX. FVIIa binds to tissue factor (TF) exposed at the site of injury

and generates through FX activation of the TF-bearing cells a limited amount of thrombin, enough to activate FVIII, FV and FXI as well as platelets. The local thrombin activated platelets provide a perfect template for binding of activated FVIII, FIX, FV, further activation of FX and thrombin generation. rFVIIa in high concentrations binds to thrombin activated platelets and is capable of activating FX, thereby generating thrombin independently of the presence of FVIII or FIX.

Treatment

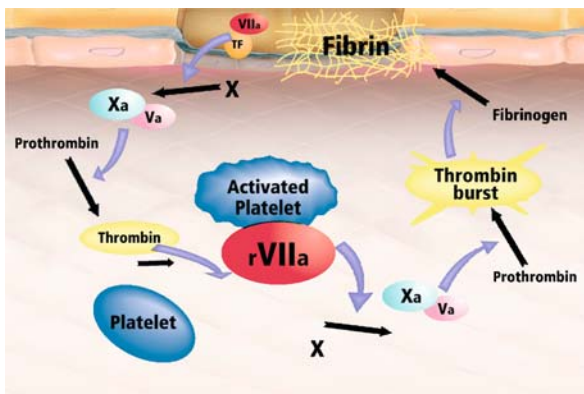
rFVIIa has been shown to have a 90% efficacy in severe hemophilia patients with inhibitors subjected to major surgery and suffering from serious limb- or life-threatening bleedings. Since rFVIIa enhances thrombin generation thereby providing the formation of tight, stable fibrin hemostatic plugs resistant to premature lysis, it should be hemostatic also in other situations characterized by an impaired thrombin generation. A hemostatic effect has accordingly been reported in patients suffering from various platelet disorders, as well as FVII- and FXI-deficiencies. Fig. 1 shows the mechanism of action of rFVIIa.

Recombinant von Willebrand Factor

A recombinant vWF preparation was developed in the mid 1990s; however, due to inconsistent pre-clinical results this program was discontinued. A combined rFVIII/vWF product is now under development.

Treatment

At present, most cases of type 1 vWD are treated with desmopressin (DDAVP) to correct the FVIII/vWF levels; for severe forms of vWD a plasma FVIII/vWF concentrate is needed.



Hereditary Hemostatic Defects and Recombinant Proteins for Treatment. Figure 1 Mechanism of Action of rFVIIa (adapted from Hoffman and Munroe). rFVIIa directly activates Fx on the surface of the locally activated platelets.

Carrier or Fetus Screening

Screening of FVIII and FIX genes for point mutations has been introduced. It is more feasible for the FIX gene than for the FVIII gene due to the difference in length of the genes (33.5 vs 186 kb).

A substantial proportion of female carriers have low concentrations of FVIII or FIX, which may equal a mild hemophilia. Factor concentrations should therefore be measured in girls and women who are definite or possible carriers; however, a reproducibly normal concentration does not exclude carrier status, which can then be identified only by mutation detection.

Gene Therapy

Hemophilia is a prime target for gene therapy for several reasons. The patient lacks only a single gene product and a small rise in coagulation factor concentration is considered to be sufficient to obtain the desired effect. Based on promising animal results, different gene therapy systems have been or are being tested in both hemophilia A and B patients and more than 25 hemophiliacs have by now been treated in phase I clinical trials. However, reliable therapeutic concentrations of FVIII or FIX have not yet been obtained. Further, the safety profile has not yet been established and concern is expressed as to the possible risk of developing inhibitory antibodies along with possible hepatotoxicity and germline gene transfer.

References

1. Bonton-Maggs PHB, Pasi KJ (2003) Haemophilias A and B. *Lancet* 361:1801–1809
2. Castaman G, Federici AB, Rodeghiero F et al (2003) von Willebrand's disease in the year 2003: Towards the complete identification of gene defects for correct diagnosis and treatment. *Haematologica* 88:94–108
3. Peyvandi F, Duga S, Akhavan S et al (2002) Rare coagulation deficiencies. *Haemophilia* 8:308–321
4. McVey JH, Boswell E, Mumford AD et al (2001) Factor VII Deficiency and the FVII Mutation Database. *Human Mutation* 17:3–17
5. Salomon O, Zivelin A, Livnat T et al (2003) Prevalence, causes, and characterization of factor XI inhibitors in patients with inherited factor XI deficiency. *Blood* 101:4783–8

Hereditary Motor Neuropathies

► Hereditary Neuropathies, Motor and/or Sensory

Hereditary Neuropathies, Motor and/or Sensory

BERND RAUTENSTRAUSS¹, JAMES R. LUPSKI²,
EVA NELIS³

¹Friedrich-Alexander-University, Institute of Human Genetics, Erlangen, Germany

²Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, USA

³University of Antwerpen, Molecular Genetics Laboratory, Peripheral Neuropathy Group, Antwerpen, Belgium,
bernd.rautenstrauss@humgenet.uni-erlangen.de
jlupski@bmc.tmc.edu
eva.nelis@ua.ac.be

Synonyms

Hereditary Motor Neuropathies Hereditary Sensory Neuropathies (HSN); Charcot-Marie-Tooth disease (CMT); Dejerine-Sottas Neuropathy (DSN); Congenital Hypomyelinating Neuropathy (CHN); Hereditary Neuropathy with Liability to Pressure Palsies (HNPP)

Definition

► **Charcot-Marie-Tooth disease (CMT)** is the most common inherited peripheral neuropathy. Since the first description in 1886, it has become apparent that CMT is not one single disease entity, but is clinically and genetically heterogeneous. Dyck et al. (1) introduced a classification of peripheral neuropathies that distinguished hereditary motor and sensory neuropathies (HMSN), hereditary motor neuropathies (HMN), hereditary sensory neuropathies (HSN) and hereditary recurrent focal neuropathies. Each of these entities is further subdivided on the basis of genetic, electrophysiological and neuropathological criteria.

Characteristics

The prevalence of CMT is 1:2500. CMT is symmetrically slowly progressive affecting at first the distal limbs. Most patients suffer from demyelinating neuropathies (CMT type 1). This group is characterized by reduced nerve conduction velocities (► **NCV**, motor median NCV ≤ 38 m/s) due to a partial or complete loss of the myelin sheath. Patients suffering from a primary axonal neuropathy (CMT type 2) are characterized by a reduced compound muscle action potential (► **CMAP**), while NCV is normal or slightly reduced. A morphological examination of sural nerve biopsies is, due to the rapid progress and implementation of molecular genetics in the diagnosis of this group of disorders, not

usually performed on a routine basis. However, it is still an important diagnostic tool in some cases with an equivocal diagnosis. Morphological analysis of sural nerve biopsies reveals typical signs of de- and remyelination with “onion bulb formation” in CMT type 1 and axonal regeneration in CMT type 2. HNPP, a related peripheral neuropathy, is characterized by temporary recurrent nerve palsies. Histopathological analysis of sural nerves reveals focal hypermyelination manifested in structures called “tomacula”(lat.= “sausage”) and segmental demyelination. With emerging genetic knowledge some overlap of demyelinating and axonal forms is apparent in that specific mutations in a given disease gene manifest themselves either as CMT 1 or CMT 2. The mixed phenotype is most likely due to different mutation-dependent disease mechanisms; alterations in the myelin sheath will almost invariably also affect the axonal partner cell and *vice versa* reflecting the intimate interactions between supporting Schwann cells and the axon of the nerves (2, 3).

Distal HMN, also called spinal CMT or distal spinal muscular atrophy (distal SMA), primarily presents with a classical CMT phenotype, since it also produces the clinical picture of peroneal muscular atrophy, although in the absence of sensory losses.

HSN presents with sensory abnormalities, sometimes leading to poorly healing ulcers and eventually amputation of distal parts of the limbs. Varying degrees of autonomic disturbances may be present; therefore HSN is also called hereditary sensory and autonomic neuropathy (HSAN).

Genetics

Autosomal dominant (AD), autosomal recessive (AR) and X-linked patterns of Mendelian inheritance are observed for inherited peripheral neuropathies (Table 1), with AD-CMT1 being the most common pattern. During the last 20 years, efforts have been made to determine the chromosomal loci, to clone genes and to identify mutations for CMT and related peripheral neuropathies. To date, 32 genes and 14 additional genetic loci have been associated with inherited peripheral neuropathies (Table 1). Over 500 distinct mutations involving these genes have been reported (4). Interestingly, distinct mutations in the same gene can result in different phenotypes (e.g. CMT1, CMT2, DSS and CHN caused by *MPZ* mutations, Fig. 1), while mutations in distinct genes can result in the same phenotype (e.g. CMT1 caused by mutations in *PMP22*, *MPZ*, *GJB1*, *EGR2* and *SIMPLE*). The genes involved in peripheral neuropathies code for proteins with very diverse function: structural myelin component (*PMP22*, *MPZ*), transport through myelin sheath (*GJB1*), axonal transport (*NEFL*, *KIF1B*, *GAN*), transcription factor (*EGR2*, *IGHMBP2*), signaling

Hereditary Neuropathies, Motor and/or Sensory. Table 1 Current genetic subtypes, loci and genes for hereditary motor and/or sensory neuropathies

Type	Inheritance	Locus	Gene
HMSN type I			
CMT1A	AD	17p11.2-12	PMP22
CMT1B	AD	1q22-q23	MPZ
CMT1C	AD	16p13.1-p12.3	SIMPLE
CMT1D	AD	10q21.1-q22.1	EGR2
CMT1	AD	8p21	NEFL
CMT4A	AR	8q13q21	GDAP1
CMT4B1	AR	11q23	MTMR2
CMT4B2	AR	11p15	SBF2/MTMR13
CMT4C	AR	5q23-q33	KIAA1985
CMT4D (HMSN-L)	AR	8q24	NDRG1
CMT4E (HMSN-R)	AR	10q21-q22	unknown
CMT4F	AR	19q13.1-q13.3	PRX
CMT4H	AR	12p11.21-q13.11	unknown
CCFDN	AR	18q23-qter	CTDP1
CMT1X	XR/XD	Xq13.1	GJB1
HMSN type II			
CMT2A	AD	1p35-p36	KIF1B MFN2
CMT2B	AD	3q12-q22	RAB7
CMT2C	AD	12q23-q24	Unknown
CMT2D	AD	7p14	GARS
CMT2E	AD	8p21	NEFL
CMT2F	AD	7q11-q21	HSPB1
CMT2	AD	1q22-q23	MPZ
AR-CMT2A (AR-CMT2B1)	AR	1q21.2-q21.3	LMNA
AR-CMT2B (AR-CMT2B2)	AR	19q13.3	Unknown
AR-CMT2C + pyramidal signs	AR	8q21.3	Unknown
AR-CMT2D + hoarseness	AR	8q21.1	GDAP1
CMT2X	XR	Xq24-q26	Unknown
HMSN type III			
DSS/CH	AD/AR	17p11.2-12	PMP22
DSS/CH	AD/AR	1q22-q23	MPZ
DSS/CH	AD/AR	10q21.1-q22.1	EGR2
DSS	AR	19q13.1-q13.3	PRX
AD-DSS	AD	8q23	Unknown

Hereditary Neuropathies, Motor and/or Sensory. Table 1 Current genetic subtypes, loci and genes for hereditary motor and/or sensory neuropathies (Continued)

Type	Inheritance	Locus	Gene
HMSN intermediate			
DI-CMTA	AD	10q24.1-q25.1	Unknown
DI-CMTB	AD	19p12-p13.2	DNM2
DI-CMTC	AD	1p34-p35	Unknown
HMSN-P	AD	3p14.1-q13	Unknown
Distal HMN			
Distal HMN II	AD	12q24	HSPB8
Distal HMN V	AD	7p14	GARS
Distal HMN VII	AD	2q14	Unknown
Congenital distal HMN	AD	12q23-q24	Unknown
Distal HMN with vocal cord paralysis	AD	2p13	DCTN1
Distal HMN/ALS4	AD	9q34	ALS4
Distal HMN/Silver syndrome	AD	11q12-q14	BSCL2
Distal HMN VI	AR	11q13-q21	IGHMBP2
Distal HMN-J	AR	9p12-p21.1	Unknown
HSN			
HSN I	AD	9q22	SPTLC1
HSN II	AR	12p13.33	HSN2
HSN III	AR	9q31	IKBKAP
HSN IV	AR	1q21-q22	NTRKI
HSN V	AR	1p13.1	NGFB
Hereditary recurrent focal neuropathies			
HNPP	AD	17p11.2-12	PMP22
HNA	AD	17q25	Unknown
Others			
GAN	AR	16q24.1	GAN
AS	AR	15q13-q15	SLC12A6

(*PRX*, *MTMR2*, *SBF2/MTMR13*, *NDRG1*, *GDAP1*) and endosome function (*RAB7*, *SIMPLE*). A summary of some of these genes and the associated phenotypes is presented.

Genes Associated with HMSN

Peripheral Myelin Protein 22 Gene (*PMP22*; ►OMIM *601097)

Seventy percent of AD-CMT1 patients carry a 1.4 Mb tandem duplication in chromosome 17p12. The

reciprocal deletion of the 1.4 Mb region leads to HNPP (5, 6). The CMT1A duplication/HNPP deletion region contains the *PMP22* gene, which was shown to be the dosage-sensitive gene. *PMP22*, a 4-transmembrane domain protein, plays a structural role in myelin formation and maintenance.

Point mutations in *PMP22* are associated with a wide variety of clinical phenotypes, such as CMT1, DSN (DSN is traditionally viewed as a recessive disease but has recently been used to describe severe cases of CMT 1 regardless of the mode of inheritance), CHN or

HNPP. Most *PMP22* mutations are located in the transmembrane domain of the protein (3, 4, 7).

Myelin Protein Zero Gene (*MPZ*; OMIM *159440)

In 1993, the first mutations in *MPZ* that were associated with demyelinating peripheral neuropathies (classified as CMT1B) were identified. Most mutations are associated with the demyelinating forms CMT1B, DSN or CHN, but in some cases signs of primary axonal involvement similar to CMT2 have been described. *MPZ* is highly expressed by myelinating Schwann cells of the PNS, where it accounts for more than half of the protein in the peripheral myelin sheath. The protein consists of a single extracellular immunoglobulin-related domain that mediates homophilic adhesion, a transmembrane domain and a short basic intracellular domain. Most *MPZ* mutations are located in the extracellular domain (3, 4, 7).

Gap Junction Protein β 1 (*GJB1*; OMIM *304040)

GJB1, a four transmembrane domain protein, is localized to the paranodes and the Schmidt-Lanterman incisures of myelinating Schwann cells and enables a shortened lateral diffusion pathway for ions and low-molecular-weight compounds between the adaxonal and abaxonal Schwann cell cytoplasm. The first mutations in *GJB1* associated with the X-linked form of CMT (CMTX) were identified in 1993 and alterations in *GJB1* have turned out to be the second most common mutations in CMT, after the CMT1A duplication. Over 250 different mutations in *GJB2* have been described. The mutations are scattered over the entire protein and are not concentrated in a specific domain. Electrophysiological findings in patients suggest a significant contribution by an axonal component associated with specific *GJB1* mutations (3, 4, 7).

Early Growth Response 2 Gene (*EGR2*; OMIM *129010)

Expression of *EGR2* starts before the onset of myelination, when Schwann cells are formed from their precursors, persists throughout life and, as a **zinc finger** transcription factor, plays a major role in the development of the peripheral nervous system. *EGR2* regulates the expression of myelin proteins (including *PMP22*, *MPZ*, *GJB1* and *PRX*) and enzymes required for the synthesis of lipids. *EGR2* mutations are found in severe CMT1, DSN and CHN. Mutations in the DNA-binding domain of *EGR2* lead to dominantly inherited CMT1. In contrast, mutations in the inhibitory R1 domain appear to be recessive. The spectrum of disease severity associated with *EGR2* mutations is broad and can vary considerably, even in patients with the same mutation (3, 4, 7).

Periaxin (*PRX*; OMIM *605725)

Mutations in *PRX* have been identified in CMT type 4F, a recessively inherited severe form of demyelinating peripheral neuropathy and autosomal recessive DSN. CMT4F patients suffer from severe neuropathic pain in addition to the common features of a pronounced neuropathy. *PRX* is specifically expressed by myelinating Schwann cells. Alternative splicing leads to two different proteins, S-periaxin and L-periaxin. L-periaxin is targeted to the nucleus in embryonic Schwann cells, in myelinating Schwann cells it redistributes to the plasma membrane. S-periaxin is restricted to the cytoplasm. L-periaxin forms a complex with dystrophin-related protein-2 (DRP2) and with dystroglycan linking the cytoskeleton of the Schwann cell with the extracellular matrix in the abaxonal Schwann cell membrane (3, 4).

Myotubularin-related-protein-2 Gene (*MTMR2*; OMIM *603557)

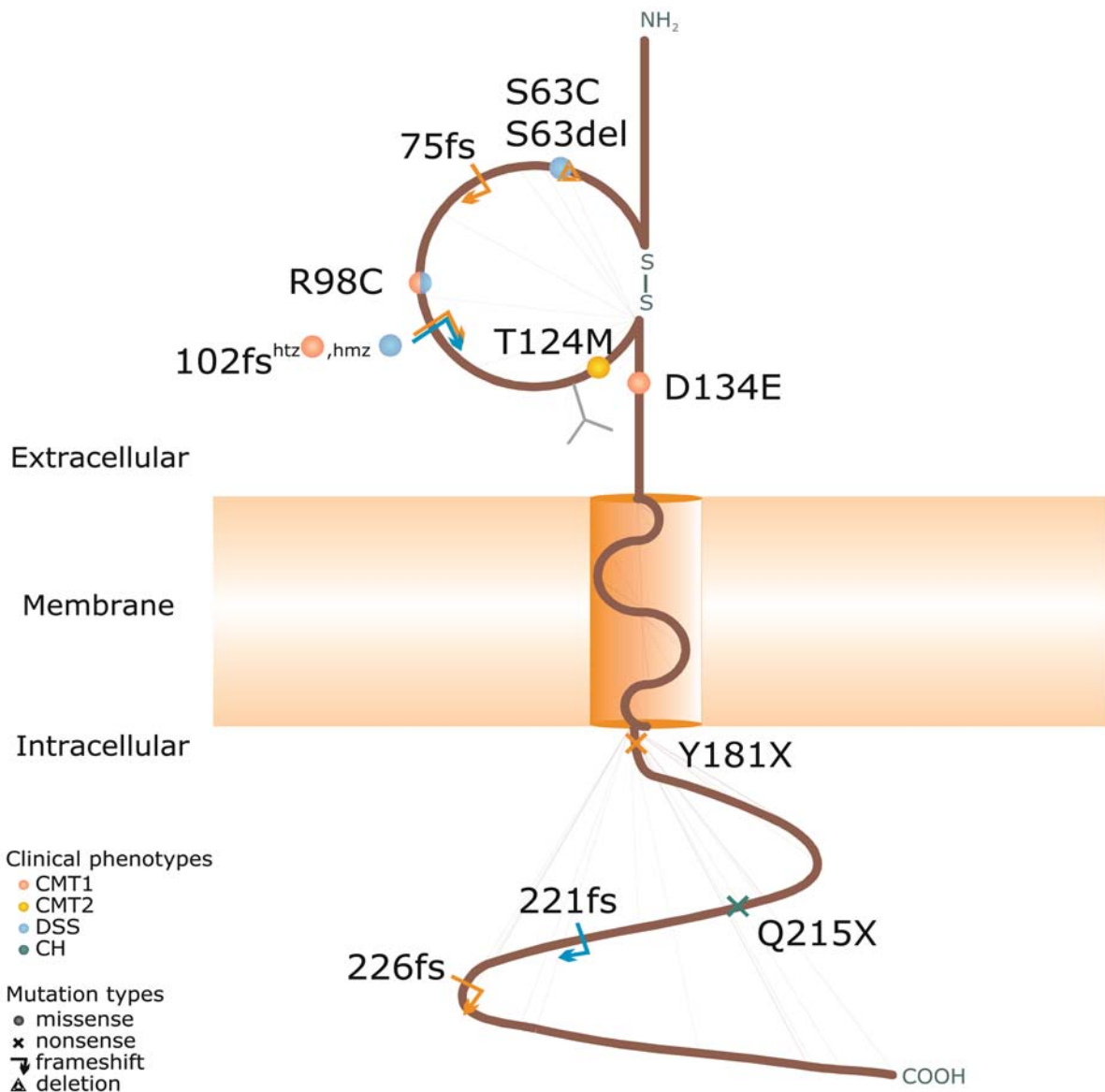
Mutations in *MTMR2* have been shown to be responsible for the rare form of CMT4B1. CMT4B1 is a demyelinating motor and sensory neuropathy characterized by focally folded myelin sheets on sural nerve biopsy. The onset of the disease is very early in life, adult patients are severely handicapped and usually wheelchair bound. *MTMR2* is a member of the myotubularin family, which contains a domain with similarity to dual-specificity phosphatases in which diverse associated mutations have been found to affect protein function (3, 4).

N-myc Downstream-regulated Gene 1 (*NDRG1*; OMIM *605262)

A nonsense mutation in *NDRG1* is associated with hereditary motor and sensory neuropathy-Lom (HMSN-L; CMT4D). HMSN-L is a severe autosomal recessively inherited neuropathy with deafness and unusual neuropathological findings. Individuals with HMSN-L carry a stop codon within the *NDRG1* gene resulting in a truncated *NDRG1* protein. Transcription of *NDRG1*, a phosphatase, is up-regulated by various chemical agents and the *NDRG1* protein contains several phosphorylation sites susceptible to substances that induce stress in the endoplasmic reticulum (3, 4).

Kinesin Family Member 1B Gene (*KIF1B*; OMIM *605995)

A missense mutation in *KIF1B*, was identified in a single family with axonal CMT type 2A. *KIF1B* is a member of the kinesin superfamily. This superfamily represents a group of proteins that are essential for the transport of organelles and materials along microtubules within cells. *KIF1B α* and *KIF1B β* , the two isoforms, result from alternative splicing. *KIF1B α* is responsible for the transport of mitochondria and shows a widespread expression pattern, whereas



Hereditary Neuropathies, Motor and/or Sensory. Figure 1 Structure of MPZ, illustrating that mutations can cause distinct clinical phenotypes.

KIF1B β is specifically expressed in neurons where it is essential for the transport of synaptic vesicles (3, 4).

Mitofusin 2 (MFN2; OMIM *608507)

In seven families linked to the CMT2A locus, mutations in *KIF1B* were excluded. Mutation analysis of other genes in the linked region revealed missense mutations in *MFN2*. Therefore, mutations in *MFN2* are the primary cause underlying CMT2A. *MFN2* is ubiquitously expressed and mRNA transcripts were detected in spinal cord and peripheral nerve. *MFN2* is localized to the outer mitochondrial membrane and regulates the mitochondrial network architecture by fusion of mitochondria.

Gigaxonin (GAN; OMIM *605379)

Mutations in the gigaxonin (*GAN*) gene lead to the rare and severe autosomal recessive giant axonal neuropathy. This neuropathy affects the peripheral and central nervous system. *GAN* is involved in the dynamic organization of the actin network. Defects in *GAN* cause the main stigmata of the disease, severe neuropathy, ataxia and curled, kinky hair (3, 4).

Neurofilament, Light Polypeptide Gene (NEFL; OMIM *162280)

Mutations in *NEFL* are found in patients with CMT2, CMT1 and DSN. Most patients have an early onset, severe phenotype with moderate to severely reduced

NCVs. NEFL codes for a subunit of neurofilaments, the major type of intermediate filaments in neurons. Most mutations are clustered in the head domain of the protein (3, 4).

Ganglioside-induced Differentiation-associated Protein 1 Gene (*GDAP1*; OMIM *606598)

Mutations in *GDAP1* were found to be associated with demyelinating CMT4A, axonal AR-CMT2 with vocal cord paresis and AR-CMT with both demyelinating and axonal features. *GDAP1* might have a function in a signal transduction pathway in neuronal development (3, 4).

Lamin A/C Gene (*LMNA*; OMIM *150330)

A specific missense mutation in *LMNA* has been linked to autosomal recessive axonal CMT (CMT2B1). Different mutations in *LMNA* cause several other disorders, including Emery-Dreyfuss muscular dystrophy and Hutchinson-Gilford progeria. *LMNA* is a nuclear membrane structural protein that can be classified with the intermediate filament polypeptides (3, 4).

SET Binding Factor 2 Gene (*SBF2*) / Myotubularin Related Protein 13 (*MTMR13*; OMIM *607697)

Mutations in *SBF2*/*MTMR13* have been associated with CMT4B2, a recessive demyelinating peripheral neuropathy with focally folded myelin associated with early-onset glaucoma. *SBF2*/*MTMR13* is a member of the pseudo-phosphatase branch of myotubularins, since it contains inactivating substitutions at the catalytic cysteine (4).

Small Integral Membrane Protein of Lysosome/Late Endosome (*SIMPLE*) (OMIM *603795)

CMT type 1C is associated with missense mutations in *SIMPLE*. The protein may play a role in protein degradation pathways (4).

Small GTP-ase Late Endosomal Protein (*RAB7*; OMIM *602298)

Mutations in *RAB7* are associated with CMT type 2B, an ulcero-mutilating neuropathy. *RAB* proteins are important regulators of vesicular transport and are located in specific intracellular compartments. *RAB7* has been localized to late endosomes and shown to be important in the late endocytic pathway (4).

Solute Carrier 12 (Potassium/Chloride Transporter), Member 6 (*SLC12A6*; OMIM *604878)

Mutations in *SLC12A6* are associated with peripheral neuropathy with agenesis of the corpus callosum, also known as Andermann syndrome. The protein, a K^+ - Cl^- cotransporter, is involved in the electroneutral movement of ions across the plasma membrane (4).

Glycyl t-RNA Transferase (*GARS*; OMIM *600287)

Missense mutations in *GARS* were found in families with CMT2D and distal HMN V (distal SMA V). Aminoacyl-tRNA synthetases perform an essential function in protein synthesis by catalyzing the esterification of an amino acid to its cognate tRNA (4).

Genes Associated with HMN

Immunoglobulin mu Binding Protein 2 Gene (*IGHMBP2*; OMIM *600502)

Distal HMN type VI, also called spinal muscle atrophy with respiratory distress type 1 (*SMARD1*), results from mutations in the *IGHMBP2* gene. *IGHMBP2* is a transcription factor with a function important to motor neuron maintenance and integrity (4).

Glycyl t-RNA Transferase (*GARS*)

See above.

Dynactin 1 (*DCTN1*; OMIM *601143)

A missense mutation in *DCTN1* was identified in patients with distal HMN with vocal fold paralysis. *DCTN1* encodes the largest subunit of the axonal transport protein dynactin. The mutation is predicted to distort the folding of dynactin's microtubule-binding domain (4).

Genes Associated with HSN

Neurotrophic Tyrosine Kinase Receptor Type 1 Gene (*NTRK1*; OMIM *191315)

HSN type IV or congenital insensitivity to pain and anhidrosis (*CIPA*), is associated with mutations in *NTRK1*. Neurotrophins and their receptors play an important role in regulating development of the nervous systems (4).

Serine Palmitoyl transferase, Long Chain Base Subunit 1 Gene (*SPTLC1*; OMIM *605712)

HSN type I is caused by missense mutations in *SPTLC1*. Serine palmitoyl transferase is the key enzyme in sphingolipid biosynthesis. The mutations in *SPTLC1* are associated with increased *de novo* glucosyl ceramide synthesis in lymphoblast cell lines in affected individuals. Increased *de novo* ceramide synthesis triggers apoptosis and is associated with massive cell death during neural tube closure, raising the possibility that neural degeneration in HSN I is due to ceramide-induced apoptotic cell death (4).

Inhibitor of Kappa Light Polypeptide Gene Enhancer in B-cells, Kinase Complex-associated Protein Gene (*IKBKAP*; OMIM *603722)

HSN type III, also known as familial dysautonomia or Riley-Day syndrome, is observed predominantly in the Ashkenazi Jewish population. The condition is characterized by widespread sensory and variable autonomic

dysfunction. HSN III is caused by mutations in IKBKAP, a scaffold protein and a regulator for 3 different kinases involved in proinflammatory signaling (4).

Molecular Diagnostics

The analysis of the CMT1A duplication is the initial step in a diagnostic evaluation of a suspected hereditary neuropathy patient. The reciprocal HNPP deletion is usually detected in this test. It is important to include the deletion analysis in the first test since HNPP patients sometimes present with symptoms similar or identical to CMT type 1. Two groups of methods are used, binary methods resulting in a “yes” or “no” answer and dosage sensitive or other methods.

Binary Methods

These methods include the detection of junction fragments by restricted genomic DNA and pulsed field gel electrophoresis (►PFGE) as well as fluorescence *in situ* hybridization (►FISH).

PCR Methods Based on Simple Tandem Repeat (STR) Markers

For STR based methods several protocols for multiplex ►PCRs are available. At least two of the markers should be positive for three alleles to be indicative of the CMT1A duplication. The HNPP deletion may be suspected when several polymorphic markers reveal only one allele (hemizygosity or homozygosity). Quantitative methods are now available using real-time quantitative PCR that allow direct quantification of the product during the amplification. These methods are highly sensitive for dosage differences but need a very careful calibration with a large number of controls.

Mutation Screening in Candidate Genes

If the CMT1A duplication/HNPP deletion is excluded, screening for mutations in candidate genes may follow. Due to the large number of mutations in different genes, a detailed clinical, genetic, electrophysiological and neuropathological record of the patient should be available to orient the genetic testing.

In order to identify mutations within coding regions of candidate genes several screening methods have been applied. The most common ones are single strand conformation polymorphism analysis (►SSCP), heteroduplex analysis and denaturing high-performance liquid chromatography (►dHPLC). Screening large numbers of patient samples is often performed by these screening methods and is useful for validation of known mutations. To overcome the difficulties with sensitivity, DNA sequence analysis is the state of the art technique for diagnostic purposes.

References

1. Dyck P, Chance P, Lebo R et al (1993) Hereditary motor and sensory neuropathies. In: Dyck P, Thomas P, Griffin J et al (eds) *Peripheral neuropathies*, 3rd edn. Saunders, Philadelphia, pp 1094–1136
2. Kuhlmann G, Young P, Hunnemann G et al (2002) Clinical features and molecular genetics of hereditary peripheral neuropathies. *J Neurol* 249:1629–1650
3. Berger P, Young P, Suter U (2002) Molecular cell biology of Charcot-Marie-Tooth disease. *Neurogenetics* 4:1–15
4. Inherited Peripheral Neuropathies Mutation Database, ►<http://molgen-www.uia.ac.be/CMTMutations/>
5. Lupski JR, Garcia CA (2001) Charcot-Marie-Tooth Peripheral Neuropathies and Related Disorders. In: Scriver CR, Beaudet AL, Sly WS (eds) *Bases of Inherited Diseases*, 8th edn. McGraw-Hill, New York, Chapter 227:5759–5788
6. Saifi GM, Szigeti K, Snipes GJ et al (2003) Molecular mechanisms, diagnosis and rational approaches to management and therapy for Charcot-Marie-Tooth and related peripheral neuropathies. *Journal of Investigative Medicine* 51:261–283
7. Nelis E, Timmerman V, De Jonghe P et al (1999) Molecular genetics and biology of inherited peripheral neuropathies: a fast-moving field. *Neurogenetics* 2:137–148

Hereditary Neuropathy with Liability to Pressure Palsies

►Hereditary Neuropathies, Motor and/or Sensory

Hereditary Nonpolyposis Colorectal Cancer

PÄIVI PELTOMÄKI

Department of Medical Genetics, University of Helsinki, Helsinki, Finland
paivi.peltomaki@helsinki.fi

Synonym

Lynch syndrome

Definition

Based on clinical features and family history, hereditary nonpolyposis colorectal cancer (►HNPPCC) (MIM No.

120435-6) is an autosomal dominant disorder, characterized by the development of cancers of the colon and rectum, endometrium, and to lesser extent other organs, at an early age (1). Molecular definition of the syndrome additionally requires that there is evidence of a DNA mismatch repair (►MMR) defect.

Characteristics

Historical Perspectives

HNPCC was first described by Dr. Aldred Warthin in 1913. Dr. Henry Lynch, working with A. Krush, updated the original family ("Family G") in 1971. In 1990, the newly established International Collaborative Group on HNPCC (►ICG-HNPCC) agreed on operational diagnostic criteria for HNPCC, referred to as Amsterdam criteria I, to provide uniformity in collaborative studies. In 1993, a Finnish-American collaborative group localized the first HNPCC susceptibility gene on chromosome 2, establishing that HNPCC is a single-gene Mendelian disorder and providing a reliable molecular marker for HNPCC susceptibility. To date, nine human DNA mismatch repair genes have been isolated, including three major genes, *MSH2*, ►*MLH1*, and *MSH6*, associated with HNPCC predisposition.

Prevalence

HNPCC is among the most common hereditary diseases in man and the most common hereditary disorder predisposing to colorectal cancer. Based on clinical criteria, it is estimated to account for up to 13% of all colorectal cancers. Molecular estimates based on germline mutation carriers among consecutive patients with colorectal cancer have arrived at somewhat lower figures (0.3–3% of the total colorectal cancer burden).

Clinical Features

The most stringent diagnostic criteria for HNPCC (Amsterdam criteria I) require that at least three close relatives (one of which is a first-degree relative - parent, sibling, child - of the other two) are affected with colon cancer in two successive generations and the age at diagnosis is below 50 years in at least one. In addition to colon cancer, HNPCC patients often have an excess of extracolonic cancers. Of these, endometrial cancer, and cancer of the small intestine, ureter, or renal pelvis are considered to be highly specific for HNPCC and are recognized in the revised, less stringent diagnostic criteria for HNPCC (Amsterdam criteria II), besides colon cancer. Other cancers that are common in HNPCC include cancers of the stomach, ovary, and hepatobiliary tract.

Hereditary Nonpolyposis Colorectal Cancer.

Table 1 Clinical syndromes associated with germline mutations in DNA mismatch repair genes

Syndrome	Tumor site	Predisposing gene
HNPCC	Colon, endometrium, small bowel, ureter/renal pelvis, stomach, ovary, hepatobiliary tract	<i>MLH1</i> , <i>MSH2</i> , <i>MSH6</i> (<i>PMS2</i> , <i>MLH3</i>)
Muir-Torre	Sebaceous glands and sites typical of HNPCC	<i>MSH2</i> (<i>MLH1</i>)
Turcot	Brain and colorectum	<i>MLH1</i> , <i>PMS2</i>

Hereditary Nonpolyposis Colorectal Cancer.

Table 2 HNPCC-associated human MMR genes and their contribution to HNPCC predisposition

Gene	Chromosome	No. of mutations (%)
<i>MLH1</i>	3p21-23	225 (50%)
<i>MSH2</i>	2p21	175 (39%)
<i>MSH6</i>	2p21	32 (8%)
<i>PMS2</i>	7p22	5 (1%)
<i>MLH3</i>	14q24.3	16 (2%)

Two variant forms of HNPCC are known, Muir-Torre syndrome and Turcot syndrome, which, like classical HNPCC, are associated with germline mutations in MMR genes (Table 1). Muir-Torre syndrome is characterized by the occurrence of sebaceous gland tumors together with HNPCC type internal malignancy. In Turcot syndrome, primary brain tumors (usually glioblastomas) are associated with multiple colorectal adenomas or carcinomas. On the molecular level, Turcot syndrome overlaps with both HNPCC (the subset with MMR gene mutations) and ►*FAP* (the subset with ►*APC* mutations).

Cellular and Molecular Regulation

Susceptibility Genes

At present, four (*MSH2*, *MLH1*, *MSH6*, ►*PMS2*), possibly five (*MLH3*) human MMR genes are known, whose germline mutations cause predisposition to HNPCC (Table 2). The international collaborative group on HNPCC, currently known as InSiGHT (International Society for Gastrointestinal Hereditary

Tumors), maintains a database of mutations associated with susceptibility to this disorder (2; ►<http://www.insight-group.org>). To date, more than 450 different predisposing mutations are known that occur in over 700 families from all over the world. Of all HNPCC-associated mutations, 50% affect *MLH1*, 39% *MSH2*, and 9% affect *MSH6*, whereas the share of *PMS2* and *MLH3* is only a few percent.

Generally, *MSH2* and *MLH1* mutations give rise to typical HNPCC. The overall lifetime cancer risk associated with germline mutations of these genes is around 90%. *MSH6* mutations are found in both typical and atypical HNPCC families with a frequent occurrence of endometrial cancer. *PMS2* mutations are mainly associated with the Turcot variant and their penetrance may vary. More data are necessary to evaluate the role of *MLH3* in HNPCC predisposition; according to preliminary experience, *MLH3* mutations mostly occur in atypical HNPCC families.

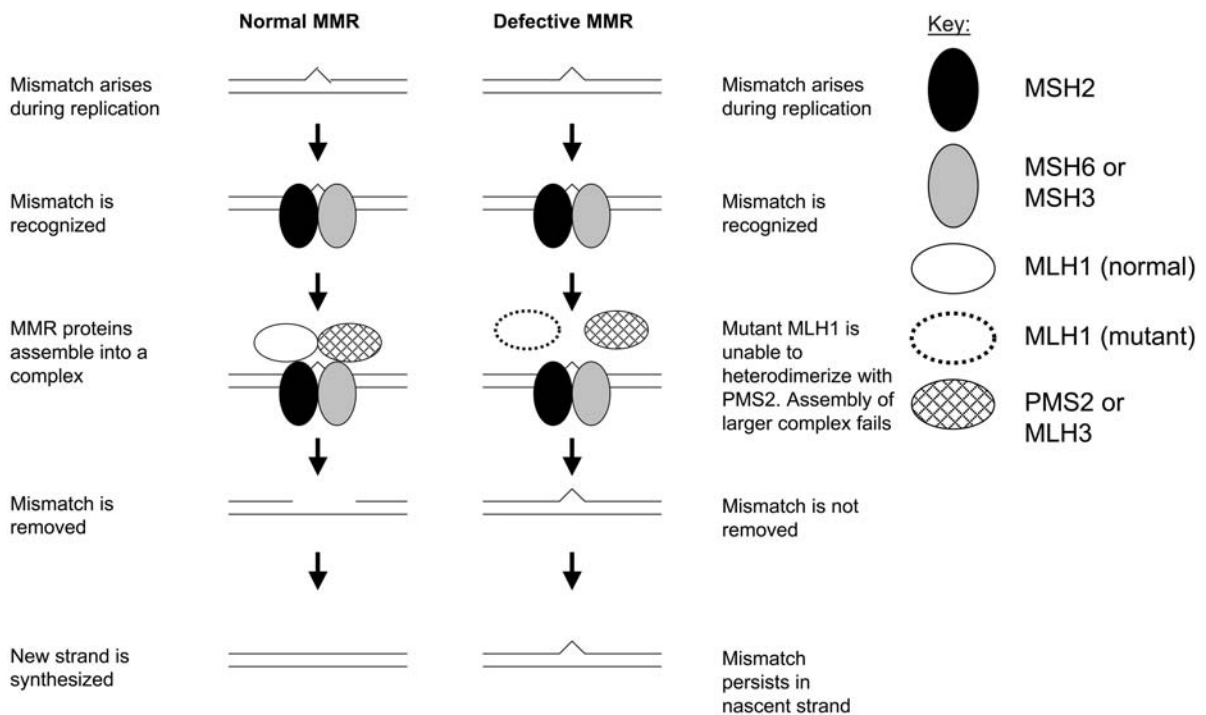
DNA Mismatch Repair Defect in HNPCC

The primary function of the protein products of the MMR genes is to correct ►single nucleotide mismatches and ►insertion/deletion loops that may arise during DNA replication. Failure to accomplish this function results in elevated (~100-fold) rates of point mutations as well as instability at repetitive tracts

(►microsatellites), termed as microsatellite instability (►MSI) (3). Increased mutational load and microsatellite instability (appearing as extra alleles as compared to paired normal tissue) are typical features of tumors from HNPCC patients.

Human MMR requires at least six different MMR proteins (4; Fig. 1, left panel). For mismatch recognition, the *MSH2* protein forms a heterodimer with either *MSH6* or ►*MSH3* depending on the type of lesion to be repaired: *MSH6* is required for the correction of single-base mispairs while both *MSH3* and *MSH6* may contribute to the correction of insertion-deletion loops. A heterodimer of *MLH1* and *PMS2* coordinates the interplay between the mismatch recognition complex and other proteins necessary for MMR. Besides *PMS2*, *MLH1* may interact with *MLH3* and *PMS1*. Recent observations suggest that *PMS2* is required for the correction of single base mispairs while both *MLH3* and *PMS2* contribute to the correction of insertion-deletion loops. The role of the *MLH1*-*PMS1* complex in MMR needs to be clarified.

A majority of HNPCC-associated MMR gene alterations are ►frameshift or ►nonsense mutations that lead to truncated proteins, although for *MLH1* and *MSH6* in particular, ►missense mutations are also common. Such mutations typically result in the loss of important interaction domains (truncating mutations) or changes



Hereditary Nonpolyposis Colorectal Cancer. Figure 1 Normal and defective DNA mismatch repair.

in the local structure or conformation (missense mutations) that impair the ability of the proteins to interact with their partners or other components of the MMR pathway, or otherwise properly accomplish the MMR function (Fig. 1, right panel).

Tumorigenesis and Pathology

HNPCC tumors, as well as sporadic colorectal carcinomas with high-degree of MSI (~15% of all colorectal cancers), show common clinicopathological features, whose molecular or cellular mechanisms are only partly understood. These characteristics include predilection for the proximal colon, diploid DNA content, high histological grade, high mucin content, marked host-lymphocytic infiltration, association with female sex and better survival.

Adenomas, the precursor lesions of colon cancers, are not more frequent in HNPCC patients than in the general population, but they are more likely to undergo malignant transformation due to the accumulation of mutations as a result of MMR deficiency. Hence, HNPCC is viewed as a disease of progression. Non-neoplastic cells (such as those from blood) that carry one normal and one defective copy of a given MMR gene are still MMR-proficient, but when the wild-type copy is lost or acquires somatic mutations, as occurs in tumor precursor cells, the cells become MMR-deficient. Apart from anonymous microsatellites, hypermutability in HNPCC affects important growth-regulatory genes, especially those containing repeat sequences as mutation targets (e.g. *TGF β RII*, *BAX*, *MSH3* and *MSH6*), thus providing a driving force for cancer development and progression ("mutator pathway") (5).

Clinical Relevance

Cancer Surveillance

Colonoscopy screening every 1–2 years, beginning at age 20–25, is recommended for mutation carriers or at-risk individuals from HNPCC families. It has been demonstrated that colonoscopy screening and removal of precursor lesions, adenomas, significantly reduces morbidity and mortality in such families. Surveillance for extracolonic cancers is also indicated, notably for endometrial cancer (through regular pelvic examination and transvaginal ultrasound or endometrial aspirate).

Diagnostics

The detection of a predisposing mutation in an HNPCC family is important since it makes it possible to focus cancer screening and surveillance on those members truly at increased risk, namely mutation carriers. MSI analysis and/or immunohistochemical staining for MMR protein expression offer useful preliminary tests

for MMR gene involvement. Loss of a specific MMR protein by immunohistochemical analysis of tumor tissue can pinpoint the defective gene and provide a basis for screening that gene for possible mutations in blood DNA (or RNA). For most cases, exon-by-exon sequencing is the method of choice for detecting mutations in *MSH2*, *MLH1*, and *MSH6*, and results in the identification of predisposing changes in up to two-thirds of families. As many as one-fifth of HNPCC families may have large genomic deletions or duplications in MMR genes, and the identification of such alterations requires more complex technology (e.g. multiplex ligation-dependent probe amplification or ▶ [Conversion technology](#)).

There is increasing evidence that a proportion of families clinically presenting with HNPCC are not associated with MMR defects and have an as yet unknown molecular etiology. The absence of MSI, combined with normal MMR protein staining, is a strong indication against MMR gene involvement.

Drug Sensitivity or Resistance

Detection of a MMR defect may provide an important piece of information to guide clinical management of the patients. For example, malfunction of the MMR system may modify the response to cytotoxic drugs; cancers with MSI may be particularly sensitive to 5-fluorouracil and other antimetabolites, whereas the same tumors show increased resistance to alkylating agents. It is speculated that 5-fluorouracil, which acts as a competitive inhibitor for substrates critical for DNA synthesis, may work in concert with MMR deficiency and enhance apoptosis in these cells. The resistance of MMR-deficient cells to alkylating agents may result from the fact that defective MMR proteins are unable to detect alkylation adducts and consequently, fail to induce apoptosis.

▶ [Colorectal Cancer](#)

References

1. Lynch HT, de la Chapelle A (1999) Genetic susceptibility to non-polyposis colorectal cancer. *Journal of Medical Genetics* 36:801–818
2. Peltomäki P, Vasen HF (2004) Mutations associated with HNPCC predisposition - Update of ICG-HNPCC/In-SIGHT mutation database. *Disease Markers* 20:269–276
3. Umar A, Boland CR, Terdiman JP et al (2004) Revised Bethesda guidelines for hereditary non-polyposis colorectal cancer (Lynch syndrome) and microsatellite instability. *J Natl Cancer Inst* 96:261–268
4. Jiricny J, Nyström-Lahti M (2000) Mismatch repair defects in cancer. *Current Opinion of Genetics and Development* 10:157–161
5. Fishel R (2001) The selection for mismatch repair defects in hereditary nonpolyposis colorectal cancer: revising the mutator hypothesis. *Cancer Research* 61:7369–7374

Hereditary Sensory Neuropathies

► Hereditary Neuropathies, Motor and/or Sensory

Hereditary Sideroblastic Anemia

Definition

Hereditary sideroblastic anemia is an anemia resulting from a genetic deficiency of the X-chromosome linked enzyme erythroid-specific δ -aminolevulinate synthase.

► Hemochromatosis

Hereditary Spastic Paraplegia

EVAN REID

Department of Medical Genetics, University of Cambridge, Cambridge, UK
ereid@hgmp.mrc.ac.uk

Synonyms

Familial spastic paraplegia; Strümpell disease

Definition

The principal neurological pathway over which signals pass to drive a deliberate movement can be divided into 2 stages. In the first stage, a descending connection is made from neurons in the motor cortex of the brain (often referred to clinically as “►upper motor neurons”) to cells in the anterior and posterior horns of the spinal cord, while in the second stage “►lower motor neurons” in the anterior horn of the spinal cord connect to the skeletal muscles at the neuromuscular junction. The axons that project from the upper motor neuron cell bodies to the spinal cord constitute the corticospinal tract and this tract is “crossed”, so that upper motor neurons in the motor cortex of one side innervate lower motor neurons on the opposite side of the spinal cord. A wide variety of insults to the corticospinal tract result in ►spastic paralysis. The ►hereditary spastic paraplegias (HSPs) are ►single gene disorders in which the axons of the corticospinal tract either fail to develop

normally, or show progressive degeneration after initial normal development.

Characteristics

All HSPs share the principal clinical feature of progressive lower limb spastic paralysis. They are conventionally subdivided into pure and complicated forms, depending on the absence or presence of additional neurological or non-neurological features (1).

Pure HSPs

Pure hereditary spastic paraplegia is probably the single largest subgroup of HSP conditions. A majority of pure HSP (PHSP) families show autosomal dominant inheritance, with autosomal recessive inheritance accounting for most of the remainder. X-linked recessive inheritance is very rare. Histopathological studies of pure HSP show a length dependent “dying back” of the terminal ends of the corticospinal tract axons, with longest axons involved first, suggesting that these conditions might provide useful models for understanding the molecular factors involved in normal maintenance of the corticospinal tracts.

Complicated HSPs

The complicated HSPs comprise a large number of conditions in which spasticity is accompanied by other features such as muscle wasting, mental retardation, epilepsy, cerebellar signs, extrapyramidal movement disorder, optic atrophy, pigmentary retinopathy or ichthyosis. Unlike PHSP, they are most commonly inherited in an autosomal recessive pattern, although autosomal dominant and X-linked inheritance can occur.

Cellular and Molecular Regulation

The hereditary spastic paraplegias show extreme genetic heterogeneity. The 20 HSP genes that have been mapped to date are shown in Table 1. Eight of these genes have now been identified and their pathological mechanisms can tentatively be divided into 4 groups.

Cell Recognition and Signalling: L1-CAM

The cell adhesion molecule L1 (L1-CAM) is a cell-surface membrane-associated glycoprotein that is a member of an immunoglobulin superfamily. Mutations in L1-CAM are found in a form of X-linked complicated (by mental retardation and absence of the extensor pollicis longus muscle) spastic paraplegia termed SPG1, in MASA (mental retardation, aphasia, shuffling gait, adducted thumbs) syndrome and in X-linked hydrocephalus (2). Over 90 different mutations, of varying forms, in the L1-CAM gene can cause these conditions.

Many of the mutations are predicted to eliminate expression of L1-CAM, although some ►**missense mutations** affect particular functions of the protein. L1-CAM is expressed throughout the nervous system on populations of developing and differentiated neurons, as well as in Schwann cells in the peripheral nervous system. *In vitro* studies have suggested that, along with several other functions, L1 interactions may mediate axon bundling and may be involved in the processes of neurite extension and axonal guidance. L1-CAM knockout mice show phenotypes similar to the human L1-CAM associated conditions, including disruption of the normal anatomy of the corticospinal tracts, where impaired axonal pathfinding results in failure of the normal crossing over of the tracts in the medulla.

L1 CAM sits at the centre of a complex set of extracellular and intracellular interactions and it binds to a wide variety of extracellular ligands (2). Binding of L1-CAM to other L1-CAM molecules (i.e. ►**homophilic binding**) promotes cell adhesion and neurite outgrowth. This may be influenced (positively or negatively) by interaction with other potential ligands. The molecule is linked to a number of intracellular signalling pathways. Key amongst these is its interaction with fibroblast growth factor receptors (FGFRs), membrane associated molecules that have intracellular tyrosine kinase signalling domains. Homophilic binding of L1-CAM stimulates FGFR tyrosine kinase activity, which in turn stimulates Ca^{2+} influx, resulting in modification of the axonal growth cone by a mechanism that is not yet entirely clear. L1-CAM also has important interactions *via* its cytoplasmic domain, notably with ankyrin, a linker protein of the spectrin-based cytoskeleton, so that homophilic binding may cause re-organisation of the cytoskeleton.

Abnormalities of Myelination: Proteolipid Protein

Mutations in the *proteolipid protein gene* (PLP1) can cause pure or complicated HSP (SPG2), or the much more severe Pelizaeus-Merzbacher disease (PMD) (3). Several missense mutations in the gene are associated with HSP, whereas PMD is associated with other missense mutations or, more commonly, with duplications of the entire gene. The PLP1 gene codes for 2 myelin proteins, PLP and DM20. There is a correlation between PLP1 gene mutation and phenotype, with point mutations that affect highly conserved DM20 amino acid residues being associated with the most severe PMD phenotypes. The PLP and DM20 proteins are thought to play a major role in oligodendrocyte maturation and also form structural components of the myelin sheath. The precise mechanism by which certain missense mutations in the PLP1 gene cause spastic paraplegia is not clear. MRI studies in human

SPG2 families demonstrate central nervous system white matter abnormalities reminiscent of the pathological changes found in the rumpshaker natural mouse mutant model of SPG2, which shows hypomyelination with reduced levels of PLP, but not DM20. It seems likely that perturbations of corticospinal tract myelination, perhaps accompanied or caused by abnormal trafficking-induced toxic effects on oligodendrocytes, may contribute to spasticity in SPG2.

Abnormalities of Mitochondrial Molecular Chaperones: Paraplegin and HSP60

Mutations in the *paraplegin* gene (SPG7), at chromosome 16q24, are associated with pure or complicated (by optic, cortical or cerebellar atrophy) autosomal recessive HSP (4). *Paraplegin* is a mitochondrial metalloprotease and a member of the AAA (ATPases associated with diverse cellular activities) protein family. Members of this large family of proteins are characterised by the presence of one or two copies of a 220–250 amino acid long sequence motif, the AAA cassette. Members of the AAA family of proteins are proven or putative ATPases that take part in many cellular functions, including cell-cycle regulation, protein degradation in the 26S proteasome, organelle biosynthesis, membrane fusion and intracellular membrane trafficking events. HSP patients with *paraplegin* mutations have defects in mitochondrial oxidative phosphorylation. The precise mechanism of action of paraplegin has not been characterised. However, it has a high degree of homology to a mitochondrial subclass of yeast and fungal ATPases that have proteolytic and ►**chaperone**-like activities at the inner mitochondrial membrane. These proteins may participate in protein “quality control”, degrading misfolded proteins in the mitochondrial intermembrane space. Yeast and fungi lacking the respective paraplegin homologues Yme1 or *iap-1* have defective respiration under certain conditions, with yeast accumulating abnormally formed mitochondria. By analogy, it has been suggested that defects in paraplegin may cause an accumulation of abnormally synthesised mitochondrial proteins, leading to mitochondrial respiratory dysfunction and eventual axonal degeneration.

A missense mutation in the mitochondrial chaperone heat shock protein HSP60 has been found in the original family used to map the chromosome 2q autosomal dominant pure HSP (SPG13) locus. Elegant complementation studies demonstrated that the mutant has a functional effect and so it is highly likely that abnormalities of this gene can cause HSP. The pathological mechanism of the mutation is not yet clear and may involve loss of function or a dominant negative effect on the multimeric chaperone complex in which HSP60 participates.

Hereditary Spastic Paraplegia. Table 1 Genetic loci for HSP. CHSP=complicated HSP, PHSP=pure HSP, XLH=X-linked hydrocephalus, PMD= Pelizaeus-Merzbacher disease, MASA= mental retardation, aphasia, shuffling gait, adducted thumbs syndrome. OMIM=online Mendelian inheritance in man database

Gene symbol	OMIM number	Chromosomal location	Gene product (see text)	Phenotype
X-linked				
SPG1	312900	Xq28	L1-CAM	CHSP; MASA; XLH (see text)
SPG2	312920	Xq22	PLP; DM20	PHSP; CHSP; PMD (see text)
SPG16	300266	Xq11.2	-	PHSP; CHSP
Dominant				
SPG3A	182600	14q12-q21	Atlastin	PHSP
SPG4	182601	2p21-p24	Spastin	PHSP; ?CHSP
SPG6	600363	15q11.2-q12	-	PHSP
SPG8	603563	8q24	-	PHSP
SPG9	601162	10q23.3-q24.2	-	CHSP (cataracts, gastroesophageal reflux, distal muscle wasting)
SPG10	604187	12q13	KIF5A	PHSP
SPG12	604805	19q13	-	PHSP
SPG13	605280	2q24-q34	HSP60	PHSP
SPG17	270685	11q12-q14	-	CHSP (Silver Syndrome- severe distal muscle wasting)
SPG18	Pending	reserved	-	-
SPG19	Pending	9q33-q34	-	PHSP
Recessive				
SPG5	270800	8q11-q13	-	PHSP
SPG7	602783	16q24.3	Paraplegin	PHSP; CHSP (optic, cerebellar and cerebral atrophy)
SPG11	604360	15q13-q15	-	PHSP; CHSP (thinning of corpus callosum)
SPG14	605229	3q27-q28	-	CHSP (mental retardation and distal muscle wasting)
SPG15	606859	14q22-q24	-	CHSP (Kjellin syndrome-pigmentary maculopathy, distal muscle wasting, mental retardation)
SPG20	275900	13q12.3	Spartin	CHSP (Troyer syndrome- see text)

Defects of Intracellular Molecular Trafficking: KIF5A, Spartin, Spastin and Atlastin

The most direct evidence for the pathological involvement of HSP proteins in intracellular membrane traffic events comes from the identification of a mutation in the kinesin heavy chain gene KIF5A in a family with autosomal dominant pure HSP (SPG10). Deletion of the [orthologous](#) kinesin heavy chain genes of several

species causes movement defects reminiscent of human spastic paraplegia. KIF5A is a member of the kinesin superfamily and forms part of a conventional kinesin-1 motor, a molecular complex that transports membrane-bound cargoes on microtubules in an ATP dependent fashion. Kinesins have been implicated in multiple membrane traffic events, including transport from the Golgi to endoplasmic reticulum, from the trans-Golgi

network to the plasma membrane or endosomes and, in neurons, in fast anterograde axonal transport (5). KIF5A is a neuronal specific kinesin and is localised diffusely in mouse motor and sensory neuronal bodies. Its precise cellular function and sites of action are not known.

Mutations in the chromosome 2p24 HSP gene *spastin* (SPG4) are responsible for approximately 40% of autosomal dominant pure HSP. More than 100 spastin mutations have been described, including numerous missense, ►nonsense, ►frameshift, and ►splice site mutations, as well as less frequent large-scale deletions. Although not entirely certain, the broad mutational spectrum observed in spastin-associated HSP suggests that the molecular pathological mechanism in most cases is likely to be loss-of-function, with disease occurring once functioning spastin levels fall below a critical threshold level. Tolerance for reduced dosage of functioning spastin may be very low, since some “leaky” (i.e. creating both wild type and aberrant splice variants) splice site mutations result in only slight reductions in wild type mRNA expression. Initial sequence analysis was not helpful in identifying a functional role for the spastin protein, since it placed it in a subgroup of the AAA family, members of which participate in many different cellular processes (see above). It has been suggested that spastin might have a microtubule severing effect, similar to related AAA proteins such as katanin. On the other hand, evidence based on recent sequence analysis has suggested that spastin may have a role in intracellular membrane traffic events (6). The identification of the *spartin* gene as the cause of Troyer syndrome, an autosomal recessive HSP complicated by dysarthria, distal amyotrophy, mild developmental delay and short stature, led to the delineation of a novel protein domain termed MIT (microtubule interacting and trafficking), present in spartin and spastin. This domain is also present in several proteins involved in membrane traffic events, including the yeast vacuolar sorting protein (VPS) VPS4 and its mammalian homologue SKD1, sorting nexin 15 (SNX15) and RPK118. Interestingly, VPS4/SKD1 is also an AAA protein, belonging to the same AAA family subgroup as spastin.

The chromosome 14q autosomal dominant pure HSP gene, *atlastin* (SPG3A) codes for a novel GTPase that has sequence homology to members of the dynamin family of large GTPases, particularly guanylate binding protein-1 (7). The dynamins are non-conventional motor molecules that take part in vesicle trafficking events, including recycling of synaptic vesicles and in the dispersion of mitochondria. *Atlastin* is widely expressed, most abundantly in brain and spinal cord, and codes for a protein of 558 amino acids. Mutations

in atlastin are responsible for approximately 10% of autosomal dominant pure HSP, and are particularly associated with childhood onset disease.

Clinical Relevance

Identification of the genes and corresponding proteins involved in HSP has significant implications for families affected by these conditions, where in some cases ►diagnostic gene testing and ►predictive gene testing can now be offered. In the long term and more widely, a full understanding of the function of molecular factors involved in corticospinal tract maintenance may eventually have implications for the treatment of more common diseases of the corticospinal tracts that cause spastic paralysis.

References

1. Reid E (2003) Science in motion: common molecular pathological themes emerge in the hereditary spastic paraplegias. *J Med Genet* 40:81–86
2. Kenwright S, Watkins A, De Angelis E (2000) Neural cell recognition molecule L1: relating biological complexity to human disease mutations. *Hum Mol Genet* 6:879–886
3. Yool DA, Edgar JM, Montague P et al (2000) The proteolipid protein gene and myelin disorders in man and animal models. *Hum Mol Genet* 9:987–992
4. Casari G, De Fusco M, Ciarmatori S et al (1998) Spastic paraplegia and OXPHOS impairment caused by mutations in Paraplegin, a nuclear-encoded mitochondrial metallo-protease. *Cell* 93:973–983
5. Allan VJ, Thompson HM, McNiven MA (2002) Motoring around the Golgi. *Nat Cell Biol* 4:E236–242
6. Crosby AH, Proukakis C (2002) Is the transportation highway the right road for hereditary spastic paraplegia? *Am J Hum Genet* 71:1009–1016
7. Zhao X, Alvarado D, Rainier S et al (2001) Mutations in a newly identified GTPase cause autosomal dominant hereditary spastic paraparesis. *Nat Genet* 29:326–331

Heritability

Definition

Heritability is defined as the proportion of the total phenotypic variance between individuals or – in the context of disease – as causation of a disease that is due to genetic effects; as opposed to environmental effects.

- Common (Multifactorial) Diseases
- COPD and Asthma, Genetics
- Genetic Epidemiology

Heritable Skin Disorders

GABRIELE RICHARD, JOUNI UITTO

Department of Dermatology and Cutaneous Biology,
and Jefferson Institute of Molecular Medicine,
Thomas Jefferson University, Philadelphia, PA, USA
jouni.uitto@jefferson.edu

Definition

Keratin Disorders

Keratin disorders are a clinically and genetically heterogeneous group of ►Mendelian disorders with abnormal differentiation and desquamation of stratifying epithelia resulting from pathogenic mutations in more than 19 different ►keratin genes (1) (Fig. 1). Keratin defects may affect ►epidermis, hair, nails, sebaceous glands and mucous membranes. The nature and distribution of clinical features generally correspond to the function and expression pattern of the faulty keratin proteins. Because keratins are expressed in pairs that form obligate ►heterodimers, mutations in either gene generally produce a comparable ►phenotype resulting in ►genetic heterogeneity. Mutations are usually ►autosomal dominant and exert a deleterious, dominant-negative effect on the assembly of ►keratin intermediate filaments (KIF), causing weakening of the ►cytoskeleton and cell fragility, the hallmarks of keratin disorders (Fig. 2). At least 7 distinct clusters of mutations have been recognized, of which mutational ►‘hot spots’ at the beginning and end of the central rod domain are most common and consistently produce a severe phenotype (1). In addition, a few ►autosomal recessive mutations leading to the ‘►knock-out’ of a keratin in the basal epidermis have been described in ►epidermolysis bullosa simplex.

Characteristics

Introduction

Heritable skin disorders comprise a group of phenotypically diverse conditions with variable degrees of skin involvement. In some diseases, there is only a minor, occasionally cosmetic, involvement of the skin, hair and/or nails, whereas in other conditions, skin manifestations can be part of severe multi-system disorders with significant morbidity or even premature demise of the affected individuals. Over the past decade or so, with the advent of molecular genetics in general, spectacular progress has been made in understanding the genetic basis of many heritable skin diseases, and in

fact, mutations have now been identified in some 200 genes in these diseases (2).

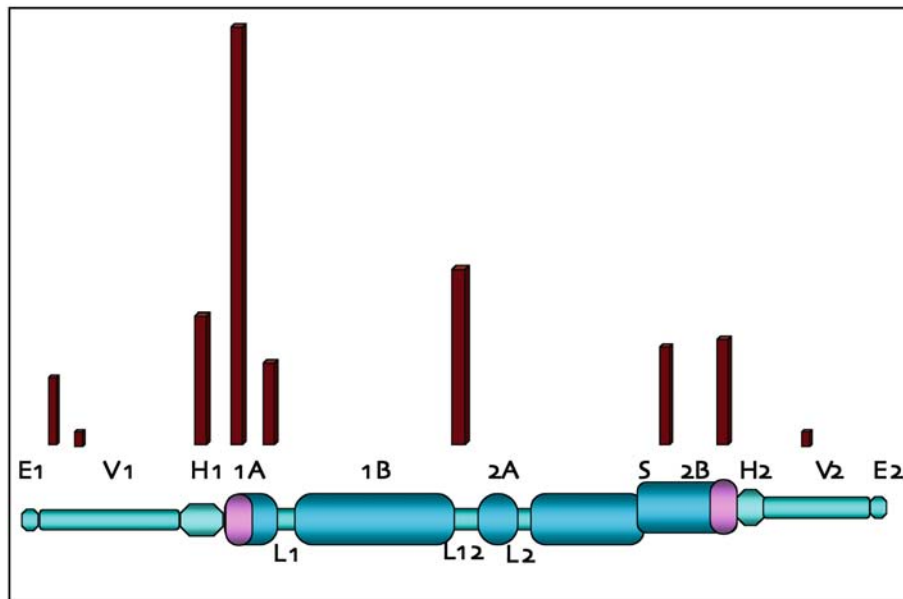
Examination of the mutation database in genodermatoses has revealed a number of both predictable and surprising genes. In several conditions, the mutated genes could have been predicted, on the basis of clinical, histopathological, immunohistochemical and/or ultrastructural analysis, to serve as the candidate gene/protein systems. For example, certain genes encoding structural proteins, such as basal cell keratins 5 and 14, were predicted to serve as candidate gene/protein systems in skin fragility syndromes and this prediction is now shown to be true in case of epidermolysis bullosa simplex (EBS). At the same time, a number of mutated genes have turned out to be rather surprising and the exact relationship of the mutations in the affected genes and their consequences at the clinical and morphological level are not well understood. An example of such conditions is ►steatocystoma multiplex characterized by dermal cysts of the pilosebaceous origin, which is known to result from mutations in the keratin 17 gene.

As an example of the progress in understanding the molecular basis of heritable skin disorders, we will highlight keratin disorders as a paradigm for cell fragility and perturbed differentiation.

Epidermolysis Bullosa Simplex (EBS)

EBS was the first human keratin disorder to be discovered in 1991. It is a clinically heterogeneous, ►congenital blistering skin disorder due to fragility of basal cells of the epidermis. Three major subtypes have been recognized and share a common disease pathology with an abnormal KIF network, i.e. ►cytolysis of basal cells leading to intraepidermal blister formation and skin fragility, while the differentiation of keratinocytes in the upper layers of the epidermis remains largely undisturbed (1). The structural abnormalities result from faulty keratin-5/keratin-14 heterodimers due to mutations in their corresponding genes, *KRT5* and *KRT14*. These are the primary keratins synthesized by undifferentiated keratinocytes in the basal layer of the epidermis, corresponding with the sites of disease pathology in EBS. Clinical features and severity of EBS subtypes are determined by the nature and location of the underlying mutations in the basal keratin genes *KTR5* and *KRT14* (3).

EBS Dowling-Meara (EBS-DM) is a common and the most severe subtype manifesting at birth with erythema, widespread blistering, erosions and areas of denuded skin. Later in life, spontaneous development of grouped blisters and severe thickening of palms and soles (►palmoplantar keratoderma) prevail. Typically, pathogenic defects are ►heterozygous ►missense



Heritable Skin Disorders. Figure 1 Schematic diagram of keratin protein depicting the structural domains and common mutation sites. The non-helical head domain consists of E1, V1 and H1 (only in type II keratins). The α -helical rod domain is composed of 4 segments, 1A, 1B, 2A and 2B (with stutter-s), which are interrupted by non-helical linker domains L1, L12 and L2. H2 (only in type II keratins), V2 and E2 form the non-helical tail domain. The highly conserved helix initiation and termination motifs, which are mutational hot spot regions, are shown in pink. Common sites of dominant mutations are depicted above the diagram. The height of bars reflects the relative frequency of mutations.

mutations clustering at 2 regions that encode the highly conserved boundaries of the alpha-helical rod domain of keratin 5 or keratin 14 (Fig. 1). An arginine codon of the ►helix initiation peptide in keratin 5 (R125) is most commonly mutated, probably because it contains a hypermutable ►CpG dinucleotide (4). While most keratin gene mutations in EBS are dominant, recessive *KRT14* mutations have been identified in 6 families, encompassing predominantly nonsense and splice site mutations leading to premature termination of protein translation and ablation of the affected keratin. While the 'knock-out' of *KRT14* usually results in severe EBS-DM, 2 mutations were noted to have a milder phenotype with minimal extracutaneous involvement. The Köbner variant of EBS (EBS-K) is characterized by milder blistering of the skin without apparent clustering, predominantly involving hands, feet and extremities. *KRT5* or *KRT14* mutations appear more widely distributed across the keratin polypeptides and may lie within and outside the highly conserved helix boundaries as well as in non-helical linker segments (1, 4) (Fig. 1).

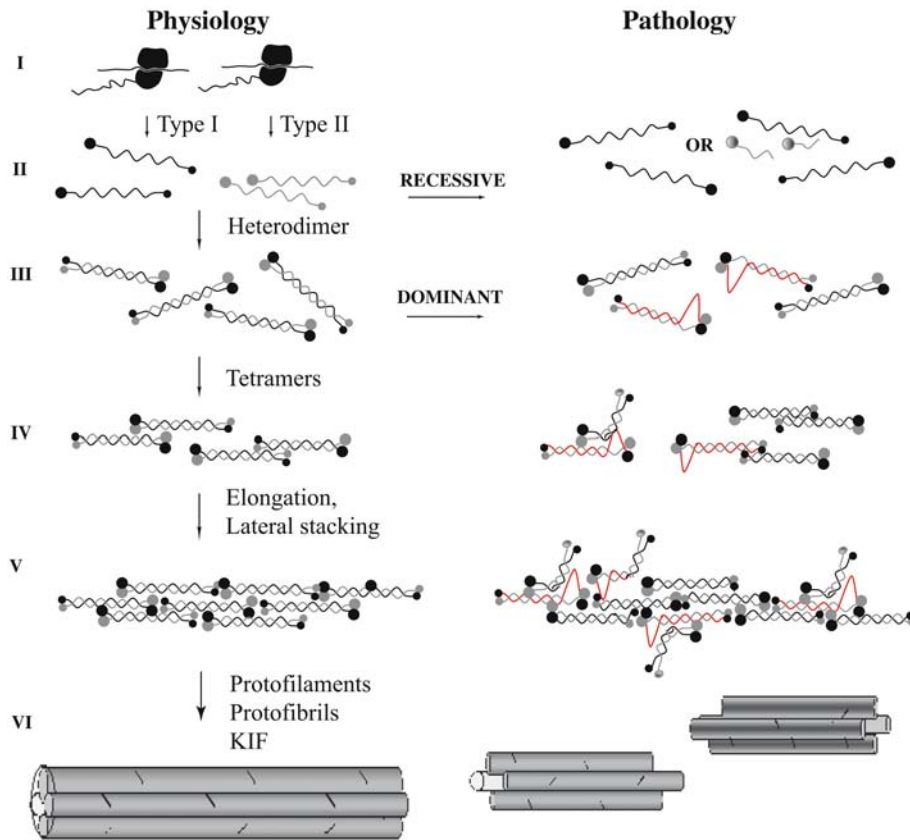
The most common Weber-Cockayne subtype (EBS-WC) is a mild, localized variant, in which serous blisters are confined to the hands, feet and areas of friction or trauma. In contrast to the other EBS subtypes, blisters are usually not present at birth but develop later

in response to mechanical irritation of the skin. In this mild form of EBS, pathogenic mutations lie in principle outside the helix boundaries in the rod domain of K5 or K14 or in the amino-terminal homologous domain of K5. Dominant point mutations usually result in amino acid substitutions, but small ►in-frame deletions in *KRT14* have also been described (1, 4).

A rare form of EBS with generalized or acral blistering and small, hyper- and hypo-pigmented spots that form a reticulate pattern has been reported as EBS with mottled pigmentation (EBS-MP). This variant is also associated with thickened, dystrophic nails and punctate PPK. The molecular basis of EBS-MP is a heterozygous missense mutation in exon 1 of *KRT5* that results in substitution of proline 25 with leucine (P25L) in the non-helical head domain of K5. This mutations was consistently detected in each of 7 unrelated families tested to date (4).

Epidermolytic Hyperkeratosis (EHK; Bullous Congenital Ichthyosiform Erythroderma of Brocq; Bullous Ichthyosis)

This autosomal dominant congenital ►ichthyosis manifests with erythroderma and widespread areas of erosions and denuded skin. While skin fragility decreases with age, patients develop progressively



Heritable Skin Disorders. Figure 2 Physiology and pathophysiology of keratin intermediate filaments (KIF). The syntheses of type I and type II keratin polypeptides and their assembly into keratin intermediate filaments are depicted on the left side of the figure. The keratin mRNAs are translated on the ribosomes of epithelial cells, which synthesize these keratins (I). In the cytoplasm, a type I and a type II keratin polypeptide align in parallel and register (II) and oligomerize to obligate heterodimers (III). Pairs of heterodimers align in an antiparallel, mostly overlapping fashion to tetramers (IV), which subsequently polymerize to elongated chains packed into keratin intermediate filaments (V). The presence of a mutation in a keratin gene can lead to different pathological processes depicted on the right side of the figure. For example, mutations that introduce a premature termination codon or affect mRNA splicing may result in the synthesis of truncated polypeptides. Subsequent nonsense-mediated mRNA decay or enhanced degradation of the truncated polypeptide results in absence of the mutant protein and lack of formation of corresponding heterodimers and KIF, usually associated with severe disease (EBS-DM) (I). Alternatively, the truncated or altered keratin polypeptides can prevent the heterodimer formation and KIF assembly (II). In autosomal dominant keratin disorders, the majority of mutations result in non-conservative amino acid replacements at sites of high sequence conservation (helix boundaries). The mutant keratin polypeptides interfere in a dominant negative manner with head-to-tail interactions and proper alignment of heterodimers (IV) as well as elongation and lateral packing during KIF assembly (V), thus producing a severe phenotype. Mutations with a less severe phenotype reside outside the helix boundaries and may have a more subtle effect on KIF assembly or affect keratin phosphorylation or interaction with other proteins.

severe hyperkeratosis with a verrucous, ridged or cobblestone surface. However, there is considerable phenotypic variability giving rise to at least six clinical subtypes. Structural and ultrastructural abnormalities of keratinocytes are distinct and described as ‘▶**epidermolytic hyperkeratosis**’. The KIF are fragmented and clumped leading to cytolysis of the upper epidermis and small intraepidermal blisters, combined with massive thickening of the stratum corneum

(hyperkeratosis) and epidermis. Epidermolytic hyperkeratosis is caused by heterozygous mutations in the genes encoding keratin 1 and keratin 10 (*KRT1*, *KRT10*). These keratins are expressed in the differentiated suprabasal and granular layers of the epidermis, which are the sites of disease pathology in this disorder. Mutations in the keratin 1 gene are usually associated with severe palmoplantar keratoderma, whereas mutations in the keratin 10 gene spare palms

and soles because the gene is not expressed at these locations. About 50% of all cases harbor ►**sporadic mutations**. As with EBS, mutations cluster at the boundaries of the α -helical keratin rod (mutational ‘hot spots’) and represent predominantly missense mutations that alter conserved amino acid residues. Mutations in the remainder of the rod are rare and generally associated with a milder and often unusual phenotype, such as ‘cyclic ichthyosis with epidermolytic hyperkeratosis’ and ‘annular epidermolytic ichthyosis’. A third category of mutations is known to affect the variable head and tail domains of keratin 1 resulting in severe palmoplantar keratoderma and ►**ichthyosis hystrix Curth-Macklin** (5). Finally, postzygotic ►**somatic mutations** in *KRT1* or *KRT10* have been detected in patients with a mosaic variant of epidermolytic hyperkeratosis, which is characterized by typical skin lesions following the lines of Blaschko. Because mutations may also involve gonadal cells (►**germline mosaicism**), they can be transmitted from the germline to the offspring resulting in full-blown, generalized disease.

► **Ichthyosis Bullosa of Siemens (IBS; Ichthyosis Exfoliativa)**

This congenital ichthyosis overlaps with mild epidermolytic hyperkeratosis. Characteristics are superficially denuded areas described as ‘molting’ or ‘Mäuserung’, which develop due to superficial blistering and shedding of the stratum corneum. Histopathological abnormalities are similar to EHK but are limited to the granular cell layers, which express keratin 2e. The disorder is due to heterozygous missense mutations in the *KRT2e* gene clustered at sites corresponding to the boundaries of the rod domain of keratin 2e, which preferentially alter a glutamic acid codon (E493K) in the helix termination peptide (1, 4).

► **Epidermolytic Palmoplantar Keratoderma (EPPK)**

In this keratinopathy, epidermolytic hyperkeratosis and occasional blistering are mostly confined to palmar and plantar skin. Typically, a sharp, erythematous border surrounds the waxy-appearing thickened skin of palms and soles. The sites of skin pathology correspond with the expression pattern of keratin 9 and disease-causing missense mutations have been identified in the *KRT9* gene, all of which cluster at the helix initiation motif (1, 4).

► **Pachyonychia Congenita (PC), Focal Non-epidermolytic Palmoplantar Keratoderma and Steatocystoma Multiplex**

PC is characterized by thickened, hard, convex nails, variably associated with follicular hyperkeratosis and plaque-like hyperkeratosis of palms and soles (focal

palmoplantar keratoderma). In PC type I, focal PPK is severe and oral leukokeratosis is common. This variant is caused by germline mutations in the genes encoding the keratin pair K6a and K16. In PC type II, patients often have natal teeth, hoarseness and pili torti as well as multiple steatocysts (dermal cysts arising from the pilosebaceous unit), which appear after puberty. Molecular changes are consistently detected in the co-expressed K6b and K17 genes. Interestingly, *KRT16* mutation may also result in isolated focal PPK and *KRT17* mutations may lead to steatocystoma multiplex, an autosomal dominant disorder with pilosebaceous cysts without other associated features, underscoring the phenotypic variability of the disease phenotype (4).

► **Monilethrix**

This congenital hair disorder manifests with fragile, easily breakable hair resulting in alopecia and short stubble of hair. On microscopic examination, hair shafts appear beaded (moniliform) due to alternating constrictions and elliptical nodes. Other features include follicular hyperkeratosis, especially on the nape of the neck, scalp, and sometimes the extremities, and nail dystrophy. Monilethrix results from missense mutations in the keratin genes *hHb1* and *hHb6*, which encode basic ‘hard’ or ‘trichocyte’ keratins specifically expressed in the hair cortex (4).

Other Keratinopathies

Following the paradigm of keratin disorders of the skin and its adnexae, mutations in other tissue-specific keratin genes have been detected. An example is ►**Meesmann corneal dystrophy**, a congenital, superficial keratitis with formation of corneal epithelial microcysts that may lead to photophobia, contact lens intolerance and recurrent decline of visual acuity. Keratin gene mutations affecting the helix boundary motifs were detected in *KRT3* and *KRT12*, both of which are co-expressed in the stratifying corneal epithelium. Another example is mutations in the mucosal keratin genes *KRT4* and *KRT13* resulting in ►**white sponge nevus**, a congenital disorder with benign, ‘spongy’ hyperkeratosis (leukokeratosis) of the mucous membranes of mouth, tongue, esophagus, nasopharynx, urogenital and anal regions (4).

Cellular and Molecular Regulation

Keratin proteins are the major constituents of epithelial cells, representing up to 85% of the cellular protein. They represent a large family of over 49 proteins that form intermediate filaments and are expressed in tissue and differentiation specific patterns (5). Keratins are expressed in pairs of acidic (type I) and basic (type II)

proteins, the genes of which cluster on chromosomes 17q12-q21 and 12q11-q13. Keratin monomers are organized as a central, alpha-helical rod domain flanked by variable, non-helical head and tail domains (Fig. 1). They form obligate heterodimers, which polymerize and assemble into intermediate filaments. The discovery of keratin gene mutations in numerous cell fragility disorders has fostered our understanding that keratins are important cytoskeletal proteins providing the structural stability and flexibility of epithelial cells. To date, over 80 distinct mutations have been identified in *KRT5* and *KRT14* in EBS, and over 36 distinct pathogenic mutations in *KRT1* and *KRT10* in EHK. This wealth of data has been instrumental in recognizing the specific functions of keratins and individual protein domains as well as for [▶genotype-phenotype correlations](#). For example, we have learned that the site and pathology of human keratin disorders are primarily determined by the tissue and differentiation-specific expression profile of the faulty keratin as illustrated above. Disease severity, however, predominantly depends on the position of the altered amino acid residue (Fig. 1). Between 60% and 80% of all mutations in EBS and EHK, respectively, alter the conserved boundaries of the rod segment, known as helix initiation peptides and helix termination peptides, with half of them involving a conserved arginine codon in the helix initiation motif of type I keratins. Functional [▶in vivo](#) and [▶in vitro](#) studies demonstrated that the helix boundaries are zones of overlap between keratin heterodimers during filament assembly and that mutant keratin molecules perturb proper keratin alignment and thereby oligomerization, filament assembly and integrity (Fig. 2). Consequently, the KIF network is fragile, disrupted and less resistant to stress, thus compromising mechanical strength and cell integrity of the epidermis, leading to cytolysis, blistering and generally a severe phenotype. The age-dependent improvement of skin fragility observed in EBS might be attributable to the compensatory overexpression of the alternate (hyperproliferative) keratins 6 and 16 in affected skin. On the other hand, blistering and cell fragility in EHK and other suprabasal keratin disorders are in general less severe than in EBS because basal keratin proteins, which may persist throughout the differentiated cell layers, provide a KIF scaffold and reinforce the cytoskeleton. Other mutation sites, such as the homologous domain in the head of type II keratins or non-helical linker and stutter motifs are associated with a milder phenotype as mentioned earlier.

In contrast, keratin mutations outside the central rod segment, in the variable head and tail domains, often differ considerably in pathological features and pathogenesis. KIF appear normal and cell fragility or

cytolysis are not observed, confirming experimental *in vitro* observations that non-helical keratin tail domains are not necessary for KIF oligomerization, alignment and assembly. Since the keratin ends vary greatly in sequence and length between different members of the keratin family, they are thought to confer unique and perhaps tissue-specific properties. The tail domain of type II keratins, for example, might be instrumental in bundling of KIF, anchorage of the cytoskeleton to the nucleus and interactions with loricrin, a major component of the cornified cell envelope (5). Certain areas of the keratin head have been shown to engage in interactions with loricrin and involucrin as well as desmosomal proteins. In addition, investigations of the proline 25 residue in K5 implicated in EBS with mottled pigmentation suggest that the keratin head domain binds directly to a cytoplasmic dynein cargo complex that transports melanosomes, thus unraveling the basis for abnormal pigment distribution in this rare disorder.

While the major molecular and cellular mechanisms leading to cell fragility and blistering in keratin disorders have been recognized, those responsible for hyperkeratosis remain to be elucidated. The mechanisms for inducing hyperproliferation and decreased desquamation of the epidermis are not well understood. Recent studies in patients and mouse models of EHK demonstrated a dramatic loss of [▶skin barrier](#) function with increased [▶transepidermal water loss](#) despite a normal-appearing stratum corneum. It is possible that the massive hyperkeratosis seen in many keratin disorders is a physical compensation to overcome this permeability barrier defect but the exact pathophysiological processes are still unknown.

Clinical Relevance

Advances in molecular genetics and biology of inherited skin disorders have impacted and benefit patient care already in different ways. Identification of specific mutations and genotype-phenotype correlations advance our ability to diagnose these rare disorders with certainty, allow accurate family counseling, offer the opportunity for reliable and early [▶prenatal testing](#) and will potentially lead to the development of targeted therapies for these disorders. DNA based mutation testing is available for numerous heritable skin disorders and has already provided means for recognizing and delineating the phenotypic spectrum of disorders, such as clinical subtypes of epidermolysis bullosa and EHK or in connexin disorders and establishing genotype-phenotype correlations (1, 2, 4). Molecular testing increasingly becomes a valuable tool to assist in early and reliable diagnosis in heritable disorders that are difficult to

distinguish solely on clinical grounds. This has important prognostic implications regarding the severity and natural course of the disease as well as preventive or therapeutic interventions where possible. Another direct benefit of molecular genetics relates to improved genetic counseling in families with increased ►[recurrence risk](#) for a disorder in the same or subsequent generations, especially when it is not possible to distinguish between autosomal dominant and recessive form of inheritance. Molecular testing in patients with sporadic, severe EBS-DM, for example, allows distinction of *de novo* occurring keratin mutations on one allele (autosomal dominant inheritance) from recessive keratin mutations on both alleles (autosomal recessive inheritance). In the first scenario, the risk of the parents having another affected child is very low whereas in the latter scenario the risk is one in four. At the same time, patients with a dominant, *de novo*, keratin gene mutation face a risk of 50% of having an affected child. Due to the low carrier frequency of EBS in the general population, the risk of having an affected child for a patient with recessive EBS is very small, barring consanguineous unions. Recognizing the specific molecular abnormalities underlying inherited skin disorders has tremendously advanced the development of DNA-based prenatal testing in families with severe forms of disease. Such testing is readily available for EBS, EHK and a wide range of other genodermatoses and can be performed in the first trimester of pregnancy, either from ►[chorionic villi samples](#) as early as the 10th week of gestation or from amniotic fluid from the 15th week on. Another newly developed method to determine the genotype of a fetus even before the pregnancy is established is ►[preimplantation genetic diagnosis](#), which is performed in conjunction with ►[in vitro fertilization](#). Together, these methods have replaced troublesome mid-trimester fetal skin biopsies and present for many disorders the only means of recognizing an affected fetus.

Finally, the elucidation of the molecular basis of inherited skin disorders provides essential information for understanding of the complex cellular mechanisms leading to disease with the ultimate goal of developing new and efficient therapeutic interventions. An important step in this process is the successful generation of animal models for these disorders. They may emulate the human disease phenotype and permit elucidation of the specific functions and contributions of individual proteins during embryological development and for physiological processes in a given tissue. In addition, animal models, such as those for EBS and EHK with targeted expression of mutant keratin chains in the epidermis, will be needed to evaluate various new therapeutic strategies. ►[Ex vivo](#) or *in vivo* gene

correction and gene and protein replacement approaches are currently under intense scrutiny for various genodermatoses, including keratin disorders, epidermolysis bullosa and recessive ichthyoses. Although much more time, effort and technological breakthrough will be needed to overcome the current problems with stem cell targeting, efficacy, long-term application, safety and other difficulties, durable gene therapy will undoubtedly become feasible for severe and debilitating genodermatoses in the future.

References

1. Lane EB (1994) Keratin diseases. *Curr Opin Genet Dev* 4:412–418
2. Uitto J, Pulkkinen L, Ringpfeil F (2002) Progress in molecular genetics of heritable skin diseases: the paradigms of epidermolysis bullosa and pseudoxanthoma elasticum. *J Invest Dermatol Symp Proc* 7:6–16
3. Fine J-D (2000) Inherited Epidermolysis Bullosa: Selected Outcomes and a Revised Classification System. *Pediatr Dermatol* 17:75–83
4. Irvine AD, McLean WHI (1999) Human keratin diseases: the increasing spectrum of disease and subtlety of the phenotype-genotype correlation. *Br J Dermatol* 140:815–828
5. Sprecher E, Ishida-Yamamoto A, Becker OM et al (2001) Evidence for novel functions of the keratin tail emerging from a mutation causing ichthyosis hystrix. *J Invest Dermatol* 116:511–519

Heroin

Definition

Heroin (“smack,” “H,” “skag,” and “junk”) is a highly addictive drug. Street heroin is associated with serious health conditions. Heroin is processed from morphine, a naturally occurring substance extracted from the seedpod of the Asian poppy plant. Heroin usually appears as a white or brown powder. Heroin is injected, smoked or snorted. It is a misconception that the latter form of administration will not lead to addiction.

►[Addiction, Molecular Biology](#)

Heterochromatin

Definition

The eukaryotic genome is divided into euchromatin and heterochromatin, the latter representing a relatively

inaccessible form of chromatin that is structurally condensed and generally transcriptionally silent. Facultative heterochromatin defines chromatin that is capable of undergoing a transition between heterochromatic and euchromatic. According to the classical definition, it is a fraction of the chromatin that can be positively stained with Gimsa. From a cytogenetic point of view, heterochromatin is found around the centromeres and/or in the G-bands of metaphase chromosomes. It is rich in highly repetitive sequences, such as satellite DNA. More recently, the term heterochromatin has been used to indicate the transcriptionally inactive and highly compacted chromatin states.

- ▶ Centromeres
- ▶ Chromatin Acetylation
- ▶ Gene Silencing
- ▶ Genomic Imprinting
- ▶ Methylation of Proteins
- ▶ Nucleosomes
- ▶ Transgene Silencing
- ▶ Transcriptional Repression

Heterodimer/Heterodimeric protein

Definition

Heterodimer defines a protein consisting of two different polypeptide chains, which are encoded by two different genes.

- ▶ Heritable Skin Disorders
- ▶ Nucleotide Excision Repair

Heterogeneous, Genetically/ Heterogeneity, Genetic

Definition

The terms describe a single phenotype or disease with different genotypes that may originate from different mutations in one particular disease gene, or mutations in different genes.

- ▶ Atopy Genetics
- ▶ Heritable Skin Disorders
- ▶ Hereditary Hemostatic Defects and Recombinant Proteins for Treatment

Heterologous Inducers

Definition

During kidney development, unknown signals derived from the ureteric bud induce metanephric mesenchymal cells to differentiate into kidney tubules. In addition to the ureteric bud, a number of other tissues, so called heterologous inducers (for example embryonic spinal cord), can induce this process *in vitro*.

- ▶ Kidney

Heterologous Promoter Control

Definition

Heterologous promoter control refers to antigen-encoding sequences in genetic immunization, which are cloned into expression plasmids in which antigen expression is driven by strong viral promoter/enhancer sequences.

- ▶ DNA-based Vaccination

Heteronuclear NMR Experiments

Definition

Heteronuclear NMR experiments describe nuclear magnetic resonance (NMR) experiments in which the frequencies, and thus chemical shifts, of two or more different types of nuclei are recorded in the several dimensions of a multidimensional experiment.

- ▶ Multidimensional NMR Spectroscopy

Heteronuclear RNP Proteins

Definition

Heteronuclear RNPs (ribonucleoproteins) comprise of a large set of proteins that assemble on newly synthesized RNA and form an RNA-protein complex.

- ▶ Splicing

Heteronuclear Single Quantum Coherence

►HSQC

parent. A heterozygous individual with a mutated allele is also called a carrier.

- Bone and Cartilage
- Drosophila Model of Cardiac Disease
- Familial Hypercholesterolemia
- Genetic Screening in Populations
- Mutagenesis Approaches in the Zebrafish

Heteroplasmy

Definition

Heteroplasmy is a mixture within a cell of mutant and wild-type mitochondria.

- Hereditary Diseases Genetic Basis

Heterotrimeric Guanine Nucleotide-Binding Protein

Definition

A heterotrimeric guanine nucleotide-binding protein (G protein) is a complex of three proteins involved in signal transduction arising from activation of classical seven transmembrane-type receptors. Ligand binding to the G protein coupled receptor induces dissociation of the α and β - γ subunits, which then activate downstream pathways, e.g. adenylate cyclase and phospholipase C, respectively.

- Autosomal Dominant (Inherited Disorder)
- Diabetes Insipidus, a Water Homeostasis Disease
- G-Proteins and G-Protein Mutations in Human Disease
- Polycystic Kidney Disease, Autosomal Dominant
- Wnt/Beta-Catenin Signaling Pathway

Heterozygote (Heterozygous)

Definition

A heterozygote (or heterozygous organism) is a diploid organism possessing two different alleles at a given gene locus under analysis, one inherited from each

Hexosamine

Definition

Hexosamine defines a six-carbon monosaccharide, where at least one of the OH groups is replaced by an NH_2 group, such as glucosamine and galactosamine.

- Glycosylation of Proteins

Hexose

Definition

Hexose is a six-carbon monosaccharide, such as glucose, galactose, and mannose.

- Glycosylation of Proteins

HFE

Definition

HFE is a gene located on human chromosome 6. The gene product of the HFE gene is homologous to major histocompatibility complex (MHC) class I proteins, but unlike MHC class I molecules, HFE lacks the usual peptide binding cleft of HLA molecules and does not present peptides to T cells. Mutations of this gene are present in homozygous form in most white patients with hemochromatosis. However, most persons who are homozygous for such mutations do not develop clinical hemochromatosis. Thus, the mutations may be thought of as risk factors, rather than the sole cause of the disease.

- Hemochromatosis

HGPRT

► Hypoxanthine-Guanine-Phosphoribosyl-Transferase

HIF

► Hypoxia Inducible Factors

HH

► Hedgehog

HH Receptor

Definition

HH (hedgehog) receptors

► Hedgehog Signalling

HIF-Prolyl-Hydroxylase

Definition

HIF-prolyl-hydroxylase is a member of the 2-oxoglutarate dependent hydroxylase family, which hydroxylates prolyl residues within the oxygen-dependent degradation domain of HIF- α , allowing binding of the von Hippel-Lindau (VHL) tumour suppressor protein (pVHL), which induces subsequent ubiquitination by an E3 ubiquitin-ligase and proteosomal degradation.

► Hypoxia Inducible Factors

HHV-8

► Human Herpesvirus Type 8

HIF- α

Definition

► HIF-Prolyl-Hydroxylase

Hidden Markov Models

► HMM

Hierarchical Clustering

Definition

Hierarchical clustering is a particular type of clustering, which results in a hierarchy of clusters, typically presented as a dendrogram of clustered objects.

► Microarray Data Analysis

High Content Screening

Definition

High content screening comprises cell based assays that collect multiple kinds of information from each well of microplates. This data can be any combination of multiple targets, phenotypes, and cell subpopulation activities or different cell types. In contrast to HCS, enabling single cell evaluations, High Throughput Screening (HTS) typically makes only one measurement per well as a cell population average.

► Automated High-Throughput Functional Characterization of Human Proteins

► High-Throughput Approaches to the Analysis

High Content Screening

► Functional Assays

High Mannose-Type

Definition

High mannose-Type is an N-glycan containing primarily mannose, which usually has multiple non-reducing terminal mannose residues.

► Glycosylation of Proteins

High Performance Liquid Chromatography

Definition

High performance liquid chromatography refers to various forms of chromatography, where sample compounds are separated at high pressure, resulting in mobile phase velocities that are 10–60 times greater than conventional chromatographic techniques. Components of a mixture are resolved according to their physical properties such as their size, shape, charge, hydrophobicity, and affinity. The compounds are partitioned between a liquid mobile phase and a stationary phase, packed within a column. The stronger a compound interacts with the stationary phase, the longer the retention time on the column.

► Affinity Chromatography and *In Vitro* Binding (Beads)

► Mass Spectrometry: ESI

► Mass Spectrometry: Quantitation

► Protein Interaction Analysis: Chemical Cross-Linking

► Shotgun Libraries

High Throughput Functional Cell-Based Screening

► Automated High Throughput Functional Characterization of Human Proteins

High-Density Lipoprotein

Definition

High-density lipoprotein (HDL) is operationally defined as the class of lipoproteins from the serum isolated between 1.063–1.21 g/ml in ultracentrifugation. Large size of HDL is called HDL2 corresponding to 1.063–1.125 g/ml. Small size of HDL is called HDL3 corresponding to 1.125–1.21 g/ml. About 50% of HDL consists of apoA-I apoA-II and apoC. Several enzymes and transfer proteins are also located on HDL. Low HDL levels may indicate an increased risk of atherosclerotic heart disease.

► Dyslipidemia

► HDL

► Tangier Disease

High-HDL Syndrome

AKIHIRO INAZU, HIROSHI MABUCHI
Molecular Biochemistry and Molecular Biology
Laboratory, Graduate School of Medical Sciences,
Kanazawa University, Kanazawa, Japan
inazua@mhs.mp.kanazawa-u.ac.jp

Synonyms

► **High-density lipoprotein (HDL)** is an established negative coronary risk factor, in contrast with low-density lipoprotein (LDL) as a positive risk factor. On a quantitative basis, there are two syndromes of hyperalphalipoproteinemia (HALP) and hypoalphalipoproteinemia; the former indicates excess HDL in plasma, the latter insufficient amounts of HDL. Alpha-lipoprotein is named after the alpha-migration characteristics of HDL in agarose gel electrophoresis. Cholesteryl ester transfer protein (CETP) is also named lipid transfer protein-I. Hepatic triglyceride lipase (HTGL) is a tri-acyl glycerol lipase located in the liver and is also named hepatic lipase. ► **Scavenger receptor class B type I**, SR-BI is also named CD36 and LIMPII analogous-1 (CLA-1).

Definition

The recent cholesterol guidelines propose that low HDL-cholesterol is <40 mg/dl. The frequency of the low HDL state is 18% in men and 12% in women in Japan. Low HDL-cholesterol is usually associated with ► **hypertriglyceridemia**, ► **insulin resistance**, and

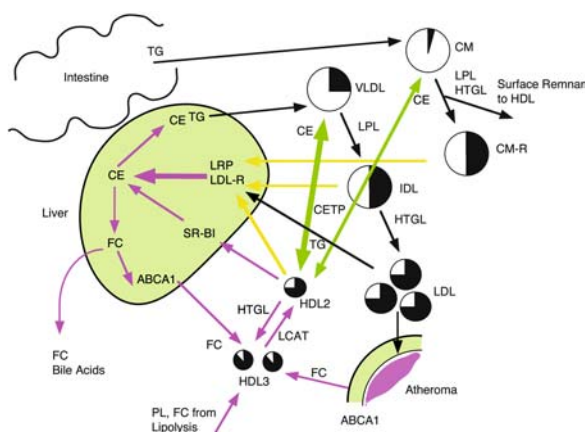
obesity. Severely low HDL-cholesterol is associated with genetic disorders such as Tangier disease, lecithin: cholesterol acyl transferase (LCAT) deficiency and apoA-I deficiency. In Japan, high HDL-cholesterol >100 mg/dl is considered to be hyperalphalipoproteinemia (HALP). The frequency of this condition is 0.8% in men and 0.5% in women. In a hospital-based study, the frequency of HALP values >100 mg/dl is 0.2%, 43 out of 46 subjects were women from the US.

Characteristics

HDL levels depend on the synthetic rate of apoA-I from liver and intestine, ►cholesterol efflux activity mediated by ►ATP-binding cassette transporter A1 from liver and peripheral tissues, and the efficacy of lipolysis during which free cholesterol (FC) and phospholipid (PL) are generated as surface remnants of chylomicrons and VLDL (see Fig. 1). The catabolic rate of HDL also affects HDL levels in plasma. The catabolic rate depends on receptor-mediated uptake of HDL through the LDL-receptor or the ►LDL-receptor related protein (LRP). These receptors can bind apoE-containing HDL as well as intermediate density lipoprotein (IDL) and ►chylomicron remnants. Both LDL-receptor and LRP actively take up chylomicron remnants by 55–75% and 20–25%, respectively.

SR-BI is a hepatic HDL-receptor, which mediates selective CE uptake from HDL, without endocytosis of HDL particles. The rate of neutral lipid transfer / exchange from HDL to other lipoproteins through cholesteryl ester transfer protein (CETP) determines the efficacy of remodeling of HDL lipids, thereby modifying the catabolic rate of HDL. Phospholipid transfer protein (PLTP) is another plasma lipid transfer protein, which mediates PL transfer from very low-density lipoprotein (VLDL) to HDL. ApoA-I is catabolized in kidney probably *via* megalin and cubilin; both are members of the LDL receptor superfamily (1). Megalin is also known as Heymann nephritis antigen or glycoprotein 330. HTGL activity hydrolyzes TG in IDL and HDL. High HDL-TG levels are accompanied by increased HDL size (HDL2) and a decreased catabolic rate of HDL.

In hyperalphalipoproteinemia (HALP), there are two genetic syndromes combined. One is CETP deficiency, and the other is HTGL deficiency. CETP deficiency is characterized by increased HDL-cholesterol > 100 mg/dl with low CETP activity. The normal level of plasma CETP is about 2 mg/l. HTGL activity or amount determined in a plasma sample after heparin injection is known as post-heparin lipolytic activity (PHLA). The differentiating features between CETP and HTGL deficiencies are HDL-TG and IDL-LDL levels; HDL-TG is low in CETP deficiency because of less TG transfer from VLDL to HDL, but it is high in HTGL



High-HDL Syndrome. Figure 1 Synthesis and catabolism of HDL including the reverse cholesterol transport pathway. The ApoE-dependent pathway, shown in yellow, involves receptor-mediated uptake of large HDL with apoE, IDL, and chylomicron remnant. CE is shown in black, and TG is shown in white in lipoproteins. CM, chylomicron; CM-R, chylomicron remnant.

deficiency. Cholesterol levels in VLDL-IDL-LDL are decreased in CETP deficiency, but increased in HTGL deficiency. HTGL deficiency is not always associated with increased levels of HDL-cholesterol, but some mutations are associated with increased TG levels in IDL-LDL and HDL and also with combined hypertriglyceridemia/ HALP and familial combined hyperlipidemia (2) or type III hyperlipoproteinemia (dysbetalipoproteinemia). Also, some cases of familial hypobetalipoproteinemia caused by truncated apoB, have HALP in addition. This condition results from a decrease in net CE transport from HDL to VLDL-IDL-LDL because of fewer CE acceptors, in contrast to normal acceptors but less carrier protein in CETP deficiency.

Secondary causes of HALP are chronic obstructive pulmonary disease, alcohol and drug-related effects of estrogen, insulin, steroids and some H₂ blockers (ranitidine). The mechanisms of secondary HALP are not yet understood.

Rare genetic causes of HALP are less characterized. Cases with excess lipoprotein lipase (LPL) activity associated multiple symmetric lipomatosis, or increased production of apoA-I, have been described. However, the molecular bases have yet to be determined. The oculocerebrorenal syndrome of Lowe is a X-linked recessive disorder, which is characterized by congenital cataracts, cognitive impairment and renal tubular dysfunction and frequently manifests HALP (3). The genetic cause of Lowe syndrome is a

deficiency of phosphatidylinositol 4,5-bisphosphate 5-phosphatase.

Another candidate gene for HALP is endothelial lipase (EL), also named endothelial-derived lipase (EDL). EL is synthesized by endothelial cells and primarily acts as a phospholipase. Increased EL activity reduces HDL levels, and its inhibition causes high HDL (large HDL) through decreased clearance of HDL-PL *in vivo*.

High HDL syndrome is generally associated with low coronary artery disease (CAD) risk, but some cases have evidence for CAD. The genotype/CAD relationship is less well-established in HALP. Premature CAD and xanthoma are not observed in homozygous CETP deficiency (4). Cross-sectional surveys of the TaqIB polymorphism frequently demonstrated that low CAD prevalence is associated with the 'low CETP genotype'. However, a protective effect of HDL, induced by CETP lowering, on CAD is not yet established in a prospective study or an intervention study in humans. HTGL deficiency produces VLDL remnant (i.e. IDL) accumulation as well as increased HDL levels, thereby making it somehow pro-atherogenic. Gene/environmental interaction between low HTGL activity and intake of saturated fat was demonstrated (5). Plasma HDL-cholesterol levels are high in low *HTGL* genotypes when saturated fat intake amounts to >30% of energy, but conversely it is low when fat intake exceeds 30%. Therefore, coronary risk may depend on fat intake.

Clinical Relevance

Genetics

Variation of plasma HDL-C levels has been shown to be heritable to ~50%. Yet it appears that haplotype blocks associated with activity changes in several genes as described below, represent multifactorial causes of HALP. In familial HALP, 2 persons with increased HDL-C >100 mg/dL within a nuclear family are not commonly found.

Autosomal recessive inheritance is usually seen in HALP caused by CETP deficiency except for a dominantly expressed mutation (e.g. D442G) (4). Half of HALP cases were diagnosed by *CETP* gene mutations. Genetic cause(s) of the remainder are not known except for HTGL deficiency, which is reported in <10 families in the world. Frequency of CETP deficiency is high in Japanese as compared to Caucasians.

Human and murine genome-wide scans suggested that quantitative traits for regulating HDL-C levels are rather overlapped between species (6).

There is no report indicating association between HALP and HDL receptors such as SR-BI and the

ectopic beta-chain of ATP synthase, except Imerslund-Grasbeck syndrome, autosomal recessive malabsorption of vitamin B12 due to cubilin deficiency (7).

Molecular Diagnostics

CETP (chromosome 16q13) consists of 16 exons, encoding 476 amino acids in the mature protein. CETP deficiency is caused by missense and nonsense mutations in addition to mutations affecting splicing. Intron 14 (+1) G→A and D442G are prevalent mutations in Japan. The allele frequency of the former mutation is 0.006 and the latter is also common in Asian populations with allele frequency 0.02–0.06 (4). Homozygosity of intron 14(+1) G→A mutation is caused by ▶[consanguinity](#) or mating between unrelated mutant carriers, resulting in severe HALP (CETP <0.1 mg/l, HDL-cholesterol 167 mg/dl). Frequency of ▶[compound heterozygotes](#) of intron 14(+1) G→A and D442G is also high (CETP 0.6 mg/l, HDL-cholesterol 134 mg/dl). Rare mutations are also reported in Caucasians.

HTGL (LIPC) (chromosome 15q21) consists of 9 exons, encoding 476 amino acids in the mature protein. V73M and L334F are susceptibility alleles for combined hypertriglyceridemia/HALP (2). A useful haplotype block involving –514 in the promoter could explain up to 30% of variability of HTGL activity in the general populations of various ethnic groups, thereby modify CAD risk (8).

EL (LIPG) (chromosome 18q21) consists of 10 exons, encoding 482 amino acids in the mature protein. Complete deficiency has not been documented. However, several polymorphisms are associated with HDL levels.

LPL (chromosome 8p22) consists of 10 exons, encoding 448 amino acids in the mature protein. S447X polymorphism lacks the last 2 amino acids (S-G), resulting in increased LPL activity by 4% and reduced CAD risk in the meta-analysis.

SR-BI (chromosome 12q24) consists of 13 exons, encoding two transmembrane regions and a large extracellular loop. A few polymorphisms are associated with both LDL and HDL levels (8).

In summary high-HDL syndrome is predominantly caused by CETP deficiency. Other genetic causes are rare or not well defined yet. Atherogenicity may depend on genetic causes; however, molecular diagnosis could be performed in only half of the HALP subjects with HDL-cholesterol >100 mg/dl.

References

1. Willnow TE (1999) The low-density lipoprotein receptor gene family: multiple roles in lipid metabolism. *J Mol Med* 77:306–315

2. Gehrisch S, Kostka H, Tiebel M et al (1999) Mutations of the human hepatic lipase gene in patients with combined hypertriglyceridemia/hyperalphalipoproteinemia and in patients with familial combined hyperlipidemia. *J Mol Med* 77:728–734
3. Charnas LR, Bernardini I, Rader D et al (1991) Clinical and laboratory findings in the oculocerebrorenal syndrome of Lowe, with special reference to growth and renal function. *New Engl J Med* 324:1318–1325
4. Inazu A, Koizumi J, Mabuchi H (2000) Cholesteryl ester transfer protein and atherosclerosis. *Curr Opin Lipidol* 11:389–396
5. Ordovas JM, Corella D, Demissie S et al (2002) Dietary fat intake determines the effect of a common polymorphism in the hepatic lipase gene promoter on high-density lipoprotein metabolism. *Circulation* 106:2315–2321
6. Wang X, Paigen B (2002) Quantitative trait loci and candidate genes regulating HDL cholesterol. *A Murine Chromosome Map. Arterioscler Thromb Vasc Biol* 22:1390–1401
7. Kozyraki R (2001) Cubilin, a multifunctional epithelial receptor: an overview. *J Mol Med* 79:161–167
8. Zambon A, Deeb SS, Pauletto P et al (2003) Hepatic lipase: a marker for cardiovascular disease risk and response to therapy. *Curr Opin Lipidol* 14:179–189

High-Mobility-Group Transcription Factor

Definition

High-mobility-group (HMG) proteins constitute a family of transcription factors that contain a structurally defined HMG domain, which binds DNA in a sequence specific manner and affects the chromatin structure. The name “high-mobility” refers to the unusual physico-chemical behaviour of these proteins.

►Neural Development

High-Throughput Analysis of Gene Function in Mammalian Cells Using Transfected Cell Arrays

DOMINIQUE VANHECKE, MICHAL JANITZ
Max Planck Institute for Molecular Genetics,
Department Vertebrate Genomics, Berlin, Germany
vanhecke@molgen.mpg.de
janitz@molgen.mpg.de

Definition

►DNA microarrays have revolutionized molecular biology and have greatly increased the rate at which genes are discovered. The functional analysis required for determining the putative causal role of newly discovered candidate genes and proteins in cellular processes however, and their potential clinical impact or suitability as drug targets, is slow and rate limiting. Functional validation is usually accomplished in molecular- and cell-based assays on a gene-by-gene basis, creating a bottleneck effect for the characterization of the plethora of targets arising from ►genomics and ►proteomics surveys. A recently described, novel cell-based microarray system, called the ►transfected cell array (TCA) (1) paved a way for high throughput gene analysis in the field of ►functional genomics. The principle of the TCA technique is based on the ►transfection of DNA or RNA molecules immobilized on a solid surface into mammalian cells with subsequent detection of the physiological effects caused by the introduction of the foreign nucleic acid on these cells. Briefly, full-length ►open reading frames of genes inserted in ►expression vectors are printed at a high density on a glass slide along with a lipid transfection reagent, using a robotic arrayer. Densities of up to 10,000 cell clusters per standard slide can be achieved. When the microarray of DNA constructs is covered with a ►monolayer of adherent cells only the cells growing on top of the DNA spots become transfected, resulting in the expression of specific proteins in spatially distinctive groups of cells. The phenotypic effects of this ►reverse transfection of hundreds or thousands of genes can be detected using specific cell-based bioassays, such as autofluorescence or immunofluorescence (Fig. 1).

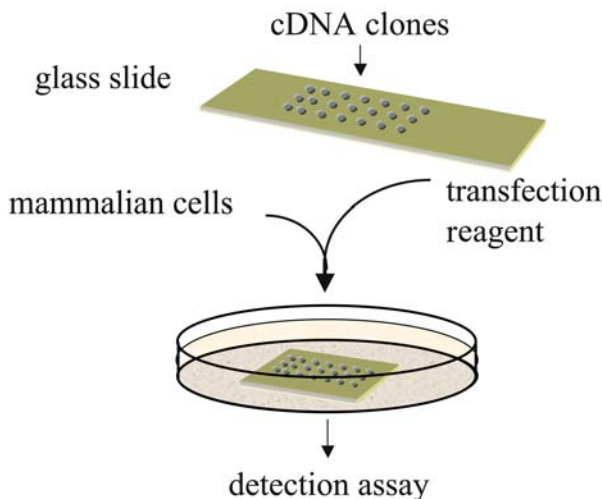
Characteristics

The transfected cell array (in short the cell array) combines microarray technology with cellular biology methods such as transfection and cell culture. The process of assay preparation and performance can be divided into three distinct steps namely (i) microarray preparation, (ii) reverse transfection and (iii) detection assay.

Microarray Preparation

In contrast to microarrays that are used for gene expression profiling studies, the DNA that has to be transfected in the cell arrays cannot be immobilized permanently on the glass surface. The reason is that in the following steps of the procedure, DNA should enter the cells growing on top of the DNA spot. Therefore the DNA to be arrayed is first dissolved in a gelatin solution, which facilitates transfer of the DNA from a solid support into the cells. The generation of arrays of DNA/gelatin solutions is performed using either stealth

Transfected Cell Array

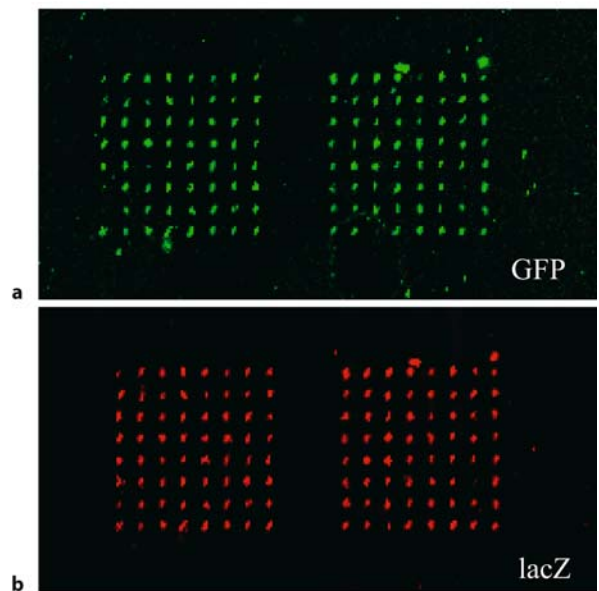


High-Throughput Analysis of Gene Function in Mammalian Cells Using Transfected Cell Arrays.
Figure 1 Schematic presentation of the principles of the transfected cell array technique.

pins (e.g. SMP4 from TeleChem) and any standard robotic arrayer or using a non-contact piezoelectric dispensing printer. The diameter of the single DNA spot can range from 100–300 μm whereas spots are separated from each other by 400–1000 μm , depending on the detection assay used. Design of the cell array, in terms of spot arrangement throughout the slide, does not differ from the classical array and is determined by the number of pins attached to the printing head of the roboter and individual settings of the robotic arrayer software. For example, in our laboratory we routinely prepare blocks of 8×8 spots that are separated by 1.5–2.0 mm (Fig. 2). 10 such blocks are prepared on a single slide, resulting in a total of 640 spots, though the density can be increased up to 10,000 spots per slide if required. Microarrays can be stored in a desiccator at room temperature for up to six months without any negative effect on reverse transfection efficiency.

Reverse Transfection

Before covering the array with a monolayer of cells the spots have to be incubated with transfection reagent so that the DNA becomes complexed with liposome-like structures and consequently can be taken up by the overlying cells. The time of incubation with the transfection reagent is crucial for successful transfection and has to be optimized for each cell line. Subsequently, the prepared array is placed in a culture



High-Throughput Analysis of Gene Function in Mammalian Cells Using Transfected Cell Arrays.
Figure 2 An example of signals detected from a transfected cell array. (a) Expression plasmid containing green fluorescence protein (GFP) has been spotted in 8×8 spot blocks and subsequently reverse transfected in an array format into the human epithelial kidney derived HEK293-T cell line. Green fluorescence signal indicates cells expressing GFP. Cells covering the area outside of the spots remain negative for GFP. (b) The same plasmid also contains the beta galactosidase (lacZ) gene, which can be detected with specific antibody labeled with Cy3 dye giving a red fluorescence. In this case lacZ detection serves as a control of reverse transfection efficiency. Detection of fluorescence signals was performed with a microarray scanner.

dish and covered with cells. The cells are allowed to grow on the slide surface for up to 3 days. Longer incubation times lead to a decline in the signals detected following transfection, because the viability of cells cultured at such a high density decreases gradually with time. Therefore, the transfected cell array is limited to those cellular events that are induced up to 72 h after reverse transfection.

Detection Assay

Depending on the nature of the molecular event to be monitored, signal detection can be performed either on living (e.g. calcium influx detection) or fixed cells (e.g. immunohistochemical protein detection). A common detection method uses signals emitted by fluorescent dyes (e.g. Cy3 or Cy5) coupled to primary or secondary [monoclonal antibodies](#). Alternatively autofluorescent reporter proteins such as [green fluorescence protein](#)

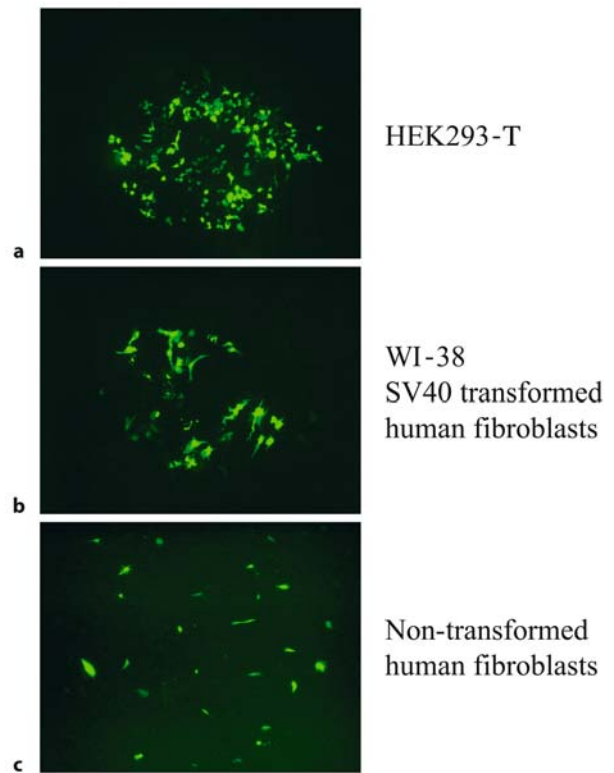
(GFP) can be readily detected. Preference in fluorescence detection is defined by the possibility of collecting signal intensities from the whole slide using standard microarray scanners. Alternatively, fluorescence microscopy allows for the detection of single or several cell clusters by collecting the signal intensities from single cells and application of statistical analysis of signal to background ratios.

In the case of studies where ►apoptosis or cell adhesion are of interest, abnormal morphology of the transfected cells can be observed as a functional read out. For example, particular cell clusters on the array can be examined for induction of cell fragmentation or expression of proteins serving as early markers of apoptosis such as annexin V.

Transfected Cell Array as a Functional Genomics Tool

Transfected cell arrays offer a number of advantages when used to study gene function. One of the most important features is that this assay provides a unique opportunity to study protein function in the context of the living cell. For obvious reasons studying mammalian cells using TCA gives several advantages over analyzing gene expression in yeast and *Drosophila melanogaster*. Not the least important is that proteins expressed in the cell array can undergo post-translational processing such as glycosylation and folding, which can be very distinct in different organisms and different cell types. Secondly, functional analysis of proteins in their natural environment guarantees that additional molecules involved in the functional process are present during the assay (e.g. cofactors, proteins involved in large complex formation etc). In fact the transfected cell arrays allow for the generation of protein arrays without the need for tedious purification of the individual proteins. Since the DNA arrays can be stored for months before subjecting them to reverse transfection, they can be converted into the ►protein arrays when needed.

There are a number of advantages in using TCA as compared to functional assays performed in microwell plate format. The array approach requires fewer cells per number of genes tested. This is especially relevant for human primary cells since only small numbers of these can be isolated out of a tissue and the *in vitro* expansion of these cells is rather limited. Although primary cells are notorious for their lack of capacity to be transfected, reverse transfection of these cells is possible but requires careful optimization of the experimental conditions for each cell type tested (Vanhecke et al., unpublished data) (Fig. 3). Transfected cell arrays also require far less DNA/RNA as well as transfection and signal development reagents as compared to assays performed in microwell plate



High-Throughput Analysis of Gene Function in Mammalian Cells Using Transfected Cell Arrays.

Figure 3 Comparison of spots of transfected cells obtained after reverse transfection of a GFP coding plasmid into three different cell types. Highest transfection efficiency (number of green cells and fluorescence intensity per cell) is obtained in the common cell line HEK293-T (a) and WI-38 human fibroblasts immortalized following transformation with SV40 virus (b). Hard-to-transfect primary human skin fibroblasts from adult healthy individuals (c) are less efficiently reverse transfected. Detection of fluorescence signals was performed with a fluorescence microscope with 10x objective.

format. Since in the case of high throughput technologies the cost of the single sample analysis must be reduced to an absolute minimum, cell arrays are at the moment the most cost-effective functional genomic tool available.

Cell arrays have already been used in a number of assays designed for high throughput gene and protein characterization, as for example in ►loss-of-function studies (2, 3) and protein-protein interactions screening (4).

Gene Silencing Studies

In the case of loss-of-function studies the cell array is combined with the technique of ►RNA interference

(RNAi) (5). ► **Small interfering RNAs (siRNAs)** are spotted in gelatin solution onto a modified glass surface using a standard robotic arrayer. The RNA array is subsequently treated with transfection reagent and covered with a monolayer of recipient cells. The decrease in expression of targeted genes is monitored using, for example, fluorescent-labeled monoclonal antibodies or autofluorescent reporter proteins. In order to control transfection efficiency, the siRNA can be labeled with rhodamine. Alternatively, the siRNA can be cotransfected with a reporter plasmid, which can be detected in parallel to the detection of the target protein. The transfection of siRNA can be performed in different experimental configurations. In one application, siRNA is co-transfected together with a plasmid that will express a ► **target gene**. This approach allows for the evaluation of the silencing capacity of different siRNAs specific for the same target gene (2). A more challenging application is the targeting of endogenous genes by reverse transfection of a genome wide collection of siRNA molecules. The cellular processes that are affected by silencing endogenous genes can be monitored by specific antibodies that can, for example, detect changes in the phosphorylation state of cell membrane bound receptors or transcription factors or that can be used to monitor changes in cellular localization of genes of interest. Alternatively, the effects of silencing genes can result in cellular processes such as apoptosis or changes in intercellular adhesion, which can be monitored by specific reagents (e.g. annexin V) or by simply analyzing morphological changes in the transfected cells.

Protein-Protein Interactions Studies

The cell array can also be used as a platform for high throughput analysis of protein-protein interactions. For example, in our laboratory we have developed the so-called THETA (two-hybrid-edition of transfected array) system, where the transfected cell array is combined with a mammalian ► **two-hybrid system** to identify proteins that interact with a protein of interest (4). The advantage of the mammalian 2-hybrid is that proteins interact in their natural environment, which is advantageous when posttranslational modifications or interactions with third party proteins are required. In the case of the THETA assay, bait and prey plasmid constructs were created whereby the bait is a known signal transduction molecule that is tested for its interaction with unknown proteins by coexpressing it with a huge number of single clones (preys) from a unigene cDNA library. Thus, thousands of human or rodent cDNAs (preys) can be screened with particular target proteins (baits) representing metabolic or signal transduction pathways.

Cellular Localization Studies

Because transfected proteins are expressed in their natural environment in the cell array, the synthesized proteins will also localize to their natural cellular compartments such as nucleus or cell membrane. Accomplishment of the human and mouse genome projects resulted in accumulation of extensive gene sequence information. However, information about the biological function of the identified genes remains a bottleneck of the post-genomic era. Thus assays providing simple functional information such as localization of the protein within the cell can be very helpful in elucidation of its function. For example, detection of protein expression on the surface of the cell will strongly suggest receptor function whereas location in the nucleus will indicate a role as transcription factor. Transfected cell array offers a robust platform for protein localization studies. Open reading frames of unknown genes can be linked to a ► **His tag** or GFP reporter protein in suitable expression vectors and are then subsequently transfected in the cell array. Cellular localization of the transfected proteins is detected either by specific anti-His antibody or directly by fluorescence of the GFP fusion protein and by counterstaining with dyes that target particular ► **organelles**. In this application the fluorescent signal is detected using a microscope equipped with a scanning platform and appropriate software for the automatic analysis of a large number of transfected spots. Although, the degree of automation of the whole procedure is limited in this application, we have been able to investigate several hundred cDNA clones derived from a human chromosome 21 library, out of which the localization of approximately 30% of expressed proteins could be unambiguously determined (Hu et al. submitted). This relatively low percentage is due to the fact that some proteins will be expressed below the detection level or will not be expressed at all as a result of improper out of frame cloning. However the high throughput of the method in terms of information provided per single experiment makes this approach still superior to the classical immunohistology methods for protein localization.

Clinical Relevance

Since the transfected cell arrays allow the analysis of many genes in parallel and furthermore can be adopted to many existing cell based assays it promises to become a robust and high-throughput means to profile disease related proteins or to study protein-protein and protein-drug interactions. The transfected cell array can be regarded as a sophisticated protein expression system where location of particular proteins on the slide can be identified by the coordinates of the arrayed cDNA. It can be used, for example, to identify

disease related proteins by analyzing the binding of sera of patients on arrays of expressed cDNA libraries, leading to identification of molecular diagnostic markers. The TCA is also well suited for screening of small molecules for their ability to specifically interact with protein targets such as G-protein-coupled receptors (GPCRs). In this case, the cell microarrays present a more advance alternative to standard protein arrays, since in TCA the proteins are *in situ* synthesized in the physiological environment of the cells, thus enabling proper post-translational protein folding and glycosylation. Furthermore, cell arrays represent a powerful functional genomics tool for target gene identification. Through the accurate selection of cell line characteristics (e.g. tumor-derived cell lines) and detection assays (e.g. induction of apoptosis) usage of cell arrays on a genome-wide scale can rapidly lead to identification of target genes related to particular diseases.

Acknowledgement

Dominique Vanhecke is supported by the German Federal Ministry of Education and Research (grant no. 0313068).

References

1. Ziauddin J, Sabatini DM (2001) Microarrays of cells expressing defined cDNAs. *Nature* 411:107–110
2. Kumar R, Conklin DS, Mittal V (2003) High-throughput selection of effective RNAi probes for gene silencing. *Genome Res* 13:2333–2340
3. Vanhecke D, Janitz M (2004) High-throughput gene silencing using cell arrays. *Oncogene* 23:8353–8358
4. Vanhecke D, Fiebitz A, Wagner F et al (2003) A high throughput mammalian two hybrid system for the detection of T helper specific signal transduction molecules. *Clin Immunol, Suppl* 1:144
5. Dykxhoorn DM, Novina CD, Sharp PA (2003) Killing the messenger: Short RNAs that silence gene expression. *Nat Rev Mol Cell Biol* 4:457–467
6. Lippincott-Schwartz J, Snapp E, Kenworthy A (2001) Studying protein dynamics in living cells. *Nat Rev Mol Cell Biol* 2:444–456

High-Throughput-Screening

- Functional Assays
- HTS

Hippocampus

Definition

The hippocampus is a part of the brain located inside the temporal lobe. It plays a central role in memory storage and also forms part of the limbic system. The name derives from its curved shape, which supposedly resembles that of a seahorse (greek: hippocampus). The hippocampus is the first region of the brain that suffers attack from Alzheimer's disease; memory problems and disorientation appear amongst the first symptoms.

- Brain
- Neutrophilic Factors

His-Tag/6xHis

Definition

His-Tag refers to a polyhistidine sequence that is used to tag recombinant proteins. The recombinant protein can then be purified using immobilized metal ions, traditionally nickel or cobalt, in a procedure termed immobilised metal chelate affinity chromatography (IMAC). Further analysis of the protein can be achieved by using an antibody that recognises a 6xHis (or 5xHis) sequence in proteins.

- Affinity Chromatography and *In Vitro* Binding (Beads)
- High-Throughput Analysis of Gene Function in Mammalian Cells Using Transfected Cell Arrays.
- Protein Tags
- Recombinant Protein Expression in Bacteria

Histocompatibility

Definition

Histocompatibility is the state in which a donor and recipient share antigens so that a graft is accepted and remains functional. It also refers to the study of factors that determine the acceptance or rejection of grafted tissues or organs. Histocompatibility testing is the performance of assays to determine whether recipient and donor are histocompatible.

Histone

Definition

Histone refers to a group of highly conserved proteins in eukaryotes with a very high proportion of positively-charged amino acids (lysine and arginine), which is involved in the packaging of the DNA into tightly packed chromosomes. The positive charge helps to tightly bind to DNA, regardless of its nucleotide sequence. There are five major types of histones in the somatic cell. H2A, H2B, H3, and H4 compose the nucleosome, the structural unit of the chromatin. H1 functions as a linker between nucleosomes. The N-terminal regions of the core histones are targets for epigenetic modifications such as acetylation, methylation, phosphorylation, poly(ADP-ribosylation), or ubiquitination that can influence the stability of the chromatin.

- ▶ [Chromatin Acetylation](#)
- ▶ [Genomic Imprinting](#)
- ▶ [Nucleosomes](#)
- ▶ [Polyglutamine Disease, the Emerging Role of Transcription Interference](#)
- ▶ [Protein/DNA Interaction](#)
- ▶ [Transcriptional Repression](#)
- ▶ [Translational Control in Eukaryotes](#)

Histone Acetylation

- ▶ [Chromatin Acetylation](#)

Histone Acetyltransferases

Definition

Histone acetyltransferases comprise of a class of enzymes that catalyze the acetylation of histones, with consequences for chromatin structure and function. The activity of these enzymes is opposed by histone deacetylases.

- ▶ [Chromatin Acetylation](#)
- ▶ [Polyglutamine Disease, the Emerging Role of Transcription Interference](#)
- ▶ [Polyglutamine Disease, the Emerging Role of Transcription Interference](#)
- ▶ [Transcription Factors and Regulation of Gene Expression](#)

Histone Code

Definition

The pattern of different posttranslational modifications that occur on the N-terminal tails of histones is referred to as the histone code. It is hypothesized that this code is “read” or bound by different proteins.

- ▶ [Methylation of Protein](#)

Histone Deacetylase

Definition

Histone deacetylases comprise of a class of enzymes that catalyze the deacetylation of histones. Acetylation of the ϵ -amino group of lysine residues of histones reduces the net positive charges of the core histones, leading to a decrease in their binding affinity for DNA. The termini are subsequently displaced from the nucleosome, the nucleosome unfolds and provides access for transcription factors. Deacetylation of histones by histone deacetylases removes the acetyl group, allowing ionic interactions between the negatively charged DNA phosphate backbone and the positively charged amino termini of the core histones. This results in a more compact chromatin structure, which is not as accessible to the transcriptional machinery. While histone acetylation and hyperacetylation has been correlated with transcriptionally active chromatin, histone deacetylation is thought to be involved in repression of transcription.

- ▶ [Chromatin Acetylation](#)
- ▶ [Histone](#)
- ▶ [Polyglutamine Disease, the Emerging Role of Transcription Interference](#)
- ▶ [Transcription Factors and Regulation of Gene Expression](#)
- ▶ [Transcriptional Repression](#)

Histone Methylation

Definition

Histone methylation is a covalent attachment of methyl groups to lysine or arginine side chains in histone amino-termini by histone methyltransferases. H3 lysine 4 methylation correlates with gene activation, whereas

methylation of lysine residue 9 of histone H4 (H3–mLys9) generates a high affinity binding site for the silencing protein heterochromatin protein 1 (HP1), which is essential for the establishment and maintenance of heterochromatin.

►Chromatin Acetylation

Histone Methyltransferase

Definition

Histone methyltransferases transfer methyl groups to arginine or lysine residues present in the N-terminal “tail” region of histones.

►Transcriptional Repression

Histone Octamer

Definition

Histone octamer describes the complex of two molecules, each of histones H2A, H2B, H3 and H4.

►Nucleosomes

HIV-1

►Human Immunodeficiency Virus Type 1

HLA

►Human Leukocyte Antigens

HLA and Disease

KATSUSHI TOKUNAGA

University of Tokyo, Graduate School of Medicine,
Department Of Human Genetics, Tokyo, Japan
tokunaga@m.u-tokyo.ac.jp

Definition

The human ►major histocompatibility complex (MHC), located on the short arm of human chromosome 6, consists of three regions, i.e. classes II, III and I from the centromeric side toward the telomeric side. The nucleotide sequences of the entire human MHC regions over 3.6 megabases long were determined in 1999, and numerous known and unknown genes were localized (1). MHC is one of the most gene-rich regions in the human genome and contains more than 130 expressed genes. The human leukocyte antigen (HLA) gene family contains several genes including HLA-A, -B and -C belonging to class I and HLA-DRB1, -DRA, -DQB1, -DQA1, -DPB1 and -DPA1 belonging to class II. It is well known that these genes are involved in immune response control by presenting antigen peptides to T cells and in NK cell activity control by functioning as NK receptor ligands. However, a majority of the expressed genes located in MHC are non-HLA genes and include not only the genes of TAP1, TAP2, PSMB9(LMP2), PSMB8(LMP7) and tapasin (TAP-binding protein) involved in the antigen peptide presentation, a gene cluster of C4A, C4B, C2 and BF belonging to the complement system and others belonging to the immune system, but also many genes unrelated to the immune system.

Characteristics

Variations Within the MHC and Their Associations with Diseases

It has been demonstrated that the HLA gene family shows the highest degree of polymorphism in functional genes, and approximately 350 alleles for the HLA-DRB1 gene and over 500 alleles for the HLA-B gene have been identified. Many of the polymorphic sites in HLA genes are of functional significance, altering the pocket structures and affinities in which antigen peptides are sandwiched, thus associations with a number of diseases have actually been reported (Table 1). A majority of these MHC-associated diseases are autoimmune or infectious diseases, and are genetically classified as multifactorial diseases in which various genetic and environmental factors are involved. The associations between MHC and some of the diseases were commonly observed in different populations, as reported in the association of HLA-B27 with ankylosing spondylitis, HLA-DRB1*1501 and HLA-DQB1*0602 with narcolepsy and HLA-DR4 shared epitope with rheumatoid arthritis. Associations with other diseases vary in different populations. These findings may be interpreted in several ways. First, there may be variations in the repertoire of HLA alleles and/or environmental factors between populations. Secondly, HLA genes themselves might not be disease susceptibility factors. Rather, polymorphisms

HLA and Disease. Table 1 Some of the diseases strongly associated with HLA genes

Disease	Population	HLA allele or haplotype
Ankylosing spondylitis	European	B*2705
	Japanese	B*2704, B*2705
Narcolepsy	European	DRB1*1501-DQA1*0102-DQB1*0602
	Japanese	DRB1*1501-DQA1*0102-DQB1*0602
Type I diabetes	European	DQA1*0301-DQB1*0302
	Japanese	DRB1*0405-DQA1*0301-DQB1*0401
Rheumatoid arthritis	European	DRB1*0401, DRB1*0404, DRB1*0101
	Japanese	DRB1*0405
Systemic lupus erythematosus	European	DRB1*0301, DRB1*1501
	Japanese	DRB1*1501
Psoriasis Vulgaris	European	Cw6
	Japanese	Cw6, Cw7
Behcet	Chinese	B51
	Japanese	B51

of non-HLA genes adjacent to HLA genes in the MHC region might be primary factors. Furthermore, it is also possible that not a single gene but multiple susceptibility genes are situated in MHC. In addition to the determination of the nucleotide sequences of the entire human MHC, two kinds of most useful polymorphic markers, microsatellites and ►single nucleotide polymorphisms (SNPs), have recently been extensively screened in MHC (2, 3). This led to the establishment of the strategy of identifying the primary susceptibility genes, on the basis of detailed linkage disequilibrium analyses and inter-population comparative analyses. The present status of MHC-disease association studies is described in the following by discussing four diseases: type 1 diabetes, ►rheumatoid arthritis, ►systemic lupus erythematosus, and ►narcolepsy.

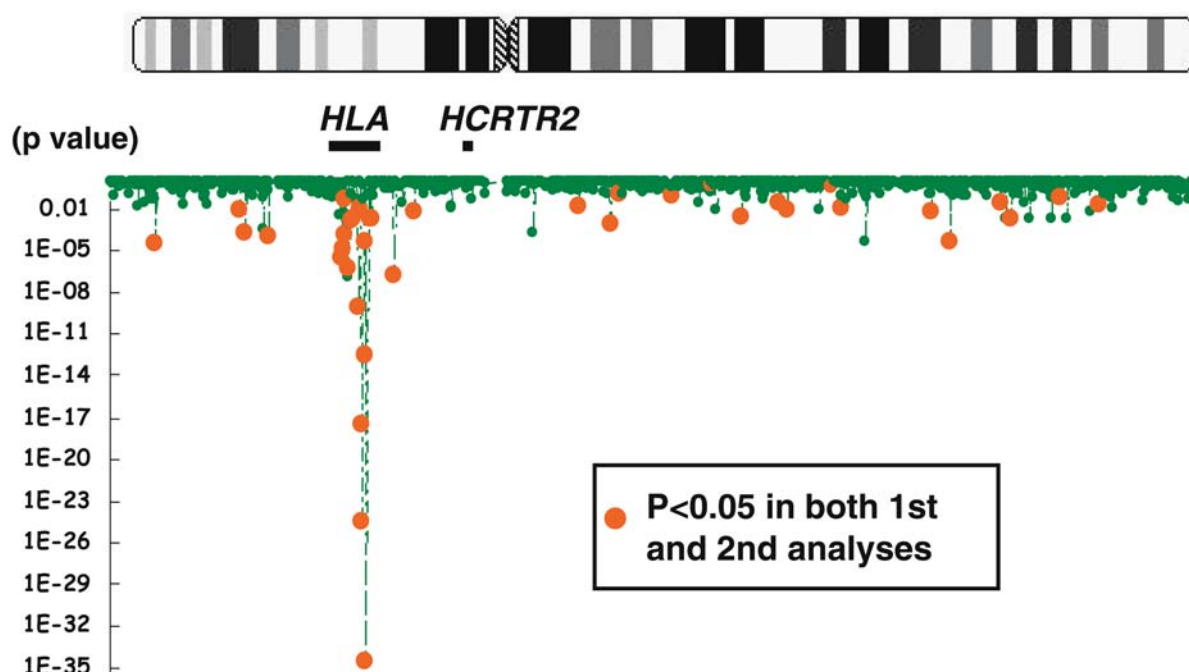
Type 1 Diabetes

MHC is a region showing the most prominent susceptibility in type I diabetes, detected by genome-wide linkage analyses with affected sib pairs (IDDM1). The strong associations with type I diabetes of HLA-DRB1 and HLA-DQB1 genes in the class II region in particular, have been established. The association of HLA-DRB1*0301 and HLA-DRB1*0401 with type I diabetes in European populations have been reported. It was proposed that the primary factor is attributable to the 57th amino acid of DQB1 gene product being non-Asp. However, this view cannot explain the

associations of DRB1*0405 and DRB1*0901 with type 1 diabetes in the Asian populations. Thus, the current view is that the degrees of susceptibility or resistance shown by various alleles of DQB1 and DRB1 genes are hierarchical. Furthermore, the potential existence of a second susceptibility gene in the class I or class III region has been reported in many publications recently. The HLA-B and TNF genes and the genes adjacent to them, in particular, were considered to be the candidates for the second susceptibility gene (4). The basis of onset of type I diabetes could be hypothesized as follows; class II genes may determine the immune response to diabetogenic peptides, while class I genes may determine the degree of target cell destruction.

Rheumatoid Arthritis

Several genome-wide linkage analyses with affected sib pairs demonstrated that MHC is a major susceptibility region for rheumatoid arthritis, and the association of HLA-DR4 with the disease has been established in most populations. When these reports were examined in detail, a strong association of HLA-DRB1*0401 and DRB1*0404 with rheumatoid arthritis was observed in the European populations, whereas there was a strong association of HLA-DRB1*0405, a different allele, with the disease in the East Asian populations and of HLA-DRB1*0101 in the Jews, who rarely have these DR4 genes. From these findings, it was



HLA and Disease. Figure 1 Chromosome-wide association analysis of human narcolepsy with 1,265 microsatellite markers on human chromosome 6.

established that the sequence around the 70th amino acid of the HLA-DRB molecule being dominated by basic amino acids (QRRAA or RRRRAA: shared epitope) is a shared susceptibility factor. In addition, the existence of a second weak susceptibility locus has been suggested recently. The probable region was narrowed down by a high-density microsatellite marker analysis, and an association analysis was then carried out using SNPs. Thus, it was reported that the *IkBL* gene in the class III region might be the second susceptibility gene (3).

Systemic Lupus Erythematosus

Non-parametric linkage studies with affected sib pairs have identified the MHC region as a candidate susceptibility region for systemic lupus erythematosus. In the European populations, both HLA-DRB1*0301 and HLA-DRB1*1501 were reported to be associated with the disease. The association of HLA-DRB1*1501 with the disease was also observed in the Japanese, whereas HLA-DRB1*0301 was very rare. In the Thai population, in contrast, another allele of the HLA-DRB1*15 group, HLA-DRB1*1502, was associated with the disease (5). Accordingly, HLA-DRB1 polymorphisms may predispose to systemic lupus erythematosus in different populations.

Narcolepsy

The association of the HLA-DRB1*1501-HLA-DQB1*0602 haplotype with narcolepsy is one of the

strongest MHC-disease associations, comparable to the association of HLA-B27 with ankylosing spondylitis. The elucidation of causative genes in dog and mouse models for narcolepsy contributed to the findings that a reduced cerebrospinal fluid level of orexin (hypocretin) leads to the onset of narcolepsy. The proposed basis of onset of narcolepsy is that an autoimmunity, in which the HLA class II haplotype is involved, destroys cells containing orexin (hypocretin). We recently performed an extensive association analysis on the entire chromosome 6 using 1,265 microsatellite polymorphic markers (6). A very strong association of the markers around the HLA class II region with narcolepsy ($P < 10^{-34}$) was noted, as had been predicted (Fig. 1). In contrast, the orexin (hypocretin) receptor 2 gene (*HCRT2*), which is also located on the chromosome 6, was not associated with human narcolepsy. We already reported that the polymorphism in the *TNFA* gene promoter region might be a weak genetic factor in the Japanese population (7). The same association was also observed in the German population, recently.

Clinical Relevance

Associations of MHC with a large number of multifactorial diseases have been reported, but the exact mechanisms of onset of the diseases have not been elucidated. Hence, few of these reports have led to the development of new therapies for the diseases until now.

However, associations between different clinical characteristics of various diseases and particular HLA alleles have been reported. Therefore, the typing of HLA gene polymorphisms is diagnostically significant and is expected to become essential to achieve personalized medicine for immune disorders in the future.

References

1. The MHC Sequencing Consortium (1999) Complete sequence and gene map of a human histocompatibility complex (MHC). *Nature* 401:921–923
2. Okamoto K, Makino S, Yoshikawa Y et al (2003) Identification of IkbL as the second major histocompatibility complex-linked susceptibility locus for rheumatoid arthritis. *Am J Hum Genet* 72:303–312
3. Geraghty DE, Daza R, Williams LM et al (2002) Genetics of the immune response: identifying immune variation within the MHC and throughout the genome. *Immunol Rev* 190:69–85
4. Zavattari P (2001) Conditional linkage disequilibrium analysis of a complex disease superlocus, IDDM1 in the HLA region, reveals the presence of independent modifying gene effects influencing the type 1 diabetes risk encoded by the major HLA-DQB1, -DRB1 disease loci. *Hum Mol Genet* 10:881–889
5. Sirikong M, Tsuchiya N, Chandanayingyong D et al (2002) Association of HLA-DRB1*1502-DQB1*0501 haplotype with susceptibility to systemic lupus erythematosus in Thais. *Tissue Antigens* 59:113–117
6. Kawashima M, Ikuta T, Tamiya G et al (2004) Genome-wide association study of narcolepsy: initial screening on chromosome 6. In: Hansen JA, Dupont B (eds) *HLA 2004: Immunobiology of the Human MHC*, IHWG Press, Seattle
7. Hohjoh H, Nakayama T, Ohashi J et al (1999) Significant association of a single nucleotide polymorphism in the tumor necrosis factor-alpha (TNF-alpha) gene promoter with human narcolepsy. *Tissue Antigens* 54:138–145

HLH

► [Helix-Loop-Helix](#)

HMG Box

Definition

HMG (high mobility group) box is a conservative DNA binding motif of ~ 75 amino acid residues that is characteristic of high mobility group proteins and

several transcription factors (► [High-mobility-group transcription factors](#)). The HMG box is capable of binding and bending DNA. The sex-determining gene on the Y chromosome (SRY) is one example of a HMG box protein belonging to the SOX subfamily.

► [Colorectal Cancer](#)
 ► [RNA Polymerase I](#)
 ► [SRY – Sex Reversal](#)

HMM

Definition

In sequence analysis, Hidden-Markov-Models represent a probabilistic model of a multiple sequence alignment, which can be used, for example, to search for homologs in sequence databases.

► [Protein Domains](#)

HMN

Definition

Hereditary Motor Neuropathy.

► [Hereditary Neuropathies, Motor and/or Sensor](#)

HMSN

Definition

Hereditary Motor and Sensory Neuropathies.

► [Hereditary Neuropathies, Motor and/or Sensor](#)

HNA

Definition

Hereditary neuralgic amyotrophy (HNA) is an autosomal-dominant inherited recurrent focal neuropathy affecting mainly the brachial plexus.

► [Hereditary Neuropathies, Motor and/or Sensor](#)

HNF4A Gene

►Hepatocyte Nuclear Factor 4a

HNPCC

►Hereditary Nonpolyposis Colorectal Cancer

HNPP

Definition

HNPP (hereditary neuropathy with liability to pressure palsies) is a rare, autosomal, dominantly inherited disorder characterized by recurrent pressure palsies. Most HNPP patients have a 1.5 Megabases deletion in chromosome 17p11.2-p12. The disease is associated with increased vulnerability of the peripheral nerve system with mostly reversible sensorimotor deficits; other designation: tomaculous neuropathy.

►Hereditary Neuropathies, Motor and/or Sensor

hnRNA

Definition

hnRNA stands for heterogeneous nuclear RNA. It refers to the large pre-mRNAs of various nucleotide sequences that are made by RNA Polymerase II, and processed in the nucleus to become cytoplasmic mRNAs.

►RNA Capping
►RNA Polymerase II Transcription

Holiday Junction

Definition

Holiday junction designates the crossing of DNA strands that belong to different DNA helices without

disruption of base pair contacts. This DNA structure is formed from two homologous double-stranded DNA molecules by the exchange of single DNA strands. The name is in honor of the scientist who first postulated the existence of these recombination intermediates.

►Cre/Lox P Strategies
►DNA Recombination

Holoprosencephaly

Definition

Holoprosencephaly is a heterogeneous disorder (1:20,000 newborns), which is frequently caused by functional loss of sonic hedgehog (SHH) at human chromosome 7q36.3, and characterized by highly variable expressivity. It results from a failure of the foetal prosencephalon (the embryonic forebrain) to divide normally into hemispheres. It is often associated with other midline defects particularly of the skull and face. These range in severity from close set eyes (hypotelorism) and a single maxillary incisor, to a single eye (cyclopia) and a missing nose or a proboscis (a tubular-shaped nose). Mildly affected individuals may only show microsigns, such as ►ptosis or single maxillary incisor, and low normal intelligence.

►Microdeletion Syndromes
►Hedgehog Signalling

HOM-C

Definition

HOM-C refers to a cluster of evolutionarily conserved homeobox genes found in insects and nematodes.

►Homeodomain Transcription Factors

Homeobox

Definition

Homeobox describes a conserved DNA sequence of 180 nucleotides that codes for the homeodomain, a

protein domain of 60 amino acids. This homeodomain is a DNA-binding domain that occurs in proteins that are usually transcription factors. The sequence of the homeobox is similar enough, even across species, to identify homeotic genes.

- [Axis Formation – Formation and Function of the Dorsal Organizer](#)
- [Drosophila Model of Cardiac Disease](#)
- [Homeobox Transcription Factors](#)
- [HOX Genes](#)

Homeobox Genes

Definition

Homeobox genes comprise of a family of genes encoding transcription factors that bind DNA via a conserved 60 amino acids domain, the homeodomain. Homeobox genes play crucial functions during embryonic development including antero-posterior axis formation, segmentation and cell identity. Some of them are required for cell specific expression in adult organs. In blood cell progenitors, the expression of individual HOX genes follows tightly regulated programs, which are specific for the stage and lineage of progenitor-cell development, with universal down-regulation of HOX gene expression as the progenitors differentiate into mature blood cells.

- [Axis Formation – Formation and Function of the Dorsal Organizer](#)
- [Leukemia](#)
- [Gut Epithelium](#)

Homeobox Transcription Factor

Definition

Homeobox genes encode transcription factors that are characterized by the presence of a homeobox. This domain forms three α -helical regions, one of which is involved in sequence specific DNA binding. The name “homeobox” derives from findings that mutations in some of these genes lead to homeotic transformations: one body structure develops in place of another.

- [Neural Development](#)

Homeodomains

MARY TRUSCOTT¹, ALAIN NEPVEU^{1,2,3}

Departments of Biochemistry¹, Medicine², and Oncology³, McGill University, Montreal QC, Canada
alain.nepveu@mcgill.ca

Definition

The [homeodomain](#) is a highly conserved 60-amino-acid protein domain that is encoded by the [homeobox](#) and is found in organisms as diverse as mammals, insects, plants and yeast. Homeodomains function as DNA binding domains and are found in many transcription factors that control development and cell fate decisions. Structurally, homeodomains are related to helix-turn-helix proteins of bacteria and consist of three α helices folded around a hydrophobic core and a flexible N-terminal arm. Homeodomain-containing proteins are often referred to as homeoproteins.

► [Homeotic](#) transformations cause one body segment to adopt the identity of another segment. Homeotic transformations were later found to result from mutations affecting the expression of genes that share a 180-nucleotide sequence, the homeobox. Some homeobox genes are found in clusters, termed [HOM-C](#) in insects and nematodes and [HOX](#) in mammals. Clustered homeobox genes specify body segment identity and positional information along the antero-posterior axis. Other homeobox genes are dispersed throughout the genome and play a role at another level in development.

Characteristics

Historical Background

Homeotic transformations, whereby a body segment exhibits the features of another body segment, were first reported over a century ago. As they were segregated in a Mendelian fashion, homeotic transformations were assumed to be caused by genetic variations, termed [homeotic mutations](#), and the affected genes were called [homeotic genes](#). Sequence analysis revealed that the coding sequence of all homeotic genes share a conserved sequence of 180 nucleotides, the homeobox. The homeobox codes for the homeodomain, a protein domain that is capable of DNA binding and can also be involved in protein-protein interactions. Although not all homeobox-containing genes are homeotic genes (i.e. control segment identity), most if not all homeobox-containing genes play a role in development at one level or another.

Classification

Homeodomain proteins can be subdivided into many classes on the basis of several criteria including organization into a gene complex, sequence identity within the homeodomain itself or in the flanking region, association with other DNA binding or protein-interaction domains and the position of introns (Table 1) (1).

The Complex Superclass

Phylogenetic analysis suggests the existence of a primordial gene cluster, which was duplicated once, or more in the various evolutionary branches. Thus, in all metazoans there are one or more clusters of homeobox genes known as HOM-C in insects and nematodes and HOX in mammals. However, the term Hox is often used loosely to describe clustered and non-clustered homeobox genes in invertebrates as well as in vertebrates. In *Drosophila*, there are two gene clusters, the ►Bithorax complex (BX-C) and the ►Antennapedia complex (ANT-C). In mammals, there are four Hox clusters, called *HOXA*, *HOXB*, *HOXC*, and *HOXD*. The corresponding genes in the four mammalian clusters are said to be paralogous. Thus, ►paralogs belong to the same species. In contrast, the term ►orthologs describes homologous genes of different species. ►Orthologous genes and ►paralogous genes not only exhibit sequence conservation within the homeobox and its flanking sequences but are also positioned similarly within the gene cluster, are expressed in an analogous manner and fulfill a similar function.

The Dispersed Superclass

A large number of homeobox genes are also found dispersed throughout the genome. Recent advances in phylogenetic analysis have revealed that some of these genes also reside within clusters, extended HOX, NKL, paraHOX, metaHOX and EHGBbox. In many cases, homeodomain-encoding exons are found associated with exons coding for a second DNA binding domain (like Paired, Pou, Cut repeat, MADS box or a zinc-finger domain) or a protein-protein interaction domain (like LIM). Homeoproteins encoded by such genes are thus classified on the basis of this additional domain.

Biological Functions

Homeoproteins are transcription factors that play regulatory roles at all stages of development (2). In yeast, the homeoproteins MAT alpha 2 and MATA1 control the switch between the two different haplotypes, a and alpha. In metazoans, homeoproteins are involved early on in the formation of the overall body plan and then later in virtually every developmental decision including the determination of cell

Homeodomains. Table 1 Classification of Homeodomain Proteins

Homeodomain only		Homeodomain and another Domain
Complex Superclass	Dispersed Superclass	
Abdominal B (AbcdB)	Bicoid (Bcd)	Paired box
Antennapedia (Antp)	Caudal (Cad)	Cut repeats
Deformed (Dfd)	Distalless (Dll)	MADS box
Labial (Lab)	Empty spiracles (Ems)	POU domain
Proboscipedia (Pb)	Engrailed (En)	LIM domain
Sex combs (Scr)	Even-skipped (Eve)	Zinc finger motifs
Ultrabithorax/Abdominal A (Ubx/AbdA)	H2.0	Zinc finger motifs
	HNF1	
	Msh	
	NK-2	
	PBC	
	TCL/NEC	
	XANF	

fate, the specification of segment identity and cell-type identity and the regulation of cell differentiation and tissue patterning. In the adult, homeoproteins continue to control cell-type specificity and participate in the establishment and maintenance of overall homeostasis.

Much of our knowledge of the role of homeoproteins in development derives from studies performed in genetically tractable organisms like the fruit fly, *Drosophila melanogaster*, or the worm, *Caenorhabditis elegans*. Conclusions drawn from these genetic studies were generalized during the last decade using the laboratory mouse as a system. In some of the most spectacular experiments, it was shown that a human or murine ortholog was able to complement the corresponding mutant in *Drosophila*, thus demonstrating the evolutionary conservation in function. Moreover, the human diseases that are caused by mutations in homeobox genes conform exactly with the function of these genes in mouse and *Drosophila*.

In *Drosophila*, earlier studies focused on the role of Hox genes in the determination of segment identity

along the antero-posterior (A-P) axis (1, 2). A number of rules regarding the expression and function of homeotic genes were found to be conserved between vertebrates and invertebrates. The order of HOM and Hox gene expression along the A-P axis reflects the order of the genes along the chromosome, a phenomenon referred to as the ►colinearity rule. The spatial distribution and differing concentrations and combinations of homeotic transcription factors control the expression of segment-specifying factors according to a ►Hox code. A larger number of genes are usually expressed in more posterior regions. The effects of the deletion or aberrant expression of Hox genes revealed ►posterior prevalence – a hierarchy of regulatory effects whereby more posteriorly expressed Hox genes ultimately determine the specification of the segment in which the Hox genes are expressed. For example, loss-of-function mutations cause loss of structures and development of anterior structures where more posterior structures should have formed. However, the homeotic transformation occurs only in the anterior-most regions where that Hox gene is normally expressed and not within regions where a more posterior Hox gene is expressed. Thus, the more posterior Hox genes are able to promote the appropriate morphological structures without the contribution of more anterior genes. In contrast, gain-of-function mutations cause the development of posterior structures where more anterior structures should be found; however, ectopic expression of most homeotic genes does not lead to the transformation of segments posterior to their normal domains of expression.

In higher organisms, paralogs in the different Hox clusters are often expressed in similar patterns, suggesting some redundancy in function of the genes. However, there are some differences in expression patterns and there is a growing amount of evidence for the existence of functions specific to individual or pairs of paralogs.

Many conserved, non-clustered *Hox* genes exist in various chromosomal locations throughout the genome and also regulate cell fate specification in many tissues and organs. Furthermore, subsets of Hox genes display patterning functions in structures that are more phylogenetically recent than those along the main A-P axis, such as limbs and epithelial appendages. Many such functions are mediated along dorso-ventral, left-right and proximo-distal axes and involve the formation and maintenance of asymmetry in all tissues and organs including the regulation of cell-type specificity and determination of cell fate. In addition, the same homeobox genes that are expressed in different cells of a developing region in early stages of development can be expressed differently at a later stage, in specific cells in a given tissue, and can control gene expression in a cell-type-specific manner.

In addition to their role in development, homeodomain proteins have also been implicated in the regulation of cell proliferation, signaling, migration and adhesion.

Biochemical Activities and Regulation

The homeodomain functions as a DNA binding domain that exhibits preference for sequences containing a TAAT core. Homeoproteins have been characterized as transcription factors that can either activate or repress transcription. Whether a homeoprotein acts as a repressor or activator is determined by the presence within the protein of a repression or activation domain, but the directionality of the regulatory effect can also be influenced by promoter context, cooperation with other transcription factors, association with co-factors and various post-translational modifications.

There is an apparent discrepancy between the highly selective biological functions of homeoproteins and their relative lack of DNA binding specificity (3). Indeed, homeoproteins exhibit promiscuous DNA binding specificities *in vitro* and it is not clear whether a given homeoprotein regulates a limited or a large number of target genes. A partial solution to the specificity problem was provided by the findings that homeoproteins encoded by HOX genes bind to DNA in cooperation with homeoproteins of the three-amino-acid-loop-extension (TALE) family. These include PBX, MEIS and PREP1/KNOX1. Interaction requires the presence within the Hox protein of a conserved sequence N-terminal to the homeodomain. The consensus sequence of this PBX-interaction motif (PIM) is IYPWMK. Cooperative DNA binding also occurs in the case of homeoproteins that contain one or more DNA binding domain(s) in addition to the homeodomain. In cells, factors that may also contribute to increased specificity include the state of the chromatin, cooperation with proteins that are bound to other regulatory elements and interaction with co-activators, co-repressors and general transcription factors. Finally, like other transcription factors, the activity and intracellular localization of homeoproteins is regulated by a whole array of post-translational modifications.

Clinical Relevance

Mutations Associated with Diseases and Syndromes

A large number of clinical conditions were found to be caused by mutations within a homeobox-containing gene (Table 2). Point mutations (missense and nonsense) within the coding sequences of a homeobox gene have been documented in several congenital diseases (4). In addition, short deletions outside of coding regions have been noted in certain cases and were assumed to affect gene expression; however, this remains to be formally demonstrated. Most diseases due to point mutations within a homeobox-gene show a

Homeodomains. Table 2 Human Diseases Caused by Mutations in Homeobox-Containing Genes

Gene (OMIM no.)	Syndrome/disease	OMIM number	Main features	Inheritance
AIRE (607358)	Autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy syndrome	240300	Variable combinations of autoimmune endocrine diseases such as Addison's disease, hypoparathyroidism and type 1 diabetes	AR
ARIX (602753)	Congenital fibrosis of extraocular muscles	602078	Nonprogressive eye movement disorder characterized by bilateral ptosis and restrictive ophthalmoplegia with globes 'frozen' in extreme abduction	AD/AR
ARX (300382)	X-linked infantile spasm syndrome (West syndrome)	308350	Early-onset generalized seizures, hypsarrhythmia, an mental retardation	X
ATRX (300032)	Alpha-thalassemia/mental retardation syndrome	301040	severe mental handicap, characteristic dysmorphic facies, genital abnormalities, and an unusual, mild form of hemoglobin H (Hb H) disease, variable features if gene is deleted	X
CHX10 (142993)	Microphthalmia	251600	Small eye and other ocular Abnormalities	AR
CRX (602225)	Cone-rod dystrophy	120970	Chorioretinal atrophy of central and peripheral retina	AD
CRX (602225)	Leber congenial amaurosis	204000	Childhood retinal degeneration	AR
DLX3 (600525) DLX4 (601911)	trichodontoosseous syndrome	190320	taurodontism and enamel defects in teeth, kinky, curly hair at birth	AD
EMX2 (600035)	Schizencephaly	269160	infolding of cortical gray matter along a hemispheric cleft near the primary cerebral fissures	AD
HESX1 (601802)	Septo-optic-dysplasia	182230	Optic nerve hypoplasia. Absence of septum pellucidum. Agenesis of corpus callosum. Panhypopituitarism	AR
HLXB9 (142994)	Currarino syndrome	176450	Partial absence of the sacrum. Anorectal anomalies. Presacral mass	AD
HNFla (142410)	MODY3	600496	MODY is a form of familial noninsulin-dependent diabetes mellitus (NIDDM; 125853) and is characterized by an early age of onset (childhood, adolescence, or young adulthood under 25 years)	AD
IPF1 (PDX1) (600733)	MODY4	606392	expression may occur at later ages in this form of MODY	AD

Homeodomains. Table 2 Human Diseases Caused by Mutations in Homeobox-Containing Genes (Continued)

Gene (OMIM no.)	Syndrome/disease	OMIM number	Main features	Inheritance
VHNF1 (142410)	MODY5	604284	both diabetes mellitus and renal cysts (hypoplastic glomerulocystic kidney disease)	AD
HOXA11 (142958)	Radioulnar synostosis with amegakaryocytic thrombocytopenia	605432	proximal fusion of the radius and ulna, thrombocytopenia	AD
HOXA13 (142959)	Hand-foot-genital Syndrome	140000	Distal limb and distal urinary tract malformations	AD
	Guttmacher syndrome	176305	Distal limb and genital abnormalities, postaxial polydactyly of the hand and short or uniphalaengeal second toes with absent nails	AD
HOXD13	Synpolydactyly	186000	Malformation of the distal limbs	AD
IPF1 (PDX1) (600733)	congenital pancreatic hypoplasia	260370	agenesis of the pancreas, early-onset insulin-dependent diabetes mellitus, intrauterine growth retardation	
	diabetes mellitus	222100 125850	polydipsia, polyphagia, and polyuria which result from hyperglycemia-induced osmotic diuresis and secondary thirst	
ISL1 (600366) IPF1 (PDX1) (600733)	Non-Insulin Dependent Diabetes Mellitus	125853	obese body habitus and manifestations of the so-called metabolic syndrome which is characterized by diabetes, insulin resistance, hypertension, and hypertriglyceridemia	
LMX1B (602575)	Nail-patella syndrome	161200	Nail hypoplasia or dysplasia. Absent or hypoplastic patellae. Other skeletal abnormalities. Nephropathy	AD
MSX1 (142983)	Familial tooth agenesis	106600	Agenesis of second premolar and third molar	AD
MSX2 (123101)	Craniosynostosis	604757	Premature fusion of calvarial sutures that causes abnormal skull shape	AD
MSX2 (123101)	Foramina parietalia permagna	168500	Headache. Scalp defects. Structural or vascular malformation of the brain	AD
ALX4 (605420)				AD
NKX2.5/ CSX (600584)	Congenital heart disease	108900	Atrial or ventricular septal defects. Atrioventricular conduction block	AD
PAX3 (193500)	Waardenburg syndromes	193500 193510 600193	Sensorineural deafness, pigmentation defects, hypoplastic alae nasi	AD

Homeodomains. Table 2 Human Diseases Caused by Mutations in Homeobox-Containing Genes (Continued)

Gene (OMIM no.)	Syndrome/disease	OMIM number	Main features	Inheritance
PAX6 (607108)	Aniridia	106210	Iris hypoplasia, other ocular abnormalities	AD
PITX2 (601542)	Rieger syndrome	180500	Ocular anterior segment dysgenesis. Dental and umbilical abnormalities	AD
POU3F4 (3000039)	Sensorineural deafness	304400	Conductive hearing loss resulting from stapes fixation and progressive sensorineural deafness	XR
PROP1 (601538)	Combined pituitary hormone deficiency	262600	Hypoplasia of pituitary gland. Deficit of LH, FSH, PRL, GH, TSH	AR
POU1F1/PIT1 (173110)				AR/ADb
RAG1 (179615)	Omenn syndrome	603554	Early-onset generalized erythrodermia, lymphadenopathy, hepatosplenomegaly, fever, protracted diarrhea, and failure to thrive, hypereosinophilia	AR
RUNX2 (600211)	Cleidocranial dysplasia	119600	Persistently open skull, hypoplasia or aplasia of the clavicles, wide pubic symphysis, short middle phalanx of the fifth fingers, dental anomalies, vertebral malformation	(haploinsufficiency)
SHOX (312865)	Leri-Weill dyschondrosteosis	127300	Skeletal dysplasia characterized by disproportionate short stature with predominantly mesomelic limb shortening and deformity of the arm	XYD
SHOX (312865)	Turner syndrome	312865	Idiopathic growth retardation and short stature	X (haploinsufficiency)
SIX3 (603714)	Holoprosencephaly	157170	From various severe cranio-facial defects to mild facial dysmorphism	AD
SOX9 (114290)	Campomelic dysplasia	114290	Congenital bowing and angulation of long bones. Other skeletal and extraskelatal defects	AR (haploinsufficiency)
ZFHX1B (605802)	Mowat-Wilson syndrome	605802	Distinct facial appearance, mental retardation	

dominant inheritance. In several cases, the dominant effect of HD mutations was demonstrated to result from [▶haploinsufficiency](#), but in a few instances a dominant negative effect of the mutated protein was suggested

from the more severe phenotype observed when the mutated protein was partially active instead of completely inactive. In several cases, mutation of a single homeobox gene causes malformation of multiple

Homeodomains. Table 3 Homeobox Genes At Sites Genomic Rearrangement

Gene (OMIM no.)	Cancer	Genomic Rearrangement
HOX11 186770	T-cell acute leukemia	t(10;14) or t(7;10) The HOX11 gene adjacent to the 10q24 region is transcriptionally activated after translocation to the proximity of either the T cell receptor D at 14q11, or B at 7q35.
PBX1 176310	Pre-B-cell acute lymphoblastic leukemias	t(1;19)(q23;p13.3) The translocation juxtaposes the 5' portion of the E2A transcription factor gene (147141) on 19p13.3 with the 3' portion of the PBX1 TALE homeobox gene on 1q23 resulting in the synthesis of a chimeric protein E2A-PBX11.
TLX3 (HOX11L2) 604640	T-cell acute lymphoblastic leukemia	t(5;14)(q35;q32) The translocation brings the TLX3 gene on 5q35 under the control of transcriptional regulation elements on 14q32 that drive expression of the CTIP2 gene during normal T-cell differentiation.
HOXA9 142956 HOXD11 142986 HOXC13 (142976) PMX1 (PRRX1)	Acute myeloid leukemia	t(7;11)(p15;p15); t(2;11)(q31;p15) t(11;12)(p15;q13) t(1;11)(q24;p15) The translocation juxtaposes the 5' portion of the NUP98 gene (601021) on 11p15 with the 3' portion of the HOXA9, HOXD11, or HOXD13 gene thereby generating a chimeric protein.
HOXA9 142956	Chronic myeloid leukemia	t(7;17)(p15;q23) The translocation juxtaposes the 5' portion of the MSI2 gene (607897) on 17q23 with the 3' portion of the HOXA9 gene on 7p15 thereby generating a chimeric protein.
PAX3 (193500)	Rabdomyosarcomas	t(2;13)(q35;q14) The translocation juxtaposes the 5' portion of the PAX3 gene on 2q35 with the 3' portion of the FKHR gene (136533) on 13q14 thereby generating a chimeric protein.
LHX4 (602146)	Pre-B-cell acute lymphoblastic leukemia	t(1;14)(q25;q32) The translocation resulted in the 5'prime regulatory region of LHX4 on 1q25 being replaced by the enhancer region of the IgH gene on 14q32.
NKX3.1	Prostate cancer	Loss-of-heterozygosity on 8p21
CDX2	Colorectal carcinoma	Loss-of-heterozygosity on 13q12.3 and mutations within coding sequences

body structures. This is probably due to the fact that a single homeobox gene can be expressed in several different cell types and control multiple developmental decisions.

There are over 200 homeobox genes in the human genome, of which 39 are members of one of the four HOX clusters. Mutations in only two HOX genes have been identified so far. Considering the importance of these genes in development, the identification of such a relatively small number of mutations in HOX genes may suggest the existence of functional redundancy among paralogs of the *HOX* clusters. Alternatively, it has been suggested that severe mutations in these genes might lead to defects early in embryonic development,

which therefore would not be found in the living population.

Homeobox Genes Implicated in Cancer

A number of homeobox genes have been implicated in various cancers. In human cancers, both gain- and loss-of-function of homeobox genes have been reported to occur. The general trend is that those homeobox genes that are over-expressed in tumors are normally expressed in proliferating cells, whereas those homeobox genes that are down-modulated are normally expressed in the terminally differentiated cells of that tissue-type (5). It is not always clear whether mis-expression of a homeobox gene is a direct cause of

cancer or is merely a consequence of the cancer phenotype. However, there is abundant experimental evidence in cell-culture models and in the laboratory mouse to support a direct role of homeobox genes in cancer. In particular, ectopic expression of some homeodomain proteins and gene inactivation of two homeobox genes were found to cause various types of neoplasias in the mouse. Moreover, for some homeobox genes that are implicated in human cancers, the homologous murine gene was found to be activated following the integration of a retrovirus in murine cancer models.

In human cancers a number of homeobox genes have been found at the site of genomic rearrangements including chromosomal translocation and allelic loss (Table 3) (6). Allelic loss has been reported for two homeobox genes, CDX2 and NKX3.1, which exhibited properties of tumor suppressor genes in mouse models. Chromosomal translocations can result in the generation of fusion proteins that acquire novel biochemical properties or the juxtaposition of tightly regulated homeobox genes next to highly active promoter regions. Deregulated expression may also be caused by more subtle regulatory events not involving changes in DNA (5). In most of these cases, however, expression studies were limited to mRNA analysis and it remains to be verified whether protein expression or function always reflects the level of mRNA expression.

- ▶ Bone and Cartilage
- ▶ Homeobox Transcription Factors
- ▶ Mutagenesis Approaches in Medaka
- ▶ Transcription Factors and Regulation of Gene Expression

References

1. Gehring WJ, Affolter M, Burglin T (1994) Homeodomain proteins. *Ann Rev Biochem* 63:487–526
2. Manak JR, Scott MP (1994) A class act: conservation of homeodomain protein functions. *Dev Suppl*:61–77
3. Biggin MD, McGinnis W (1997) Regulation of segmentation and segmental identity by *Drosophila* homeoproteins: the role of DNA binding in functional activity and specificity. *Development* 124:4425–4433
4. D'Elia AV, Tell G, Paron I, Pellizzari L et al (2001) Missense mutations of human homeoboxes: A review. *Hum Mutat* 18:361–374
5. Abate-Shen C (2002) Deregulated homeobox gene expression in cancer: cause or consequence? *Nat Rev Cancer* 2:777–785
6. Owens BM, Hawley RG (2002) HOX and non-HOX homeobox genes in leukemic hematopoiesis. *Stem Cells* 20:364–379

Homeostasis

Definition

Homeostasis describes the ability to maintain an equilibrium/stable condition of the internal systems in a changing environment.

- ▶ Bone and Cartilage
- ▶ Diabetes Insipidus, a Water Homeostasis Disease
- ▶ Limb Girdle Muscular Dystrophies

Homeotic Genes

Definition

Homeotic genes were originally identified because of the fact that mutations in these genes can cause extraordinary phenotypes (▶ homeotic mutations) in *Drosophila*. Homeotic genes regulate, for example, whether the segment develops a halter or a wing, by regulating a large number of downstream genes required for the development of segment specific structures. Accordingly, different homeotic genes are active in different segments. Thus, “homeotic” is a functional description for genes that cause homeotic transformations (see also ▶ HOX genes). Homeotic genes include a conserved DNA sequence called the ▶ homeobox.

- ▶ Axis Formation – Formation and Function of the Dorsal Organizer
- ▶ Bone Disease and Skeletal Disorders, Genetics
- ▶ Gut Epithelium
- ▶ Homeodomain Transcription Factors

Homeotic Mutations

Definition

Homeotic mutations are mutations that are responsible for the transformation of an organ into another organ, normally expressed in a different part or segment of the organism.

- ▶ Homeobox Transcription Factors

Homogeneous Assay

Definition

Homogeneous assay refers to an assay format allowing to make an assay-measurement by a simple mix and read procedure without the necessity to process samples by separation or washing steps.

► [Automated High-Throughput Functional Characterization of Human Proteins](#)

Homologous Recombination

Definition

Homologous recombination describes the recombination of DNA sequences bearing significant homology on the same chromosome, on homologous chromosomes or on non-homologous chromosomes. This term was previously restricted to the exchange of DNA sequences between homologous chromosomes, and accompanied by a crossover event, as during meiosis. It occurs widely in bacteria and viruses. The recombination reaction is equivalent to highly specific cutting and ligation reactions. Homologous recombination is distinguished from other types of recombination (site-specific recombination). Homologous recombination in mouse ► [ES-cells](#) is a commonly used method to introduce targeted mutations into a host genome (► [gene targeting](#)).

► [Chromosomal Instability Syndromes](#)

► [Cre/Lox P Strategies](#)

► [Mouse Genomics](#)

► [Recombinant Protein Expression in Bacteria](#)

► [Transgenic and Knockout Animals](#)

Homologous Recombination Repair

Definition

Homologous Recombination Repair (HRR) describes a DNA repair mechanism using DNA double strand breaks with the help of an intact DNA duplex provided by the sister chromatid or the homologous chromosome.

► [DNA-Repair Mechanisms](#)

Homology

Definition

Homology refers to a similarity in sequence (amino acids, nucleotides), based on descent from a common ancestor.

► [Proteomics in Human-Pathogen Interactions](#)

► [Protein Databases](#)

Homology Modeling

MICHAEL NILGES

Unité de Bio-Informatique Structurale, Institute

Pasteur, Paris, France

nilges@pasteur.fr

Synonyms

Comparative modeling

Definition

Homology modeling or comparative modeling is the prediction of the tertiary structure of an unknown protein using a known three-dimensional (3D) structure of a homologous protein. This is based on the observation that protein 3D structure is better conserved than the amino acid sequence and that therefore similar sequence implies similar 3D structure. The homology modeling method is based on sequence alignment and molecular modeling techniques (such as Monte Carlo or molecular dynamics simulations). Homology modeling has important applications in structural biology (for solving experimental X-ray crystal or ► [NMR](#) structures) and for the prediction and characterization of functional aspects of proteins of unknown structure.

Description

Our knowledge about the primary sequences of proteins grows much faster than that of their 3D structures. There are approximately 20,000,000 protein sequences known (1), but only around 20,000 3D structures (2). As a complement to the experimental structure determination methods, homology modeling plays an indispensable role in obtaining structural information. It relies on the fact that the three-dimensional structure of proteins is, in most cases, better conserved than their sequence and that therefore similar amino acid sequence implies similar 3D

structure. (Homology between two proteins strictly speaking indicates only that two proteins have a common evolutionary ancestor and does not therefore directly imply a similar fold.) The sequence of the unknown (query) protein is compared systematically to the sequences of all proteins with known 3D structures; those with high sequence similarity then serve as templates to model the 3D structure of the query protein. The development of better homology modeling methods is still an active field of research.

Among the 20,000 known 3D structures there are only 300 distinct protein folds. Structural genomics projects aim specifically at “filling the fold space”, at determining experimentally specifically those protein structures that cannot be modeled by homology and would in turn allow many more protein structures to be predicted.

If the sequence similarity is too weak to be detected by standard methods alone, threading (or fold recognition) techniques can be used, which compare the query sequence directly to all known protein folds and evaluate the “fitness” of a particular sequence for a fold. Recently, decisive advances have also been made in the field of “*ab initio*” protein fold prediction, that is prediction of the 3D structure without a structural template.

Methodology

In many cases, homology modeling has become routine work and automated or semi-automated procedures and even WWW services exist. There are two basic difficulties. The alignment between the query sequence and the template structure(s) has to be of very good quality since errors in the alignment cannot be corrected by the modeling procedure and parts of the query protein that have no direct counterparts in the template (for example loops) need special attention. Independently of the particular algorithm or program that one is using, the overall procedure usually consists of a series of steps: (1) identification of protein template(s) with known three-dimensional structure; (2) alignment of the query sequence with the template(s); (3) building a three-dimensional model of the query protein; (4) validation (evaluation of the quality of the model).

Identification and Selection of Template Structures

The principal resource for homology modeling is the protein data bank (PDB, ►<http://www.rcsb.org/pdb>) (2), which contains the 3D structures of all experimentally determined structures of biological macromolecules. Each entry in the PDB has two sections. The first contains general information about the molecule and its structure determination, such as the organism, the amino acid sequence, the experimental method (X-ray crystallography, nuclear magnetic resonance, electron

microscopy), the authors, related scientific publications, data base links and the resolution in the case of X-ray crystal structures. The second contains the atomic coordinates themselves, with some information on the dynamics of each atom. The latter can serve as a rough measure of the reliability of different regions of the structure.

In the simplest case, the query protein has relatively high sequence homology with at least one protein in the PDB such that it can be identified with a standard pairwise sequence alignment method such as BLAST (3). More sensitive methods use multiple sequence alignments. A multiple sequence alignment can be built from a database search of a sequence database with the query sequence. The program PSI-BLAST allows automated iterative creation of multiple sequence alignments. Once a multiple sequence alignment is generated, it can be used to search the PDB. Alternatively – and faster – one can search existing databases of multiple sequence alignments (such as PFAM, ►<http://www.sanger.ac.uk/Software/Pfam/>, or SMART, ►<http://smart.embl-heidelberg.de/>, which contain alignments of protein or protein domain families). These databases contain links to structures in the PDB. Among the most sensitive searches are those against multiple sequence alignments derived from the alignment of similar structures (for example, 3DPSSM, ►<http://www.sbg.bio.ic.ac.uk/~3dpssm/>) If several templates can be identified, one needs to choose the ones that are most appropriate for modeling. The choice can be based on the sequence identity or similarity, the score given by the alignment program or the location of query and template on a phylogenetic tree of the protein family. But the choice should also depend on other information, for example a similar biological function or a similar environment (e.g. if the template has a ligand or is in the free form). The quality of the structure, which is also important, can be estimated from the resolution and the R-values for X-ray crystal structures.

Alignment of the Query Sequence with the Template(s)

The most critical phase of the homology modeling procedure is the alignment of the query sequence with the identified templates. A homology model can only be as good as the sequence alignment. Unless the sequence identity between query and template is very high, this step should always be based on multiple sequence alignment. In contrast to a simple alignment between two sequences, a multiple sequence alignment contains evolutionary information (for example which residues are conserved in the course of evolution and therefore important for structure or function). Even though only one model is to be built, this alignment would contain several or many similar sequences, among them one or several templates. Some programs

can improve the quality of the alignment by using structural information, e.g. if several templates are known, an independently obtained structural alignment of the templates can be used to guide the sequence alignment. New and better multiple sequence alignment methods and strategies are still being developed. As a rule of thumb, a good alignment is easy to obtain once one has around 60% identity. However, it is possible to have 50% identity between two carefully designed protein sequences with a different fold (the so called JANUS protein). To obtain a correct alignment may become challenging for 40% and manual revision of the alignment may be necessary. In the so-called twilight zone of only 20–30% sequence identity, the information from multiple alignments becomes particularly important, since in a pair-wise alignment only 20% of residues can be expected to be correctly aligned if the sequence identity is below 30%.

Building a Three-Dimensional Model of the Query Protein

In building the 3D structure of the query protein based on the template, one can distinguish two different tasks. The first, easier task is to build the parts of the protein that could be well aligned. This consists mostly of replacing the amino acid side chains of the template protein with those of the query protein, using the main chain as a scaffold. The other, more difficult task concerns building those parts of the backbone where no good template can be used, the loops on the surface of the protein. These are unfortunately often the most interesting parts of the protein, containing parts of active or interaction sites of the protein.

Even modeling side chains on the main chain is, in principle, a very complex task, since the number of conformations that has to be evaluated grows exponentially with the number of side chains (it is computationally a so-called NP-hard problem). Side-chain conformations in the interior of a protein are largely determined by packing interactions, which simplifies the energetic evaluation of different conformations. Among the many methods available are interactive data base searches, the use of libraries with preferential conformations for each amino acid (rotamer libraries) and specific algorithms that circumvent the exponential explosion of conformations to be evaluated (dead-end elimination). Heuristic algorithms are based on minimization of a cost function and can be based on Monte Carlo, molecular dynamics, self-consistent mean field or distance geometry. At the end of the placement of the side chains, the backbone may need to be adjusted slightly. This is usually done by minimization with a molecular dynamics force field.

Modeling protein loops can be seen as a small “*ab initio*” prediction problem, more difficult because a

large number of possible conformations for the loop have to be evaluated and because the conformation of the loop depends on many more factors than simply packing. Searches of a database of loop conformations and the use of specialized cost functions that incorporate knowledge of likely loop conformations are most valuable for this task.

A popular homology-modeling package is SwissModel, based on a World Wide Web service (► <http://www.expasy.org/swissmod/SWISS-MODEL.html>) and interactive molecular graphics (SwissPDBViewer) to validate the alignment of the sequence with the 3D template and build the loops by database search. The service and program is associated to the SwissProt database (1). Another program is MODELLER (<http://salilab.org/modeller/modeller.html>) (4), a fully automated program that uses minimization of a cost function to build the structure of the query protein. The cost function is made up of several parts coming from sequence homology, a molecular dynamics force field and a special scoring function for loops. Sequence homology is incorporated by spatial restraints derived from a statistical analysis of protein structures, which bias the predicted structure towards that of the template with weights depending on the local degree of sequence similarity.

Validation and Evaluation of Quality

The criteria for quality assessment are the same as those used for experimental structures and range from simple geometric criteria (such as deviations from covalent bond lengths) to the assessment of the probability of environments of amino acids (e.g. a hydrophobic amino acid is more likely to be found in the interior of the protein and a hydrophilic amino acid in contact with the solvent). In general, properties derived from the homology models should not deviate much from those found in the experimental structures in the PDB. Examples of programs evaluating the quality of protein structures are WHATIF (► <http://www.cmbi.kun.nl/whatif/>), PROCHECK (► <http://www.biochem.ucl.ac.uk/~roman/procheck/procheck.html>) or PROSA (► <http://lore.came.sbg.ac.at/>). Larger errors in models of protein structures are usually due to sequence alignment errors or to parts of the protein where the template is not similar to the query protein (e.g. loops). If several different alignments between query and template are similarly possible one can build several models and evaluate all of them.

Clinical Applications and Therapeutic Consequences

Understanding the 3D structure of a macromolecule is crucial for understanding its function. Many properties of the 3D structure cannot be deduced directly from the

primary sequence, for example surface properties and electrostatic potentials. These properties are important for understanding the interactions of the protein with other molecules, for example with other proteins. Single nucleotide polymorphisms can be mapped on 3D structures of proteins in order to elucidate specific structural causes of disease.

Receptor-based or structure-based drug design exploits knowledge of the 3D structure of the binding site to construct potential drugs, for example inhibitors of a receptor protein. This can be an important first step in the development of a new therapeutic agent. Good homology models can serve as basis for drug design. The 3D structure of the protein can be used in several ways. Based on a detailed model of the protein-ligand interaction, one can propose modifications of the ligand to modify its binding characteristics. These modifications, proposed and characterized in the computer, can then be chemically synthesized and tested experimentally. The structure can also be used in virtual screening applications to screen a library of (sometimes millions of) compounds, for example to find likely inhibitors of the protein. In this way, the number of molecules that need to be screened chemically can be reduced.

- ▶ [Molecular Dynamics Simulation in Drug Design](#)
- ▶ [Protein Databases](#)
- ▶ [Structure-Based Drug Design](#)
- ▶ [Structural Genomics: Structure-to-Function Approaches](#)

References

1. Bairoch A, Apweiler R (2000) The SWISS-PROT protein sequence database and its supplement TrEMBL in 2000. *Nucleic Acids Res* 28:45–48
2. Berman HM, Westbrook J, Feng Z et al (2000) The Protein Data Bank. *Nucleic Acids Res* 28:235–242
3. Altschul SF, Madden TL, Schaffer AA et al (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
4. Sali A, Blundell TL (1993). Comparative protein modelling by satisfaction of spatial restraints. *J Mol Biol* 234:779–815

Homology-Derived Secondary Structure of Proteins Database

- ▶ [HSSP \(Homology-Derived Secondary Structure of Proteins\) Database](#)

Homonuclear NMR Experiments

Definition

Homonuclear NMR experiments are nuclear magnetic resonance (NMR) experiments in which the frequencies, and thus chemical shifts, of only one type of nucleus is recorded in all dimensions of a multi-dimensional experiment.

- ▶ [Multidimensional NMR Spectroscopy](#)

Homophilic Binding

Definition

Homophilic binding refers to the binding of a receptor molecule to an identical molecule.

- ▶ [Adhesion Molecules](#)
- ▶ [Hereditary Spastic Paraplegias](#)

Homozygosity Mapping

Definition

Homozygosity mapping refers to mapping of a gene by virtue of the identity-by-descent (▶ [IBD](#)) of both alleles.

- ▶ [Bloom Syndrome](#)

Homozygote/Homozygous

Definition

Homozygote/homozygous refers to an individual that has two identical alleles at a genetic locus.

- ▶ [Heritable Skin Disorders](#)
- ▶ [Mutagenesis Approaches in the Zebrafish](#)

Horizontal Transmission

Definition

Horizontal transmission refers to the lateral transfer of a gene, or other DNA sequence, from one genome to

another. Horizontal transmission is distinct from the normal mode of vertical transfer by which genetic information is passed from parent to offspring.

► [Transposons](#)

Hormone(s)

Definition

Hormone(s) is/are biochemical structures (steroid protein/peptide or amine) that are released into the circulation, or act locally and produce specific effects at a target organ.

► [Hypothalamic and Pituitary Diseases Genetics](#)

Horseradish Peroxidase

Definition

Horseradish peroxidase (HRP) is an enzyme commonly used in immunohistochemistry for substrate detection, which oxidizes diaminobenzidine to produce a brown precipitate.

► [Immunochemical Method, Localization](#)

Host Genome

► [Genetic Background](#)

Hotspot Mutation

Definition

The term describes mutations occurring at a chromosomal region, which is more susceptible to genetic damage/change than average sequences, or contains sequences that are more likely to be altered by spontaneous mutations, such as ► [CpG dinucleotides](#).

► [Heritable Skin Disorders](#)

► [Tumor Suppressor Genes](#)

Housekeeping Genes

Definition

Housekeeping genes refer to a set of genes that are expressed in all cells regardless of type, and which serve to maintain the basic functions of the cell.

► [CpG Islands](#)

► [Mutagenesis Approaches in Yeast](#)

Hox Cluster

Definition

Hox cluster refers to an array of ► [Hox genes](#) that arose from rounds of tandem duplication. In insects and vertebrates, the Hox genes are arranged into a cluster in the genome. Invertebrates have one Hox cluster, while mammals have four. The expression of these genes along the anteriorposterior axis usually correlates with their arrangement in the genome. Genes close to the 3' end of the cluster are expressed more anterior while genes towards the 5' end more posterior.

► [Mutagenesis Approaches in Medaka](#)

Hox Code

Definition

The Hox code describes a number of rules regarding the expression of ► [Hox](#) (homeobox) genes and their effects on segment identity. For example, a larger number of homeobox genes are usually expressed in more posterior regions. Loss-of-function mutations cause loss of structures and development of anterior structures where more posterior structure should have formed. However, the homeotic transformation occurs only in the anterior-most regions, where that Hox gene is normally expressed, and not within regions where a more posterior Hox gene is expressed.

► [Homeobox Transcription Factors](#)

Hox Genes

Definition

Hox genes comprise of a subset of ►[homeobox genes](#). In mammals, they are typically organized into 4 clusters (A–D), each spanning approx. 100 kb of DNA and containing 8–11 Hox genes. HOX genes have a primitive role in the regulation of anterior-posterior axis development. They regulate patterning during limb development, probably functioning in the determination of the timing and extent of local growth rates. A polyalanine stretch in the amino-terminal region of the HOXD13 gene causes synpolydactyly.

- [Bone Disease and Skeletal Disorders, Genetics](#)
- [Homeobox Transcription Factors](#)
- [Hox Cluster](#)
- [Gene Duplications](#)
- [Repeat Expansion Diseases](#)

marker. The *Hprt* gene is required for purine biosynthesis, and ES cells containing a functional *Hprt* can survive in the presence of the drug ‘HAT’ (hypoxanthine, aminopterin and thymidine). Thus, following Cre expression in ES cells that have been targeted using both these HPRT vectors, a functional *Hprt* is formed (by the joining of the 5′ HPRT and 3′ HPRT cassettes), and these cells can be selected for by growth in the presence of HAT.

- [Mouse Genomics](#)

HRR

- [Homologous Recombination Repair](#)

HpaII Tiny Fragments

- [CpG Islands](#)

HS(A)N

Hereditary Sensory (and Autonomic) Neuropathy

- [Hereditary Neuropathies, Motor and/or Sensor](#)

HPLC

- [High-Performance Liquid Chromatography](#)

HPRT Vector

Definition

There are two ‘HPRT [hypoxanthine phosphoribosyl-transferase] vectors’ that are used in chromosomal engineering: the 5′ HPRT vector and the 3′ HPRT vector. The 5′ vector consists of the 5′ portion of the HPRT cassette that contains exons 1–2 of the *Hprt* gene, a single *loxP* site and the selectable marker NeoR. The 3′ vector consists of the 3′ portion of the HPRT cassette (containing exons 3–9 of the *Hprt* gene), a single *loxP* site and the PuroR selectable

HSCs/HPCs

- [Hematopoietic Stem and Progenitor Cells](#)

HSF

Definition

HSF stands for a transcription factor (Heat Shock Factor) and is activated in response to an elevated temperature or other stress treatment.

- [Protein Folding](#)
- [Sumoylation](#)

HSN

- Hereditary Sensory Neuropathies

HSP

- Heat Shock Protein
- Hereditary Spastic Paraplegias

Hsp90

Definition

Hsp90 proteins constitute an abundant and conserved heat shock protein family. Hsp90 proteins function as molecular chaperones, and can be found involved in regulating signal transduction pathways as part of steroid receptor heterocomplexes, and by interacting with certain protein kinases.

- Peptidyl Prolyl Cis/Trans Isomerases
- Protein Folding

HSPGs

- Heparin Sulfate Proteoglycans

HSQC

Definition

The large variety of multidimensional NMR experiments are distinguished by abbreviations of their names which reveal something about the kind of NMR sequence employed. The HSQC (Hetero-nuclear Single Quantum Coherence) is a two-dimensional NMR experiment that correlates the frequencies of either

^{15}N and ^1H nuclei in amide groups, or ^{13}C - ^1H pairs. In the case of proteins, HSQC spectra require isotopic labelling. HSQC spectra yield a pattern of signal characteristic for a particular protein, and are often used as fingerprint spectra to quickly assess the complexation state of a protein.

- NMR-Based Screening

HSRs

Definition

HSRs (homogenously staining regions) is a cytogenetic term that describes extended regions with uniform staining in G-banded chromosomes. Homogenously staining regions result from amplification of small parts (a few genes) of chromosomes.

- DNA Amplification

HSSP Database

Definition

HSSP-database displays a database of homology-derived secondary structure of proteins (HSSP), generated by aligning to each protein of known structure all sequences deemed homologous on the basis of the threshold curve.

- Protein Databases

HTF Islands

Definition

HTF island refers to HpaII tiny fragments, or CpG Islands.

- CpG Islands

HTH

- Helix-Turn-Helix

HTLV

► Human T-Cell Leukemia Virus

Htr

Definition

Halobacterial transducer of rhodopsin (Htr) is a two membrane spanning helix protein which binds in a 2:2 complex to cognate sensory rhodopsins. The cytoplasmic domain is highly homologous to the corresponding domain of chemotaxis receptors. Together with an adaptor protein CheW CheA, a His-kinase is activated.

► Photoreceptors

HTS

Definition

HTS stands for High-throughput-screening. It denotes an automated method that allows the analysis of a manifold of species from a molecular library with regard to a certain property. For example, the interaction of ligands from a library of small organic compounds to a cellular receptor may be probed by a fluorescence-based binding assay.

► Protein-Ligand-Interaction by NMR

HuCAL[®]

Definition

HuCAL[®] stands for Human Combinatorial Antibody library. It represents a fully synthetic phage display antibody library which consists of fully human Fab sequences for the generation of human antibodies.

► Immunological Methods, the Use of Antibodies to Discover the Function of Novel Gene Products

Human Genome (Sequence) Project

Definition

The human genome (sequence) project was an international effort that aimed to identify all the approximately 30,000 genes in human DNA, to determine the sequences of the 3 billion chemical base pairs that make up human DNA, and to store this information in databases (<http://www.gene.ucl.ac.uk/hugo/>). The Human genome project ended in 2003 with the completion of the human DNA sequence.

► Repeat Expansion Diseases

► Rheumatism Related Genes, Identification

Human Herpesvirus Type 8

Definition

This member of the gamma subfamily of herpesviruses is causally linked to the Kaposi's sarcoma, and two rare lymphoproliferative diseases, Castelman's disease and body cavity based lymphoma. The virus encodes several cell derived proteins, among them k-cyclin, a

► cyclin D homologue.

► Viral Oncogenesis

Human Immunodeficiency Virus Type 1

Definition

Human immunodeficiency virus type 1 (HIV-1) is a retrovirus that is the etiological agent responsible for HIV disease/► AIDS (acquired immune deficiency syndrome).

► Reverse Transcriptase

► Translational Frameshifting, Non-standard Reading of the Genetic Code

Human Leukocyte Antigens

Definition

These polypeptides are genetically encoded in the ► major histocompatibility complex (MHC) on the

short arm of chromosome 6. These highly polymorphic antigens are not only expressed on leukocytes, but on almost all cells containing a nucleus. HLA molecules are expressed on the surface of cells, and are important for the recognition of the body's own and foreign cells by the immune system. Their main function is to help the immune system to defend against invaders such as bacteria, viruses and parasites. The immune system can also recognize as foreign the histocompatibility antigens of other people's cells and will attend them, causing rejection of grafts. There are many different HLA antigens (e.g. class I, II; HLA-A, HLA-B, HLA-DR, rendering the individual immunologically unique). Therefore, they are particularly important for transplantation medicine and play an important role in autoimmune disorders.

► **Diabetes Mellitus, Genetics**

► **Narcolepsy**

► **Rheumatoid Arthritis**

Human Repetitive DNA

ADAM PAVLICEK, VLADIMIR V. KAPITONOV,
JERZY JURKA

Genetic Information Research Institute,
Mountain View, CA, USA
jurka@girinst.org

Definition

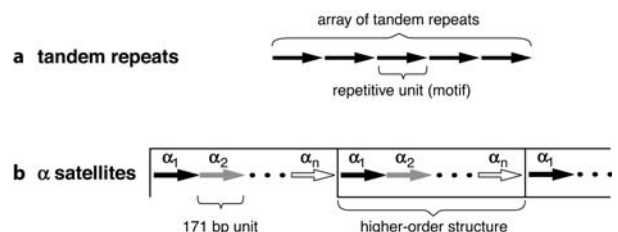
Repetitive DNA is usually defined as DNA present in multiple copies in the genome, without a clearly defined biological function. There are two main groups of ► **repetitive elements**, tandem repeats and interspersed repeats. ► **Tandem repeats** are repetitions of the same sequence motifs arrayed in a head-to-tail fashion. They occupy a significant fraction of heterochromatin and centromeric regions. ► **Interspersed repeats** are remnants of ► **transposable elements** (TEs), and constitute approximately 45% of the human DNA (1). TEs are either autonomous, i.e. encoding proteins necessary for their amplification, or nonautonomous, amplified by proteins encoded by autonomous elements. Most interspersed repeats are highly mutated ancient TEs, incapable of transposition. However, there are a few families of non-LTR retrotransposons including Alu and LINE1 (L1) elements that still actively proliferate in the human genome and occasionally insert into human genes (2). Interspersed repeats can also promote

chromosomal instabilities by homologous recombination between dispersed copies, causing various congenital defects and cancer (3). On the protein level, antigens produced by the translationally competent elements, particularly by endogenous ► **retroviruses**, can trigger an autoimmune reaction against affected tissues (4).

Characteristics

Tandem repeats are divided into microsatellites, minisatellites, centromeric satellites, and telomeric and subtelomeric repeats (Fig. 1). ► **Microsatellites** are tandem repetitions of short (1–9 bp) units. ► **Minisatellites** are composed of longer (10 and more bp) units. The 10 bp boundary separating micro- and minisatellites is somewhat arbitrary as it ranges from 6–13 bp in the literature. Micro- and minisatellites are traditionally also referred to as simple sequence repeats (SSRs) or variable number of tandem repeats (VNTRs), respectively. In the human genome, microsatellites composed of short 1–2 bp units are most frequent. Minisatellites, particularly those with longer units, are relatively rare. The total length of micro- and minisatellite arrays varies from a dozen or hundreds of bp for the shortest microsatellites to dozens of kb for minisatellites. ► **Satellites** are minisatellite-like units spanning up to 10^6 bp and located in well-defined chromosomal regions such as heterochromatin and centromeric regions (5). Alpha satellites representing repetitions of 171 bp units assembled into higher-order structures (Fig. 1b) are most abundant.

► **Telomeric repeats** are located at the ends of human chromosomes called telomeres. They are composed of the tandemly repeated conserved TTAGGG hexamer creating 3–20 kb long arrays (6). Telomeres play an essential role in protecting chromosomes from damage and degradation. In addition, telomeres can act as regulators of cell life span and affect cancer predisposition.



Human Repetitive DNA. Figure 1 Structure of human tandem repeats. (a) Schematic structure of an array of tandem repeats. (b) Higher-order structure of centromeric alpha satellites. The human genome contains about 12 distinct types of alpha satellite units (small arrows with different intensity of gray). A repetition of specifically ordered different units is called a higher-order structure (marked by boxes).

►**Subtelomeric repeats** are satellite-like sequences located at the 100–300 kb long boundary between the telomere and the rest of the chromosome. These repeats are generally derived from the telomeric TTAGGG hexamer, but are much less conserved and may differ between chromosomes (6).

Interspersed repeats are created by germline insertions of transposable elements (TEs). Copies of different TEs share founder-specific mutations and are classifiable into families or subfamilies of repeats. Most human interspersed repeats were active 30–150 million years ago; the oldest known repeats are older than 200 million years. Currently there are only a few active families of non-LTR retrotransposons in the human genome (1, 2, 7).

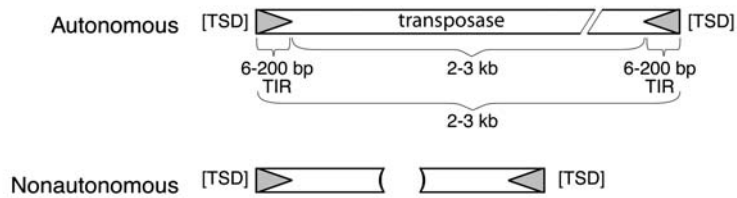
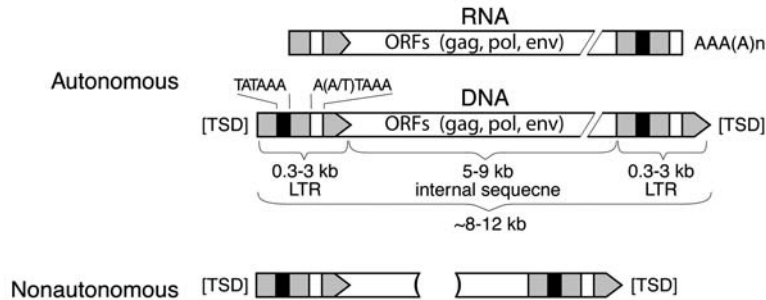
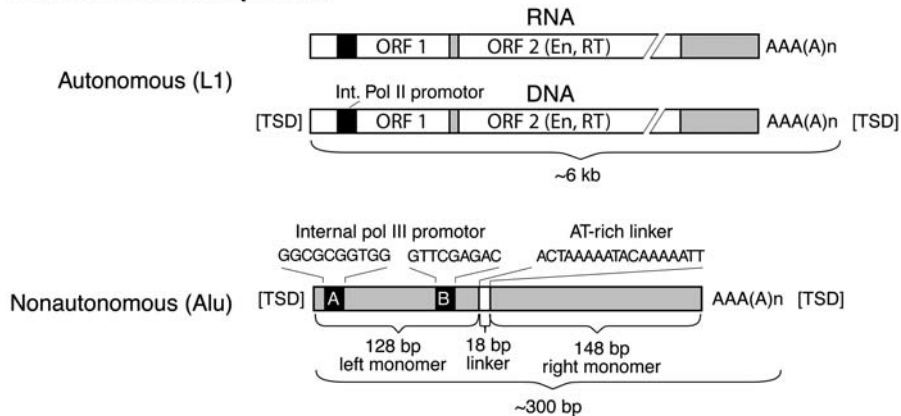
The human genome harbors three major groups of interspersed repeats derived from DNA transposons, non-LTR retrotransposons and retrovirus-like LTR-retrotransposons (1, 7). ►**DNA transposons** amplify via so-called cut-and-paste transposition, i.e. excision from one (usually freshly replicated) genomic site and insertion of the excised DNA into a new location. Autonomous DNA transposons code for an enzyme transposase that binds terminal inverted repeats and catalyses transposition (Fig. 2). The vast majority of human DNA transposons are inactive relics of ancient insertions with no elements capable of ►**transposition** detected (Table 1).

Non-LTR ►**retrotransposons** and retrovirus-like LTR-retrotransposons proliferate via an RNA intermediate reverse transcribed into complementary DNA (►**cDNA**). The ►**retrotransposition** process is catalyzed by ►**reverse transcriptase** (RT), an RNA-dependent DNA polymerase encoded by autonomous elements. LTR-retrotransposons, including human endogenous retroviruses (HERVs), structurally resemble exogenous retroviruses (Fig. 2). Autonomous LTR-retrotransposons contain several open reading frames (ORFs) homologous to retroviral *gag*, *pol* (*pro*) and, in some cases, *env* and other auxiliary genes (4). A tRNA primer encoded by the host cell initiates reverse transcription of LTR-retrotransposons, which takes place in cytoplasmic retrovirus-like particles. The retroviral reverse transcription produces a DNA copy that differs from the original RNA by the presence of long terminal repeats (LTRs) containing pol II promoters and other regulators of transcription. The DNA copy inserted into the genome is called a ►**provirus**. Currently, there are no retrotransposition-competent LTR-retrotransposons in the human genome and no insertion of LTR-retrotransposons is reported in human mutation databases. However, one member of the HERVK family with intact ORFs is polymorphic in the human population. In addition, several elements of young HERV families may contain individual ORFs

capable of translation and production of functional proteins.

The main characteristics distinguishing non-LTR from LTR retrotransposons is the lack of long terminal repeats and their utilization of DNA target-primed reverse transcription. The DNA genome of non-LTR retrotransposons is structurally collinear with RNA (Fig. 2). Transcription of these elements is initiated from internal pol II promoters. Autonomous non-LTR retrotransposons are called ►**long interspersed repeats** (LINEs). The human genome contains three families of LINEs: L1 (LINE1), which includes active subfamilies, and the extinct L2 and L3 LINEs similar to CR-1 elements from birds, fishes, and reptiles. Active L1 elements are around 6 kb long and contain two open reading frames, *ORF1* and *ORF2* (2). *ORF1* encodes a nucleic acid binding protein of uncertain function. *ORF2* encodes a single protein with reverse transcriptase and endonuclease domains (Fig. 2). The coupled nuclear process of reverse transcription and integration is primed by the 3' end of the initial nick created during cleavage of the genome by L1 endonuclease. The vast majority of the ~500,000 L1 copies in the haploid human genome (Table 1) are inactive, and there are only about 50 copies with intact reading frames capable of retrotransposition. The active copies are mostly from the L1-Ta subfamily.

Nonautonomous non-LTR retrotransposons are called ►**short interspersed elements** (SINEs). SINEs are transcribed by pol III polymerase (Fig. 2). The most abundant human SINEs are Alu elements, mobilized by the reverse transcriptase/endonuclease of L1 elements. Human Alu elements are approximately 300 bp long, dimeric sequences derived from 7SL RNA, which is a component of the ►**signal recognition particle** (SRP). Over one million Alu elements are present in the human genome (Table 1) and some AluY subfamilies still continue to retrotranspose (2). Apart from Alu elements, the human genome contains another active family of SINEs called SVA. SVA repeats are composite elements containing one complete Alu, one Alu monomer, fragments of HERVK endogenous retrovirus including one LTR and a microsatellite region. SVA elements are probably L1-dependent, since they share similar structural features, including the TTAAAA target site, 5–20 bp long target site duplication and the 3' polyA tail. In addition to SINEs, L1 elements can also occasionally retrotranspose cellular RNAs, such as intronless copies of human genes called ►**processed pseudogenes**. Due to the lack of promoters in mRNA, processed pseudogenes are not functional. Two other families of human SINEs are remnants of extinct MIR and MIR3 families that coamplified with L2 and L3 LINEs respectively, 200–300 millions years ago.

a DNA transposons**b LTR retrotransposons****c Non-LTR retrotransposons**

Human Repetitive DNA. Figure 2 Structure of human interspersed repeats. (a) DNA transposons. Human DNA transposons contain a single ORF flanked by terminal inverted repeats (TIRs) and by 2-10 bp long target site duplications (TSDs). (b) LTR retrotransposons. Autonomous elements contain an internal coding sequence with two or more open reading frames. Retroviral RNA and DNA differ from each other by the presence of 0.3-3 kb long terminal repeats (LTRs) in DNA. LTR harbors an RNA polymerase II promoter (black TATA box) and polyadenylation signal (white AATAAA or ATATAA signal). Elements are flanked by 4-6 bp long TSDs. (c) Non-LTR retrotransposons. Active autonomous L1 contains an internal RNA polymerase II promoter (black box) and two ORFs. In contrast to LTR retrotransposons, DNA of non-LTR retrotransposons is a perfect copy of RNA. Nonautonomous Alu elements are dimers composed of two monomeric units separated by short AT-rich linker. An internal RNA polymerase III promoter is composed of conserved boxes called A and B (in black). Elements are flanked by 5-20 bp long TSDs.

Detection of repetitive elements is a routine part of annotation and analysis of DNA sequences. Several specialized programs and databases have been developed for this purpose (Table 2).

Clinical Relevance

Expansions of trinucleotide repeats and other microsatellites may cause several genetic disorders (see Repeat Expansion Diseases). In addition to repeat

Human Repetitive DNA. Table 1 Main groups of human interspersed repeats

Repetitive group	Families	Active families	Density (%)
DNA transposons	~100 (29 autonomous)	none	5
LTR retrotransposons	~200	none (HERVK?)	9
non-LTR retrotransposons			
a. LINEs	3 (L1, L2, L3)	L1	21
b. SINEs	4 (Alu, SVA, MIR, MIR3)	Alu, SVA	13

Human Repetitive DNA. Table 2 Programs and databases for analysis of human repeats

Database/Program	Description
Repbase Update	Database of eukaryotic repetitive sequences, including more than 600 families of human repeats. ► (http://www.girinst.org/Repbase_Update.html)
HERVd	Database of human endogenous retroviruses, maps all known endogenous retroviruses in the human genome. ► (http://herv.img.cas.cz/)
Human TE Insertion	The database collects reported insertions of repetitive elements into the human genome associated with human genetic diseases. ► (http://www.med.upenn.edu/genetics/labs/kazazian/human.html)
Censor	Program for identification of interspersed and tandem repeats in both nucleic and amino acid sequences. Both on-line and local versions are available. ► (http://www.girinst.org/Censor_Server.html)
RepeatMasker	Program for identification of interspersed and tandem repeats in nucleic acid sequences. Both on-line and local versions are available. ► (http://www.repeatmasker.org/)
Tandem Repeat Finder	Specialized program for detection of tandem repeats in DNA sequences. Both on-line and local versions available. ► (http://tandem.bu.edu/trf/trf.html)
EMBOSS	A general package for analysis of biological sequences includes programs etandem and quicktandem for detection of tandem repeats. ► (http://www.hgmp.mrc.ac.uk/Software/EMBOSS/index.html)

expansions, tandem repeats can induce genetic disorders by recombination between different arrays. Recombinations between both centromeric and telomeric satellites are involved in chromosomal rearrangements including chromosome fragility or jumping chromosomal translocations.

TE-derived interspersed repeats can also contribute to genetic instability. Active retrotransposons can occasionally insert into genes and inactivate them (► [insertional inactivation](#)). The estimated rate of genomic

retrotransposition is around 1 insertion per 4–100 individuals (2). New inserts often introduce frame-shifts, premature stop codons, or both. RNAs with ► [premature termination codons](#) (PTCs) are usually destroyed by ► [nonsense-mediated RNA decay](#) (NMD) and the corresponding defective protein is not produced. The only recently inserted elements are either L1 elements (15 cases), or L1-dependent Alu (24 cases) and SVA SINEs (4 cases). Table 3 lists insertions of TEs connected with human diseases. Many cases listed

Human Repetitive DNA. Table 3 Insertions of transposable elements and human diseases

Gene	Symbol	Chr.	Phenotype
<i>L1 insertions</i>			
adenomatosis polyposis coli	APC	5q22.2	colon cancer
fukutin	FCMD	9q31.2	Fukuyama-type congenital muscular dystrophy
b-Globin	HBB	11p15.4	beta thalassemia
retinitis pigmentosa 2 gene	RP2	Xp11.3	X-linked retinitis pigmentosa 2
cytochrome b-245, beta polypeptide	CYBB	Xp11.4-Xp21.1	chronic granulomatous disease
dystrophin	DMD	Xp21.1-2	3x muscular dystrophy, 1x X-linked dilated cardiomyopathy
choroideremia gene (Rab escort protein 1)	CHM	Xq21.2	choroideremia
coagulation factor IX	F9	Xq27.1	hemophilia B
coagulation factor VIII	F8	Xq28	hemophilia A
<i>Alu insertions</i>			
human homologue of MMLV retrovirus insertion locus	MLVI-2	n.a.	associated with leukemia
alpha 4 type IV collagen precursor	COL4A3	2q36.3	Alport syndrome
calcium-sensing receptor	CASR	3q21.1	familial hypocalciuric hypercalcemia
butyrylcholinesterase	BCHE	3q26.1	acholinesterasemia
adducin 1, transcript variant 1	ADD1	4p16.3	Huntington disease
protein tyrosine phosphatase, non-receptor type 13	PTPN13 (APO-1/CD95)	4q21.3	autoimmune lymphoproliferative syndrome
adenomatosis polyposis coli	APC	5q22.2	hereditary desmoid disease
CMP-N-acetylneuraminic acid (CMP-Neu5Ac) hydroxylase	CMAH	6p22.3	inactivated in humans, fixed
eyes absent homolog 1 (Drosophila)	EYA1	8q13.3	branchio-oto-renal syndrome
fibroblast growth factor receptor 2 isoform 1	FGFR2	10q26.13	Apert syndrome
serine/cysteine proteinase inhibitor	C1NH	11q12.1	C1 inhibitor deficiency
porphobilinogen deaminase (hydroxymethylbilane synthase)	PBGD/HMBS	11q23.3	acute intermittent porphyria
breast cancer 2	BRCA2	13q13.1	hereditary breast cancer
neurofibromin 1	NF1	17q11.2	neurofibromatosis
angiotensin I converting enzyme isoform 1	ACE	17q23.3	protection against myocardial infarction and age-related macular degeneration?
glycerol kinase	GK	Xp21.2	glycerol kinase deficiency
interleukin 2 receptor, gamma	IL2RG	Xq13.1	X-linked severe combined immunodeficiency
Bruton agammaglobulinemia tyrosine kinase	BTK	Xq22.1	X-linked agammaglobulinemia

Human Repetitive DNA. Table 3 Insertions of transposable elements and human diseases (Continued)

Gene	Symbol	Chr.	Phenotype
coagulation factor IX	F9	Xq27.1	hemophilia B
coagulation factor VIII	F8	Xq28	hemophilia A
<i>SVA insertions</i>			
alpha spectrin	SPTA1	1q23.1	hereditary elliptocytosis
fukutin	FCMD	9q31.2	Fukuyama-type congenital muscular dystrophy
low density lipoprotein receptor	LDLR	19p13.2	familial hypercholesterolemia
Bruton agammaglobulinemia tyrosine kinase	BTK	Xq22.1	X-linked agammaglobulinemia

in Table 3 are *de novo* insertions, i.e. detected in children of parents without the insertion. L1 insertions are exclusively derived from the most active subfamilies L1-Ta (or L1-preTa) and Alu insertions are from the most active AluY subfamilies. Except for one somatic L1 insertion into the *adenomatous polyposis coli* (*APC*), all other cases listed in Table 3 are germline insertions. Not all insertional inactivations cause a strong negative phenotype. Insertions into nonessential genes may be under weak negative selection and eventually become fixed. For example, an AluYb8 insertion that inactivated *CMP-N-acetylneuraminic acid hydroxylase* (*CMAH*) is fixed in the human lineage, but other primates including chimp contain a functional copy of *CMAH* without the Alu insert. Distribution of TE insertions is nonuniform across the genome (2). Some genes such as the X-linked *coagulation factors VIII* and *XI* and the *dystrophin* gene have been targeted by several independent events. As much as 49% (21 out of 43) of all documented insertions have occurred in genes located on chromosome X, which represents only 5% of the human genome. This bias is even stronger for L1 insertions of which 86% (12/14) occur on chromosome X. The bias is probably caused by the dominant character (loss of function) of mutations on chromosome X, since only one of its functional copies is present in somatic cells. Males have only one X copy and in females the second copy is silenced by **X-chromosome inactivation**. Thus X-linked mutations have a dominant effect, while autosomal gene inactivation can be compensated by the copy on the sister chromosomes and no disease is detected in heterozygotic individuals.

Insertional inactivations listed in Table 3 target both exons and introns, indicating that even intronic insertions can be damaging. Insertions into splicing regulatory regions can cause exon skipping, i.e. deletion within mRNA or insertion of noncoding

intronic sequences into mRNA (intron inclusion). Intronic inclusions can also be produced by activation of **cryptic splice signals** in existing intronic repeats. Notably, Alu elements contain cryptic splice donor and acceptor sites in the antisense orientation that, after specific mutations, can give rise to new Alu-derived exons. Such Alu cassettes can be alternatively spliced and typically produce premature termination codons. However, if the new exon becomes constitutively spliced into mRNA, the NMD control prevents efficient synthesis of the protein. So far, there are two known cases of disease-related Alu splicing into cellular mRNAs. An Alu inclusion into mRNA coding for ornithine-delta-aminotransferase (OAT) led to OAT deficiency. The second case was reported for a patient with autosomal recessive Alport syndrome, caused by an Alu inclusion into the precursor of alpha 3 type IV collagen mRNA.

Similar copies of TEs interspersed with genomic DNA may promote homologous recombinations leading to genetic instabilities. Table 4 lists loci with reported deletions, duplications and chromosomal translocations related to human diseases that are caused by recombination between repetitive elements. While for some loci only a single event of recombination between interspersed repeats is detected, other loci represent hotspots for recombinations and many independent cases have been reported. For example the *complement component 1 inhibitor precursor* (*C1NH*), *low-density lipoprotein receptor* (*LDLR*), *alpha-globin* and *adenomatous polyposis coli* (*APC*) genes are frequently affected by TE-induced recombinations. Analogously to insertions, detected recombinations are biased towards X-linked genes; 12 out of 54 (22%) deletion/duplication events were detected on chromosome X. Recombinations listed in Table 4 occurred in both germline and somatic cells. Somatic recombinations are involved in many leukemia- or other cancer-related

Human Repetitive DNA. Table 4 Repeat-induced recombinations and human diseases

Locus	Symbol	Chr.	Phenotype
<i>Deletions by Alu-mediated recombination</i>			
alpha(1,2)fucosyltransferase	FUT2	19q13.33	ABO-Bombay phenotype
ATP-binding cassette, sub-family D, member 1	ABCD1	Xq28	adrenoleukodystrophy
Bruton's tyrosine kinase	BTK	Xq22.1	agammaglobulinemia, X-linked
Albinism ocular type 1 gene	OA1	Xp22.22	albinism ocular type 1
complement component 1 inhibitor precursor	C1NH/ SERPING1	11q12.1	angioedema, hereditary
tapasin	TAPBP	6p21.32	bare lymphocyte syndrome, type I
breast cancer 1, early onset	BRCA1	17q21.31	hereditary breast and/or ovarian cancer
epidermal growth factor receptor	EGFR	7p11.2	cancer, glioblastomas
retinoblastoma	RB1	13q14.2	cancer predisposition
adenomatous polyposis coli	APC	5q22.2	colorectal cancer, adenomatous polyposis
mismatch repair protein	MSH2	2p21	colorectal cancer, nonpolyposis
mutL homolog 1	MLH1	3p22.3	colorectal cancer, nonpolyposis
insulin-receptor	INSR	19p13.2	diabetes, insulin-resistant
Ewing sarcoma breakpoint region 1	EWSR1	22q12.2	protection against Ewing's sarcoma
Alpha-galactosidase A	GLA	Xq22.1	Fabry disease
Fanconi anemia, complementation group A	FANCA	16q24.3	Fanconi anemia
integrin alpha 2b precursor, platelet glycoprotein lib, CD41B	ITGA2B	17q21.31	Glanzmann s thrombasthenia
alpha-glucosidase	GAA	17q25.3	glycogen storage disease type II
heme oxygenase-1	HMOX1	22q12.3	heme oxygenase-1 deficiency
coagulation factor VIII	F8	Xq28	hemophilia A
caspase-activated DNase, beta polypeptide	CAD	1p36.32	hepatoma
alpha chain of complement component 3	C3	19p13.3	hyperlipidemia, familial combined
apolipoprotein B precursor	APOB	2p24.1	hypobetalipoproteinemia
tumor protein p53	TP53	17p13.1	Li-Fraumeni syndrome
solute carrier family 7	SLC7A7	14q11.2	lysineuric protein intolerance
platelet-derived growth factor beta	PDGFB	22q13.1	Meningioma
proteolipid protein gene	PLP1	Xq22.2	Pelizaeus-Merzbacher disease
ATP-binding cassette, sub-family C, member 6	ABCC6	16p13.11	pseudoxanthoma elasticum
retinoschisis factor	RS1	Xp22.13	retinoschisis
hexosaminidase B, beta polypeptide	HEXB	5q13.3	Sandhoff disease
adenosine deaminase	ADA	20q13.12	severe combined immunodeficiency (SCID)

Human Repetitive DNA. Table 4 Repeat-induced recombinations and human diseases (Continued)

Locus	Symbol	Chr.	Phenotype
-		del(5)(q12.3)	small-cell lung cancers
ATP-binding cassette transporter, sub-family A member 1	ABCA1	9q31.1	Tangier disease
hexosaminidase A, alpha polypeptide	HEXA	15q24.1	Tay Sachs disease
alpha globin gene cluster	-	del(16)(p)	thalassemia, alpha
beta -globin gene cluster	-	del(11)(p)	thalassemia, delta, gamma
serine (or cysteine) proteinase inhibitor, clade C	SERPINC1	1q25.1	thrombophilia
tuberous sclerosis 2	TSC2	16p13.3	tuberous sclerosis
sclerosteosis	SOST	17q21.31	van Buchem disease
Wiskott-Aldrich syndrome protein	WAS	Xp11.23	Wiskott-Aldrich syndrome
presenilin-1	PSEN1	14q24.2	Alzheimer's disease, early onset
von Willebrand factor A	VWF	12p13.31	von Willebrand disease
breast cancer 2	BRCA2	13q13.1	breast cancer
low density lipoprotein receptor	LDLR	19p13.2	hypercholesterolemia, familial
hypoxanthine phosphoribosyltransferase	HPRT	Xq26.2	Lesch-Nyhan syndrome
iduronate-2-sulfatase	IDS	Xq28	Hunter disease
N-acetylgalactosamine-6-sulfatase precursor	GALNS	16q24.3	mucopolysaccharidosis type IVA
<i>Deletion by nonhomologous Alu-L1 recombination</i>			
Dystrophin	DMD	Xp21.1-2	muscular dystrophy
<i>Deletion by L1-mediated recombination</i>			
type IV alpha 6 collagen	COL4A6	Xq23	Alport syndrome and diffuse leiomyomatosis
phosphorylase kinase, beta subunit	PHKB	16q12.1	glycogen storage disease
mutL homolog 1	MLH1	3p22.3	hereditary nonpolyposis colorectal cancer
<i>Duplication by Alu-mediated recombination</i>			
low density lipoprotein receptor	LDLR	19p13.2	hypercholesterolemia, familial
hypoxanthine phosphoribosyltransferase	HPRT	Xq26.2	Lesch-Nyhan syndrome
lysyl hydroxylase 1/procollagen-lysine, 2-oxoglutarate 5-dioxygenase	PLOD	1p36.22	Ehlers Danlos syndrome
<i>Deletion/Duplication by HERV-mediated recombination</i>			
DEAD/H (Asp-Glu-Ala-Asp/His) box polypeptide	DBY USP9Y	Yq11.21	male infertility
<i>Alu-mediated translocations</i>			
t(11;22)	-		Ewing's sarcoma
t(1;19)(q21.3;q13.2)	PAFAH1B3-CLK2 fusion		mental retardation, ataxia and atrophy of the brain
t(X;18)	-		synovial sarcoma

Human Repetitive DNA. Table 4 Repeat-induced recombinations and human diseases (Continued)

Locus	Symbol	Chr.	Phenotype
t(Y;X)	DXYS5 locus		XX maleness
t(9;11)(p24;q23)	-		bipolar affective disorder (manic depression)
<i>Alu-mediated rearrangements associated with various forms of leukemia</i>			
7q22-q34 inversion dup(11)(q23).	many genes including MLL, MOZ CPB, etc.		leukemia, various acute/chronic forms
Philadelphia translocation t(9;22) and variants			
t(10;16)(q22;p13)			
t(11;18)(q21;q21)			
t(11;19) (q23;p13.3)			
t(14;18)			
t(14;19)			
t(2;8)			
t(4;11)			
t(8;16)(p11;p13)			
t(9;11)			

cases. Crossovers between repeats leading to DNA deletions have been reported in 54 genes, whereas duplications have been reported for only five loci.

Alu elements are by far the most active repeats involved in known rearrangements (Table 4). Alu-mediated recombinations have been reported for 55 out of 61 genes, L1 recombinations for three and endogenous retroviruses for one. One deletion and one translocation were induced by nonhomologous Alu-L1 recombinations. Alu elements are the most abundant repeats in GC- and gene-rich regions and frequently form clusters of highly identical copies prone to recombination in the gene-rich regions.

Alu repeats have also been implicated in the etiology of recurrent somatic chromosome translocations, particularly in several forms of acute and chronic leukemias (3). Repeat-induced recombinations associated with large rearrangements can lead to the **loss of heterozygosity** (LOH) in cancers and, therefore, to cancer progression. In the germline, constitutional congenital disorders can be caused by nonlethal cytogenetically balanced chromosomal translocations. For example, a patient with cranial deformations and mental retardation had a t(1;19)(q21.3;q13.2) congenital translocation, stimulated by nonhomologous Alu-L1 recombination. In another case, a familial form of the

bipolar affective disorder (manic depression) was caused by the congenital Alu-mediated translocation t(9;11)(p23;q23).

In rare situations, rearrangements induced by TEs may even protect against harmful recombinations. For example, primitive neuroectodermal tumors, including Ewing's sarcoma, are strongly linked to the reciprocal translocation t(11;22)(q24;q12) fusing the *Ewing's sarcoma (EWS)* gene on 22q.12 with the *Friend leukemia virus integration region 1 (FLII)* gene on 11q.24. A small (2.4 kb) Alu-mediated deletion inside *ESW* intron 6 located close to the translocations breakpoints seems to inhibit the t(11;22)(q24;q12) translocations. African human populations that carry this deletion have a tenfold decreased incidence of Ewing's sarcoma compared to other populations.

So far, it is estimated that insertions and recombinations between interspersed repeats are responsible for 0.1% and 0.3% of human genetic disorders, respectively. However, the mechanism of large DNA rearrangements representing ~7.4% of all human mutations remains to be elucidated.

Apart from obvious genetic rearrangements caused by repetitive elements, proteins encoded by transposable elements can also induce pathological processes in human cells and tissues. In particular, TE-encoded

protein antigens may trigger an immune reaction against affected tissues leading to autoimmune abnormalities (4). The relevant research was done mostly on human endogenous retroviruses. HERV expression has been linked to several autoimmune diseases such as multiple sclerosis, rheumatoid arthritis, systemic lupus erythematosus or insulin-dependent diabetes mellitus. A link to schizophrenia and some types of cancer has also been proposed. Retroviral mRNAs, and in some cases even entire retroviral particles were detected in affected tissues. It is possible that production of retroviral antigens such as the *gag* (group-specific antigen) may provoke an autoimmune response. However, since increased expression of TEs is associated with many pathological processes this relationship may be coincidental.

References

1. Lander ES, Linton LM, Birren B et al (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860–921
2. Ostertag EM, Kazazian Jr (2001) Biology of mammalian L1 retrotransposons. *Annu Rev Genet* 35:501–538
3. Kolomietz E, Meyn MS, Pandita A et al (2002) The role of Alu repeat clusters as mediators of recurrent chromosomal aberrations in tumors. *Gene Chromosome Canc* 35:97–112
4. Lower R, Lower J, Kurth R (1996) The viruses in all of us: characteristics and biological significance of human endogenous retrovirus sequences. *Proc Natl Acad Sci USA* 93:5177–5184
5. Wellinger RJ, Sen D (1997) The DNA structures at the ends of eukaryotic chromosomes. *Eur J Cancer* 33:735–749
6. Lee C, Wevrick R, Fisher RB et al (1997) Human centromeric DNAs. *Hum Genet* 100:291–304
7. Kapitonov VV, Pavlicek A, Jurka J (2004) Anthology of Human Repetitive DNA: In: Meyers RA (ed) *Encyclopedia of Molecular Cell Biology and Molecular Medicine*, vol 1. Wiley-VCH, Weinheim, pp 251–305

Human T-Cell Leukemia Virus

Definition

Human T cell leukemia virus (HTLV) is a member of the retrovirus family. HTLV type 1 causes adult T-cell leukemia/lymphoma, a malignancy of T-lymphocytes. HTLV-2 is less pathogenic. The virally encoded Tax oncoprotein can transform T-cells *in vitro*. Biochemically, ►Tax acts as a transcriptional transactivator of NF- κ B and CREB pathways, and stimulates the cell

cycle in the G1-phase by binding to cyclin-dependent kinase holoenzymes.

►Viral Oncogenesis

Humanization

Definition

Humanization refers to a process where mouse DNA sequences are replaced with equivalent human DNA sequences; e.g. in ES cells and hence in mice.

►Large-Scale Homologous Recombination Approaches in Mouse Embryonic Stem Cells

Humanized Antibodies

Definition

Humanized antibodies are hybrid immunoglobulins in which only the murine residues of the complementary determining regions, and others of possible structural relevance, are preserved, while all other sequences are exchanged by those of a human antibody framework. Such antibodies are produced by recombinant methods.

►Monoclonal Antibodies

►Protein Interaction-Phage Display

Humoral Immune Response

Definition

Humoral immune response describes the induction of antigen-specific antibodies.

►DNA-based Vaccination

Huntington Chorea

►Huntington's Disease

Huntington's Disease

STEFAN WIECZOREK, JÖRG T. EPPLER
Human Genetics, Ruhr-University, Bochum, Germany
stefan.wieczorek@rub.de
joerg.t.epplen@rub.de

Synonyms

HD; Huntington Chorea; Chorea major

Definition

Huntington's Disease (HD) is a progressive neurodegenerative disorder with autosomal **▶dominant** inheritance and usually adult onset (4th to 5th decade of life). Clinical findings include motor, psychiatric and cognitive abnormalities. HD is caused by an abnormal elongation of a polymorphic (CAG)_n trinucleotide block beyond a critical threshold within the first exon of the *huntingtin* gene (*htt*, *IT15*). Postmeiotic instability of the (CAG)_n block, leading to further expanded **▶alleles**, accounts for the clinical phenomenon of **▶anticipation**, i.e. earlier onset and possibly more severe disease course in subsequent generations. As in related disorders, e.g. spinobulbar muscular atrophy (**▶SBMA**), dentatorubral-pallidoluysian atrophy (**▶DRPLA**) and different types of spinocerebellar ataxia (**▶SCA**), the resulting elongated polyglutamine (polyQ) stretch within the protein causes a complex alteration of its interactions with crucial factors of cell integrity, many of which are still not completely understood. Morphological studies in HD reveal substantial loss of neuronal cells, predominantly occurring in the caudate, subthalamic nucleus and putamen. On a microscopic level characteristic cytoplasmic and intranuclear huntingtin-positive inclusion bodies are detectable.

Characteristics

Clinical Features

HD, first described by George Huntington in the 19th century, is a complex neurological disorder providing variable clinical manifestations. The most impressive symptom is deteriorating loss of motor control, often reported as beginning slightly with awkwardness or fidgetiness, difficulties in swallowing or control of eye movements. With disease progress, motor symptoms often worsen to continued, uncontrollable jerking movements of numerous body parts involving limbs, trunk, neck and head while the muscle tone is generally decreased. These movements cannot be repressed voluntarily; they advance with physical or mental stress and abate only during sleep. Weight loss, caused by immense energy consumption, disturbed food intake

and possibly other factors directly linked to the pathogenesis of the disease, is a complicating factor in HD and often a critical issue in late stages. As the disorder advances, more rigid motor dysfunction with increased muscle tone and bradykinesia may come to the fore and patients are usually confined to bed. As a second aspect, psychiatric disturbances are a common feature of HD and may be the first symptom to be presented. Affective or schizophrenic psychosis may be preceded by behavioural changes like aggressiveness and impulsive behaviour, abjection and apathy, unsociability and alcohol abuse. The risk of suicide is increased among HD patients. The third major clinical feature in HD is a cognitive decline, which, although not obligatory, is present in most patients starting mildly at early stages of the disease. Deficits in memory and concentration as well as decreased ability to perform complex thought processes are reported. It may result in distinct dementia in late stages of the disorder. Patients usually die after 10 to 15 years of disease progression from complications of immobility and body wasting, with aspiration pneumonia being a typical cause of death. The prevalence of HD among Caucasian populations is about 3–5 per 100,000 and the age of onset lies within the fourth or fifth decade in most patients. There is, however, major variability and first symptoms are reported in adolescents as well as in elder people.

Genetics

HD is inherited in an autosomal dominant manner, usually with full **▶penetrance**. By linkage analysis the HD causing **▶mutation** was mapped to chromosome 4p16.3 in 1983. In 1993 the pathological expansion of the polymorphic (CAG)_n block in the first **▶exon** of the *huntingtin* (*htt*, *IT15*) gene was defined as molecular cause of the disorder (1). The *htt* gene consists of 67 exons spanning 180,000 nucleotides with the mature huntingtin protein comprising 3,144 amino acids (350 kD). In the normal population the size of the trinucleotide motif varies between 10 and 35 CAG copies with most chromosomes comprising up to 26 CAG copies and the (CAG)₁₇ allele being most common. In contrast, pathologically expanded (CAG)_n stretches between 36 and >120 copies are found in patients with HD. While alleles \geq (CAG)₄₀ show full penetrance, reduced penetrance has been observed for the rare intermediate alleles between (CAG)₃₆ and (CAG)₃₉. Since the genetic defect is known, **▶predictive genetic testing** is available for persons at risk for HD. There is a strong statistical correlation between the size of the (CAG)_n stretch and the average age of onset. Large alleles predispose to an earlier onset of first symptoms and a more severe disease course. However, within the most common group comprising 40–45 CAG copies the age of onset varies considerably and

the number of CAG copies must not be used for predicting clinical features in most cases. Yet, patients with alleles $>(\text{CAG})_{60}$ usually present first symptoms in adolescence and rare cases with childhood onset and trinucleotide blocks $>(\text{CAG})_{200}$ copies have been reported. The clinical features in adult onset usually differ from the ones in early onset HD, the latter showing more Parkinson-like **▶phenotype** with high, rigid muscle tone and bradykinesia or akinesia, severe cognitive decline and rapid disease progression. The origin of large $(\text{CAG})_n$ stretches is likely to be caused by postmeiotic expansion of an already enlarged allele in one of the patient's parents, usually the father. This mechanism constitutes the genetic correlate of the clinical finding of anticipation, i.e. an average decrease of age of disease onset in subsequent generations. In rare cases this may lead to the situation that a child shows first symptoms before its parent carrying the HD mutation, implying a case of a new mutation or skipping of a generation. This phenomenon is found far more often in families with paternal inheritance and is thought to have its origin in a higher mutation rate during spermatogenesis. Patients homozygous for the HD mutation do not show an earlier mean age of onset compared to heterozygous ones but recent studies indicate more severely affected phenotypes in homozygous patients.

Cellular and Molecular Regulation Huntingtin and Associated Proteins

Since the discovery of the molecular genetic defect in HD several animal (e.g. mice, rats, *Drosophila*) and cellular models have been developed to examine the possible pathways leading to cell death. These models and additional examinations in humans have revealed ample data not only of great value for understanding of HD but also several other disorders. HD is one of the most intensively studied diseases and has become a veritable model disease.

Located in the first exon of *htt* the $(\text{CAG})_n$ block is translated into a polyQ stretch close to the N-terminus of the mature huntingtin protein. The *htt* gene is expressed ubiquitously. Little is known about its physiological function but it is likely to be of great importance during neuronal development as homozygous *hdh* (the mouse homolog of human *htt*) knockout mice die early during embryogenesis shortly after gastrulation. There are numerous proteins associated with huntingtin, including HIP1, HIP2, HAP1 and GAPDH, a major component of cell glycolysis. HIP1 (huntingtin interacting protein) is a membrane-associated protein linked to the cytoskeleton, which is probably involved in mechanisms of endocytosis. Its interaction with huntingtin is impaired by the expansion of the polyQ stretch, indicating disturbed membrane-cytoskeletal integrity in HD.

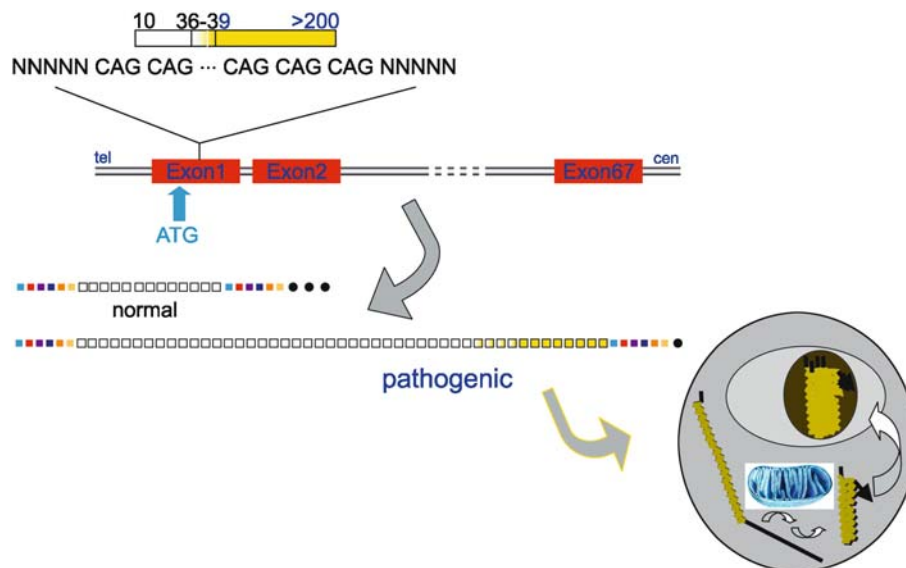
Over-expression of HIP1 results in activation of apoptosis *via* caspase 3. While HIP2 may be involved in huntingtin degradation, HAP1 (huntingtin associated protein) is also associated with the cytoskeleton. Recently, the interaction of another protein, HIP14, with huntingtin has been shown to be dependent on the length of the polyQ stretch. HIP14 is likely to play a role in cell trafficking and its interaction with huntingtin may contribute to the pathogenesis of HD.

Protein Aggregation—A Crucial Step in the Pathogenesis

The polyQ expansion causes a pathological **▶gain-of-function** in the huntingtin protein, reflecting the dominant mode of inheritance. A hallmark of the pathological processes underlying HD is the occurrence of cytoplasmic and, in particular, of intranuclear inclusions comprising huntingtin and ubiquitin. These are assumed to emanate from the intranuclear aggregation of N-terminal huntingtin fragments, possibly generated by caspase cleavage, to amyloid fibrils *via* the formation of oligomers. This process may be mediated either by transglutaminases or, more probably, by the potential of polyQ stretches to self-assemble, e.g. acting as polar zippers. While full-length wild-type huntingtin harbours a nuclear export signal at the C-terminus of the protein, intranuclear accumulation of N-terminal fragments of mutant huntingtin may be promoted by the lack of this motif. Notwithstanding, the impact of visible intranuclear inclusions has not been clarified sufficiently in the pathogenesis of HD. The original hypothesis that intranuclear inclusions constitute the major pathogenetic factor in HD has recently been disputed, and models of these inclusions have been discussed as being an epiphenomenon or even a protective factor established by the cell. Although the presence of intranuclear inclusions has been shown to lack correlation with huntingtin induced cell death, expanded polyQ stretches and even those of physiological huntingtin proteins have the potential to cause cell death when delivered to the nucleus (2). Thus, the toxic potential of intranuclear polyQ elements is generally accepted, but the stage of aggregation initiating the cascade to cell death remains to be elucidated. Oligomerization of polyQ elements may be crucial (3).

From Aggregation to Pathology

Various complementary models suggest how the intranuclear occurrence of polyQ elements causes cell death. Mutant huntingtin is not only able to build complexes with itself but also with other proteins including transcription factors like cAMP-response element binding protein (CREB) binding protein (CBP) and TATA binding protein (TBP), both containing polyQ elements, and SP1, harbouring a glutamine-rich



Huntington's Disease. Figure 1 The polymorphic (CAG)_n motif within the first exon of the *huntingtin* gene is translated into a polyglutamine stretch at the N-terminus of the huntingtin protein. Aggregation of polyglutamine elements beyond a critical threshold results in altered interaction of the huntingtin protein with various cellular elements including the mitochondria. Impaired degradation and nuclear accumulation of polyQ elements causes cell death predominantly in striatal neurons.

domain. However, in mouse models these transcription factors are not trapped within the intranuclear inclusions to a great extent, emphasizing the potential importance of the intermediate stages of polymerization. There is evidence for a reduced expression of signalling genes in a mouse model of HD (4). Mutant *htt* exon 1 protein binds to the acetyltransferase domain of CBP and p300/CBP associated factor (P/CAF) resulting in decreased levels of acetylated histones H3 and H4. Thus, impairment of protein expression controlled by huntingtin-interacting transcription factors may be important. Wild-type huntingtin up-regulates transcription of brain derived neurotrophic factor (BDNF) whereas mutant huntingtin lacks this ability (5). BDNF was found to be reduced in the putamen of HD patients. As BDNF is required for both differentiation and survival of neuronal cells, decreased BDNF activity may represent a ►**loss-of-function** pathway in HD pathogenesis. There is evidence that pathogenesis in HD is partly related to mitochondrial dysfunction. Enzymatic activities of different mitochondrial proteins have been shown to be decreased in HD neurons, affecting different complexes of the respiratory chain. Decreased metabolic capacity is proposed to enhance the susceptibility of the cell to chronic damage caused by oxidative stress and excitotoxicity subsequently leading to apoptosis. Huntingtin fragments are degraded by the proteasome complex in a polyQ length-dependent manner with fragments containing shorter polyQ stretches being

metabolized more efficiently. The ability of proteasomes to react with other proteins may be impaired by the binding of mutant huntingtin. It has been shown that proteasomal dysfunction is associated with disrupted mitochondrial membrane potential, cytochrome c release from the mitochondria, and activation of caspase-9 and -3-like proteases resulting in apoptotic cell death.

Point of Origin – Expansion Mechanisms in HD

Intermediate alleles carrying 27–35 CAG units show a greater tendency to expand into a pathological range, i.e. $\geq(\text{CAG})_{36}$, than the common alleles $\leq(\text{CAG})_{26}$, probably accounting for the rare cases of true new mutations in HD. Mechanisms by which this expansion is generated are not fully understood, nor is the process of how pathologically enlarged (CAG)_n blocks undergo further expansion during gametogenesis. Probably closely linked to these mechanisms is the occurrence of age-dependent somatic ►**mosaicism** with larger (CAG)_n stretches found in the striatum of HD patients (6). Recently a model has been established suggesting that expansions in germ cells are not due to mitotic mechanisms like homologous ►**recombination** or replication slippage of the DNA polymerase. In contrast, the expansion mechanism appears to be located within the postmeiotic haploid cell, i.e. during the maturation of spermatids to spermatocytes, and is likely to be mediated by a complex containing Msh2, a DNA mismatch repair protein. This complex may

stabilize DNA loops, which emerge after single-strand or double-strand breaks in the DNA, and facilitate their integration into the DNA by gap filling. As neuronal cells are generally post-mitotic and HD Msh2-knock-out mice do not show somatic instability of the (CAG)_n stretch, expansion mechanisms within tissues other than germ cells are likely to underlie a similar pathogenetic model.

Clinical Relevance

Due to its fatal course and the lack of a sufficient causal therapy, HD constitutes a great challenge for interaction with patients and their families. Predictive genetic testing for persons at risk for HD requires profound genetic counselling including psychological and social support in order to provide the individual with the best possible ability to cope with the test result, be it positive or negative. Current medication protocols for patients with manifest HD improve psychiatric symptoms using antidepressants, neuroleptics or anxiolytic drugs, respectively. Motor symptoms may be reduced by antidopaminergic neuroleptics like tiaprid or tetra-benazine whereas patients with hypokinetic features (e.g. Westphal variant) may even benefit from low doses of dopaminergic drugs. Additionally, physical, occupational and logopedic therapy constitute fundamental concepts in HD care.

Based on the pathogenetic mechanisms in HD, several new therapeutic approaches have been suggested. Prevention of aggregation is an obvious principle to overcome the probable basis of HD pathology. Certain heat shock proteins (HSP40 and HSP70) are able to bind polyQ proteins and alter their solubility. Over-expression of HSP40 and HSP70 results in reduced *htt* exon 1 protein aggregation in mammalian cells. Histone deacetylase inhibitors (HDAC) have been shown to slow down neurodegeneration in a Drosophila model of HD. Caspase inhibitors like minocycline, inhibiting the pathological cleavage of huntingtin, slow down disease progression in a HD mouse model. A recent finding is the neuroprotective potential of tauroursodeoxycholic acid (TUDCA), a hydrophilic bile acid, in HD mice, which is of particular interest as it may provide a therapeutic concept which is likely to show only minor side effects.

It is highly remarkable that HD symptoms and occurrence of nuclear inclusion bodies was shown to be reversible in mice expressing exon 1 of mutant *htt* after the mutant gene was switched off. Thus, the aforementioned approaches may be a basis not only for the prevention but also for a cure of HD symptoms in humans in future times.

►Polyglutamine Disease, the Emerging Role of Transcription Interference

►Repeat Expansion Disorders

References

1. The Huntington's Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 72:971–983
2. Yang W, Dunlap JR, Andrews RB et al (2002) Aggregated polyglutamine peptides delivered to nuclei are toxic to mammalian cells. *Hum Mol Genet* 11:2905–2917
3. Sanchez I, Mahlke C, Yuan J (2003) Pivotal role of oligomerization in expanded polyglutamine neurodegenerative disorders. *Nature* 421:373–379
4. Luthi-Carter R, Strand A, Peters NL et al (2000) Decreased expression of striatal signaling genes in a mouse model of Huntington's disease. *Hum Mol Genet* 9:1259–1271
5. Zuccato C, Ciammola A, Rigamonti D et al (2001) Loss of huntingtin-mediated BDNF gene transcription in Huntington's disease. *Science* 293:493–498
6. Telenius H, Kremer B, Goldberg YP et al (1994) Somatic and gonadal mosaicism of the Huntington disease gene CAG repeat in brain and sperm. *Nat Genet* 6:409–414

HUPO

The Human Proteome Organization (<http://www.hupo.org>).

►Protein Databases

Hybridization Probes

Definition

In the context of molecular biology, a probe traditionally refers to the labelled, soluble partner in a nucleic acid hybridization experiment such as a Southern, northern or dot blot. More recent usages in the field of microarrays confusingly invert the usage, and call the partner bound to the solid support the “probe” and the soluble, labelled partner the “target”.

►YAC and PAC Maps

Hybridization

Definition

Hybridization is the process of binding complementary single-stranded DNA molecules (or DNA and RNA)

resulting in a helical structure (Watson-Crick base pairing). The hybridization process is temperature and salt dependent.

- ▶ DNA Chips
- ▶ Multifactorial or Common Diseases
- ▶ PNA Chips
- ▶ Rheumatism Related Genes, Identification
- ▶ Thermodynamic Properties of DNA

Hydrogel

Definition

A hydrogel is a cross-linked polymer network that absorbs water. These materials are being explored as scaffolds in tissue engineering and as carriers for drug delivery.

- ▶ Proteomics in Microfluidic Systems

Hydrogen Bond

Definition

An electrostatic interaction between an electronegative atom (e.g. O, S, N, P) and a hydrogen covalently linked to a second electronegative atom. For instance, the amide nitrogen of the peptide bond is the hydrogen-donor, whereas the oxygen of the carbonyl group is the hydrogen-acceptor. The hydrogen-bond energies are in the range of 3-7 kcal/mol.

- ▶ Amino Acids: Physicochemical Properties
- ▶ Protein/DNA Interaction

Hydrolysis

Definition

Hydrolysis refers to a chemical reaction in which a compound reacts with water causing decomposition, and the production of two or more compounds.

- ▶ Proteases and Inhibitors

Hydropathy

Definition

Hydropathy is a measure of the hydrophobic character of an amino acid. It represents a continuous spectrum, at one end of which is hydrophobicity of a molecule or a chemical group, and at the other end, hydrophilicity. Hydrophobic chemical groups tend to be dissolved in non-polar organic solvents, whereas hydrophilic groups tend to be dissolved more readily in water.

- ▶ Amino Acids: Physicochemical Properties

Hydrophilic-/Hydrophobic Cross-Linkers

Definition

Hydrophilic-/Hydrophobic cross-linkers are cross-linking reagents that are water soluble or water insoluble. They covalently link two different or equal molecules.

- ▶ Protein Interaction Analysis: Chemical Cross-Linking

Hydrophobic

Definition

Hydrophobic defines the lacking affinity of a substance for water; i.e. the tendency not to dissolve in or mix with water.

- ▶ Nuclear Hormone Receptors: Regulators of Gene Transcription and Cellular Signaling

Hydrophobic Interaction Chromatography

Definition

Hydrophobic interaction chromatography refers to a form of chromatography that exploits the hydrophobic properties of molecules as a basis of separation on a hydrophobic matrix.

- ▶ Protein Interaction Analysis: Chemical Cross-Linking

Hydrophobic Interactions

Definition

Hydrophobic interactions are molecular interactions that are driven by the exclusion of water from non-polar groups of atoms that are not able to participate in hydrogen bonding.

► [Protein/DNA Interaction](#)

Hydrophobicity

Definition

Hydrophobicity characterizes a thermodynamic scale that indicates whether a considered molecule has affinity for water or not. Due to their atomic composition, hydrophobic molecules are incapable of forming substantial hydrogen bonds with water. In such polar solvents, these molecules tend to bind together, because this phenomenon is accompanied by a decrease of the Gibbs free-energy.

► [Genetic Code](#)
 ► [Protein-Protein Interaction](#)
 ► [QSAR](#)
 ► [Two Hybrid System](#)

Hydroxyl

Definition

Hydroxyl is a negative ion formed by the attachment of an oxygen atom with a hydrogen atom.

► [Proteases and Inhibitors](#)

Hydroxylase

Definition

Hydroxylase is a specific example of an oxygenase, in which the formation of a hydroxyl group is catalysed by reductive splitting of molecular oxygen.

► [Enzyme Catalyzed Post-translational Hydroxylation of Proteins](#)

Hydroxyproline-Rich Glycoprotein

Definition

Hydroxyproline-rich glycoprotein is typically a plant glycoprotein that contains abundant amounts of the modified amino acid hydroxyproline; often glycans are found attached to the hydroxyproline residues.

► [Glycosylation of Proteins](#)

Hyper- and Hypoparathyroidism

GEOFFREY N. HENDY

Calcium Research Laboratory, Royal Victoria Hospital, and Departments of Medicine, Physiology and Human Genetics, McGill University, Montreal, Canada
geoffrey.hendy@mcgill.ca

Definition

Hyperparathyroidism

► [Primary hyperparathyroidism](#) – the increased secretion of parathyroid hormone (PTH) – is due to neoplasms or idiopathic hyperplasia of the parathyroid glands. ► [Secondary hyperparathyroidism](#) associated with hyperplasia of the parathyroids arises as a result of disordered metabolism producing hypocalcemia, as in chronic uremia due to renal disease, malabsorption, ► [rickets](#) or ► [osteomalacia](#). The focus of this review is on primary hyperparathyroidism.

Hypoparathyroidism

Hypoparathyroidism is due to diminution or absence of the secretion of PTH. Pseudohypoparathyroidism is a disorder resembling hypoparathyroidism, with high serum phosphate and low calcium levels, but normal or elevated serum PTH levels due to lack of end-organ responsiveness to PTH. The focus of this review is on hypoparathyroidism.

Characteristics

Hyperparathyroidism

The important biochemical features of primary hyperparathyroidism are elevated serum calcium, decreased fasting serum phosphate, increased serum ► [parathyroid hormone](#) (PTH) and hypercalciuria. Eighty-five percent of cases involve a single benign parathyroid adenoma, while hyperplasia (multiple hypercellular glands) is present in 15%. Parathyroid carcinoma is seen in less than 1% of cases and ectopic secretion of PTH by nonparathyroid tumors is extremely rare.

Although primary hyperparathyroidism is most often detected by its biochemical features, ►nephrolithiasis due to hypercalciuria is a frequent clinical manifestation. Nephrolithiasis with kidney stones occurs in 20% of patients. Other organ systems can be involved, including skeleton, gastro-intestinal tract and central nervous system. Formerly a classic feature, overt bone disease (►osteitis fibrosa cystica, ►brown tumors or pathological fractures) is now rare. Gastrointestinal manifestations (nausea, vomiting or constipation) are common but nonspecific, while ►pancreatitis and peptic ulcer disease are now rare. Hypercalcemia and hypercalciuria lead to increased urinary frequency and progressive nocturia is common. Even mild elevations of calcium can be associated with weakness and a feeling of lassitude. Major mood disturbances and psychotic behavior can be seen with more severe hypercalcemia.

Primary hyperparathyroidism is one of the common causes of hypercalcemia in adults. With the advent of multichannel autoanalyzer screening, the population incidence was shown to be 1 in 1,000. Although the disorder can occur at any age, it is most frequent in the sixth decade. More women than men are affected, the ratio being about 3:1. When found in children, primary hyperparathyroidism is likely to be a component of a ►familial endocrinopathy and a sex bias is not observed.

Hypoparathyroidism

In hypoparathyroidism, serum calcium concentrations are decreased, serum phosphate levels are increased and serum PTH concentrations are low or undetectable. The signs and symptoms of hypoparathyroidism include evidence of latent or overt neuromuscular hyperexcitability. There is wide variation in the severity of the symptoms, and patients may have ►circumoral numbness, ►paresthesias of the distal extremities or muscle cramping, which can progress to ►carpopedal spasm or ►tetany. ►Laryngospasm or ►bronchospasm and seizures may occur. Other less specific manifestations include fatigue, irritability and personal disturbance. Severe hypocalcemia may be associated with a prolonged ►Q-Tc interval on ►electrocardiography, which reverses with treatment. Patients with chronic hypocalcemia may have calcification of the ►basal ganglia or more widespread intracranial calcification. Also seen are extrapyramidal neurological symptoms, subcapsular cataracts, ►band keratopathy and abnormal dentition.

Cellular and Molecular Regulation

Hyperparathyroidism

Molecular analyses have determined that most, if not all, parathyroid ►adenomas and ►carcinomas are ►monoclonal cell expansions. Alterations in two

genes, ►cyclin D1 and the ►multiple endocrine neoplasia type 1 (MEN1) gene, have been implicated in the development of some sporadic parathyroid adenomas. Cytogenetic loss of 13q containing the ►retinoblastoma (RB) and other tumor suppressor genes has also been implicated in parathyroid carcinoma. ►Karyotypic abnormalities involving several other chromosomal regions have been observed in parathyroid adenomas, but the genes remain to be identified.

One in ten cases of primary hyperparathyroidism is hereditary, occurring as an isolated form or associated with other abnormalities. Genetic linkage studies have confirmed the hereditary nature of these syndromes and in some cases identified specific mutations responsible for parathyroid hyperfunction. These entities include familial hypocalciuric hypercalcemia (FHH)/neonatal severe hyperparathyroidism (NSHPT), MEN1 and multiple endocrine neoplasia type 2 (MEN2), the hyperparathyroidism-jaw tumor syndrome (HPT-JT) and familial isolated hyperparathyroidism (FIHP).

Hypercalcemia and elevated serum PTH levels are the biochemical hallmarks of primary hyperparathyroidism. It is recommended that more than one measurement of these parameters be made, as values may be within the normal range early in the disease. Changes in serum proteins affect the ionized (biologically active) calcium fraction and total calcium values should be corrected for albumin concentration or direct measurement of the ionized calcium made. Serum PTH should be assayed by a ►two-site immunoradiometric assay that detects the intact molecule. Elevated PTH levels occur in 90% of primary hyperparathyroid patients, but a high normal PTH level in the face of hypercalcemia indicates lack of suppression and is consistent with hyperparathyroidism. Serum phosphate is usually at the lower end of the normal range because of the phosphaturic action of PTH. Hypophosphatemia occurs in about 25% of patients. To be diagnostically useful, serum phosphate measurements should be made in the fasting state to avoid postprandial fluctuations. Serum bone-specific alkaline phosphatase is a practical measure of skeletal involvement. Mild hyperchloremic acidosis is common because of the effects of PTH on renal chloride and bicarbonate handling. Urinary calcium is frankly elevated in one-third of patients. The effect of PTH on the renal 25-hydroxyvitamin D-1 α -hydroxylase enzyme is reflected in serum 1,25-dihydroxyvitamin D levels at the upper end of the normal range or frank elevations in a third of cases.

Hypoparathyroidism

Primary hypoparathyroidism most commonly results from surgical excision of, or damage to, the parathyroid glands. However, genetic forms of hypoparathyroidism are known. Familial isolated hypoparathyroidism may show autosomal dominant, autosomal recessive or

X-linked inheritance. In a few instances of either autosomal dominant or recessive disease, mutations in the PTH gene have been found. Gain-of-function mutations in the CASR gene have been identified in several families with autosomal dominant hypocalcemia or hypoparathyroidism. In the parathyroid gland, the activated CASR suppresses PTH secretion and in the kidney, it induces hypercalciuria, which may contribute to the hypocalcemia. A patient with neonatal hypoparathyroidism was homozygous for a partial deletion of the glial cell missing-2 gene (GCMB). This gene, which is expressed in the PTH-secreting cells of the developing parathyroid glands, encodes a transcription factor that is critical for their development. Hypoparathyroidism with multiple malformations can occur, for example, as part of a 22q11 microdeletion. This is a common cause of the DiGeorge syndrome as well as isolated congenital heart disease and velocardiofacial (VCF) syndrome. With DiGeorge syndrome, patients may present with neonatal hypocalcemic seizures due to hypoparathyroidism, severe infections due to thymic hypoplasia and conotruncal heart defects. However, it is not unusual for the hypoparathyroidism to remain asymptomatic until adolescence or require provocative testing to confirm the decreased parathyroid reserve in normocalcemic adults. Craniofacial abnormalities include cleft palate, pharyngeal insufficiency and mildly dysmorphic facies. The syndrome complex arises from a failure to develop the derivatives of the third and fourth pharyngeal pouches leading to agenesis or hypoplasia of the parathyroid glands and thymus. The identification of novel developmental genes in the 22q11 region is being keenly pursued. The Tbx1 transcription factor has been implicated although the full expression of DiGeorge syndrome may involve loss of other contiguous genes.

Hypoparathyroidism is part of the Barakat or HDR (hypothyroidism, nerve deafness and renal dysplasia) syndrome. Deletions of two non-overlapping regions of chromosome 10p contribute to a DiGeorge-like phenotype—the DiGeorge critical region II on 10p13-14 and the HDR syndrome (10p14-10pter). Haploinsufficiency of the transcription factor GATA3 has been implicated by the finding of heterozygous loss-of-function mutations in HDR patients. Thus, GATA3 appears essential for normal embryonic development of the parathyroids, auditory system and kidney.

Hypoparathyroidism is found variably associated with the typical picture of Kenny-Caffey syndrome of growth retardation, ▶osteosclerosis, cortical thickening of the long bones and delayed closure of the anterior fontanel. Both dominant and recessive modes of inheritance have been observed but it is likely that some of the latter cases are examples of Sanjad-Sakati syndrome. This is a recessive condition characterized by congenital hypoparathyroidism, seizures, growth

and developmental retardation and characteristic dysmorphic features. The defect had been localized to 1q43-44 and recently mutations in the ▶TBCE gene have been identified in affected individuals. TBCE encodes a chaperone protein important for the proper functioning of the tubulin assembly pathway important for development of the parathyroid.

Hypoparathyroidism can be due to metabolic disease. It is, for example, a variable component of the neuromyopathies caused by mitochondrial gene defects. Among these are the Keams-Sayre and Pearson marrow pancreas syndromes and mitochondrial encephalomyopathy. Long-chain hydroxyacyl-CoA dehydrogenase deficiency is a metabolic disorder of fatty acid oxidation that may be accompanied by hypoparathyroidism. Parathyroid insufficiency and symptoms of hypocalcemia are occasionally seen in inherited metabolic disorders leading to excess storage of iron (thalassemia, Diamond-Blackfan anemia, hemochromatosis) or copper (Wilson disease).

▶Autoimmune parathyroid gland ablation or destruction can occur. Antibodies directed against parathyroid tissue have been detected in over one-third of patients with either isolated hypoparathyroid disease or combined with other endocrine deficiencies. Antibodies that bind to and activate the parathyroid CASR have been found in patients with either type 1 autoimmune polyglandular syndrome (APS-1) or acquired hypoparathyroidism associated with autoimmune hypothyroidism. In APS-1, the most common associated manifestations are hypoparathyroidism with ▶mucocutaneous candidiasis and ▶Addison's disease. The phenotype is highly variable and patients may not express all elements of the basic triad. The disease, also known as the autoimmune polyendocrinopathy-candidiasis-ectodermal dystrophy syndrome (APECED), presents in infancy or early childhood. It may show sporadic or autosomal-recessive inheritance. The mutated gene, called the autoimmune regulator (AIRE), maps to chromosome 21q22 and encodes a putative transcriptional regulator.

Clinical Relevance

Hyperparathyroidism

With the frequent detection of hypercalcemia in biochemical screening programs, the differentiation of primary hyperparathyroidism from other conditions causing hypercalcemia has become increasingly important. Causes of non-parathyroid hypercalcemia include humoral ▶hypercalcemia of malignancy, vitamin D or A intoxication, ▶milk-alkali syndrome, granulomatous disorders (especially sarcoidosis), immobilization of patients with a pre-existing high bone turnover state such as adolescence, thyrotoxicosis, ▶Paget's disease and treatment with thiazide diuretics or lithium. ▶Parathyroid hormone-related protein

(PTHrP) is the major causative agent in the humoral hypercalcemia of malignancy, although many other circulating factors may contribute. New immunoassays for PTH and PTHrP have greatly facilitated the differential diagnosis of primary hyperparathyroidism from malignancy-associated hypercalcemia.

Study of the relatives of patients with hypercalcemia can contribute to establishing the diagnosis in the 10% of all cases of primary hyperparathyroidism that prove to be hereditary. The finding of another relative with hypercalcemia furnishes evidence of primary hyperparathyroidism if FHH is not suggested by a relatively low urinary calcium-to-creatinine clearance ratio or definitively diagnosed by the identification of a mutation of the calcium-sensing receptor (CASR) gene. The finding of a hypercalcemic relative also requires investigation of the patient for manifestations of the MEN or HPT-JT syndromes.

Criteria for surgery in hyperparathyroidism have been established by a consensus conference and a follow-up workshop of the National Institutes of Health. Candidates for surgery are those having one or more of the following: hypercalcemia >11.6 mg/dL; hypercalciuria >400 mg/day; kidney stones; reduced bone density or age >50 years). Asymptomatic patients who are managed conservatively with twice yearly serum calcium and urinary calcium excretion determinations and annual bone densitometry generally do well since the progression of the disease is usually quite slow.

Hypoparathyroidism

Acute hypocalcemia can be life threatening and present with seizures, tetany or cardiac arrhythmias. Intravenous calcium can alleviate these symptoms rapidly although this should be done cautiously to minimize risks associated with this route of administration. With chronic hypocalcemia, oral supplementation with calcium and vitamin D analogues is usually employed. Careful monitoring of the patient's calcium homeostatic status is necessary as, with under replacement, cataracts and symptoms of numbness and tingling can occur. With over replacement, there is the risk of ►nephrocalcinosis and nephrolithiasis.

References

1. Hendy, GN, Arnold, A (2001) Molecular basis of PTH overexpression. In: Bilezikian JP, Raisz LG, Rodan GA (eds) *Principles of Bone Biology*, 2nd edn. Academic Press, San Diego, pp 135–148
2. Bilezikian, JP, Potts, JT Jr, Fuleihan, Gel-H et al (2002) Summary statement from a workshop on asymptomatic primary hyperparathyroidism: a perspective for the 21st century. *J Bone Miner Res* 17(S2):N2–N11
3. Thakker, RV (2001) Molecular basis of PTH underexpression. In: Bilezikian JP, Raisz LG, Rodan GA (eds) *Principles of Bone Biology*, 2nd edn. Academic Press, San Diego, pp 1105–1106

Hypercalcemia of Malignancy

Definition

Hypercalcemia of malignancy is the most common metabolic consequence of skeletal metastases and/or tumor. HCM is based on humoral factors released by the tumors that mediate accelerated bone resorption.

►Hyper-and Hypoparathyroidism

Hyperferritinemia Cataract Syndrome

Definition

Hyperferritinemia cataract syndrome refers to a disease that is characterized by increased serum ferritin levels and the early formation of cataracts. The molecular defect is explained by mutations in the iron responsive element (IRE), which blocks iron regulatory protein (IRP) binding and results in aberrant translation control.

►Translational Control in Eukaryotes

Hyperglycemia

Definition

Hyperglycemia is an increase in plasma glucose ("blood sugar") above normal. It can turn into a complex medical condition, for example, diabetic ketoacidosis and coma. Hyperglycemia is usually the first sign of diabetes mellitus.

►Diabetes Mellitus, Genetics

Hyperkeratosis

Definition

Hyperkeratosis is a term which denotes abnormal cornification and thickening of a stratified squamous epithelium, usually the epidermis. An increase in cornification usually arises as a defence against chronic abrasion and hyperkeratosis is often a reflection of cell fragility, possibly resulting from a chronic wound response. It is unrelated to the collagen accumulation of scarring, which is a dermal response.

►Desmosomes

►Heritable Skin Disorders

►Intermediate Filaments

Hypermethylation

Definition

Hypermethylation denotes an overall increase in DNA methylation of a certain CpG island.

►CpG Islands

Hyperparathyroidism

►Hyper- and Hypoparathyroidism

Hyperphagia

Definition

Hyperphagia designates increased food intake.

►Prader Willi and Angelman Syndromes

Hyperphosphorylation

Definition

Hyperphosphorylation refers to phosphorylation of a given protein at multiple phosphorylation sites leading to altered function. For example, phosphorylation of serines 63/73 and threonines 69/71 of c-Jun and ATF-2, respectively, results in enhanced transactivation function in response to extracellular signals. Hyperphosphorylation may be involved in the pathogenesis of neurodegenerative disorders. For example, several proteins such as tau, neurofilaments, β -tubulins or β -catenins are hyperphosphorylated in Alzheimer's disease brains.

►Jun/Fos

Hyperpolarization

Definition

Hyperpolarization denotes the overshoot response of the repolarization phase of the action potential in a neuron.

►Ion Channels/Excitable Membranes
►Neurons

Hypersensitivity Drug Reactions

►Idiosyncratic Drug Reactions

Hypertension

Definition

Hypertension is a chronic elevation of blood pressure, which is a major risk factor for cardiovascular and renal disease. Hypertension is strongly associated with stroke, myocardial infarction, aneurysm, progression of renal disease and other vascular disturbances. Current opinion is that blood pressures $>140/90$ mm Hg are labeled "hypertensive", while blood pressures $>135/85$ but $<140/90$ mm Hg are termed "prehypertensive".

Cases of hypertension can be classified as primary (►essential hypertension) when no secondary causes such as renovascular diseases, renal failure, hyperaldosteronism or pheochromocytoma are present. Those cases of hypertension account for 95% of all cases of hypertension. This condition is a heterogeneous disorder with different patients having different causal factors leading to high blood pressure, and usually requires pharmacological treatment. Essential primary or idiopathic hypertension is defined as high blood pressure in which causes of ►secondary hypertension or monogenic (mendelian) forms are not present.

►Mendelian Forms of Human Hypertension and Mechanisms of Disease

Hyperthyroidism

Definition

Hyperthyroidism is caused by excessive production of thyroid hormones. Symptoms of hyperthyroidism include: nervousness, irritability, sweating and too warm skin, fatigue, rapid heartbeat, and hand tremors.

►Cardiac Signaling: Cellular, Molecular and Clinical Aspects
►Thyroid Disorders, Genetic Basis

Hyperthyroxinemia

► Thyroid Disorders, Genetic Basis

Hypertriglyceridemia

Definition

Hypertriglyceridemia describes a state of excess triglyceride level in plasma. Plasma lipoproteins of chylomicron VLDL (very low density lipoprotein; 0.950 – 1.006 g/ml) and IDL (intermediate density I; 1.006 – 1.019 g/ml) contain a large amount of TG (triglyceride). Overproduction of VLDL (Type IV hyperlipoproteinemia in WHO classification), decreased lipolysis of chylomicron or VLDL (Type I and type V hyperlipoproteinemia), or decreased receptor-mediated uptake of IDL (Type III hyperlipoproteinemia), result in hyper-triglyceridemia.

► High-HDL Syndrome

Hypertrophic Chondrocyte

Definition

Hypertrophic chondrocyte is a late stage differentiation form of the chondrocyte, the cartilage-forming cell.

► Bone Disease and Skeletal Disorders, Genetics

Hypertrophy

Definition

Hypertrophy designates an enlargement in an organ, or part of it, due to an increase in cell size.

► Bone and Cartilage

► Cardiac Signaling: Cellular, Molecular and Clinical Aspects

Hypnagogic Hallucinations

Definition

Hypnagogic hallucinations are dreamlike, but realistic, perceptions occurring around sleep onset. They typically have a frightening character and represent one feature of narcolepsy.

► Narcolepsy

Hypocholesterolemia

Definition

Hypocholesterolemia refers to a lower than normal plasma cholesterol levels.

► Dyslipidemia

► Tangier Disease

Hypochondroplasia

Definition

Hypochondroplasia is a skeletal dysplasia characterized by short stature with disproportionately short arms and legs. The phenotype is like achondroplasia but milder. Hypochondroplasia is inherited in an autosomal dominant pattern. About 70 percent of hypochondroplasia cases are caused by mutations in the FGFR 3 (fibroblast growth factor receptor 3) gene.

► Bone Disease and Skeletal Disorders, Genetics

Hypocretins

► Orexins

Hypogammaglobulinemia

Definition

Hypogammaglobulinemia describes an inherited or acquired disorder that is caused by low levels of immunoglobulins (antibodies) in the blood.

► Hypothalamic and Pituitary Diseases Genetics

Hypoglossal Cord

Definition

Hypoglossal cord refers to a transient structure in the embryo harboring cells that migrate as a cohesive mass to form intrinsic tongue and pharyngeal muscles.

► [Muscle Development](#)

Hypogonadism

Definition

Hypogonadism (also known as gonadal deficiency) is a condition in which the sex glands (ovaries, testes) do not function properly, and secretion of hormones is reduced. As a result, normal sexual development is disturbed.

► [Prader Willi and Angelman Syndrome](#)

Hypoinsulinemic

Definition

Hypoinsulinemic describes a condition of having lower than normal levels of insulin.

► [Diabetes Mellitus, Genetics](#)

Hypokalemia

Definition

Hypokalemia is defined as a reduction in the plasma potassium (K⁺) concentration below 3.5 mM. Hypokalemia can result due to a reduction in intake of dietary K⁺, or from a shift of K⁺ into the intracellular space. The most common cause of hypokalemia, however, is renal loss of K⁺.

► [Mendelian Forms of Human Hypertension and Mechanisms of Disease](#)

Hypomethylation

Definition

Hypomethylation denotes an overall decrease in DNA methylation in normally methylated CG dinucleotides. Hypomethylation can occur within a CpG island or in the bulk genome.

► [CpG Islands](#)

Hypomorph/Hypomorphic Mutation

Definition

Hypomorphic mutation describes any mutation of a gene/allele that leads to the reduced activity of a protein, rather than no activity (amorphic mutation or null mutation). The term was originally introduced by Muller in 1932 working with *Drosophila* mutants.

In another context, the term hypomorph is also applied to describe a person whose standing height is short in proportion to their sitting height, owing to shortness of limbs.

► [Chromosomal Instability Syndromes](#)

► [Double-Strand Break Repair](#)

► [Jun/Fos](#)

► [Large-Scale Homologous Recombination Approaches in Mice](#)

► [Mouse Genomics](#)

► [Mutagenesis Approaches in Yeast](#)

Hypopituitarism

► [Hypothalamic and Pituitary Diseases, Genetics](#)

Hypospadias

Definition

Hypospadias is a relatively common birth defect, which is characterized by an abnormal positioning of the urethra opening on the underside of the phallus.

► [SRY – Sex Reversal](#)

Hypothalamic and Pituitary Diseases, Genetics

SALLY RADOVICK, ELIZABETH A. ROCHOWICZ
Section of Pediatric Endocrinology, The University of
Chicago, Chicago, IL, USA
sradovici@peds.bsd.uchicago.edu

Synonyms

Hypopituitarism; Panhypopituitarism; Combined Pituitary Hormone Deficiencies

Definition

Genetic diseases of the hypothalamus and pituitary gland, or hypopituitarism, are characterized by one or more pituitary ►**hormone** deficiencies. The hormone deficiencies result in target organ dysfunction and disease. The pituitary gland is a tiny gland (0.5 g and 10–15 mm in adult humans) that has an anterior and a posterior lobe called the adenohypophysis and the neurohypophysis, respectively. The pituitary gland is located in the sella turcica (Turkish saddle) of the sphenoid bone of the skull. The hypothalamus is a small wedged-shaped gland located superiorly to the pituitary gland and has multiple complex functions in addition to regulation of the pituitary gland.

The pituitary hormones that can be deficient in hypopituitarism are synthesized in and released from five cell types in the anterior pituitary gland. These hormones include somatotropin (growth hormone, GH), thyrotropin [►**thyroid stimulating hormone** (TSH)], ►**prolactin** (PRL), ►**adrenocorticotropin** (ACTH), ►**follicle stimulating hormone** (FSH), and ►**luteinizing hormone** (LH). These anterior pituitary hormones are regulated by stimulatory/inhibitory hormones and neurotransmitters produced in the neurons

of the hypothalamus. These hormones are transported from the hypothalamus *via* a capillary plexus in the pituitary stalk to the anterior pituitary gland (see Table 1).

The posterior pituitary gland stores hormones, ►**oxytocin** and ►**vasopressin** (ADH), originally produced in the supraoptic and paraventricular nuclei in the hypothalamus. Then, the hormones are transported to the posterior pituitary gland *via* neural tracts to be released into the vasculature.

The etiology of the congenital pituitary hormone deficiencies is usually idiopathic but the molecular and/or genetic basis of some of them has been clarified. The genetic and/or molecular defects affect the pituitary gland during embryogenesis and can result in pituitary aplasia or hypoplasia. The anterior and posterior lobes of the pituitary gland are derived embryologically from different sources, and are usually affected separately by different molecular defects. The molecular abnormalities that have been discovered to cause hypopituitarism include pituitary transcription factor mutations, GHRH receptor mutations, GH-1 gene mutations, vasopressin gene mutations (neurophysin II component) and Wolfram syndrome gene mutations (WFS1 gene). The mutations cause deficient/decreased function of the target organ (depending on which hormones are deficient): ►**hypothyroidism** (TSH deficiency), ►**growth hormone deficiency**/short stature (GH deficiency), ►**gonadotropin deficiency**/delayed puberty/infertility (LH and/or FSH deficiency), ►**adrenal insufficiency** (ACTH deficiency), and ►**diabetes insipidus** (ADH deficiency).

Characteristics

Pituitary Transcription Factor Mutations

There are multiple transcription factors (Pit-1, Hesx-1, Ptx-2, Lhx-3, Lhx-4 and Prop-1) known to be involved in the determination of the five anterior pituitary hormone producing cells: thyrotrophs (TSH), somatotrophs (GH), corticotrophs (ACTH), gonadotrophs

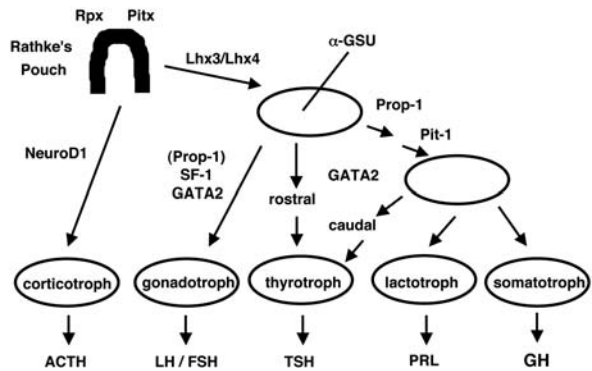
Hypothalamic and Pituitary Diseases, Genetics. Table 1 Hypothalamic Regulation of Anterior Pituitary Hormones

Hypothalamic Regulation of Anterior Pituitary Hormones	Action – Anterior Pituitary Gland
Corticotropin-releasing hormone (CRH)	Stimulates ACTH synthesis and release
Thyrotropin-releasing hormone (TRH)	Stimulates TSH & PRL synthesis and release
Gonadotropin-releasing hormone (GnRH)	Stimulates LH & FSH synthesis and release
Growth hormone-releasing hormone (GHRH)	Stimulates GH synthesis and release
Growth hormone-release-inhibiting hormone (Somatostatin)	Inhibits GH & TSH synthesis and release
Prolactin-releasing factor (PRF) (possibly VIP or TRH)	Stimulates PRL release
Prolactin-inhibiting hormone (PIH) (actually Dopamine)	Inhibits PRL synthesis and release

(LH/FSH), and lactotrophs (PRL). These factors initiate a cascade of developmental events in a temporal and spatial pattern that results in mature anterior pituitary cell types (see Fig. 1). A mutation in one of the genes encoding these transcription factors results in anterior pituitary hormone deficiencies (see Table 2). Pit-1 (also called GHF-1 and now officially called POUF1) is a **POU transcription factors** that has a POU-specific (POU-S) and a POU-homeo (POU-H) protein domain. Pit-1 has been shown to be essential for somatotroph, lactotroph and thyrotroph development, cell specific gene expression and regulation. Thus, Pit-1 mutations result in GH deficiency, PRL deficiency and TSH deficiency. There have been at least 17 different point mutations found in the Pit-1 gene in humans resulting in GH, PRL and TSH deficiency (see Fig. 2) with the majority the result of an amino acid substitution (missense mutation), but deficiencies have also been found to be due to a nonsense mutation or a base pair deletion. The arginine to tryptophan mutation at codon 271 (R271W) is the most common Pit-1 mutation. Most of these mutations have been found to be sporadic, although some are inherited in an autosomal recessive manner.

Hesx-1 (also called Rpx-Rathke's pouch homeobox) is one of the paired-like homeobox genes and is an acronym for homeobox gene expression in embryonic stem cells. Since it is the earliest known marker for the pituitary primordium, it may have a role in the differentiation of the pituitary gland. Hesx-1 transcripts also disappear as the pituitary cell continues to differentiate. At least four different Hesx-1 mutations discovered in humans are known to cause hypopituitarism in the form of isolated GH deficiency combined pituitary hormone deficiency, panhypopituitarism or septo-optic dysplasia (S.O.D.). S.O.D. is a heterogeneous condition with any combination of optic nerve hypoplasia, pituitary gland hypoplasia (resulting in pituitary hormone deficiency), and midline brain abnormalities (such as absence of the **corpus callosum** or **septum pellicidum**). Hesx-1 mutations have been found to be heterozygous and homozygous missense mutations secondary to amino acid substitutions (arginine to cysteine (R53C and R160C), serine to leucine (S170L), and asparagine to serine (N125S)).

Ptx-2 (also called Pitx2, P-otx2, or RIEG) is a paired-like **homeodomain** transcription factor that is present in the fetal and adult anterior pituitary gland. It is expressed in thyrotrophs, gonadotrophs, somatotrophs and lactotrophs, but not in corticotrophs. It is also expressed in other adult organs including kidney, lung, testis, and tongue. There have been at least 18 mutations found in the Ptx-2 gene (RIEG gene) inherited in an autosomal dominant pattern. These mutations have resulted in variable manifestations of hypopituitarism, anomalies of the anterior chamber of



Hypothalamic and Pituitary Diseases, Genetics.

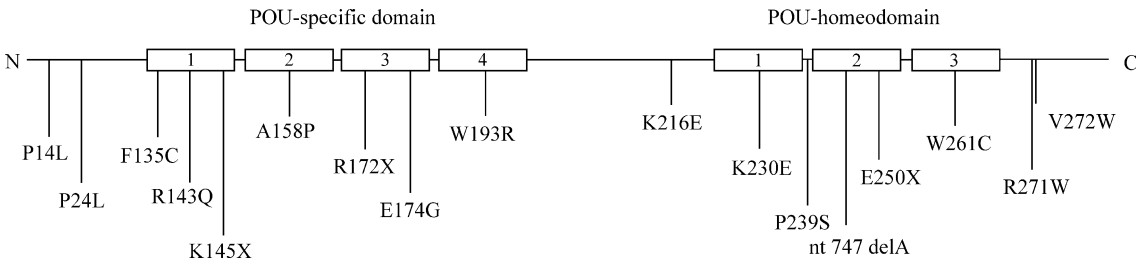
Figure 1 Pituitary-specific transcription factors involved in anterior pituitary development. The homeodomain factors play a role in the determination of specific pituitary cell lineages. Rpx is the earliest known specific marker for the pituitary primordium. Pitx is present in the fetal pituitary and in most cells of the adult pituitary, albeit at different levels. Lhx3 and Lhx4 are expressed in gonadotrophs, thyrotrophs, somatotrophs, and lactotrophs. Prop-1 determines Pit-1 cell lineages and may play a role in directing some of the precursors of the Pit-1 cell lineage into the gonadotroph lineage, prior to terminal differentiation events. Pit-1 has been shown to be essential for the development of somatotrophs, lactotrophs, and thyrotrophs. There is a Pit-1-independent population of thyrotrophs in the rostral tip of the developing anterior pituitary gland that disappears by the day of birth. There are other specific transcription factors involved in the determination of the anterior pituitary cell lineages abbreviated in the figure that have no known mutations resulting in GH deficiency or CPHD. These transcription factors are beyond the scope of this review (including SF-1 (steroidogenic factor-1), α -GSU (alpha subunit of common G-protein), NeuroD1, and GATA-2 (a member of the GATA family of transcription factors with a common core nucleotide sequence - first found to be W-GATA-R with W = A or T and R = A or G)).

the eye, dental hypoplasia, protuberant umbilicus and mental retardation collectively called Rieger's syndrome. Some of the Ptx-2 mutations were missense with amino acid substitution in the homeodomain, some were splicing mutations in the intron dividing the homeobox sequence and one was an in-frame duplication of 21 base pairs (causing a seven amino acid duplication (7aakup of residues 6–12) threonine to lysine in the homeodomain).

Lhx-3 (also called **LIM-3** and P-Lim) is a LIM-type homeodomain protein. The LIM proteins contain two repeat cysteine/histidine LIM domains and a homeodomain. The LIM domains are thought to be involved in regulation of transcription but do not bind DNA. Lhx-3 is expressed in fetal and adult anterior pituitary cells (gonadotrophs, thyrotrophs, somatotrophs and

Hypothalamic and Pituitary Diseases, Genetics. Table 2 Hormone Deficiencies associated with the respective Transcription Factors

Deficient Factor	ACTH Deficiency	GH Deficiency	LH/FSH Deficiency	Prl Deficiency	TSH Deficiency
Rpx	yes	yes	yes	yes	yes
Ptx-2	no	yes	yes	yes	yes
Lhx-3	no	yes	yes	yes	yes
Prop-1	+/-	yes	+/-	yes	yes
Pit-1	no	yes	no	yes	yes



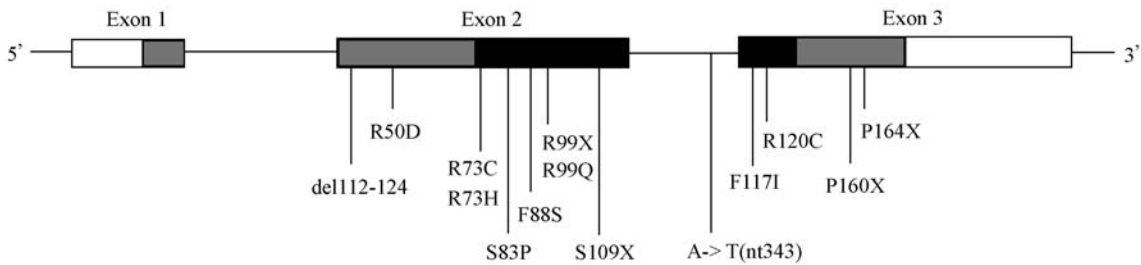
Hypothalamic and Pituitary Diseases, Genetics. Figure 2 Schematic representation of the Pit-1 protein with mutations. The 4 α -helices of the POU-specific domain and the 3 α -helices of the POU-homeodomain are depicted as boxes, and mutations are noted at their locations. The E250X mutation results from both a missense mutation at codon 250 and from a 1 bp deletion at nt 747, which also causes a missense mutation at codon 249.

lactotrophs), as well as in the ventral hindbrain and spinal cord during embryogenesis. Lhx-3 mutations have been found to cause hypoplasia or hyperplasia of the anterior pituitary gland (with secondary LH/FSH deficiency, TSH deficiency, GH deficiency and prolactin deficiency) and a rigid cervical spine (limiting head rotation). Mutations caused by a missense mutation with an amino acid substitution of tyrosine for cysteine at codon 116 (Y116C) result in a hypoplastic anterior pituitary gland, whereas an intragenic deletion of 23 amino acids results in an enlarged pituitary gland of unknown mechanism and a severely truncated protein without the homeodomain. Lhx-4 is another LIM-type homeodomain protein involved in early pituitary development. One Lhx-4 mutation was found in a family cohort (penetrated over 3 generations) that results in a phenotype consisting of CPHD (GH deficiency, TSH deficiency and ACTH deficiency), small sella turcica, hypoplastic anterior pituitary gland, ectopic posterior pituitary gland and herniation of the cerebellar tonsils (called Arnold-Chiari malformation type1). This mutation was inherited in an autosomal dominant pattern. The mutation was a heterozygous point mutation (G to C substitution) in the intronic region before exon 5.

Prop-1 is a paired-like homeodomain transcription factor expressed in the anterior pituitary only during embryogenesis and thus, is named as the prophet of pit-1. Several prop-1 mutations (at least 14) have been discovered that result in hypopituitarism with GH deficiency, PRL deficiency, TSH deficiency and LH/FSH deficiency (see Fig. 3). Most patients have a hypoplastic pituitary gland. The gonadotropin deficiency (LH/FSH deficiency) in some humans with prop-1 mutations prevents the start of puberty, however, in others, puberty is delayed with diminished LH/FSH secretion with increasing age. The mutations found include intronic point mutations, one base pair deletions, two base pair deletions and 12 base pair deletions that result in amino acid substitutions and stop codons.

GHRH Receptor Mutations

GHRH receptor mutations have been found in at least two separate groups. A consanguineous Indian Moslem family demonstrated a nonsense mutation with a substitution of glutamate at codon 72 for a stop codon. The group described as “Dwarfism of Sindh” showed a nonsense mutation with a substitution of glutamate at codon 50 for a stop codon.



Hypothalamic and Pituitary Diseases, Genetics. Figure 3 Schematic representation of the *Prop-1* gene. Exons are depicted as shaded areas, and the homeodomain is noted in black. Mutations are noted at their locations. The S109x mutation results from both a 2 bp deletion between nt 296-301, and a 2 bp deletion at codon 50 (nt 149-150) that causes divergence of amino acid sequence from codon 52. The P164X mutation results from a 1 bp deletion at codon 50, which causes divergence of amino acid sequence from codon 53. The P160X mutations results from a 13 bp deletion between nt 112-124, which causes divergence of amino acid sequence from codon 38.

GH-1 Gene Mutations

GH-1 gene mutations result in isolated GH deficiency and are classified as Type IA, Type IB, Type II and Type III.

Type IA GH deficiency is caused by a complete GH-1 gene deletion that is the result of autosomal recessive inheritance. Thus, there is complete absence of GH with growth retardation present in infancy.

Type IB GH deficiency is caused by point mutations of the GH-1 gene that are also inherited in an autosomal recessive manner. Type IB GH deficiency mutations include G to C or G to T transversion in intron 4 (causing splice deletion of one-half of exon 4 and frameshift within exon 5) and other nonsense, splicing and frameshift mutations. These individuals demonstrate a partial GH deficiency.

Type II GH deficiency is the result of intronic transitions (T to C) in intron 3 of the GH-1 gene that inactivate the donor splice site and delete exon 3. Type II GH deficiency has an autosomal dominant inheritance and results in partial GH deficiency.

Type III GH deficiency is caused by mutations of the X chromosome, and thus, is inherited in an X-linked manner. These mutations result in partial GH deficiency. Some individuals may also have [hypogamaglobulinemia](#).

Vasopressin Gene Mutations

Vasopressin gene mutations result in central diabetes insipidus (ADH deficiency) and are inherited in an autosomal dominant pattern, although one point mutation (amino acid substitution proline for leucine at codon 7 (P7L)) was found to be the result of autosomal recessive inheritance. Mutations (nonsense, missense, deletion and frameshift) have been found at more than 30 different sites in the precursors of the vasopressin hormone (vasopressin pre-prohormone).

The majority of mutations have been found in the neurophysin portion of the precursor of vasopressin, although some have been found in the signal peptide and vasopressin peptide region of the gene.

WFS-1 (Wolframin) Gene Mutations

WFS-1 gene mutations result in DIDMOAD syndrome (diabetes insipidus (ADH deficiency), [diabetes mellitus](#), optic atrophy and sensorineural deafness) that is also referred to as Wolfram syndrome (hence, the gene name Wolfram Syndrome-1). The WFS-1 gene encodes a transmembrane protein consisting of 890 amino acids with unknown function. Multiple single nucleotide mutations in the WFS-1 gene have been found to cause Wolfram syndrome or DIDMOAD. The mutations included stop, frameshift and amino acid deletions and insertions, suggesting loss-of-function mutations as the cause of Wolfram syndrome.

Cellular and Molecular Regulation

Pituitary Transcription Factor Mutations

The Pit-1 gene mutations vary in the molecular mechanisms that cause hypopituitarism. In the most common Pit-1 mutation, R271W, Pit-1 binds normally to DNA, but the mutant Pit-1 inhibits transcription in a dominant manner and may impair dimerization. In the K216E Pit-1 mutation, Pit-1 binds to DNA, does not inhibit basal activation of the GH and PRL genes, but is unable to produce retinoic acid induction of the Pit-1 gene distal enhancer. The A158P Pit-1 mutation results in a protein with a slightly decreased ability to bind DNA and a lack of (or minimal) transcriptional activation of the GH, PRL, and Pit-1 promoters. The P239S Pit-1 mutation binds to DNA normally but has decreased activation of the GH promoter. The F135C mutation is located in the hydrophobic core of the POU-S domain near the dimer interface of the POU-S

domain. Since the normal F135 residue appears to be the key in assembling and stabilizing the first, third, and fourth α -helices of the POU-S domain according to molecular models, the mutant F135S Pit-1 prevents proper assembly and stabilization of α -helices of the POU-S domain, which prevents normal interaction with other pituitary transcription factors. There are some Pit-1 mutations (E174G, R172X, W193R, nt747delA, E250X) located at amino acids critical for DNA binding, which thus impair transcriptional activation. The other Pit-1 mutations that have been found have yet to be characterized.

Rpx-1 mutations have not been well characterized. The R53C Rpx-1 mutation has been shown to decrease DNA binding. Other Rpx-1 mutations were inherited from an unaffected parent and may represent ►poly-morphisms that do not affect normal Rpx-1 function or low penetrance of the phenotype.

Ptx-2 mutations result in several different molecular causes of hypopituitarism. The L54Q Ptx-2 mutation in helix 1 of the homeodomain results in an unstable protein. The T68P Ptx-2 mutation has decreased affinity for DNA and does not permit enhanced binding to Pit-1. The V45L has a slight decrease in DNA binding but actually increases transcriptional activity by 200-fold. This increased transcriptional activity is postulated to cause hypopituitarism but this has not been proven. The Ptx-2 mutation that is secondary to 7aa duplication of residue 6–12 of the homeodomain has significantly decreased DNA binding (100-fold) and no transactivation of a target promoter construct. The K88E Ptx-2 mutation has reduced DNA binding and transactivational activity which suppressed its synergism with Pit-1. This may represent a dominant-negative effect.

The two Lhx-3 mutations result in reduced gene activation. However, the Y116C Lhx-3 mutation binds DNA, but the intragenic 23 amino acid deletion does not bind DNA.

The Lhx-4 mutation abolishes normal Lhx-4 splicing and activates two exonic cryptic splice sites. The first splice site leads to an in-frame deletion of four highly conserved amino acids in the third helix of the homeodomain (deletion of VWFQ at positions 47–50 in the homeodomain). The second splice site alters the reading frame in the homeodomain at position 47 that results in a premature stop codon in exon 5. Thus, two different proteins with deletions in their homeodomain are produced and have altered function.

Prop-1 mutations characterized reveal various molecular defects. S83P Prop-1 mutation results in an 8-fold decrease in DNA binding at two sites in the Pit-1 early enhancer required for Pit-1 gene activation. This indirectly causes diminished transactivation at these

two DNA binding sites. The 296 delGA Prop-1 mutation changes the reading frame and results in a stop codon (S109X). This truncated product does not have promoter binding or transcriptional activation. There are other Prop-1 mutations that have resulted in a truncated product from a premature stop codon ((P160X, S109X, P164X, R99X). The F117I and R120C Prop-1 mutations cause diminished DNA binding and activation of target gene reporter constructs. The A-T nt343 Prop-1 mutation in the intronic region preceding exon 3 abolishes normal splicing and ultimately causes loss of the first 12 nucleotides of exon 3.

GHRH Receptor Mutations

GHRH-R mutations result in a truncated GHRH receptor without the membrane spanning regions and a G-protein site.

GH-1 Gene Mutations

The GH-1 gene is flanked by long stretches of highly homologous DNA that predispose it to mutations that result in Type IA GH deficiency. Mutations that cause Type IB deficiency have been characterized as affecting the stability and biological activity of the mutant GH protein that may derange targeting of the GH peptide into secretory granules. The mechanism of the autosomal dominant Type II GH deficiency is unknown. Type III GH deficiency is the result of interstitial (q13.3-21.1) deletions or microduplications (certain regions) on the X chromosome.

Vasopressin Gene Mutations

The multitude of vasopressin gene mutations (in the pre-prohormone) cause vasopressin deficiency by an unknown mechanism.

WFS-1 Gene Mutations

Studies suggest the *wolframin* gene encodes a novel transmembrane protein and mutations result in loss of function.

Clinical Relevance

There have been many mutations discovered that cause pituitary and hypothalamic diseases originally regarded as “idiopathic.” Present characterization of these mutations and correlation with the phenotypes have started the journey to unravel the genetic basis for hypopituitarism. Further understanding of the genetic basis of pituitary and hypothalamic diseases along with advances in medicine show promise to eventually lead to earlier diagnosis, treatment or prevention and ultimately decrease the morbidity and mortality associated with these diseases.

References

1. Radovick S, Mac Gillivray M (2003) Pediatric Endocrinology-A Practical Clinical Guide. Humana Press, Totowa, NJ, pp 3–25 and pp 155–172
2. Sperling M (2002) Pediatric Endocrinology, 2nd edn. Saunders, Philadelphia, pp 211–288 and pp 289–322
3. Becker KL (2001) Principles and Practice of Endocrinology and Metabolism, 3rd edn. Lippincott Williams & Wilkins, Philadelphia pp 84–90 and pp 90–97
4. Cohen L, Radovick S (2002) Molecular Basis of Combined Pituitary Hormone Deficiency. *Endocr Rev* 23:431–442

Hypothyroidism

Definition

Hypothyroidism refers to a condition in which the body lacks thyroid hormone (T4 and T3), which is produced by the thyroid gland (underactive thyroid gland). It results in fatigue, weight gain, dry skin, cold intolerance, constipation, depression and menstrual cycle irregularity. It is termed primary, secondary or tertiary depending on where the thyroid hormone deficiency stems from (primary-thyroid gland itself; secondary-pituitary gland/TSH deficiency; tertiary-hypothalamus/TRH).

- Hypothalamic and Pituitary Diseases Genetics
- Thyroid Disorders, Genetic Basis

Hypoxanthine-Guanine-Phosphoribosyl-Transferase

Definition

Hypoxanthine-guanine-phosphoribosyl-transferase (HGPRT) is an enzyme involved in the salvage pathways of nucleotide synthesis. The enzyme catalyzes conversion of ►hypoxanthine to IMP, and guanine to GMP. When HGPRT is present in low levels, affected persons usually experience gout, due to precipitation of uric acid in the joints. Complete absence of HGPRT causes Lesch-Nyhan syndrome, characterized by increased uric acid concentration in the serum and by neurological problems.

- Bloom Syndrome

Hypoxia Inducible Factors

TILL ACKER

Edinger Institute, Neuropathology, Johann-Wolfgang Goethe University, Frankfurt, Germany
till.acker@med.uni-frankfurt.de

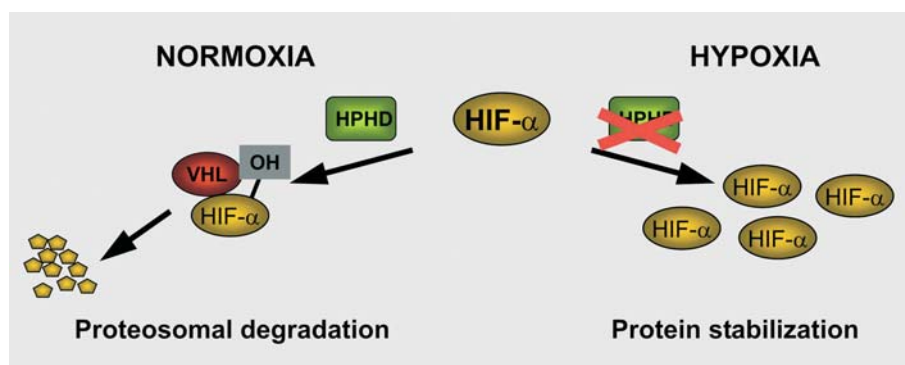
Definition

The maintenance of oxygen homeostasis, i.e. the balance between oxygen consumption and delivery, represents an essential cellular and systemic function, requiring the coordinate regulation of a wide array of genes. It is only within the past few years that the underlying molecular mechanisms have been elucidated. The identification of the HIF transcription system in 1995 constituted a milestone in our understanding of oxygen physiology (1). Since then the HIF system has emerged as a key regulatory system of responses to hypoxia both on a local as well as a systemic level in a variety of developmental, physiological and pathological conditions.

Characteristics

The HIF transcriptional complex is a heterodimer composed of HIF- α and HIF- β subunits belonging to the ►bHLH (basic helix loop helix)-PAS family of transcription factors, conserved among mammalian cells and invertebrate model organisms such as *Drosophila melanogaster* and *Caenorhabditis elegans*. The bHLH domain contains the basic DNA-binding domain and the primary dimerization interface HLH. The adjacent PAS-domain with two conserved regions termed PAS A and PAS B contains the second dimerization motif. Both HIF- α and HIF- β proteins exist as isoforms (HIF-1 α , HIF-2 α , HIF-3 α and ARNT, ARNT2 and ARTN3, respectively). Whereas HIF- β subunits are constitutively expressed and partly involved in other transcriptional responses, specificity to hypoxia-mediated responses seems to be conferred by HIF- α subunits. At least two mammalian α subunits, HIF-1 α and HIF-2 α , are regulated by oxygen in a similar fashion.

Transcriptional activity of HIF- α subunits seems to be mainly regulated by cellular O₂ concentrations involving two mechanisms, namely protein levels and transactivation domain functions. Under normoxic conditions HIF- α subunits are subject to rapid ubiquitination and proteasomal degradation. Upon hypoxia, ubiquitination is dramatically reduced and protein levels increase rapidly (Fig. 1). Deletion analysis revealed that HIF- α contains two transactivation regions, termed the amino- and carboxy-terminal transactivation domains (►N-TAD, ►C-TAD), which



Hypoxia Inducible Factors. Figure 1 Schematic representation of structure, function and posttranslational modification of HIF-1 α , affecting protein stability (ODD) or transactivation activity (N-TAD).

upon reduced oxygen levels are relieved of a negative control. Thus, decreasing oxygen levels lead to a progressive stabilization of HIF- α subunits, which translocate into the nucleus and dimerize with HIF- β subunits, allowing binding to the conserved consensus DNA-binding motif RCGTG residing in the hypoxia-responsive elements (HRE) of many oxygen regulated genes. Transactivation is subsequently initiated by recruitment of co-activators such as CBP (CREB binding protein)/p300, SRC (steroid receptor coactivator)-1 and TIF2 (transcriptional intermediary factor), which is promoted by the redox regulatory protein Ref (redox effector factor)-1. Though oxygen-dependent regulation seems to provide the prevailing control mechanism of HIF function, receptor mediated signals *via* binding of various growth factors and cytokines including angiotensin II, epidermal growth factor (EGF), platelet derived growth factor (PDGF), tumour necrosis factor (TNF) α , insulin and insulin-like growth factors (IGF) 1 and 2, represent an alternative way to enhance HIF activity under normoxia. This induction is generally less intense than that mediated by reductions in oxygen tension.

Cellular/Molecular Regulation

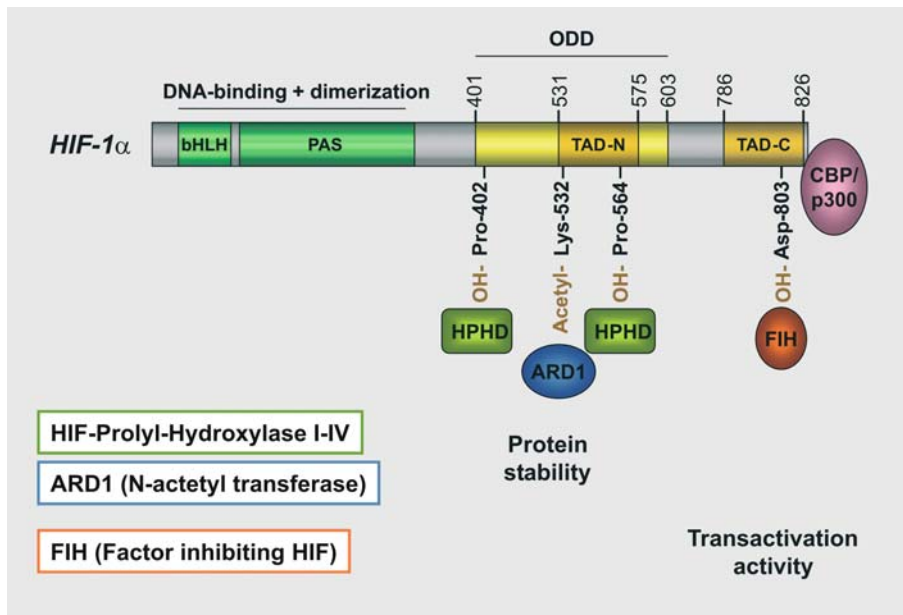
Regulation of HIF activity is complex, involving multiple mechanisms of control at the level of mRNA expression, protein stability, nuclear translocation and transactivation activity, which combine co-operatively to activate HIF to maximal levels under decreasing oxygen concentrations. On the molecular level this is mediated by subjecting HIF- α subunits to multiple modes of modification including two different types of hydroxylation, acetylation and phosphorylation (Fig. 2).

Oxygen Dependent HIF Activation

O₂-regulated degradation of HIF- α subunits is mediated by a functional domain of approximately

200 amino acids, termed the oxygen-dependent degradation domain (\blacktriangleright ODD). This domain confers hypoxic stabilization on HIF-1 α and HIF-2 α , the feature being transferable to various fusion partner proteins. pVHL (von Hippel Lindau protein), loss of function of which is implicated in \blacktriangleright VHL-disease, acts as the recognition component of an E3 ubiquitin-protein ligase that binds to subsequences within the ODD, thus targeting HIF- α subunits for proteasomal degradation. HIF- α protein stability seems to be also regulated by ubiquitin ligases other than VHL, e.g. p53 has been implicated in promoting HIF degradation and decreasing transactivation of HRE-bearing genes, possibly by promoting binding of the ubiquitin ligase MDM-2 and competing for the co-activator p300.

Interaction of VHL with HIF- α requires an O₂- and iron-dependent hydroxylation of specific prolyl residues (Pro 402, Pro 564) within the HIF- α ODD. This posttranslational modification is conferred by a distinct, conserved subclass of 2-oxoglutarate-dependent-oxygenases termed \blacktriangleright HIF-prolyl hydroxylases (HPHD) (2). So far, four isoforms of HPHD have been described (HPHD I-IV). A second oxygen dependent switch involves an iron and 2-oxoglutarate dependent hydroxylation of an asparagine residue within the C-TAD of HIF- α subunits by a recently identified HIF asparaginyl hydroxylase called \blacktriangleright factor-inhibiting HIF (FIH-1). Asparagine hydroxylation apparently interferes with recruitment of the coactivator p300 resulting in reduced transcriptional activity. Both, HPHD and FIH, belong to a superfamily of \blacktriangleright 2-oxoglutarate dependent hydroxylases that employ non-haem iron in the catalytic moiety. They require oxygen in the form of dioxygen with one oxygen atom being incorporated in the prolyl or asparagyl residue and the other into 2-oxoglutarate yielding succinate and CO₂ (3). Thus, the hydroxylation reaction is inherently dependent on ambient oxygen pressure, providing a direct link between oxygen-dependent enzymatic activity and



Hypoxia Inducible Factors. Figure 2 Regulation of HIF-protein stability by oxygen dependent activity of HIF-prolyl-hydroxylases and VHL-mediated ubiquitination and proteosomal degradation.

the regulation of hypoxia-inducible responses, a crucial criterion for an oxygen sensor.

Recently, a novel mode of posttranslational modification of HIF- α subunits under normoxia has been identified, which involves acetylation of a lysine residue (Lys 532) within the ODD domain by an acetyl-transferase termed ARD1. Lysyl acetylation has been shown to modulate HIF- α protein stability by promoting VHL-binding and subsequent proteasomal degradation. With decreasing oxygen tensions acetylation is gradually reduced due to decreased ARD1 mRNA levels and decreased affinity of ARD1 for HIF- α subunits.

Alternative Mechanisms of HIF Activation under Normoxic Conditions

Receptor mediated enhancement of HIF activity seems to involve two major intracellular signal transduction pathways, namely the phosphatidylinositol-3 kinase (PI3K-PTEN-AKT-FRAP) and the mitogen activated protein kinase (RAS-RAF-MEK-ERK) pathways, the latter including the p42/p44 MAP kinase (4). Both signalling cascades have been found to influence HIF activity *via* mechanisms differing from those induced by reductions in oxygen tension with the former directly inducing HIF- α protein synthesis and the latter HIF- α transcriptional activity. As a consequence of PI3K-AKT-FRAP pathway signalling, translational regulatory proteins such as eIF-4E and p70S6K are activated, enhancing translation of HIF- α mRNA. Thus, AKT leads to an increase in HIF transcriptional

activity by increasing the pool of available HIF- α protein. In contrast, direct phosphorylation of HIF- α subunits within the \blacktriangleright C-TAD region by MAPK has been reported to increase HIF transcriptional activity by relieving transcriptional repression, possibly due to interference with FIH-binding and FIH-mediated hydroxylation. There seems to be some cross talk between both pathways as e.g. H-RAS transformation appears to increase HIF- α levels under normoxia and hypoxia *via* recruitment of the PI3K signalling cascade.

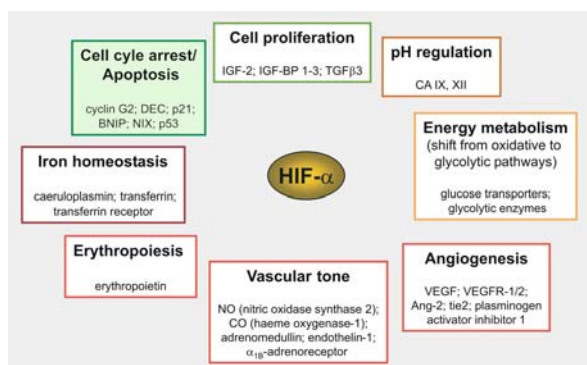
Regulatory Mechanisms

The HIF transcriptional system acts as a master regulator of oxygen-regulated gene expression inducing adaptive mechanisms that serve the common purpose of maintenance of oxygen homeostasis. To date more than 30 HIF-target genes have been identified (Fig. 3).

HIF-regulated Genes and Function

\blacktriangleright Angiogenesis and Erythropoiesis

The vascular system is tightly regulated by oxygen tension. Various molecular players have been identified which are involved in orchestrating specific stages and mechanisms of vascular growth in response to developmental, physiological and pathological hypoxia. Among these, members of the VEGF (vascular endothelial growth factor) and the angiopoietin (Ang) family seem to have a predominant role. In addition, different mechanisms have been identified which control local blood flow by regulating the vascular



Hypoxia Inducible Factors. Figure 3 HIF-regulated target genes and function.

tone through production of NO, CO, adrenomedullin and endothelin-1, all of which involve up-regulation of HIF target genes. Red blood cell production (erythropoiesis) by the bone marrow underlies tight feedback control by oxygen-regulated EPO (erythropoietin) production by the liver and kidney. Taken together, coordinated up-regulation of different sets of genes enhances O₂ delivery to the cell by increasing vascular density, thus decreasing oxygen diffusion distance and increasing oxygen transport capacity by boosting red blood cell mass.

Iron Homeostasis

Iron is required for haem formation, being the most common limiting factor for erythropoiesis. Interestingly, HIF-target genes regulate different steps in iron homeostasis from iron uptake to iron transport and iron storage. The implication of iron in [▶oxygen sensing](#) by 2-oxoglutarate dependent hydroxylases and the involvement of iron in oxygen toxicity through the Fenton reaction gives an additional need for tight iron regulation, making the interaction between oxygen and iron homeostasis physiologically highly appropriate.

Energy Metabolism and pH Regulation

Hypoxia leads to a metabolic adaptation of the cell by inducing a shift from oxidative to glycolytic pathways with anaerobic glycolysis becoming the predominant mode of cellular ATP generation. HIF mediates coordinate up-regulation of genes of the glycolytic pathway ranging from glucose uptake to lactate production. Enhanced lactate production by up-regulation of glycolysis is thought to be the major source of protons under limited oxygen conditions, resulting in decreases in pH. However, pH homeostasis is tightly controlled by various proton extrusion mechanisms. Among these the transmembrane carbonic anhydrases (CA) 9 and 12, recently identified as a new class of HIF- α regulated genes, provide a potential link between

metabolism and pH regulation. They catalyze the reversible hydration of carbon dioxide to bicarbonate and protons. Taken together, these findings suggest that HIF controls fundamental metabolic changes in response to hypoxia by inducing a shift from oxidative to glycolytic pathways whilst at the same time providing means to deal with the resulting increase in proton load by up-regulation of CA.

Proliferation/Apoptosis

Recent studies support the view that apart from inducing pro-proliferative proteins, such as IGF (insulin-like growth factor)-2, IGF-BP (binding proteins) 1-3 and TGF (transforming growth factor) β 3, the HIF pathway includes responses with adverse effects on cell function by inducing cell-cycle arrest specific and pro-apoptotic proteins such as DEC (defective chorion)-1, BNIP (Bcl2/adenovirus E1B 19kD-interacting protein)-3, its homologue NIX (Nip3-like protein X) and cyclin G2. In addition, direct stabilization of the proapoptotic protein p53 has been suggested by studies demonstrating physical and functional interaction between HIF-1 α and p53.

Thus, the HIF system transactivates an extended physiological pathway, which encompasses a wide array of physiological responses to hypoxia ranging from mechanisms that increase cell survival to those inducing cell-cycle arrest or even apoptosis.

Clinical Relevance

Disruption of oxygen homeostasis and activation of HIF have been implicated as essential features in the pathophysiology in a multitude of diseases including cardio- and cerebrovascular disease, chronic obstructive lung disease and tumour formation (5, 6). Though HIF-1 α is ubiquitously expressed in all cells and HIF-2 α in a wide range of cells, the level of HIF- α protein detectable in normoxia and the capacity of HIF- α up-regulation by hypoxia varies in different cell types, making the disease specific contribution of each subunit difficult to discern. To date, current belief suggests that while HIF might play a major role in the prevention of ischemic disease it is involved in the pathogenesis of pulmonary hypertension and cancer; e.g. myocardial ischemia is a potent inducer of HIF-activity and VEGF expression, eliciting an angiogenic response, which, however, is often not sufficient to prevent infarction. In contrast, essential HIF target genes such as endothelin-1, angiotensin II or PDGF have been linked to pulmonary artery remodelling leading to pulmonary hypertension. Similarly, widespread HIF activation in tumours as a consequence of tumour hypoxia and genetic alterations has been associated with tumour progression, invasion and increased resistance to chemo-/radiotherapy. Thus, small molecule inducers of HIF activity might provide

protection against ischemic disease, while small molecule inhibitors might be employed to prevent pulmonary hypertension and cancer. However, with the above outlined dual functions of HIF in cell survival and death it becomes clear that HIF action is far more complex. Depending on the individual components that are induced in response to tissue hypoxia, activation of the HIF system can result in deleterious or beneficial effects, arguing for a cautious application of therapeutic strategies aimed at the general manipulation of the HIF pathway.

References

1. Wang GL, Semenza GL (1995) Purification and characterization of hypoxia-inducible factor 1. *J Biol Chem* 270:1230–1237
2. Epstein AC, Gleadle JM, McNeill LA et al (2001) C. elegans EGL-9 and mammalian homologs define a family of dioxygenases that regulate HIF by prolyl hydroxylation. *Cell* 107:43–54
3. Lando D, Gorman JJ, Whitelaw ML et al (2003) Oxygen-dependent regulation of hypoxia-inducible factors by prolyl and asparaginyl hydroxylation. *Eur J Biochem* 270:781–790
4. Bilton RL, Booker GW (2003) The subtle side to hypoxia inducible factor (HIF α) regulation. *Eur J Biochem* 270:791–798
5. Semenza GL (2001) Hypoxia-inducible factor 1: oxygen homeostasis and disease pathophysiology. *Trends Mol Med* 7:345–350
6. Acker T, Plate KH (2002) A role for hypoxia and hypoxia-inducible transcription factors in tumor physiology. *J Mol Med* 80:562–575

Hypoxia/Normoxia

Definition

Hypoxia is a condition of low oxygen tension, typically in the range 1–5% O₂, and is often found in the central region of tumours due to poor vascularisation. Normoxia is used to describe oxygen tensions between 10–21%, and hyperoxia for those above 21%.

► [Hypoxia Inducible Factors](#)

► [Enzyme Catalyzed Post-Translational Hydroxylation of Proteins](#)

Hypoxic Response

Definition

Hypoxic response designates the adaptations that organisms make (both long- and short- term) to adapt to the effects of hypoxia. In higher animals, specific processes include angiogenesis, vasculogenesis and changes in energy metabolism (associated with a change in the expression of glycolytic enzymes).

► [Hypoxia Inducible Factors](#)

► [Enzyme Catalyzed Post-Translational Hydroxylation of Proteins](#)

I/A Domain

Definition

I/A domain refers to “inserted” sub-domain of a protein. It is involved in ligand binding and present in the extracellular domains of several α integrin subunits.

► [Integrin Signalling](#)

IAP

► [Inhibitor of Apoptosis Proteins](#)

IAP

Integrin-Associated Protein

► [Signal Transduction: Integrin-Mediated Pathways](#)

IBD

► [Identity-by-Descent](#)

ICAM

Intercellular Adhesion Molecule

► [Signal Transduction: Integrin-Mediated Pathways](#)

ICAT

Definition

ICAT designates a method of quantitative proteome analysis (proteomics) based on a class of reagents termed isotope-coded affinity tags (ICAT). The ICAT reagent incorporates three elements: a specific chemical reactivity (usually cysteine-specific), two different isotopically coded linkers (light form and heavy form), and an affinity tag. Proteins of cell state 1 and 2 are derivatized with the isotopically light and heavy forms, respectively. The two samples are combined enzymatically, cleaved (trypsin), and the tagged peptides are isolated and analyzed by capillary liquid chromatography mass spectrometry. The relative amounts of the original proteins are determined by isotope peak ratios of identical peptide sequences which differ in mass only due to different isotopically coded modifications. Identification of the proteins is performed by tandem mass spectrometry which yields amino acid sequence information.

► [Mass Spectrometry: MS/MS](#)

► [Proteomics in Cancer](#)

ICAT

► [Inhibitor of \$\beta\$ -Catenin and TCF-4](#)

I-Cell Disease

Definition

I-cell disease (Mucopolipidosis Type II) is an autosomal recessively inherited lysosomal storage disorder with

clinical manifestations at birth or in the first few months of life. The disease is characterized by the presence of inclusion bodies in cultured fibroblasts (I-cells). Lysosomes lack hydrolases due to a deficiency of N-acetylglucosaminyl-1-phosphotransferase. The functional deficiency of lysosomal enzymes results in abnormal cell architecture. The most severely affected system is the skeletal system, in which trabeculation of bone and cartilage structures are abnormal. Muscular tissue, including cardiac muscle, is relatively spared; however, significant vacuolization is present in the connective tissue cells that are in the heart valves. This leads to thickening of the valves, which results in clinically significant valvular disease. Affected individuals usually die from pneumonia or congestive heart failure within the first decade of life.

► [Glycosylation of Proteins](#)

ICG-HNPCC

Definition

ICG-HNPCC stands for International Collaborative Group on HNPCC.

► [Hereditary Nonpolyposis Colorectal Cancer](#)

Ichthyosis

Definition

Ichthyosis is a skin disorder that is characterized by generalized scaling of the skin.

► [Heritable Skin Disorders](#)

Ichthyosis Bullosa of Siemens

Definition

Ichthyosis bullosa of Siemens refers to an autosomal dominant skin disorder characterized by superficial peeling and hyperkeratosis of the skin.

► [Heritable Skin Disorders](#)

Ichthyosis Hystrix Curth-Macklin

Definition

Ichthyosis Hystrix Curth-Macklin describes an autosomal dominant ichthyosis, with mutilating palmo-plantar keratoderma and characteristic ultrastructural abnormalities.

► [Heritable Skin Disorders](#)

Identity-by-Descent

Definition

The term describes the identity of alleles due to descent from a common ancestor.

► [Bloom Syndrome](#)

Idiopathic

Definition

Idiopathic means without any identifiable cause for a disease.

► [Familial Dilated Cardiomyopathy](#)

Idiosyncratic Drug Reactions

JACK UETRECHT

Pharmacy and Medicine, University of Toronto,
Toronto, Canada

jack.uetrecht@utoronto.ca

Synonyms

Hypersensitivity drug reactions; type B reactions; allergic drug reactions

Definition

Idiosyncratic drug reactions are adverse drug reactions that do not occur in most patients at any dose and do not involve the known pharmacological properties of the drug. Using this definition, adverse reactions, such as

cardiac arrhythmias caused by drugs that affect potassium channels, would not be idiosyncratic because they involve known pharmacological properties of the drug. However, many people would reasonably call these latter reactions idiosyncratic because they only occur in a small fraction of the population whose polymorphisms in the potassium channel make them more susceptible.

Characteristics

There are an almost infinite number of adverse drug reactions because every drug can interact with the body in many different ways. It would be too difficult to describe them all and the focus of this entry will be on idiosyncratic drug reactions. These reactions share common characteristics and although the exact mechanism is probably different for different reactions, most such reactions appear to be immune-mediated. Characteristically there is a delay between starting a drug and the onset of an idiosyncratic drug reaction on initial exposure to the drug. The length of delay varies with the type of adverse reaction. Idiosyncratic drug reactions can affect virtually any organ but skin, liver and bone marrow as well as generalized reactions are the most common. With most skin rashes and antibody-mediated reactions, the delay is usually from a week to a month, with agranulocytosis and hepatic necrosis the delay is usually 1 month to 6 months and with drug-induced lupus the delay is often more than a year. In contrast, on reexposure to the offending drug, the adverse reaction is usually immediate (minutes to a few days) although there are exceptions. Idiosyncratic drug reactions are usually more common in women than in men, but the basis for this difference is unknown, although other types of immune-mediated diseases, such as thyroid disease and lupus, are also more common in women. The reactions are usually said to be dose-independent, but all effects are dose dependent; it is just that the relationship is made more complex by the fact that most people will not have an idiosyncratic reaction at any dose of the drug. In addition, the dose required to cause an idiosyncratic reaction is often smaller than the dose required to exert the desired pharmacological effect. This is especially true in a sensitized patient where the dose required to cause an idiosyncratic reaction is often lower than the dose required on initial exposure. However, the dose-response curve for idiosyncratic drug reactions in susceptible patients is often similar to that for the desired effect and a dose can always be found that does not cause an idiosyncratic reaction. In fact, drugs that are given at a dose of 10 mg/day or less are not usually associated with a significant incidence of idiosyncratic drug reactions.

Cellular and Molecular Regulation

With a few notable exceptions, the mechanism of most idiosyncratic drug reactions is not known with certainty. One notable exception is the allergic reactions that some people have to penicillin. These reactions are known to be due to the covalent or irreversible binding of penicillin to protein. In some patients this induces IgE antibodies in response to the modified or “foreign” protein and these antibodies are responsible for the allergic reaction. Some other idiosyncratic drug reactions are also mediated by antibodies. Some antibodies are directed against the drug, such as aminopyrine-induced agranulocytosis, while others are autoantibodies, as in the case of α -methyldopa-induced hemolytic anemia, which recognize normal red blood cells. Other idiosyncratic drug reactions, such as severe skin rashes (e.g. toxic epidermal necrolysis) are probably mediated by cytotoxic T lymphocytes. Some idiosyncratic drug reactions may be mediated by the innate immune system, which is made up of cells such as [macrophages](#) and natural killer (NK) cells (1).

Penicillin is chemically reactive and can directly bind to proteins. Most drugs are not chemically reactive, but many are metabolized to chemically reactive species, which can bind to protein and there is a large amount of circumstantial evidence that such reactive metabolites are responsible for many idiosyncratic reactions (2). This is the basis for the classic hapten hypothesis, i.e. small molecules are not immunogenic but if they bind to biological macromolecules, usually proteins, these modified proteins can be immunogenic. This binding must be essentially irreversible or covalent in order to survive antigen processing. A small molecule that modifies a protein leading to an immune response is referred to as a hapten. However, there are many drugs that form reactive metabolites and yet are not associated with a significant incidence of idiosyncratic reactions. There are several possible reasons for this observation. One is that the amount of covalent binding is simply too low to induce an immune response. Although covalent binding can be quantified under certain circumstances, it is not possible to quantify covalent binding *in vivo* in humans in the usual target organs of idiosyncratic drug reactions. Therefore, we do not know how good a correlation exists between the degree of covalent binding and the risk that a drug will cause idiosyncratic drug reactions. Another possible reason is that covalent binding to certain types of proteins may be more likely to cause an idiosyncratic reaction than binding to others and we still do not know what proteins are modified by most drugs. Finally, it is possible that covalent binding is necessary but not sufficient, or possibly not even necessary, for the induction of an idiosyncratic drug reaction.

It has been known for some time that foreign proteins do not usually induce a significant immune response in the absence of an ►[adjuvant](#). The purpose of the adjuvant is to up-regulate costimulatory molecules on antigen presenting cells and this is necessary for the induction of an immune response. Matzinger proposed that it is not the foreignness of something that determines whether it will induce an immune response but rather whether it causes some type of danger to the organism. This is referred to as the danger hypothesis (3). Applying this hypothesis to idiosyncratic drug reactions, it may be that a reactive metabolite must induce some type of danger signal in order to induce an immune response (1). By their very nature, reactive metabolites usually cause some type of cell damage and this could act as a danger signal. In fact, the ability of the reactive metabolite to act as a hapten may not even be important (4). These hypotheses have yet to be adequately tested. At the present time there is no accurate way to test drug candidates for their potential to cause idiosyncratic drug reactions. However, if there are biological markers of cell damage or cell stress that predict that a drug will cause idiosyncratic drug reactions, this would have a major impact on drug development. It is unlikely that such effects could be accurately mimicked in cell cultures and *in vivo* tests would be required. Furthermore, there is likely to be more than one pattern of cell stress that is associated with an increased risk of adverse reactions and in some cases there will probably be differences between animals and humans, so such predictive tests are unlikely to be simple.

As suggested by the name, most people do not have an idiosyncratic drug reaction to a drug that is associated with such reactions. If “danger” plays an important role in the mechanism of idiosyncratic drug reactions, then environmental factors, such as infections or surgery would probably increase the risk of such reactions. This appears to be the case. Patients who have HIV infections are at increased risk of idiosyncratic drug reactions. Furthermore, it appears that patients who are started on procainamide immediately after open-heart surgery have a 10 fold higher incidence of procainamide-induced agranulocytosis than those started on the drug under more usual circumstances. However, it is important to point out that even in the presence of open-heart surgery, 95% of the patients do not develop agranulocytosis. It appears that the immune systems of individuals who do not have an idiosyncratic reaction do respond to the drug but the response is ►[immune tolerance](#) (5). The risk of idiosyncratic reactions to some drugs can be decreased by treating a patient with a lower dose of the drug for 2 weeks; this promotes tolerance. Although not proven, surely genetic factors must play a dominant role in the interindividual differences in the risk of an idiosyncratic drug reaction. Many of the

enzymes responsible for metabolizing drugs, including the conversion of drugs to reactive metabolites, are genetically polymorphic and this could be the basis for interindividual differences in risk (6). The effect of genetic polymorphisms in drug metabolizing enzymes has received much attention as a basis for the large differences observed in the response of different individuals to drugs. However, genetic polymorphisms in drug transporters and the therapeutic targets of drugs are also very important (7). In the case of idiosyncratic drug reactions, there are several examples where genetic polymorphisms in drug metabolizing enzymes have been shown to be important (8). For example, the slow acetylators are at higher risk of idiosyncratic reactions to sulfonamides and isoniazid-induced liver necrosis. However, in most cases when such associations are sought they are usually not found. It is important to note that many of the genes coding for specific components of the immune system are extremely polymorphic. Our survival as a species may have been very much dependent on such polymorphism because if the immune system were monomorphic we might have been wiped out by some pathogen. For the same reason the immune system must also be complex and redundant. Therefore there are many more polymorphic genes involved in the control of the immune system than there are involved in drug metabolism and, on a statistical basis, there will probably be more “immune” polymorphic genes that are associated with idiosyncratic drug reactions than polymorphic genes involved in drug metabolism. The genes that have been studied the most as a risk factor for idiosyncratic reactions are the ►[major histocompatibility complex](#) genes (8), which are highly polymorphic and are also involved in the presentation of antigens to T lymphocytes. For example, clozapine-induced agranulocytosis has been linked to the several ►[human leukocyte antigen](#) (HLA) genes and the associations were different in different ethnic groups. However, these genes are also linked to genes controlling expression of ►[tumor necrosis factor](#) and ►[heat shock proteins](#) and this type of gene linkage makes mechanistic inferences difficult. Although there may be cases in which one gene polymorphism plays a dominant role in the risk of an idiosyncratic drug reaction, it is likely that in most cases hundreds if not thousands of genes will be involved, each making only a small contribution to the risk. If this is the case, it will be very difficult to determine the risk of a specific individual having an idiosyncratic drug reaction to a specific drug. In most cases the history of an idiosyncratic reaction to one drug does not predict an increased risk of an idiosyncratic reactions to other drugs. Although some individuals believe they are “allergic” to all drugs, this does not appear to be true. However, there are some examples where an idiosyncratic

reaction to one drug is associated with an increased risk of an idiosyncratic reaction to another drug, presumably because the structures are very similar, such as different penicillins, or because the mechanism is similar. An example of the latter is that if a patient has a reaction to one aromatic anticonvulsant, i.e. phenytoin, carbamazepine or phenobarbital, they are at increased risk of having a similar idiosyncratic reaction to the others in this group. It was initially believed that there was a type of reactive metabolite, specifically an arene oxide, common to these three anticonvulsants and an increased risk of an idiosyncratic reaction was due to defective detoxication of this reactive metabolite by epoxide hydrolase. However, this does not appear to be the case.

Clinical Relevance

It has been estimated that adverse drug reactions in general, are the 4th to 5th leading cause of death in North America. This is probably somewhat misleading because it is easy to argue about the cause of death in a patient. For example, if someone has cancer, is treated with anticancer agents that suppress the bone marrow and make the patient susceptible to infections and they develop pneumonia and die, what was the cause of death? If the death is blamed on cancer or pneumonia, then adverse drug reactions are not the 4th leading cause of death, but if the death is blamed on drugs, i.e. the anticancer drugs, then adverse drug reactions may be the 4th leading cause of death. Most of the deaths in these studies involved anticancer drugs and drugs used to treat severe heart problems and these patients were at very high risk of dying no matter what the treatment. Even though I suspect that the incidence can be overstated, idiosyncratic drug reactions are certainly very important. Most adverse drug reactions are not idiosyncratic, but about 20% are and these are very difficult to deal with because of their unpredictable nature. They are often severe. They also greatly impede the development of new drugs because a major part of the expense of drug development is trying to ensure drugs are safe and yet idiosyncratic reactions are usually not detected by preclinical testing and clinical trials. In the last quarter of the 20th century, 10.2% of drugs that were released on the market either had to be withdrawn from the market or received a “black box” warning that severely limited the use of the drug. As indicated above, if biological markers could be found that predicted the risk of idiosyncratic drug reactions, it would have a major impact on drug development.

References

1. Uetrecht JP (1999) New concepts in immunology relevant to idiosyncratic drug reactions: the “danger hypothesis” and innate immune system. *Chem Res Toxicol* 12:387–395
2. Park BK, Pirmohamed M, Kitteringham NR (1998) Role of drug disposition in drug hypersensitivity: A chemical, molecular, and clinical perspective. *Chem Res Toxicol* 11:969–988
3. Matzinger P (1994) Tolerance, danger and the extended family. *Annu Rev Immunol* 12:991–1045
4. Pichler WJ (2002) Pharmacological interaction of drugs with antigen-specific immune receptors: the p-i concept. *Curr Opin Allergy Clin Immunol* 2:301–305
5. Ju C, Uetrecht JP (2002) Mechanism of idiosyncratic drug reactions: reactive metabolite formation, protein binding and the regulation of the immune system. *Curr Drug Metab* 3:367–377
6. Phillips KA, Veenstra DL, Oren E et al (2001) Potential role of pharmacogenomics in reducing adverse drug reactions: a systematic review. *J Am Med Assoc* 286:2270–2279
7. Ingelman-Sundberg M (2001) Pharmacogenetics: an opportunity for a safer and more efficient pharmacotherapy. *J Intern Med* 250:186–200
8. Pirmohamed M, Park BK (2001) Genetic susceptibility to adverse drug reactions. *Trends Pharmacol Sci* 22:298–305

IEF

Definition

Isoelectric focusing refers to the separation of molecules according to their isoelectric point. A pH gradient of molecules (ampholytes) with low Mr (relative molecular mass) is built up by an electric field. The analytes migrate until they reach their isoelectric point, where they are uncharged and therefore focus at this position.

► [Two-Dimensional Gel Electrophoresis](#)

Ig Domain

Definition

Ig (immunoglobulin) domain is a common structural motif of many proteins. The Ig domain consists of a compact structure with two cysteine residues, separated by 55–75 amino acids that are arranged as two antiparallel beta sheets.

► [Adhesion Molecules](#)

IGF 1

Definition

IGF I (insulin-like growth factor 1) is a growth factor with structural similarities to the hormone insulin, and

is produced not only in the liver but also locally in tissues. It participates in bone formation during development.

- Cytokines
- Growth Factors

IgV_H

Definition

Variable region of the immunoglobulin heavy chain gene. Chronic B-cell lymphocytic leukaemias with an unmutated IgV_H locus are associated with a poor prognosis, whereas a hypermutated IgV_H locus is associated with a better prognosis and a more indolent clinical course.

- Leukemia

IHC

- Immunohistochemistry

IL

Interleukins

- Cytokines
- Growth Factors
- Interleukins and Their Receptors, a Role in Regulating Hematopoietic Stem Cells

IL-6 Family

Definition

Cytokines of the IL-6 family include leukemia inhibitory factor (LIF), cardiotrophin-1 (CT-1), ciliary neurotrophic factor (CNTF), oncostatin M (OSM), and IL-11. This family shares a common co-receptor gp130. In the heart, all cytokines of this family have been

reported to cause hypertrophy (although with different potencies).

- Cardiac Signaling: Cellular, Molecular and Clinical Aspects
- Growth Factors
- Interleukins and Their Receptors, a Role in Regulating Hematopoietic Stem Cells

IMAC

- Immobilized Metal-Affinity Chromatography

Image Analysis

Definition

Image analysis refers to an algorithmic process where a digital image is transformed into a spot quantitation matrix.

- Microarray Data Analysis

Imatinib Mesylate

Definition

Imatinib mesylate is a selective small molecule inhibitor of a small family of tyrosine kinases including the Bcr-Abl fusion protein. Imatinib mesylate induces high remission rates in patients with chronic myelogenous leukemia (CML).

- Leukemia

Imidazolium

Definition

Imidazolium is a five member heterocyclic compound in histidine.

- Proteases and Inhibitors

Immediate Early Genes

Definition

Immediate early genes are genes that are characterised by a rapid and transient activation of transcription in response to changes of environmental conditions, such as growth factors, cytokines, tumour promoters, carcinogens and expression of certain oncogenes. Immediate early genes can also describe viral genes that are expressed first after infection.

- ▶ Jun/Fos
- ▶ Transcription Elongation

Immobilized Metal-Affinity Chromatography

Definition

IMAC is based on the interaction between a transition metal ion (Co^{2+} , Ni^{2+} , Cu^{2+} , Zn^{2+}) immobilized on a matrix and specific amino-acid side chains.

- ▶ Protein Tags

Immobilized pH Gradients

Definition

Immobilized pH gradients (IPG) are based on the principle that the pH gradient within the gel is generated by a limited number (6–8) of well-defined chemicals (immobilines), which are co-polymerized with the acrylamide matrix. The gradient is built up by a gradient mixer. IPGs allow the generation of pH gradients of any desired range (broad, narrow or ultra-narrow) between pH 3 and 12.

- ▶ Two-Dimensional Gel Electrophoresis

Immortalisation

Definition

Immortalisation describes the loss of the intrinsic limit in proliferation potential characteristic of normal

primary mammalian cells. Immortalised cells can be propagated indefinitely in culture.

- ▶ Ras Signalling
- ▶ Senescence

Immune Cells

Definition

Cells of the immune system that is activated upon the entry of microorganisms or foreign substances into the organism.

- ▶ Immunity/Immune System
- ▶ Microvilli

Immune System

- ▶ Immunity/Immune System

Immune Tolerance

Definition

Immune tolerance is the specific failure of a normally responsive individual to make an immune response to a known antigen. It results from previous contact with the antigen by an immunologically immature individual (fetus or neonate), or by an adult who has been exposed to extremely high- or low-dose antigen, ▶ radiation, antimetabolites, antilymphocytic serum, etc.

Immunity

Definition

Immunity is the ability of an organism to repel and stop the spread of disease-causing agents such as viruses, bacteria, and fungi or other foreign materials, and protect the body from infection. The immune system is a complex of organs – highly specialized cells and a

lymph circulatory system – all of which work together to ensure immunity and clear infection from the body. As a first line of defense, non-specific responses block the entry and spread of disease-causing agents. As a second line of defense, highly specific (antibody- and cell-mediated) responses, generate specific responses to an individual threat. Dysregulation/defects of the immune system can cause severe diseases like autoimmune diseases, allergies, immune deficiency syndromes and transplant rejection.

► [Innate Immunity](#)

► [Interleukins and Their Receptors, a Role in Regulating Hematopoietic Stem Cells](#)

Immunization, Active and Passive

Definition

Passive immunization is the treatment of a non-immune individual with antibodies from an immunized donor. This results in a state of transient immunity in contrast to active immunization which activates the host's own immune system and results in long-lasting immunity.

► [Monoclonal Antibodies](#)

Immunoblot

Definition

Immunoblot defines a method by which the transferred proteins on a western blot are probed with an appropriate antibody and the resulting bands visualised using fluorescence, chemiluminescence or autoradiography.

► [Immunochemical Methods: Localization](#)

► [Recombinant Protein Expression in Bacteria](#)

Immunochemical Methods, Localization

LEE BERGMAN, STEPHANIE BECHTEL,
STEFAN WIEMANN

Department of Molecular Genome Analysis, German Cancer Research Center, Heidelberg, Germany
s.bechtel@dkfz.de, s.wiemann@dkfz.de

Synonyms

Immunohistochemistry; immunocytochemistry; immunostaining; immunofluorescence; subcellular localization; cellular localization; protein localization

Definition

Immunochemical methods are processes utilizing the highly specific affinity of an ► [antibody](#) for its antigen to detect the distribution of a given protein (antigen) in tissues or cells. This methodology is highly important in the clinical context, as the normal localizations of a large number of proteins are known to be altered in various disease states, particularly in cancer.

Characteristics

Important ideas regarding the function of newly identified or uncharacterized proteins are often obtained by identifying the normal environment in which a given protein resides, by determining in which species or tissue it is expressed, in which cell type and ► [subcellular compartment](#) this protein can be found or whether there is some variation in expression of the protein during development or in different stages of, for example, the cell-cycle. In addition, some alteration in the normal expression pattern of a particular protein may be indicative of a particular disease state, information that may be useful for diagnosis and decisions regarding treatment.

Because many proteins have highly specialized functions within an organism, identifying the particular organs and tissues in which a protein is expressed provides valuable insight into the normal functions of the protein. For example, a protein expressed only within the brain is possibly involved in some aspect of nervous or sensory system regulation or in signaling pathways directing higher processes such as cognition or memory. Likewise, determining the specific cell type within a tissue in which a given protein is expressed further defines its normal function. For example, proteins only found within pancreatic islet cells may play some role in endocrine hormone production (proinsulin, glucagon, gastrin etc), whereas proteins found exclusively within pancreatic exocrine cells are more likely to be associated with pathways for digestive enzyme production. Similarly, identifying the subcellular compartment in which a protein normally resides provides more detailed information regarding its molecular function. For example, this may involve structural roles, such as gap junction or cytoskeletal components or being part of various metabolic pathways like DNA replication/maintenance and protein trafficking. This also provides an indication of the direction of future analyses.

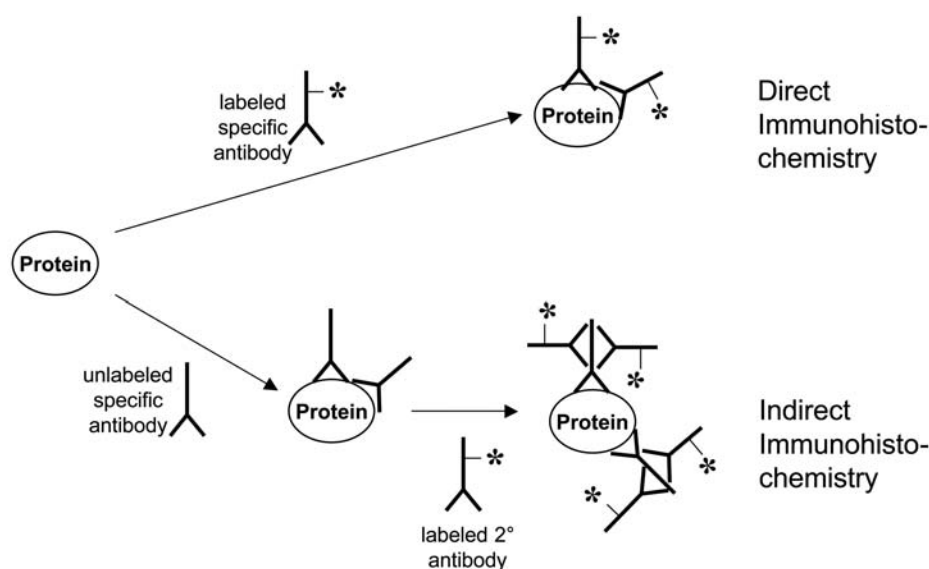
Numerous methodologies have been developed to enable the localization of proteins within tissues and cells. Most rely upon immunochemical techniques, whereby a directly or indirectly labeled antibody (generated by exposing animals such as rabbits to doses of the protein of interest, or part thereof) specific for the protein of interest binds to this protein. The localization of bound antibody is then detected microscopically. Immunohistochemical methods, as the name suggests, are principally involved in the analysis of proteins within tissue sections, whereas immunocytochemical methods are employed for analyzing proteins within a cellular context.

Either direct or indirect immunochemical techniques are employed for localization of proteins within tissues or cells. Direct immunohistochemistry involves using an antibody specific for the protein of interest coupled to a particular label, whereas indirect immunohistochemistry involves the use of an unlabeled primary antibody specific for the protein of interest, which is then detected using a labeled secondary antibody recognizing immunoglobulin (IgG) from that species in which the primary antibody was generated. Indirect immunohistochemistry is more widely used, as this approach results in amplification of the resulting signal and does not raise the potential problem of inhibition of antibody binding as a result of the label. Additionally, individual labeling of each primary antibody is not necessary. Instead, most users have a series of labeled species-specific secondary antibodies that are used to detect many primary antibodies, e.g. a labeled anti-mouse-IgG antibody generated in goat will detect any murine-specific primary antibody (1). A comparison of

direct versus indirect immunochemistry is given in Fig. 1.

Three different types of labels are primarily used for immunochemistry, ►enzymes, metallic compounds and ►fluorochromes. Enzymatic labeling of antibodies is applied almost exclusively to tissue sections. This approach provides the advantage of increased sensitivity of detection and permanence of staining. The basic principle of enzymatic immunohistochemistry involves coupling an enzyme such as ►alkaline phosphatase or ►horseradish peroxidase to a secondary antibody. These enzymes catalyze the conversion of a colorless substrate to a colored precipitate, which is then readily detected using a regular transmitted light microscope, in an extra reaction step following the immunostaining procedure. When highly detailed information is required regarding the localization of a protein within a subcellular compartment, electron microscopy may be the method of choice. This highly specialized technique requires the use of electron dense labels, such as colloidal gold or silver coupled to a primary or secondary antibody for detection of proteins.

Fluorochromes are the usual label of choice for most immunochemical methods, particularly immunocytochemistry. Many different fluorescent markers with different emission spectra are available, allowing considerable flexibility in experimental design. Furthermore, simultaneous detection of multiple proteins within the same cell or tissue section using antibodies labeled with different fluorochromes is possible, provided primary antibodies from different species are used to prevent "cross-reactivity". Table 1 shows a list of commonly used fluorochromes.



Immunochemical Methods, Localization. Figure 1 Comparison of direct and indirect immunochemistry.

Immunochemical Methods, Localization. Table 1 Fluorochromes commonly used in immunochemical techniques. This list is not exhaustive

Fluorochrome	Excitation wavelength (nm)	Emission wavelength (nm)
5-Carboxyfluorescein (5-FAM)	492	518
5-Carboxytetramethylrhodamine (5-TAMRA)	542	568
5-Hydroxy Tryptamine (HAT)	370-415	520-540
Acid Fuchsin	540	630
Acridine Red	455-600	560-680
Acridine Yellow	470	550
APC (Allophycocyanin)	630, 645	655, 660
Astrazon Brilliant Red 4G	500	585
Astrazon Orange R	470	540
Astrazon Red 6B	520	595
Astrazon Yellow 7 GLL	450	480
Beta Lactamase	409	447, 520
Calcein	494	517
Calcium Green	501, 506	531
Calcium Orange	549	575, 576
Chromomycin A	436-460	470
CMFDA	494	520
Cy3™	514	566
	552	570
	554	
Cy5™	649	666, 670
Dansyl Chloride	372	518
DAPI	359	461
Eosin	524	545
Ethidium Bromide	510, 523	595, 605
Fast Blue	360	440
Fluorescein isothiocyanate (FITC)	490, 494	520, 525
Hoechst 33258	345	487
Hoechst 33342	347	483
Hoechst 34580	392	440
Lissamine Rhodamine	572, 577	591, 592
Lucifer Yellow	425, 428	528, 536, 540
Phycoerythrin R [PE]	565	578
Propidium Iodide (PI)	(305), 536, 538	617
Rhodamine	550	573
Tetramethylrhodamine (TRITC)	555	576
Texas Red™	595	620

In addition, the protein can be detected directly with a fluorescence microscope or perhaps with a ►fluorescence-activated cell sorter (FACS), without the need for additional, sometimes inconsistent enzymatic reactions (2). The primary disadvantage of fluorochromes is their sensitivity to light, which means that their signal is only detectable for a certain time period, depending on the properties of the individual molecule. Some more unstable dyes require special handling to prevent inactivation by light exposure.

When no antibody against a protein of interest is available, over-expression of the protein fused to a peptide or protein tag, either at the beginning (N-terminal) or end (C-terminal) of the protein provides a practical alternative for localizing the protein at a subcellular level. This methodology is especially appealing for high-throughput approaches. Two key types of ►fusion tag are available, ►epitope tags and fluorescent proteins, such as ►green fluorescent protein (GFP). An epitope tag consists of a short, specific stretch of amino acids known to be immunogenic and often derived from well-characterized proteins for which primary antibodies are readily available. Because these sequences are short, they are less likely to affect the normal folding or function of the protein of interest. Table 2 shows a summary of some of the common epitope tags currently in use (3).

GFP and its family of spectral variants – for example blue, cyan, red and yellow fluorescent proteins (BFP, CFP, DsRed and YFP respectively) – have become widely used in recent years as fusion tags. These tags have a number of useful properties, such as high stability, low toxicity and fluorescence without additional experimental steps. The existence of these different spectral variants, which were originally generated by substitution of certain amino acids within GFP, allows concurrent analysis of several different proteins. Many researchers now utilize these fluorescent proteins as fusion tags for the localization of novel

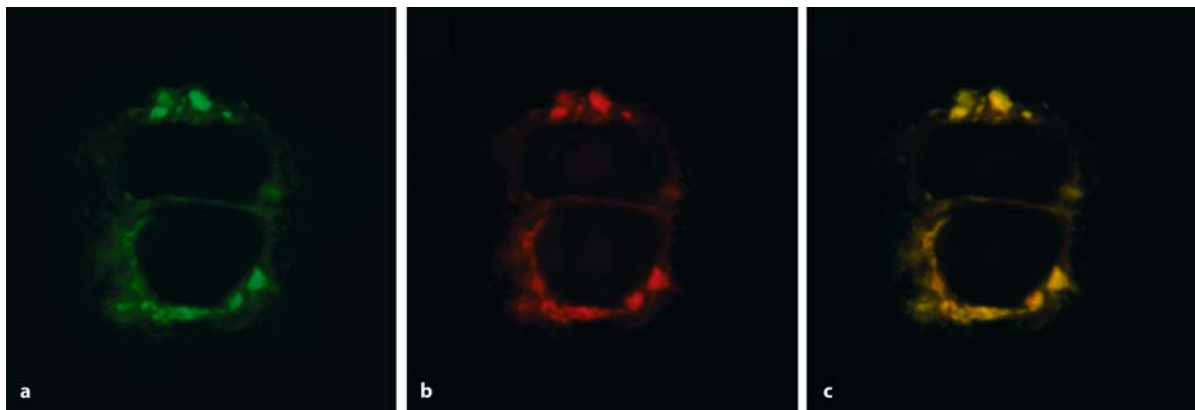
proteins (4). An example of this is shown in Fig. 2, where a novel protein was shown to be located in the mitochondria, due to the co-localization of a C-terminal YFP fusion of this protein with Hsp60, a known mitochondrial protein. It must be noted however, that the presence of GFP fused to a protein may cause incorrect localization results if the necessary sequence motifs are hidden within the tertiary structure of the protein by the presence of GFP.

As no antibody staining steps are required to view GFP fusion proteins within cells, these fluorescent proteins are also highly suitable for use in live cell imaging. This has become an important field of research in recent years, providing valuable information about the natural environment of proteins. Highly specialized techniques, such as ►fluorescence resonance energy transfer (FRET), and ►fluorescence recovery after photo bleaching (FRAP), have been developed in order to analyze properties such as interactions between proteins and intracellular protein movement.

Due to the highly structured, three-dimensional nature of organs and tissues and indeed cells, certain techniques are required to enable good detection and visualization of proteins. Primarily, the tissue must be fixed for long-term storage and use, which involves preservation of tissue morphology and immobilization of the various cellular components while still maintaining the antigenicity of proteins. Various advantages and disadvantages are associated with each of the many available fixation techniques and it is common practice to optimize the technique to suit the antibody and antigen of interest. Fixation procedures for tissues and cells commonly involve chemical alteration of cellular components to preserve their structural integrity, using ethanol to precipitate cellular proteins or formaldehyde and its derivatives, which cross-link various structural proteins. Another technique commonly applied to whole tissue sections, particularly in the clinical environment, involves snap freezing of the tissue in

Immunochemical Methods, Localization. Table 2 List of common epitope tags available for localization of fusion proteins using immunochemical methods

Tag	Peptide sequence	Origin
c-myc	EQKLISEEDL	Amino acids 411-420 of human c-myc protein
FLAG	DYKDDDDK	Synthetic peptide
HA	YPYDVPDYA	Amino acids 98-106 from human influenza virus hemagglutinin, HA1
His6	HHHHHH	Synthetic peptide
HSV	QPELAPEDPED	Peptide from herpes simplex virus glycoprotein D
Protein C	EDQVDPRLIDGK	Human protein C



Immunochemical Methods, Localization. Figure 2 Example of immunochemical methods used to co-localize two mitochondrial proteins. **A.** A novel protein expressed in HEK 293 cells is detected via a C-terminal YFP fusion tag. **B.** Expression pattern of Hsp60, a known mitochondrial protein, in the same cells. The protein was detected by incubation with a primary antibody specific for Hsp60 generated in mouse, followed by an anti-mouse IgG antibody coupled to Texas Red. **C.** Overlay of the two images to demonstrate co-localization of the two proteins within the mitochondria. Transfected cells were fixed on glass slides with 4% paraformaldehyde and permeabilized with a solution of 0.1% Triton X-100 to allow uptake of antibody. After antibody staining, slides were mounted with Mowiol to preserve structures. Images were obtained using a Zeiss LSM510 Meta laser-scanning microscope.

liquid nitrogen. After fixation of an entire tissue or organ, it is necessary to prepare ultra-thin sections from the fixed tissue for subsequent immunohistochemical staining and visualization, usually using a microtome.

Clinical Relevance

Immunochemical methods, particularly immunohistochemical techniques have been shown to be indispensable for the correct identification of the origin and stage of many tumor types, for distinguishing benign from malignant neoplasms and for diagnosing and understanding many other disease states, particularly immune system disorders, by virtue of the fact that the localization of key proteins is known to change in a highly tissue- and cell type-specific manner during these disease processes. Detailed knowledge of the precise nature of an illness is essential for a clinician in directing treatment and determining a patient's prognosis. For example, the presence of κ or λ immunoglobulin light chain in lymphocytes is indicative of either malignant B-cell lymphoma or myeloma/plasmacytoma. An important marker for distinguishing certain thyroid tumors is thyroglobulin. Tumors staining positive for this antigen are most likely to be papillary or follicular thyroid tumors, whereas medullary and metastatic thyroid tumors generally do not express thyroglobulin. Often, a panel of antigens is tested in clinical samples in order to diagnose a condition correctly. For example, specific expression patterns of IgG, IgA, IgM, C3 and κ and λ light chains are used to differentiate between numerous kidney diseases such

as glomerulosclerosis, glomerulonephritis, forms of lupus and amyloidosis (6).

As a result of detailed functional and structural information about proteins being increasingly available following the sequencing of the human genome, as well as with improved markers and detection systems for immunochemistry, the importance of these techniques as diagnostic and prognostic markers will undoubtedly increase in the future.

► **Immunological Methods, the Use of Antibodies to Discover the Function of Novel Gene Products**

References

1. Thorpe SJ, Kerr MA (1994) Common immunological techniques: ELISA, blotting, immunohistochemistry and immunocytochemistry. In: Kerr MA, Thorpe R (eds) *Immunochemistry Labfax*. Bios Scientific Publishers, United Kingdom, pp 175–207
2. Allan VJ (2000) Basic immunofluorescence. In: Allan VJ (ed) *Protein localization by fluorescence microscopy: a practical approach*. Oxford University Press, United Kingdom, pp 1–26
3. Jarvik JW, Telmer CA (1998) Epitope Tagging. *Annu Rev Genet* 32:601–618
4. Simpson JC et al (2000) Systematic subcellular localization of novel proteins identified by large-scale cDNA sequencing. *EMBO Rep* 1:287–92
5. Chamberlain C, Hahn KM (2000) Watching proteins in the wild: fluorescence methods to study protein dynamics in living cells. *Traffic* 1:755–62
6. Tubbs RR et al (1986) *Atlas of Immunohistology*. American Society of Clinical Pathologists Press, USA

Immunocytochemistry

► Immunochemical Methods, Localization

Immunofluorescence

► Immunochemical Methods, Localization

Immunogenic

Definition

Immunogenic characterizes a molecule that is capable of inducing an immune response and eliciting the formation of antibodies.

► Camel as a Model for Functional Genomics
► Proteomics in Human-Pathogen Interactions

Immunoglobulin

Definition

Immunoglobulins or ► **antibodies** are a class of related serum proteins synthesized in B-lymphocytes as part of the immunological defense system, which bind specifically to foreign molecules (antigens). Immunoglobulins consist of two identical heavy chains and two identical light chains; both contain a constant region that is specific for a certain class of immunoglobulins, and a variable region that constitutes the antigen binding site and which is specific for each immunoglobulin.

► Affinity Chromatography and *In Vitro* Binding (Beads)
► Autoimmune Diseases
► Camel as a Model for Functional Genomics
► Double-Strand Break Repair

Immunoglobulin-Fold

Definition

Immunoglobulin-fold designates a protein domain structure, first discovered in immunoglobulin constant

and variable domains, which consists of two β -sheets packed against each other.

► Protein/DNA Interaction

Immunohistochemistry

Definition

Immunohistochemistry comprises the staining of proteins inside thin tissue sections using antibodies, coupled for example to a fluorescent molecule or

► horseradish peroxidase.
► Immunochemical Methods: Localization

Immunological Dysbalance

Definition

Immunological dysbalance means the overexpression of proinflammatory and type 1 cytokines (such as IL-2 and IFN γ (► **Interferon**) produced by Th1 helper cells). Immunological dysbalance is a typical finding in psoriasis.

► Psoriasis, Molecular Basis

Immunological Methods, the Use of Antibodies to Discover the Function of Novel Gene Products

MARKUS ENZELBERGER, THOMAS VON RÜDEN
MorphoSys AG, Martinsried, Germany
enzelberger@morphosys.com,
vonRueden@morphosys.com

Definition

Antibodies are specific molecules of the vertebrate immune system, which are produced upon contact with an ► **antigen**. They bind the antigen with both high affinity and specificity and have thus become the reagents of choice for identifying and quantifying single proteins in complex protein mixtures. Antibodies are an essential tool in understanding the role of the individual gene and its function in complex

networks; they enable the detection and quantification of the gene product, they can reveal the localization of gene products *in situ* and they can be used for functional characterization of gene products (1). The combined benefit of high specificity and broad applicability in a variety of biological assays makes antibodies valuable tools. High throughput methods for antibody generation, antigen production and antibody characterization have been developed.

Characteristics

The primary questions when elucidating the function of novel proteins are where, when, in what amount, in which modification and for what purpose is the protein expressed?

How is this pattern altered under pathological circumstances?

Antibodies specific for the protein of interest can answer these questions by their use in various methods.

Detection and Quantification of Soluble Proteins

► **Enzyme Linked Immuno-Sorbent Assay (ELISAs)** are used for quantifying the protein of interest. The assays might be multiplexed by using ► **antibody arrays** on planar or bead surfaces (3); on planar arrays, the antigen is identified by the position of the signal on the array, on bead-based arrays, the antigen is detected by either the color or the size of the bead running by a detection module.

Four different ELISA protocols have been described for quantification and detection of antigen on these arrays (Fig. 1). (a) The protein of interest is labeled and detected in the lysate by binding to the antibody

directly; (b) the first antibody binds the antigen in the lysate and the second antibody carries the label; (c) the lysate is immobilized and different labeled antibodies bind to the lysate or (d) to speed up the assay development process drastically one can detect antibody/antigen binding to large antibody arrays by ► **surface plasmon resonance (SPR)** without labeling either of the binding partners.

In all cases, the challenge is to find an antibody that has a very low potential to cross react with other proteins in the lysate that are also applied to the chip.

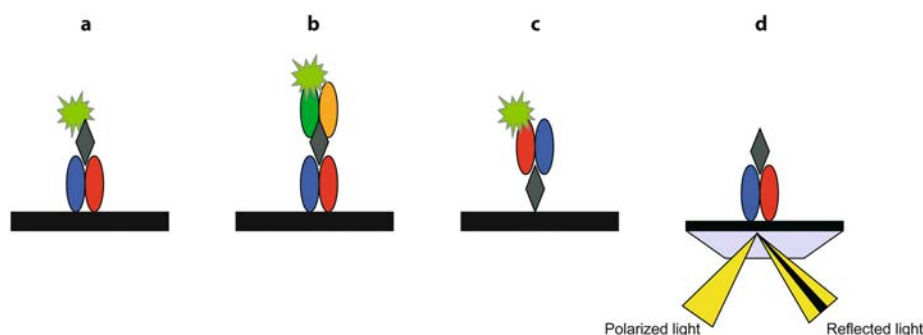
There is a trend for using specific antibodies against posttranslational modifications of proteins, e.g. phosphorylation, instead of using radioactive methods. Besides their ease of use, this offers the possibility to multiplex and analyze complex pathways in a single experiment.

Detection and Quantification of Proteins in Tissue Extracts

General protein expression of the novel protein of interest is studied by ► **Western blot (WB)** analysis. This method usually denatures the protein of interest, thus, the antibody used for detection has to be able to detect linear epitopes. Laborious sample preparation, difficulties in quantification and the comparatively large amount of protein needed for Western blot analysis make it only moderately suited for high throughput analysis of protein expression.

In Situ Localization of the Protein

► **Immunohistochemistry (IHC)** is used to identify the site of protein expression in tissue sections or on



Immunological Methods, the Use of Antibodies to Discover the Function of Novel Gene Products.

Figure 1 Protein and antibody array formats: (a) Antigen capture; antibodies are immobilized on a solid support in a known pattern. All the proteins of the lysate are labeled with a fluorescent dye. The lysate is incubated with the solid support and, after washing, the proteins contained can be identified. (b) Sandwich ELISA: (capture) antibodies are immobilized on a support in a known pattern, after addition of lysate, a second (detection) fluorescently labeled antibody and washing, the proteins contained are identified. (c) Reverse phase: the lysate is immobilized on a surface and the proteins contained are detected using differently labeled antibodies. (d) Surface plasmon resonance (SPR) detection: (Capture) antibodies are immobilized on an SPR surface in a known pattern; lysate is added and, after washing, binding can be detected by a change in the resonance angle. No labeling is needed for this method.

► **tissue microarrays** (2). The latter allow the parallel study of expression in tissues from different sites, diseases or developmental stages in parallel.

► **Fluorescence activated cell sorting (FACS)** analysis is used to determine the distributions of cell surface proteins and their variations in different cell populations.

For both methods, IHC and FACS, the antibody must be able to detect the protein in its native conformation.

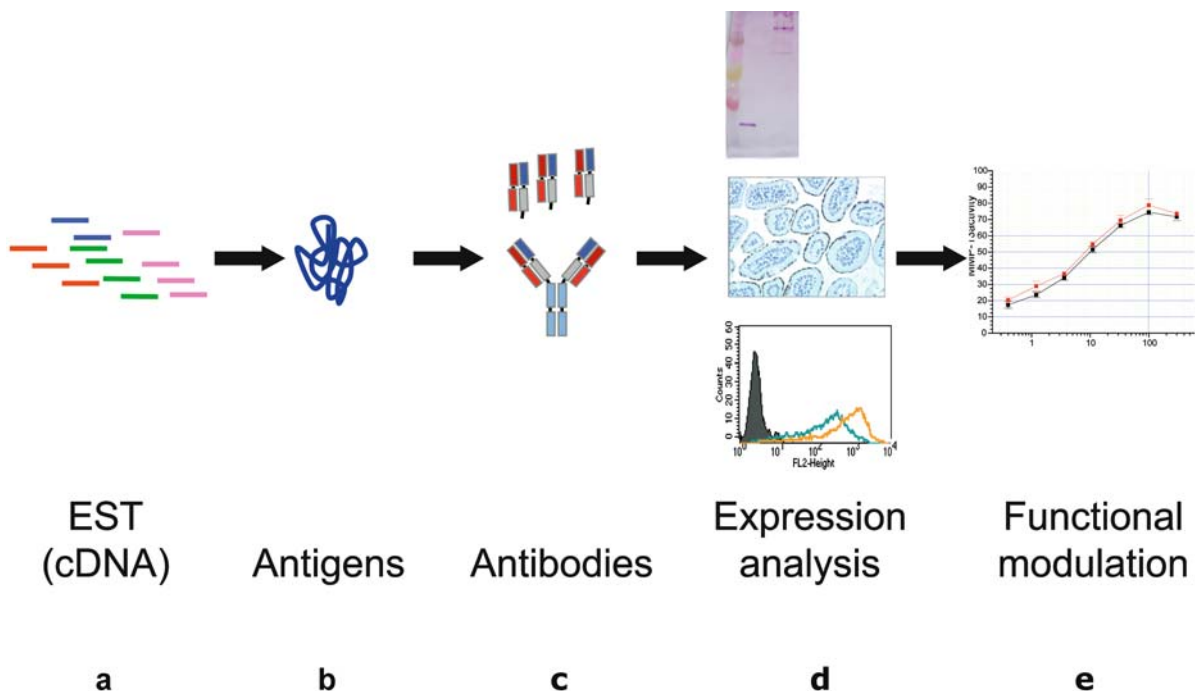
Functional Characterization of the Protein

The binding of an antibody to an antigen can disrupt the functionality of the antigen, thus leading to an understanding of the protein function. Antibodies can neutralize soluble mediators, block the interaction between ligand and receptor, bind to the active site of an enzyme or change the conformation of a biocatalyst reducing its enzymatic activity. Antibodies can also block the binding of inhibitors of a biocatalyst leading to neutralization of enzyme inhibitors. Most critical for all these applications are antibodies that are similar to the natural ligand and hence act as competitors.

Antibodies can further be used to trigger the signaling function of unknown proteins. Binding of the antibody to cell surface receptors can lead to dimerization of the receptor molecules and thus activate cell signaling cascades leading to cell proliferation or cell death. Antibodies can be expressed inside the cell and inhibit protein function intracellularly by interacting with the protein of interest in the native *in vivo* environment (4). However, the difficulty of expressing functional antibodies in the reducing environment of the cytoplasm and nucleus limits the application of this technique.

Traditional Generation of Antibodies

The classical way of producing antibodies was discovered over a hundred years ago when Behring and Kitasato observed the presence of antigen-specific antibodies in the blood of animals (mostly mice) that had been immunized with the antigen. Such an “antiserum” contains a mixture of different antibodies binding to different ► **epitopes** on the antigen. The non-defined epitope binding of the mixture and its



Immunological Methods, the Use of Antibodies to Discover the Function of Novel Gene Products.

Figure 2 Typical workflow for in vitro antibody generation for proteome or target research antibodies: Subsequent to identifying a differentially expressed gene (a) peptides, polypeptide encoded ESTs or purified proteins are produced (b). In multi selection rounds, binders from a diverse pool are enriched and afterwards screened for binding against the antigen used for selection. The format of the antibodies can now be changed to the one (e.g. IgGs or dimeric Fabs) best suited for the consecutive assays (c). To check expression of the protein of interest, Western blot, IHC, ELISA or FACS analysis is used (d) before functional assays are performed (e). Following characterization and validation, human antibodies can enter the optimization procedure for a therapeutic antibody seamlessly.

dependence on the individual animal used for immunization raise a problem with regard to the reproducibility and specificity of the desired antibody. Since the mid 1970's, this problem has been overcome with hybridoma technology. Here, one single epitope-specific antibody-producing B-cell is isolated from the spleen and fused to cancer cells to achieve immortality. Each "hybridoma" cell can then divide and produce a single epitope-specific antibody.

High Throughput Generation of Antibodies

High throughput methods for antibody generation use recombinant antibody libraries with *in vitro* selection methods like ►phage display or ►ribosomal display. Advantages of this approach are the potential availability of antibodies against "self antigens" and the ease of modifying antibodies, i.e. the addition of specific detection and purification tags.

In all cases, the antibodies are directly coupled to the gene sequence. The fusion protein is exposed to the antigen under selective pressure, e.g. binding to the antigen of interest. Non-binding members of the library are removed. Then, the remaining binding members are amplified. Methods such as PCR are used when employing ribosomal display, whereas infection of bacteria is used when using phage display methods. The amplification can be repeated several times, yielding a polyclonal pool of highly specific and high affinity binders. The polyclonal pools can be cloned and clones screened for specific binders e.g. by ELISA. Short cycle times and standardized protocols make *in vitro* methods well suited for automation and high throughput. A typical workflow for antibody generation is shown in Fig. 2 (6, 7).

A major challenge for the high throughput generation of antibodies is the need for sufficient amounts of high quality antigen. One approach for overcoming this obstacle is the use of synthetic peptides derived from the primary structure of the protein of interest as antigen (8). An alternative is the so-called ►HuCAL® EST-technology, a direct expression of the EST sequences in bacterial systems (9). Stretches of 40–400 amino acids derived from cDNA are fused to a His-tagged protein and expressed as inclusion bodies in *E. coli*. After purification and refolding of the inclusion bodies, the antigen can be used for the antibody generation process. Like the peptide-based approach, EST-encoded polypeptides can be used as surrogate substrates in the antibody generation process, yielding binders with good affinities recognizing the full length protein in its denatured form (Western blot), as well binders recognizing the native protein (IHC & FACS) (10).

Clinical Relevance

Rapid elucidation of the function, distribution and pathological relevance of a novel gene product yields data necessary for an early judgment on the potential therapeutic value of a target protein. An intriguing possibility exists that human antibodies may serve as lead compounds in the development of a drug directed against such a target. Antibodies will also be used as biosensors for patient selection and/or for monitoring specific biomarkers to improve diagnosis of human disease.

References

1. Bradbury A, Velappan N, Verzillo V et al (2003) Antibodies in proteomics II: screening, high-throughput characterization and downstream applications. *Trends Biotechnol* 21:312–317
2. Kallioniemi O P, Wagner U, Kononen J et al (2001) Tissue microarray technology for high-throughput molecular profiling of cancer. *Hum Mol Genet* 10:657–662
3. de Wildt RM, Mundy CR, Gorick BD et al (2000) Antibody arrays for high-throughput screening of antibody-antigen interactions. *Nat Biotechnol* 18:989–994
4. Cattaneo A, Biocca S (2003) *Intracellular Antibodies: Developments and Applications*. Landes Bioscience, distributed by Academic Press
5. Bradbury A, Velappan N, Verzillo V et al (2003) Antibodies in proteomics I: generating antibodies. *Trends Biotechnol* 21:275–281
6. Hallborn J, Carlsson R (2002) Automated screening procedure for high-throughput generation of antibody fragments. *Biotechniques Suppl*:30–37
7. Krebs B, Rauchenberger R, Reiffert S et al (2001) High-throughput generation and engineering of recombinant human antibodies. *J Immunol Methods* 254:67–84
8. Van Beijnum JR, Moerkerk PT, Gerbers AJ et al (2002) Target validation for genomics using peptide-specific phage antibodies: a study of five gene products over-expressed in colorectal cancer. *Int J Cancer* 101:118–127
9. Frisch C, Brocks B, Ostendorf R et al (2003) From EST to IHC: human antibody pipeline for target research. *J Immunol Methods* 275:203–212
10. Sun C, Kilburn D, Lukashin A et al (2003) Kirrel2, a novel immunoglobulin superfamily gene expressed primarily in beta cells of the pancreatic islets. *Genomics* 82:130–142

Immunophilins and Cyclophilins

Definition

Immunophilins are a family of soluble receptors capable of binding to one of two major immunosuppressant agents; cyclosporin A (CsA) or FK506.

Proteins that bind to FK506 are termed FK506 Binding Proteins (FKBPs), and those that bind to cyclosporin A are called cyclophilins (CyP). Immunophilins function as cis-trans peptidyl-prolyl isomerases (PPIase), whose activity is inhibited by their respective immunosuppressant compounds. Thus, immunophilins accelerate folding of some proteins both *in vivo* and *in vitro*, by catalyzing slow steps in the initial folding and rearrangement of proline-containing proteins.

► [Nuclear Hormone Receptors: Regulators of Gene Transcription and Cellular Signaling](#)

Immunoprecipitation

Definition

Immunoprecipitation refers to an experimental technique to identify the subunits of oligomeric proteins. It requires a specific antibody of high affinity, which captures its antigen together with any bound proteins from a cell or tissue extract. Protein A of *Staphylococcus aureus* is used to cross-link and precipitate the antibody-protein complexes which are then isolated by centrifugation. Alternatively, Protein A bound to agarose beads is used.

► [Affinity Chromatography and *In Vitro* Binding \(Beads\)](#)

Immunoproteomics

Definition

Immunoproteomics describes a technology to analyse the antigen composition within a pathogens 2-DE pattern. The antigens are detected by reaction with patients sera on a blot of a 2-DE gel.

► [Two-Dimensional Gel Electrophoresis](#)

Immunostaining

Definition

Immunostaining refers to the detection of individual proteins in the cell via the use of specific antibodies that are generally detected by fluorescently labelled secondary antibodies.

► [Immunochemical Methods, Localization](#)

Immunosuppressant

Definition

Immunosuppressant agents suppress the immune system and thereby reduce the body's natural defenses against foreign invaders or materials. Immunosuppressants are used in transplant patients to prevent rejection of transplanted organs.

► [Morbus Wegener](#)

Impedance

Definition

Impedance is a measure for the manner and degree a component resists the flow of electrical current if a given voltage is applied. It is usually denoted by the symbol Z and is measured in Ohm $[\Omega]$.

► [PNA Chips](#)

Importin

Definition

Importins are soluble transport receptors mediating macromolecular transport processes across the nuclear envelope. They recognize their substrates in the cytoplasm by a nuclear localization signal (NLS) and transport them through nuclear pores into the nucleus. In the nucleoplasm, RanGTP binds to importins, inducing the release of import cargoes. For comparison, see ► [exportins](#), which interact with their substrates only in the nucleus in the presence of RanGTP, and release them after GTP hydrolysis in the cytoplasm, causing disassembly of the export complex.

► [Nuclear Import and Export](#)

► [Nuclear Pore Complex](#)

Imprinting

Definition

Imprinting refers to an epigenetic process by which the male and the female germline confers a specific mark

onto certain chromosomal regions, so that only the paternal or the maternal copy of a gene is active in somatic cells.

- ▶ CpG Islands
- ▶ Genomic Imprinting
- ▶ Prader Willi and Angelman Syndromes
- ▶ X-Chromosome Inactivation

In Silico (Procedures)

Definition

In silico means “generated by computer.” *In silico* procedures describe the simulation and prediction of (biological) processes by use of software.

- ▶ Proteomics in Human-Pathogen Interactions
- ▶ SNP Detection and Mass Spectrometry

In Situ Hybridization

Definition

In situ hybridization describes the detection of nucleic acids using radioactive or fluorescently-tagged probes to specific target sequences. Spatial information inherent to the sample is preserved as detection at the precise location where the nucleic acid appears.

- ▶ Single Cell Gene Expression Profiling

In Vitro

Definition

In vitro means outside a living organism, e.g. in a test-tube or culture dish (*in vitro* = in a glass).

- ▶ Heritable Skin Disorders
- ▶ Proteomics in Human-Pathogen Interactions

In Vitro Binding (Beads)

- ▶ Affinity Chromatography and In Vitro Binding (Beads)

In Vitro Fertilization

Definition

In vitro fertilization is a procedure in which an ovum is fertilized by sperm outside the body, and the resulting embryo is implanted in the uterus for gestation.

- ▶ Heritable Skin Disorders

In Vivo

Definition

In Vivo means inside the body of a living organism. Some researchers use it loosely to refer to events happening inside a living cell as opposed to in test tubes.

- ▶ Proteomics in Human-Pathogen Interactions

In Vivo Imaging of Transgenic Mice with Fluorescent Protein Expression

FRITJOF HELMCHEN¹, FRANK KIRCHHOFF²

¹Brain Research Institute, University of Zurich, Switzerland

²Max Planck Institute of Experimental Medicine, Göttingen, Germany
helmchen@hifo.unizh.ch

Definition

Understanding the development of the nervous system, the establishment and functioning of neural circuits, as well as the cure of brain diseases are probably the most exciting questions for a neuroscientist. The generation and use of appropriate animal models provides new vistas not only on brain function, but also on therapeutic approaches of CNS diseases. Recently, genes encoding for ▶fluorescent proteins (FPs) have been cloned from sea animals like jellyfish or corals, and optimized for the expression in mammals. The combination of cell-type specific promoters with FP genes allows the generation of transgenic mice, in which distinct cell populations are selectively labelled, and can be visualized in their complexity within their native environment (1). Many spectral variants of FPs

are now available, so that different cell-types can be co-labelled in different colors. Applied to the nervous system, FP expression in transgenic mice provides a variety of advantages. Living brain cells can be identified in acutely isolated tissue slices by conventional ►fluorescence microscopy, without prior tissue fixation and the need for laborious immunohistochemical staining techniques. Labelled cells can be selected and analyzed by patch-clamp recording, or isolated by ►fluorescence-activated cell-sorting. Since FPs outline the cellular morphology, long-term recordings can reveal structural dynamics of the cells, e.g. cell migration or motility of their processes. Common to all these paradigms is the unequivocal identity of cells under study, neurons, astrocytes, microglia or oligodendrocytes.

In parallel to the establishment of FPs as cellular markers, recent advances in microscopy technology now make it possible to visualize FP-expressing cells in living animals (*in vivo*). In particular, application of two-photon laser-scanning microscopy (2P-LSM) (2) allows fluorescence imaging at high spatial resolution several hundred microns deep in intact tissue, enabling the visualization of structural dynamics and physiological activity in the living brain. In the following, we describe several advantages of the combination of 2P-LSM and transgenic mice with FP expression.

Characteristics

Two-Photon Microscopy

A major breakthrough during recent years have been imaging techniques, which allow the direct observation of the brain in living laboratory animals, particularly in rats, but also in mice. The most common and least invasive technique is functional magnetic resonance imaging (►fMRI). For *in situ* and *in vivo* analysis, the animals are anesthetized and placed into the observation chamber equipped with the appropriate magnets. A major limitation of fMRI is, however, the relatively low spatial resolution (about 150 µm), which prevents resolving individual cells. Compared to fMRI, 2P-LSM provides a more than hundred-fold higher spatial resolution (<1 µm), and therefore is well suited for cellular imaging, in particular in the intact brain (3). 2P-LSM is based on the simultaneous absorption of two infrared excitation photons by a fluorophore, instead of one photon of visible wavelength (2). Following excitation, the molecule fluoresces in the usual way, irrespective of the mode of excitation. As 2-photon absorption depends non-linearly (quadratically) on light intensity, fluorescence is only generated where the photon density is high, and thus is typically confined to the focal volume. These special features give 2P-LSM several advantages over conventional (one-photon) confocal microscopy for biological imaging (4):

- a) a better depth-penetration of infrared excitation photons into the specimen, due to the reduced scattering of infrared light by tissue;
- b) intrinsic optical sectioning and better image contrast, due to the localisation of excitation and the ineffectiveness of scattered infrared photons to generate fluorescence;
- c) reduced out-of-focus photobleaching, permitting long-term observations, and
- d) effective collection of scattered fluorescence from deeper tissue regions.

2P-LSM does, however, require an expensive pulsed laser source, most commonly an externally pumped Ti:Sapphire laser oscillator, which emits laser pulses of ~100 femtosecond (10⁻¹⁵ s) length at a frequency of ~80 MHz, with a tunable wavelength between 750 and 1000 nm. Pulsed laser light is of paramount importance to generate very high peak intensities.

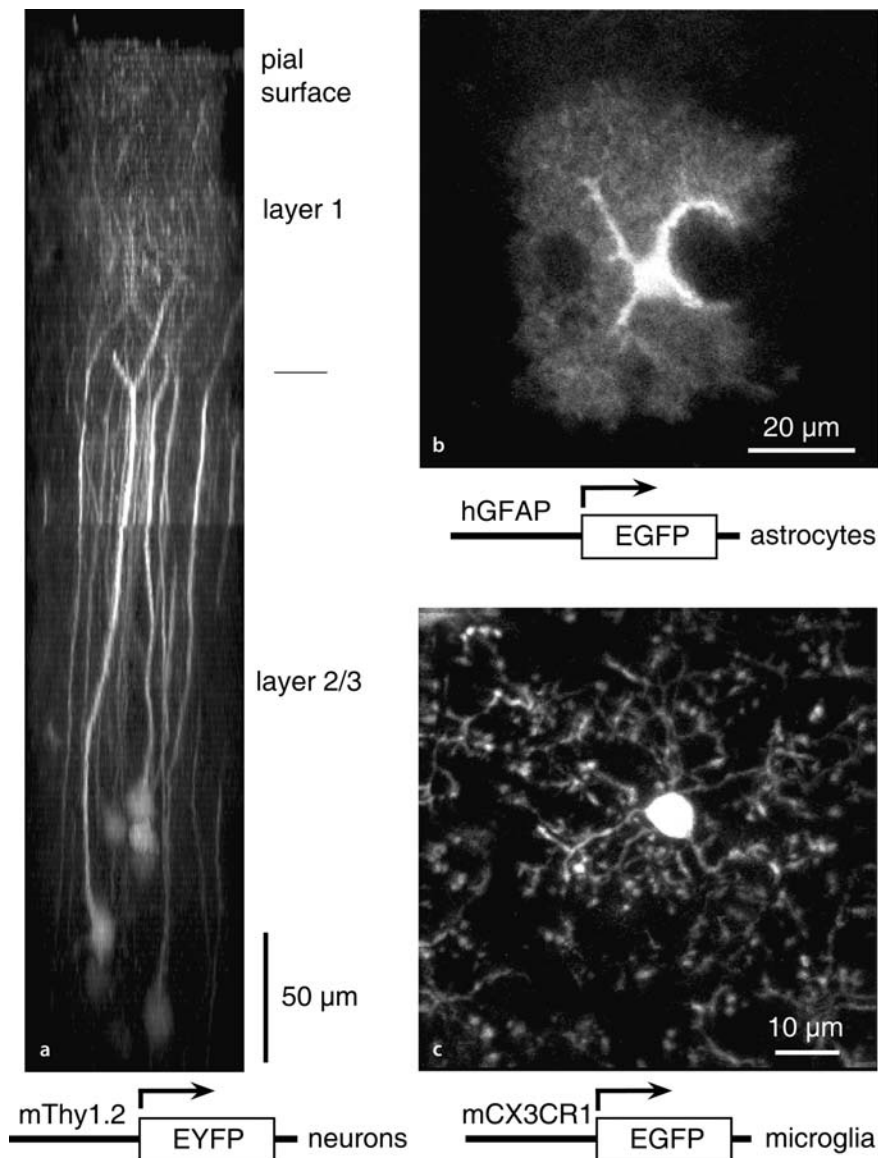
2P-LSM has become the principal technique for high-resolution imaging deep inside intact biological tissues, in particular for structural and functional analysis of cellular dynamics in the brain of living mice. Typically, imaging is performed in anesthetized animals, either through a small craniotomy of a few millimeters diameter (4) or through the thinned skull (5,6). A prerequisite of *in vivo* 2-photon imaging is specific labeling of the structures of interest with a fluorescent marker. Various fluorescent labeling techniques have been introduced, including staining of the blood plasma by tail-vein injection of fluorescent dyes (7) intracellular loading of individual neurons with calcium indicators (8) or bulk-loading of neuronal and glial cell populations with membrane-permeable calcium indicators (8). Alternatively, and perhaps most conveniently, 2P-LSM can be applied to transgenic mice expressing FPs in specific cell types (either neurons or glial cells). These animals are particularly well suited to investigate structural changes of cells and cellular processes both during development and adulthood. By multiple repeated imaging of the same cells in the same animal structural dynamics can be followed on various time scales ranging from minutes to several months (1,6).

Transgenic Mice with Fluorescent Protein Expression

Transgenic mice with cell-type specific FP expression represent an important advancement for the analysis *in situ* and *in vivo*. Cell-type specific promoters are used to target FPs with different spectral properties into cells of interest (Fig. 1). For example, in one mouse line a Thy1.2 promoter induces expression in a variety of neurones in several brain regions including cortex and hippocampus (9); in other mouse lines the human glial fibrillary acidic protein ►promoter causes expression in ►astrocytes (10) and the mouse proteolipid protein

promoter in ►**oligodendrocytes** (11). In addition, by using a knock-in strategy to place the enhanced green fluorescent protein (EGFP) gene into the fractalkine receptor (CX3CR1) locus, a mouse line was generated in which virtually all ►**microglia** cells of the brain parenchyma were fluorescently labelled (12).

These and similar transgenic mouse lines are currently being used to investigate whether cell morphologies are stable or plastic in adult animals, and under what conditions they might change. Imaging of neuronal dendrites and axons showed that overall neuronal morphologies were stable, but that synaptic structures



***In Vivo* Imaging of Transgenic Mice with Fluorescent Protein Expression.** **Figure 1** Cell-type specific *in vivo* imaging of brain cells in transgenic mice with fluorescent protein expression. All images were acquired using 2P-LSM. The cartoons below the images depict the combination of cell-type specific promoter and fluorescent protein gene to generate the genetically modified mouse lines. (a) Pyramidal cells in the mouse neocortex expressing the yellow fluorescent protein YFP under control of the mouse Thy1.2-promoter. The image is a side-projection of a stack of fluorescent images acquired from about 400 µm below the brain surface up to the pial surface. (b) A single astrocyte enwrapping a brain capillary. 2P-LSM recording in the cortex of a TgN(hGFAP-EGFP) mouse. (c) Visualization of a resting microglia cell with its numerous branched processes sensing the parenchymal environment in the cortex of a TgH(CXC3R1-EGFP) mouse.

(dendritic spines and presynaptic endings) were turned over at a certain rate, indicating plastic (possibly experience-dependent) changes of neuronal connectivity. Similarly, astrocytes appear to preserve their gross morphology, at least over the time course of weeks. In marked contrast, recent findings demonstrate that microglial cells display an extraordinary high level of structural dynamics. The cell processes of these immune cells of the brain are highly motile, apparently continually surveying the brain tissue (6). Consistent with a protective function, the motile processes immediately targeted damaged areas, e.g. shielding local disruptions of blood vessels.

Another application of 2P-LSM for the analysis of transgenic mice is the visualization of amyloid plaques in a transgenic mouse model of [▶Alzheimer's Disease](#) (5). To this end, a fluorescent [▶thioflavin T](#) derivative was injected into the tail vein of an anesthetized mouse. After 20 min. the dye had crossed the blood-brain barrier, and had exclusively stained amyloid plaques at the vessels and within the brain parenchyma. This method permits the study of plaque growth and disease progression, as well as the analysis of potential treatments (13). Crossbreeding of these mice with other mice showing specific FP expression may also permit detailed analysis of the involvement of surrounding cells.

In summary, during the past years, the fields and numbers of applications of 2P-LSM in living animals have expanded enormously, including *in vivo* studies on brain tissue, but also on skin, tumors, lymph nodes and kidney. The combination of 2P-LSM with transgenic mice expressing FPs has proven particularly well suited for studies of short- and long-term morphological changes during normal and pathological conditions.

References

1. Young P, Feng G (2004) Labeling neurons in vivo for morphological and functional studies. *Curr Opin Neurobiol* 14:642–646
2. Denk W, Strickler JH, Webb WW (1990) Two-photon laser scanning fluorescence microscopy. *Science* 248:73–76.
3. Svoboda K, Denk W, Kleinfeld D et al (1997) *In vivo* dendritic calcium dynamics in neocortical pyramidal neurons. *Nature* 385:161–168
4. Zipfel WR, Williams RM, Webb WW (2004) Nonlinear magic: multiphoton microscopy in the biosciences. *Nat Biotech* 21:1369–1377
5. Christie RH, Bacskai BJ, Zipfel WR et al (2001) Growth arrest of individual senile plaques in a model of Alzheimer's disease observed by *in vivo* multiphoton microscopy. *J Neurosci* 21:858–864
6. Nimmerjahn A, Kirchhoff F, Helmchen F (2005) Microglial cells are highly dynamic surveillants of brain parenchyma in vivo. *Science* 308:1314–1318
7. Kleinfeld D, Mitra PP, Helmchen F et al (1998) Fluctuations and stimulus-induced changes in blood flow observed in individual capillaries in layers 2 through 4 of rat neocortex. *Proc Natl Acad Sci* 95:15741–15746
8. Stosiek C, Garaschuk O, Holthoff K et al (2003) *In vivo* two-photon calcium imaging of neuronal networks. *Proc Natl Acad Sci* 100:7319–7324
9. Feng G, Mellor RH, Bernstein M et al (2000) Imaging neuronal subsets in transgenic mice expressing multiple spectral variants of GFP. *J Neurosci* 28:41–51
10. Nolte C, Matyash M, Pivneva T et al (2001) GFAP promoter-controlled EGFP-expressing transgenic mice: a tool to visualize astrocytes and astrogliosis in living brain tissue. *Glia* 33:72–86
11. Fuss B, Mallon B, Phan T et al (2000) Purification and analysis of *in vivo*-differentiated oligodendrocytes expressing the green fluorescent protein. *Dev Biol* 218:259–274
12. Jung S, Aliberti J, Graemmel P et al (2000) Analysis of fractalkine receptor CX(3)CR1 function by targeted deletion and green fluorescent protein reporter gene insertion. *Mol Cell Biol* 20:4106–4114
13. Bacskai BJ, Hickey GA, Skoch J et al (2003) Four-dimensional multiphoton imaging of brain entry, amyloid binding, and clearance of an amyloid-beta ligand in transgenic mice. *Proc Natl Acad Sci* 100:12462–12467

Inbred Strain/Inbreeding

Definition

An inbred strain is one that has been maintained by sibling (sister x brother) matings for 20 or more consecutive generations (inbreeding). Except for the sex difference, individuals of an inbred strain are as genetically alike as possible, being homozygous at virtually all of their loci.

▶[Genetic Epidemiology](#)

▶[Mutagenesis Approaches in Medeka](#)

Incidence

Definition

Incidence is the frequency of the occurrence of a condition in a population over a period of time.

▶[Crohn's Disease](#)

▶[Duchenne Muscular Dystrophy](#)

Inclusion Body

Definition

An inclusion body is a proteinaceous structure in the cytoplasm or the cellular nucleus demonstrable by histopathological methods. Often, cytosolic overexpression of recombinant proteins leads to formation of inclusion bodies.

- [Protein Tags](#)
- [Repeat Expansion Disorders](#)

Indifferent Embryonic Gonad

Definition

The indifferent embryonic gonad is the undifferentiated fetal sex gland, which may become testis or ovary, depending on the presence or absence of the Y chromosome.

- [SRY - Sex Reversal](#)

Indirect Readout

Definition

Indirect readout describes the recognition of DNA through sequence-dependent DNA deformability, which does not involve the formation of direct base contacts.

- [Protein/DNA Interaction](#)

Induced Fit

Definition

When a ligand or a substrate comes into contact with the binding site of a protein or the enzyme's active site, the active site slightly changes or moulds itself around the substrate/ligand for an effective fit. In some cases only few atoms move, sometimes whole secondary structures move, leading to conformational changes of the protein. In other cases the quaternary structure is changed. The conformational changes are often used to transport

information. Other synonymous terms for induced fit are structural plasticity and conformational changes.

- [Classification of Active Centers](#)

Inducible Expression

Definition

Inducible expression describes the expression of a gene in an organism which is dependent on the presence of specific inducer (e.g. a metabolite) that allows to trigger expression by experimental manipulation.

- [Recombinant Protein Production in Mammalian Cell Culture](#)

Ineffective Erythropoiesis

Definition

Ineffective erythropoiesis refers to the imperfect formation of red blood cells in the marrow, with destruction within the marrow of many of the defective cells that are produced.

- [Hemochromatosis](#)

Infantile Refsum Disease

Definition

Infantile Refsum Disease (IRD) is a peroxisome biogenesis disorder, with clinical signs and symptoms being less severe as compared to ► [NALD](#) and ► [Zellweger syndrome](#) (ZS).

- [Peroxisomal Disorders](#)

Infantile Spasm

Definition

Infantile spasm is a seizure that occurs in infants, characterized by either extension or flexion posturing of brief duration.

- [Tuberous Sclerosis](#)

Infantile Spinal Muscular Atrophy

► Spinal Muscular Atrophy

Inflammation

Definition

Inflammation is a series of reactions, which bring cells and molecules of the immune system to sites of infections and physical injury. This appears as an increase in blood supply, increased vascular permeability resulting in local accumulation of fluid and plasma proteins, and increased transendothelial migration of leucocytes.

► [Inflammatory Response](#)

► [Interleukins and Their Receptors, a Role in Regulating Hematopoietic Stem Cells](#)

Inflammatory Response

AHMED SHERIFF, MARTIN HERRMANN,
REINHARD E. VOLL, UDO S. GAJPL,
JOACHIM R. KALDEN

Institute of Clinical Immunology and Rheumatology,
Department of Internal Medicine III,
Friedrich-Alexander University of
Erlangen-Nuernberg, Erlangen, Germany
sheriff@genethor.com

Definition

What is an inflammatory response? ► [Inflammation](#) is a general term for the local accumulation of fluid, plasma proteins and white blood cells that is initiated by physical injury, infection or a local immune response. More extensive tissue damage or infection may result in additional systemic reactions including fever and induction of certain genes mounting the acute phase response. These largely uniform responses of higher multicellular organisms to various injuries/infections are also known as the inflammatory response. The cells that invade tissues undergoing inflammatory responses are often called inflammatory cells or inflammatory infiltrates.

Acute inflammation is the term used to describe early and often transient episodes, while chronic inflammation occurs if the infection persists or during atherosclerosis or autoimmune responses. Many different forms of inflammation are seen in different diseases.

Characteristics

► Innate Immunity and Inflammation

Barriers keep harmful materials from entering your body. Some of these barriers are the skin, stomach acid, mucus (traps microorganisms and small particles), the cough reflex and enzymes in tears and skin oils. If a pathogen gets past the external barriers, it is attacked and destroyed by the immune system. Host defense mechanisms of the innate immune system represent the first immunological protection line.

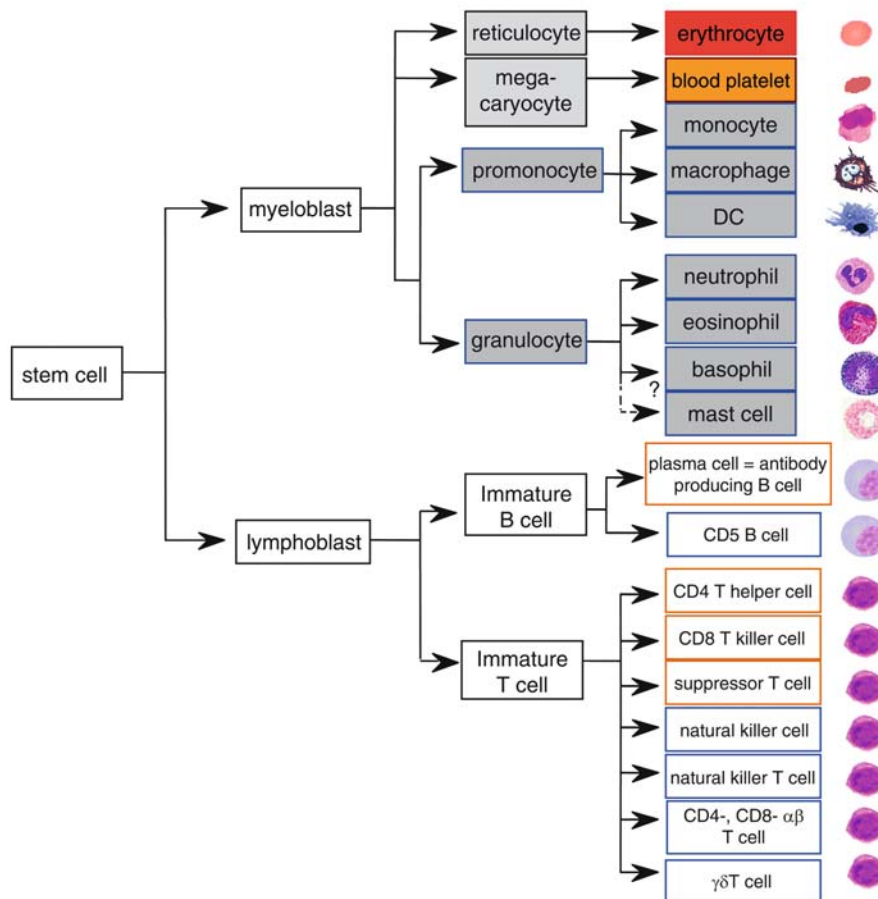
The immune system consists of certain types of white blood cells. It also includes biochemicals and proteins in the blood such as complement proteins, interferons and cytokines. Some of these directly attack foreign substances in the body and others work together to help the immune system.

The inflammatory response (inflammation) is part of innate immunity. It occurs when tissues are injured by bacteria, trauma, toxins, heat or any other cause. Proinflammatory mediators including histamine, bradykinin, serotonin and others are released by damaged tissue. These chemicals cause blood vessels to leak fluid into the tissues, resulting in localized swelling. This helps to isolate the invading substance or pathogen from further contact with tissues.

Certain proinflammatory mediators also attract white blood cells that “eat” microorganisms and dead or damaged cells. The process whereby these white blood cells surround, engulf, and destroy foreign substances is called ► [phagocytosis](#) and the cells are collectively referred to as phagocytes. Some of the phagocytes eventually die. Pus is formed from a collection of dead tissue, dead bacteria and live and dead phagocytes.

But not everything that enters the body is harmful, as can easily be understood by the example of food. So, in this line of defense, differentiation of pathogens from non-pathogens is crucial. The first line of defense is always directed by the innate immune system (Fig. 1). The innate immune system expresses a large amount of so called “pattern recognition receptors” which detect basic pathogen molecules or patterns (pathogen-associated molecular pattern [► [PAMP](#)]; Table 1). Only the activation of one or more of these pattern recognition receptors activates the second line of defense, which is performed by the adaptive immune system consisting of T and B cells (Fig. 1).

Initially, the adaptive immune system generates its receptors randomly. So, it does not know what it will bind. This leads to enhanced flexibility but also the



Inflammatory Response. Figure 1 The tree of blood cell types starting from hematopoietic stem cells. Stem cells are multipotent and may differentiate into each cell type in the figure. A strict borderline is demarked after differentiation either to lymphoblasts or myeloblasts. No way back! All adaptive immune cells (brown lining) are derived from lymphoblasts. Cells belonging to the innate immune system (blue lining) are from both branches, lymphoblasts and myeloblasts.

possibility that it might bind self-structures is high. Therefore, T and B cells undergo stringent selection processes to avoid unwanted destruction of self. If these selection processes fail, ►[autoimmune diseases](#) or allergies prevail.

Much is known about the first line of selection, especially about T cell selection, whereas less is known about B cell selection. The immunoreceptor selection processes and the whole of ►[adaptive immunity](#) are not covered by this essay. In the second line of selection processes, the innate immune system governs the fate of adaptive immunity by means of its pattern recognition mechanisms.

There are many ways to die. Cells, however, usually meet their destiny in one of two ways – death by ►[necrosis](#) or death by ►[apoptosis](#).

The decision to initiate an inflammatory response also has to be made for dying cells, which arise by hundreds

of billions each day. They carry valuable information with respect to the possible influence of pathogens. Dying cells are usually phagocytosed by specialized macrophages (tingeable body macrophages, Kupffer cells, alveolar macrophages, tissue macrophages) or less frequently dendritic cells (DC). If the differentiation between apoptosis and necrosis fails, for instance when apoptotic cells are not swiftly phagocytosed and, therefore, reach a state of secondary necrosis, autoimmune reactions against components of the lysed cells may arise. Therefore, impaired engulfment of dead cells is supposed to play an important role in the etiopathogenesis of autoimmune diseases such as ►[systemic lupus erythematosus](#) (SLE).

Necrosis

Necrosis can be viewed as a sort of violent cell death. Cells that may not have yet reached their full life span

Inflammatory Response. Table 1 Selected innate recognition receptors

Receptor	Ligand (innate response)
TLR 1 - 10	LPS, flagellin, CpG DNA, dsRNA, and many others
CD14	LPS, peptidoglycan
Complement receptor 3 (Mac-1, integrin $\alpha_M\beta_2$, CD11b-CD18)	iC3b-opsonized particles
Complement receptor 2 (CD21)	iC3b-opsonized particles
Fc-receptor (FcR)	"Natural antibodies"
Scavenger receptor AI/II	LPS, lipoteichoic acid, bacteria
MARCO	Bacteria
Integrin $\alpha_v\beta_3$ (macrophages), integrin $\alpha_v\beta_5$ (DC)	Fibronectin
Dectin-1	β -glucan
Mannose receptor	Mannan
DC-SIGN	ManLAM
FMLP receptor	N-formylated bacterial peptides

are hit by an external noxa that interrupts some of their vital functions or disrupts their physical integrity and the intracellular contents are spilled into the micro-environment. This process is often due to the influence of a pathogen or mechanical or oxidative stress.

Cells do not die violent deaths unnoticed. The intracellular fluids, proteins and organelles released into the extracellular spaces exert a proinflammatory function, attracting inflammatory cells and cytokines from the nearby vessels and tissues. What ensues is an inflammatory reaction that may be more or less damaging to the tissues.

The Apoptotic Program

Apoptosis is seen as a form of graceful death. The name itself suggests it – it is the ancient Greek term for the falling of leaves. Once the internal program is activated, the cell shuts down its internal networks and activates a series of enzymatic reactions that lead to an autoprolytic breakdown. Enzymes, proteins and DNA are cleaved internally in a "silent" fashion. In contrast to necrotic cells, apoptotic cells maintain their membrane integrity for a relatively long time, thereby

preventing the release of intracellular components that may directly damage the tissue or may induce inflammatory or immune responses. In addition, apoptotic cells undergo very early membrane changes to ensure immediate recognition and timely engulfment of the dying cells prior to secondary necrosis and cells lysis (see below).

Apoptosis plays a critical role not only during embryogenesis, pregnancy, delivery and development, when the rate of tissue modeling and remodeling is at its highest. Normal tissue homeostasis relies substantially on removal of old cells by apoptosis (eg granulocytes, gut epithelium) to allow new, young cells to take their place. Apoptotic check points are also critically involved in preventing overgrowth, once adult tissues start proliferating to repair a cell loss or extensive cell damage.

Apoptotic and Necrotic Signals to Phagocytes

Cells undergoing apoptosis subtly signal their demise to the extracellular environment. They expose "eat me" signals and bind a variety of adaptor molecules, features important for recognition and uptake by specialized phagocytes (Table 2). The receptors are highly redundant and are ordered hierarchically, reflecting the necessity for fine-tuning of the recognition process. Some recognize apoptotic cells in the early phase of cell death, others rather represent backup systems. Some recognition pathways differentiate among apoptosis, necrosis and cellular debris, while others do not. A variety of specialized phagocytes have developed in humans due to organ-specific requirements. The macropinocytosis of apoptotic cells by macrophages triggers the production of **transforming growth factor β** , a cytokine that suppresses inflammatory processes. When monocytes or immature DC encounter apoptotic cells they are triggered by CD36 to produce IL-10, which is also an anti-inflammatory cytokine. In addition to the enhanced secretion of anti-inflammatory cytokines, the production of proinflammatory cytokines such as TNF α , IL-1 and IL-12 is markedly inhibited in the presence of apoptotic cells. Phospholipids are asymmetrically distributed among the inner and outer leaflets of the intact cell membrane; the outer leaflet primarily contains sphingomyelin and phosphatidylcholine (PC), whereas the inner leaflet harbors most of the phosphatidylserine (PS) and phosphatidylethanolamine (PE). Maintenance of this asymmetry is energy dependent and requires adenosine triphosphate (ATP). In the absence of ATP as in necrosis, the phospholipids of the inner and outer leaflets undergo exchange (so-called 'flip-flop'). Another condition that induces a membrane flip-flop is apoptosis. Phosphatidylserine can be recognized by phagocytes, mostly macrophages, leading to quick

Inflammatory Response. Table 2

Receptor		Ligand and adaptor (innate response)
Phosphatidylserine receptor (PSR) => triggers the production of transforming growth factor β		Phosphatidylserine (PS)
Integrins	CR3, CR4	iC3b
	$\alpha_v\beta_3$ / CD36	PS , Tsp
	$\alpha_v\beta_5$	
	$\alpha_3\beta_5$	MFG-E8
Collectin receptors		C1q, C3, C4, MBL, SP-A, SP-D
CD14		PS , ICAM-3
CD21/35		Complement
ABC1		PS
β_2 -GPI receptor		PS , β_2 -GPI
Scavenger receptors	CD36	PS , oxLDL
	SR-A	PS , oxPL
	CD68	PS , oxLDL
	LOX-1	PS
	SR-B1	PS
Lectins		Sugar moieties, altered carbohydrates
FcγR		IgG, CRP, SAP, PTX3, HRG; Anti-S,-M,-L, oxPL, CRP
c-mer		Gas-6
CD91 / calreticulin		Calreticulin, collectins, heat shock proteins
gas-6		PS
asialoglycoprotein receptor		Glycoproteins
CD31-R (PECAM-1-R)		CD31 (PECAM-1)

removal by phagocytosis. The recognition of PS by the PS receptor (PSR) is of major importance for phagocytosis of early apoptotic cells.

The Last Order of an Apoptotic Cell: “Anti-Inflammation, Please”

Cells dying by apoptosis are not immunogenic because they are self-material that is not subject to pathogen attack. It must be finely differentiated from necrosis by apoptotic cell-associated molecular patterns (►ACAMP). The mechanisms of this differentiation are a matter of intense discussion.

In contrast to apoptotic cells, necrotic cells initiate inflammatory responses and can mediate maturation of DC. As described above, the fine-tuning of responses against dying cells is of major importance. Even primary and secondary necroses involve the display of

different inflammatory signals. Primary necrotic cells release the inflammatory high mobility group B1 (HMGB1) protein that is “frozen” on the chromatin of apoptotic cells and remains immobilized even under conditions of secondary necrosis. This might be considered one reasonable mechanism of fine-tuning. In addition, the release of heat shock proteins, substances of mitochondrial origin (which still display certain bacterial features), and other molecules appear to contribute to the proinflammatory response.

Secretory Phospholipase and ►C-Reactive Protein in Apoptosis and Necrosis

In humans, another relevant secretory protein that strongly binds PS floats freely in the serum. Its name is secretory phospholipase A2 (sPLA2) IIA. sPLA2 IIA is not able to hydrolyze the phospholipids of the outer

membrane leaflet of normal intact cells efficiently. Thus, sPLA2 IIA is only able to react with phospholipids in the outer leaflet of normal cells if they have undergone a flip-flop from the inner leaflet as in apoptosis or necrosis. After interaction with sPLA2 IIA, cells are left with an increased proportion of lysophospholipids (phospholipids that have lost free fatty acids at the *sn*-2-ester bond) like lysophosphatidylcholine (lyso-PC) in the outer membrane leaflet. This modification disturbs the packing of the phospholipids and generates binding sites for the pentraxin ►C-reactive protein (CRP) in the outer leaflet.

CRP is capable of binding to cells provided that they contain a substantial amount of lyso-PC in the outer leaflet of their membranes. Thus, in inflamed tissues, it is proposed that binding of CRP initially occurs to cells containing a significant proportion of lysophospholipids in the outer leaflet of their membranes due to the activity of sPLA2 IIA. Once bound, CRP induces complement activation *via* the classical pathway, which in turn triggers the influx of neutrophils, decorates the surface of the ligand with opsonising complement fragments and enhances phagocytosis of the cells that have bound CRP and complement. CRP also interacts with Fc receptors on phagocytic cells and acts as an opsonin. Because the occurrence of lyso-PC is dependent on the exposure of PS, CRP, in this chronological view, efficiently decorates late apoptotic cells.

In addition to the membrane of intact injured cells, CRP also binds to membranes and nuclear constituents of necrotic cells. Several nuclear constituents, including histones, small nuclear ribonucleoproteins and ribonucleoprotein particles have been shown to bind CRP in a calcium-dependent fashion. Deposition of CRP onto nuclei of necrotic cells at sites of inflammation has been observed, while CRP does not cross the plasma membrane of apoptotic cells. This might be considered a second reasonable mechanism for fine-tuning the differentiation between apoptosis and necrosis.

The ►Acute Phase Reaction: Secretory Phospholipase and C-Reactive Protein in Infection

CRP is an acute-phase protein featuring a homopentameric structure and Ca-binding specificity for phosphocholine and lyso-PC. The wide distribution of phosphocholine in polysaccharides (C-polysaccharide) of pathogens and in cellular membranes allows CRP to recognize a range of pathogenic targets as well as membranes of damaged and necrotic host cells. Thus, the main biological function of CRP appears to be host defense against bacterial pathogens and clearance of apoptotic and necrotic cells.

The acute phase response (APR) of inflammation is a well-known clinical phenomenon characterized by leukocytosis, fever, alterations in the metabolism of many organs and changes in the plasma concentrations of various so-called acute phase proteins (APPs). The prototypal APP in humans is CRP. Plasma levels of this APP start to rise as early as about 8–10 h after challenge with the inciting stimulus. In addition, circulating levels of another APP, the enzyme sPLA2 IIA start to rise approximately 6–8 h after challenge.

CRP is widely used as a biomarker of systemic inflammation due to the fact that its concentration correlates considerably with the inflammatory state. Examples are infection, sepsis, severe wounds and vast areas of necrotic cells as after myocardial infarction or stroke. The swift rise of its serum concentration during the acute phase, the magnitude of the response approaching 1000-fold increase within 24–48 h, and the equally quick return to the very low normal concentration of a few µg/ml are the most impressive biological characteristics of CRP as an inflammation marker. CRP-synthesis within the liver accounts for its blood concentration. Hepatic synthesis of CRP occurs upon stimulation by inflammatory cytokines (IL-1β, IL-6) produced in response to tissue injury. In an analogous fashion, inflammation induces a 10–100-fold increase in circulating sPLA2 IIA, starting within 3–4 h and reaching peak levels after approximately 10 h.

Transcriptional Regulation of the Inflammatory Response

Multiple cell types and genes are involved in the complex inflammatory process, whereas only a few transcription factors govern the transcriptional regulation of the inflammatory response. The promoter/enhancer sequences of most genes that are involved in inflammatory and immune responses contain binding sites for ►NF-κB, activating protein 1 (AP-1), ETS-domain transcription factors and CAAT/enhancer-binding proteins (C/EBP). Binding of at least one, often several, of these transcription factors is usually required for the efficient transcription of most genes involved in the inflammatory response. Among these transcription factors, NF-κB appears to play a key role, since it can be rapidly activated and is essential for the efficient expression of many inflammatory genes. Therefore, NF-κB can be regarded as a main switch of the inflammatory response and host defence. Pharmacological inhibition of NF-κB might represent a powerful therapeutic target in inflammatory diseases.

Concluding remark

As a basic view, an inflammatory response is governed by paranoia. This paranoia is two-fold.

Paranoia 1 is the well founded suspicion that an enemy might enter or might have entered the body.

Paranoia 2 is that adaptive immunity will accidentally react to self and so raise an autoimmune disease. This is also well founded and the reaction of the immune system to this fact is that it uses the differentiation between friend and foe to raise, upon contact with self, anti-inflammatory signals. These signals are transported by DC to T cells and establish a cohort of regulatory or suppressor T cells to gain tolerance for self.

References

1. Gaip US, Brunner J, Beyer TD et al (2003) Disposal of Dying Cells: A Balancing Act Between Infection and Autoimmunity. *Arthritis Rheum* 48:6–11
2. Gershov D, Kim SJ, Brot N et al (2000) C-Reactive Protein Binds to Apoptotic Cells, Protects the Cells from Assembly of the Terminal Complement Components, and Sustains an Antiinflammatory Innate Immune Response: Implications for Systemic Autoimmunity. *J Exp Med* 192:1353–1363
3. Hack CE, Wolbink G-J, Schalkwijk C et al (1997) A role for secretory phospholipase A2 and C-reactive protein in the removal of injured cells. *Immunol Today* 18:111–115
4. Janeway Jr, Travers P (2001) *Immunobiology: The Immune System in Health and Disease*. Garland Publications, New York
5. Kracht M, Saklatvala J (2002) Transcriptional and post-transcriptional control of gene expression in inflammation. *Cytokine* 20:91–106
6. Volanakis JE (2001) Human C-reactive protein: expression, structure, and function. *Mol Immunol* 38:189–197

Informational Property Right

Definition

Informational property right is a legally protected personal right, similar to other property rights; individuals may share properties with others for personal benefit and within legal or regulatory frameworks or contract.

► [Ethical Issues in Medical Genetics](#)

Informed Consent

Definition

Informed consent of patients and probands is an essential moral, regulatory, and legal prerequisite for medical examination and treatment, as well as in

clinical research; researchers are morally and legally obligated to inform adequately and to seek consent without pressure or promises; patients and probands must be given the right to ask for termination of treatment or involvement in research.

► [Ethical Issues in Medical Genetics](#)

Informed Contract

Definition

Informed contract between citizens seeking health care guidance, medical treatment or volunteering in clinical research, and health care professionals, scientists, and their sponsors, outlines the rights and obligations of stakeholders involved.

► [Ethical Issues in Medical Genetics](#)

In-Frame Deletion

Definition

An in-frame deletion is a mutation in the coding region of a gene, resulting in the loss of normal DNA sequence, which does not alter the normal triplet reading frame between the mutation site and the carboxyterminus of the polypeptide.

► [Heritable Skin Disorders](#)

Inherited Mental Retardation Syndromes

NIELS TOMMERUP

Department of Medical Biochemistry and Genetics, Wilhelm Johannsen Center for Functional Genome Research, University of Copenhagen, Copenhagen, Denmark
tommerup@imbg.ku.dk

Definition

Mental retardation (MR) is defined as “significant subaverage intellectual functioning and limitation in adaptive skills with onset before 18 years”. In relation

to the intelligence quotient (IQ), MR is defined as IQ below 70, which is 2 standard deviations below the mean IQ of 100. MR can be divided into moderate to severe MR (IQ <50) with a prevalence of 0.3–0.5%, and mild forms (IQ = 50–70), affecting an estimated % 1–2% of the population.

Genetic factors, caused by visible chromosomal defects, cryptic genomic alterations or gene mutations may underline most MR. MR can occur as an isolated trait (non-►**syndromic** MR) or in association with other features (syndromic MR). Both types may be inherited. However, the distinction between syndromic and non-syndromic MR is only clear-cut in a few cases. Once an apparently non-syndromic form is defined by a specific molecular defect, it often leads to the detection of additional clinical features and hence to a new syndromic form of MR. Furthermore, MR has common genetic origins with other symptoms of brain dysfunction, e.g. autism and epilepsy.

Characteristics

How Many Syndromes with MR, How Many Genes in MR?

There is an extreme ►**genetic heterogeneity** of MR. There are more than 900 registered entities associated with MR (1), and loss or gain of almost any segment of the genome can lead to MR, indicating that a proper dosage of hundreds to thousands of genes is needed for normal brain development, maturation and function.

Etiologies

Chromosomal and Genomic Disorders

Robertsonian Translocations

►**Down syndrome** (trisomy 21), the most common cause of MR is usually sporadic, caused by an extra, free chromosome 21 due to ►**chromosomal non-disjunction**, with a strong association with elevated maternal age. However, in ~4% of the cases, the underlying cause is an unbalanced Robertsonian translocation involving chromosome 21 and another acrocentric chromosome, most frequently a 14;21-translocation. Clinically, there is no difference whether the extra chromosome 21 is a free one or is part of a Robertsonian translocation. In ~40% of the cases the translocation is inherited from a balanced carrier parent, usually the mother. The risk of carriers of Robertsonian translocations having a Down syndrome child varies with gender; males have only a slightly increased risk, whereas it ranges from 10–20% in female carriers. However, if the Robertsonian translocation involves both homologous chromosomes 21, then the carrier cannot produce normal gametes and can only expect to have live born children with Down syndrome.

Reciprocal Translocations

Carriers of balanced reciprocal ►**translocations** have a high risk of producing unbalanced gametes due to aberrant segregation of the derivative translocation chromosomes during meiosis. The clinical symptoms associated with unbalanced reciprocal translocations vary, depending on the chromosomal segments involved, but MR is the most frequent single trait, often associated with multiple congenital anomalies (MCA). MR may occur sporadically in translocation families, but the mode of inheritance may occasionally mimic either an autosomal dominant (with reduced penetrance) or an autosomal recessive trait. Since the same translocation may lead to different types of imbalances in the gametes, two or more MR/MCA syndromes may occur in the same family.

An increased frequency of apparently balanced *de novo* reciprocal translocations are seen in association with mental retardation. One reason for this is, that some of these breakpoints disrupt genes where ►**haploinsufficiency** results in abnormal brain development. Another explanation is that some of these apparently balanced translocations may turn up to be associated with microdeletions/-duplications (see below).

Submicroscopic Genomic Disorders

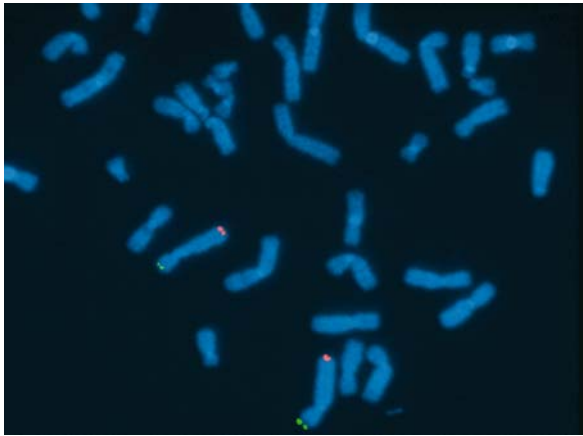
Technical improvements, especially in fluorescence *in situ* hybridization (►**FISH**) and lately array based Comparative Genomic Hybridization (CGH) have resulted in the description of a growing number of cryptic translocations and inversions, microdeletions and duplications which may be responsible for inherited MR syndromes.

Subtelomeric Cryptic Translocations

The chromosomal subtelomeric regions are gene rich and associated with a high meiotic recombination. This recombinatorial property also involves illegitimate recombination and subtelomeric deletions or duplications can be detected by FISH in 5–8% of patients with idiopathic MR (2). The subtelomeric imbalance can often be traced back to a parental carrier of a cryptic balanced reciprocal translocation, with an associated recurrence risk. Some of the subtelomeric deletions, e.g. involving chromosome 1pter, 22qter and 17pter can be recognized as specific MR/MCA syndromes, and screening for subtelomeric integrity (e.g. by FISH, Fig. 1) is now a routine procedure in MR/MCA syndromes.

Micro-Deletion and -Duplication Disorders

Almost 5% of the human genome is composed of duplicated sequences that share more than 95% identity at sequence level. These segmental duplications may facilitate recurrent deletion/duplications of the



Inherited Mental Retardation Syndromes.

Figure 1 FISH with a probe for the subtelomeric region on the short arm of chromosome 4 (4p) labelled green and a probe for the subtelomeric region on the long arm of chromosome 4 (4q) labelled red. By systematic hybridization of probes specific for all 43 unique subtelomeric regions, minute deletions, duplications and translocations involving these regions can be revealed. Courtesy Isabel Carreira, Coimbra.

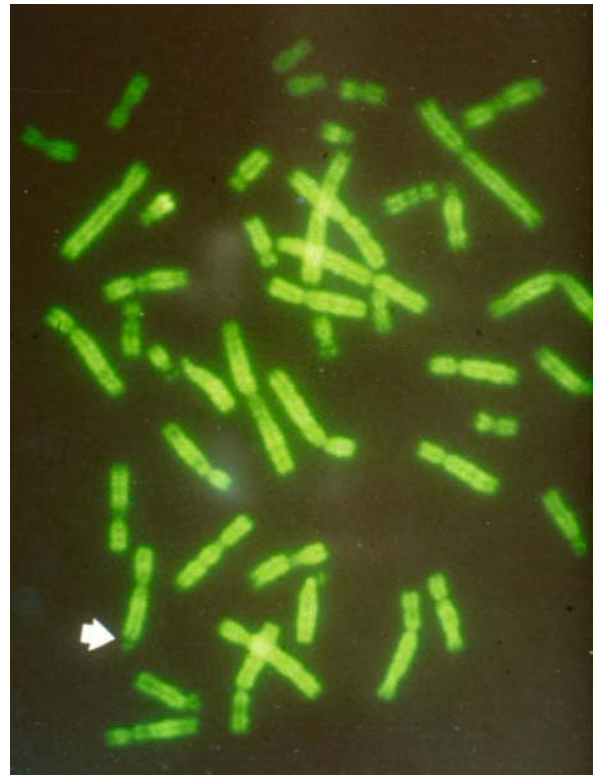
intervening sequences by non-homologous crossing-over. The resulting clinical symptoms, determined by the imbalance of the genes located between the flanking duplicons, often include MR. Most of these deletions are sporadic, but a few may be inherited, including the most common deletion of 22q11 associated with DiGeorge and velo-cardio-facial syndromes, with an estimated prevalence of 1:4,000. The list of known microdeletion and duplication syndromes is growing; the first results from genome-wide screening by array based Comparative Genomic Hybridization (CGH) suggest that submicroscopic microdeletions and –duplications scattered all over the genome will be detected in a substantial proportion of patients with idiopathic MR.

Monogenic Disorders

There are numerous autosomal recessive, autosomal dominant, X-linked and mitochondrial disorders where MR is the only symptom, part of a more complex clinical spectrum or seen in a fraction of the affected individuals. The underlying genetic defects have been identified in many of these disorders, providing direct molecular proof of the extreme genetic heterogeneity in MR. Furthermore, inherited MR diseases illustrate an abundance of basic genetic mechanisms, and have revealed several novel ones.

X-linked Mental Retardation

More males than females are mentally retarded, with a M:F sex ratio of 1.5, a skewing which is likely due to



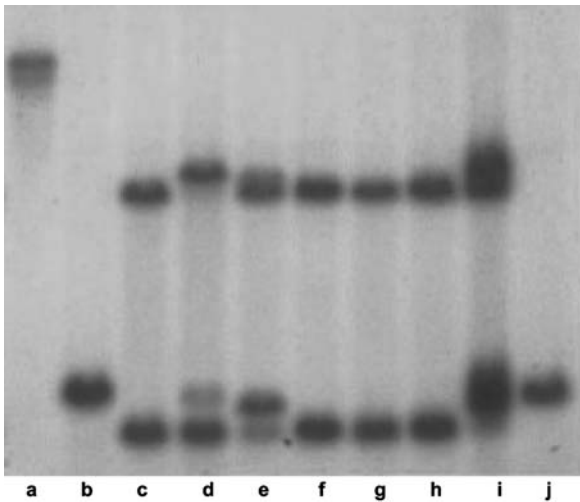
Inherited Mental Retardation Syndromes.

Figure 2 Metaphase from a male with fragile X syndrome. Arrow on the fragile site at Xq27.3 (FRAXA).

the many X-linked forms of MR (XLMR). A current estimate is that of the total genetic morbidity 30% are X-linked and 70% are autosomal forms. There are more than 150 known X-linked syndromes associated with MR and an unknown but growing number of non-specific XLMR loci, many of which have been dissected at the molecular level (3).

The ►Fragile X Syndrome

The fragile X syndrome, characterized by early hyperactivity, MR, speech defect, an elongated facies with large ears, prominent jaw and macroorchidism, is the most common cause of inherited MR. It has its name from the association with a chromosomal fragile site at Xq27.3 which can be seen in a proportion of the metaphases from affected males and heterozygous carriers (Fig. 2). This fragility is caused by elongation of a polymorphic CGG trinucleotide repeat at the 5' UTR region of the *FMR1* gene (Fig. 3). A remarkable feature of this X-linked disorder is that males with only a slight elongation of the CGG repeat (up to ~200 CGGs = a premutation) have normal intelligence and will transmit the premutation unaltered to their likewise phenotypically normal obligatory carrier daughters; however, in the gametes of these daughters, the



Inherited Mental Retardation Syndromes.

Figure 3 Southern blot of the fragile X locus: (a) A full mutation in an affected male; (b, j) premutations in phenotypically normal transmitting males; (c, f, g, h) the normal female pattern with two bands, corresponding to the active X chromosome (lower band) and inactive X chromosome (upper band); (d, e) female carriers of a premutation; (i) female carrier of a full mutation.

premutation has a high risk of expanding many fold into a full mutation, which becomes hypermethylated, resulting in transcriptional inactivation of *FMR1*. These daughters then have a high risk of having affected sons and daughters. At least some of the normal male transmitters with premutations may later develop Parkinson-like neurodegenerative symptoms. The mutation behind the fragile X syndrome is the prototype of a dynamic mutation and it provided the first molecular explanation for anticipation, i.e. that the severity and/or onset of some disorders worsen from generation to generation. All male carriers and almost half of the female carriers of a full mutation are affected, indicating that the fragile X syndrome is an X-linked dominant disease (with incomplete penetrance in the female carriers).

Rett Syndrome and Factors Affecting Male:Female Ratios in X-linked Disorders

Mutations in genes associated with X-linked dominant disorders may be male lethal *in utero*, in which case only affected females are observed and there will be an excess of females born in such families. Rett syndrome (RS), which is caused by mutations in the X-linked gene *MECP2*, almost exclusively affects females. In its classic form, RS has a stage of normal development for $\frac{1}{2}$ – $1\frac{1}{2}$ years then, after developmental stagnation, rapid progressive deterioration to dementia, autism, ataxia and loss of motor skills, followed by a stable plateau which can last for many years. Some females with

MECP2 mutations may have a phenotype resembling Angelman syndrome (see below) with mild MR and autism. *MECP2* mutations in males have been reported in association with severe encephalopathy, but there is no evidence that mutations in *MECP2* are male lethal. One explanation for the almost exclusive occurrence of RS in females could be that almost all cases are due to new mutations and that these mutations occur during spermatogenesis and hence will be transmitted to females exclusively. However, there may be other causes of occasional female excess in MR. In a family with X-linked epilepsy and MR limited to females with normal carrier males, an equal sex ratio was seen among live births. In this family, male sparing rather than male lethality seems to be involved. The mechanism behind this is unknown.

The Prototype Autosomal Recessive MR Disorders: Inborn Errors of Metabolism

In some disorders, MR is due to a primary developmental defect, in others it is caused by a progressive neuronal degeneration, which may ultimately be lethal. The latter is a frequent course in many inborn errors of metabolism. These are typically rare, autosomal recessive conditions caused by mutations in genes encoding for enzymes. Based on the biochemical and cellular defects, these disorders are classified e.g. into mucopolysaccharidoses, lysosomal storage disorders, peroxisome biogenesis disorders, etc. For each of these enzymopathies there is often a wide variability in onset and progression of clinical symptoms, which has led to the description of different clinical forms, e.g. neonatal, infantile and adult forms. The molecular basis of this clinical variability can frequently be related to different mutations in the specific disease gene, some being associated with a milder course, others with a more severe form of the disease. Homozygosity for a severe mutation will lead to severe, maybe early onset and lethal disease, homozygosity for milder mutations to a milder course, whereas a combination of two different mutations (compound heterozygosity), one a severe and one a mild form, may be observed in patients with an intermediate form of the disease.

Linking an Autosomal Recessive Enzyme Defect with Dysmorphology

Smith-Lemli-Opitz syndrome (SLO) is an autosomal recessive, multiple malformation and MR syndrome, caused by mutations in the gene encoding 7-dehydrocholesterol reductase (*DHCR7*). *DHCR7* deficiency interferes with cholesterol biosynthesis, and cholesterol is necessary for the activation of the protein Sonic hedgehog (SHH). SHH is an important secreted molecule, a ►morphogen, involved in a variety of developmental processes, including normal brain formation. Thus, SLO illustrates how an enzymatic

defect merges with dysmorphology. SLO may be more common than anticipated, and even a major unrecognized cause of fetal loss. Other defects in the SHH-pathway are implicated in other developmental disorders, some of which are associated with MR, the most striking being ►**holoprosencephaly**. Holoprosencephaly can be caused by mutations in *SHH* itself, in *PTCH*, encoding the receptor for SHH and in several downstream target genes in the SHH-signaling cascade.

Modification of Autosomal Recessive Inheritance: Bardet-Biedl syndrome (BBS)?

BBS is characterized by MR, obesity, retinopathy, polydactyly, renal malformations, hypogenitalism and predisposition to diabetes mellitus. There is extensive genetic heterogeneity in BBS, with at least eight genes, all of which have been identified. Although BBS is an autosomal recessive disorder, some patients have been reported to carry three mutated alleles (two at one locus and a third at a second locus). A likely explanation is that BBS proteins which are involved in cilia assembly or function may interact with each other at the cellular level. Although this inheritance pattern, termed triallelic or oligogenic inheritance, has not been confirmed in other studies, it is intriguing to speculate that the clinical variability observed in many monogenic disorders, even within the same family, may reflect the concerted action of variation in additional, modifying genes.

Autosomal Dominantly Inherited Diseases

Since individuals with severe MR rarely reproduce, autosomal dominant MR syndromes are less common than autosomal recessive and X-linked recessive syndromes. In some autosomal dominant disorders, mildly affected individuals may reproduce, revealing the dominant nature of the condition. ►**Tuberous sclerosis** (TSC), affecting 1 in 10,000 individuals, is an autosomal dominant disorder characterized by the growth of ►**hamartomas** in many organs, including the brain, where it is associated with MR, epilepsy, learning difficulties and behavioral problems, but with a high degree of clinical variability. Affected children may be severely retarded, whereas a carrier parent may only display microsigns (variable expressivity). Two TSC genes have been identified, *TSC1* on chromosome 9 and *TSC2* on chromosome 16. More than half of the affected children are sporadic cases, which indicates a high frequency of new mutations in TSC. The high mutation rate in *TSC1* and *TSC2* is also illustrated by a high degree of ►**gonadal mosaicism** in TSC and part of the clinical variability can be explained by a frequent presence of ►**mitotic mosaicism**. Both *TSC1* and *TSC2* are ►**tumor suppressor genes** and part of the clinical variability could also be related to

the need for a second mutational hit to occur, in accordance with Knudson's two-hit hypothesis for oncogenesis.

Modification of Autosomal Dominant Inheritance by Genomic Imprinting

►**Angelman syndrome** (AS) is characterized by severe MR, lack of speech, ataxia, a happy disposition and dysmorphism. Most cases (~75%) are sporadic, caused either by a microdeletion of 15q11, which always involves the maternal chromosome 15, or by ►**uniparental disomy** of paternal origin. A similar microdeletion of paternal origin, or uniparental disomy of maternal origin, causes Prader-Willi syndrome with a completely different phenotype. The reason is that only part of the deleted region is active on the paternal chromosome 15 (corresponding to the genes that are responsible for Prader-Willi syndrome), whereas only another part, corresponding to the single gene *UBE3A*, is active on the maternal chromosome 15 (genomic imprinting). In AS, 10% of the patients have mutations within *UBE3A*, in which case the disease can be inherited. AS illustrates that dominant transmission of a phenotype can be explained by the inheritance of one mutated allele which is expressed into an inactive product, while the other allele is silenced by genomic imprinting, in effect creating functional nullisomy. One consequence of the gender specificity of genomic imprinting is that if the mutated allele in *UBE3A* is inherited from a father, who in any case would have donated an inactive allele, the offspring will be phenotypically normal. However, if this is a daughter, her children will have a 50% risk of being affected since she reactivates the mutated allele in her gonads.

Cellular and Molecular Regulation

At the functional level, the identified genes involved in MR confirm that the extreme genetic heterogeneity reflects an extreme functional heterogeneity. However, some interesting trends have emerged: Several of the identified genes in nonspecific XLMR encode for proteins involved in regulation of small GTP-binding proteins of the Ras superfamily cycle, which mediate organization of the cytoskeleton, cell shape and migration and outgrowth of axons and dendrites. The fragile X syndrome protein FMR1 may interact with the small GTPase Rac1 and there is evidence that tuberlin (*TSC2*), defective in tuberous sclerosis, is a putative GTPase-activating protein for Rap1 and Rab5. It is likely that others of the more than 150 genes encoding for small GTP-binding proteins may be associated with MR.

The Sonic hedgehog pathway may not be the only developmental pathway involved in MR. Studies in *Drosophila* indicate that developmental pathways are

particularly sensitive to dosage effects. Thus, chromosomal imbalance and haploinsufficiency syndromes may provide links to other pathways involved.

Several MR-associated syndromes have defects in genes involved in chromatin structure or modification. These include X-linked dominant conditions (RS in females with defects in *MECP2*), X-linked recessive conditions (Coffin-Lowry syndrome (*RSK2*) and the various allelic disorders caused by defective *ATRX*), autosomal dominant (Rubinstein-Taybi syndrome (*CBP*)) and autosomal recessive syndromes (ICF syndrome (*DNMT3B*)). In these disorders, the primary defect probably leads to dysregulation of other genes, which will also be candidate MR genes.

Dissection of the genetic causes of MR may tell us how the human brain evolved so rapidly. In autosomal recessive primary microcephaly (MCPH), a grossly normal but small cerebral cortex is associated with mild to moderate MR. In the most common type, MCPH5, the affected gene *ASPM* is involved in normal mitotic spindle function, showing that mitotic spindle activity is an important factor for determining brain size. Intriguingly, there is evidence suggesting that during evolution the numbers of a specific repetitive “IQ” domain in *ASPM* have increased along with the size of the brain.

Clinical Relevance

The importance of the identification of a specific molecular defect in MR should not be underestimated. It allows for a specific diagnosis, improves the genetic counseling of the individual family, provides options for prevention of recurrence in the family and is often the first step towards an understanding of the pathophysiology of the disorders and hence hope for future therapy. In the post-genomic era, with molecular techniques that can reveal even the tiniest needle (mutation) in the genomic haystack, the importance of careful clinical investigations should not be underestimated, either. One important aspect of the clinical studies in MR will be to identify those patients who can provide clues to the finding of novel MR genes. This will include patients from the rare, large families, from inbred populations with an increased rate of consanguinity, from geographically isolated populations with founder effects and the sporadic patients with chromosomal breakpoints that have inactivated specific MR genes.

References

1. OMIM, Online Mendelian Inheritance in Man. A catalog of human genes and genetic disorders. ► <http://www.ncbi.nlm.nih.gov/entrez/query.fcgi?db=OMIM>
2. Flint J, Wilkie AO, Buckle VJ et al (1995) The detection of subtelomeric chromosomal rearrangements in idiopathic mental retardation. *Nat Genet* 9:132–140
3. Chelly J, Mandel JL (2001) Monogenic causes of X-linked mental retardation. *Nat Rev Genet* 2:669–680
4. Ropers HH, Hamel BC (2005) X-linked mental retardation. *Nat Rev Genet* 6:46–57

Inherited Neurodegenerative Disease

► Polyglutamine Disease, the Emerging Role of Transcription Interference

Inhibitor of β -Catenin and TCF-4

Definition

Inhibitor of β -catenin and TCF-4 (ICAT) defines a 9 kd nuclear protein with no known homology to other proteins. It may compete for binding of β -catenin for TCFs (T cell factors).

► Wnt/Beta-Catenin Signaling Pathway

Inhibitor of Apoptosis Proteins

Definition

The inhibitor of apoptosis proteins (IAPs) can block apoptosis by binding and inhibition of members of the caspase family. In addition, IAPs have been implicated in signal transduction and cell cycle control.

► Apoptosis, Regulation and Clinical Implications
► TNF Receptor/Fas Signaling Pathways

Initiation Complex

Definition

The Initiation complex consists of only a few base pairs and is formed upon collision of complementary DNA strands.

► Thermodynamic Properties of DNA

Initiator

Definition

Initiator is a protein that recognizes a replication origin and activates bidirectional DNA synthesis.

► [Replication Origins](#)

involve receptors encoded in the germline (e.g. Toll-like receptors; phagocytosis receptors, etc.), antimicrobial peptides in the skin and gastrointestinal tract (defensins, cathelicidins; pepsin), and secreted proteins that facilitate an adaptive immune response (complement components). Innate immunity can trigger ► [adaptive immunity](#).

► [Autoimmune Diseases](#)

► [Autoimmunity](#)

► [Inflammatory Response](#)

INK4a/ARF

Definition

INK4a/ARF describes a genetic locus that encodes the cell-cycle regulator p16INK4a and the p53 regulator ► [ARF](#). This locus is very frequently mutated in human cancer.

► [Senescence](#)

► [Tumor Suppressor Genes](#)

Inner Boundary Membrane

Definition

Inner boundary membrane describes a region of the mitochondrial inner membrane that is in continuous and close apposition to the outer membrane. Together with the outer membrane it forms a kind of envelope of the mitochondria.

► [Mitochondria – Biogenesis and Structural Organization](#)

Innate Adjuvants

Definition

An “innate adjuvant” is a substance that enhances or modulates the innate immune response to antigen. Effector mechanisms of the innate immune system, activated within hours after contact with a microbial agent, are mediated by cells that recognize conserved microbial structures through germ-line encoded receptors.

► [DNA-based Vaccination](#)

Inner Ear

Definition

The inner ear is the sensory organ composed by the cochlea and the vestibule. It contains mechanosensory cells specialised in sound and movement detection.

► [Microvilli](#)

Innate Immunity

Definition

Innate immunity is an inborn, early host, defense mechanism of a multicellular organism to defend itself against invasion by pathogens (bacteria, fungi, viruses, etc.). These mechanisms are generally mediated by ► [antigen presenting cells](#) that recognize unique molecular patterns of the invading pathogens (e.g. lipopolysaccharide of gram negative bacteria; bacterial DNA; viral double-stranded RNA). The defense mechanisms

Innexins

Definition

Innexins are proteins unrelated to connexins that form intercellular gap junction channels in invertebrates.

► [Gap Junctions](#)

Inr

Definition

Inr designates the pyrimidine-rich core promoter element, which can either act independently or in

conjunction with the TATA box. Inr spans across the transcription start site.

► [Core Promoters](#)

Insertion

Definition

Insertion refers to a type of chromosomal abnormality in which a DNA sequence is inserted into a gene, disrupting the normal structure and function of that gene.

► [COPD and Asthma Genetics](#)

Insertion/Deletion Loop

Definition

Insertion/deletion loop consists of one or more unpaired nucleotides in a double-stranded DNA, such as that typically arising through DNA polymerase slippage in the nascent or template strand during the replication of a microsatellite repeat.

► [Hereditary Nonpolyposis Colorectal Cancer](#)

Insertional (Oncogen) Activation

Definition

Insertional (oncogen) activation describes a deregulation of cellular proto-oncogene expression by the insertion of viral transcriptional control sequences (enhancer, promoter, polyA signal). The insertion results in a growth stimulus, due to the over-expression of the proto-oncogen product mutant protein or upregulation of its expression. It is frequently accompanied by genomic mutagenesis, which could result in the expression of mutant oncoproteins. Insertional c-onc activation is the most frequent form of tumour induction of murine and avian retroviruses.

► [Viral Oncogenesis](#)

Insertional Activation/Inactivation

Definition

Insertional activation/inactivation refers to either activation of an endogenous gene which is located near an integrated transgene, or to disruption of a gene or other functional sequence by insertion of a transposable element.

► [Repetitive DNA](#)

► [Transgene Silencing](#)

Insulator

Definition

Insulator designates a DNA element that blocks activating or repressing position effects, or blocks the action of enhancers and silencers.

► [Transgene Silencing](#)

Insulin

Definition

Insulin is an endocrine peptide hormone of 51 amino acids, which is secreted by the islet cells of the pancreas. The major function of insulin is to counter the concerted action of a number of hyperglycemia-generating hormones and to maintain low blood glucose levels.

► [Diabetes Mellitus, Genetics](#)

Insulin Dependent Diabetes

► [Diabetes Mellitus, Genetics](#)

Insulin Receptor

Definition

The insulin receptor is a transmembrane tyrosine kinase receptor on the cellular membrane of skeletal muscle, fat tissue and liver cells. It consists of four polypeptide

chains. Two transmembrane subunits are connected to two extracellular insulin binding domains. Upon binding of insulin, the intracellular protein kinase domain of the receptor is activated to phosphorylate a set of target proteins, which leads to the metabolic effects of insulin.

► [Affinity Chromatography and *In Vitro* Binding \(Beads\)](#)

► [Tyrosine Kinases](#)

Insulin Resistance

Definition

Insulin resistance is an impaired metabolic response to the body's own insulin. An increased insulin amount is needed to maintain low plasma glucose levels. Insulin resistance is frequently associated with obesity and type 2 diabetes.

► [Diabetes Mellitus, Genetics](#)

► [High-HDL Syndrome](#)

Integrale Membrane Proteins

Definition

Integrale membrane proteins contain at least one hydrophobic sequence of amino acids long enough to span a membrane (transmembrane region). Their removal from the membrane requires disruption of the bilayer, for example, by treatment with detergent.

► [Biological Membranes](#)

► [Fatty Acid Acylation of Proteins](#)

Integrase

► [Gammaretroviruses IN \(Integrase\)](#)

Integration

Definition

The term integration refers to the integration of a DNA element (for instance a retrovirus DNA) into its host cell genomic DNA. In the context of site-specific

recombination, integration is often defined more narrowly as a reaction similar to the lambda bacteriophage integration.

► [DNA Recombination](#)

► [Retroviruses](#)

Integrin Ligand

Definition

An Integrin ligand is an extracellular matrix protein, e.g. fibronectin, laminin, that binds the extracellular head domain of integrins.

► [Integrin Signalling](#)

Integrin Signaling

LAWRENCE E. GOLDFINGER, MARK H. GINSBERG
Department of Medicine, Division of Rheumatology,
University of California, San Diego, La Jolla, CA, USA
mhginsberg@ucsd.edu

Definition

► [Integrins](#) are transmembrane receptors that promote cellular attachment to the ► [extracellular matrix](#) (ECM) and to counter-receptors on the surfaces of neighboring cells. Integrins are responsible for mediating cell adhesion and migration, and for maintaining tissue integrity. They are expressed at the surface of virtually all multi-cellular animal cells as non-covalent heterodimers of α and β subunits. Integrins are named based on their α and β subunit composition, e.g. $\alpha_1\beta_1$, $\alpha_{IIb}\beta_3$. The integrin family in humans consists of eighteen α and eight β subunits, which are each encoded by unique genes. Each α subunit pairs with a restricted set of β subunits, resulting in a total of twenty-four known functional integrin heterodimers. Each pairing, in turn, is expressed on a limited array of cell types. Thus, each cell has its own particular repertoire of integrins. Since different integrins display distinct ligand-binding preferences, these repertoires specify adhesive preferences of different cell types. In some cases, multiple integrins can recognize the same ligand. For example, many integrins bind to the extracellular matrix protein, fibronectin. Even within the same ECM protein, integrins can vary in recognition specificity. Integrin $\alpha_v\beta_3$, for example, binds to an Arg-Gly-Asp (RGD)

motif in fibronectin, while integrin $\alpha_4\beta_1$ binds to a Leu-Asp-Val (LDV) motif present in a different part of fibronectin. Furthermore, these short peptide motifs can occur in evolutionarily unrelated proteins, where they can serve as recognition sites for many different integrins. The **RGD motif** occurs in many proteins and can bind to many integrins (e.g. $\alpha_{IIb}\beta_3$, $\alpha_v\beta_3$, $\alpha_v\beta_5$, $\alpha_5\beta_1$). This integrin-binding motif interaction is conserved in phylogeny; for example, the *Drosophila* PS2 integrin recognizes an RGD sequence in tigrin, a matrix protein. The occurrence of integrin-binding motifs, such as RGD, is an example of convergent evolution dictated by the recognition specificity of adhesion receptors.

All integrin subunits possess a single transmembrane domain, which separates the large extracellular, N-terminal ligand-binding domain from the typically short (less than 60 amino acids) C-terminal cytoplasmic domain or **cytoplasmic tail**. Besides acting as cellular “glue” to support adhesion and “feet” to facilitate migration, integrins initiate signaling pathways, which regulate processes such as embryogenesis, cell development, proliferation, apoptosis and the immune response.

Integrins derive their name from their role of integrating the extracellular matrix with the insoluble cellular skeleton (**cytoskeleton**). Binding of ligands to the extracellular portion of integrins leads to both conformational changes in the receptors and to receptor clustering. This combination of events initiates intracellular signals such as protein tyrosine phosphorylation, activation of small GTPases and changes in phospholipid biosynthesis. This process is commonly referred to as “outside-in” integrin signaling. In addition, integrins are subject to so-called “inside-out” integrin signaling. Integrins can fluctuate between three functional states, an “active” state that manifests high affinity for binding ligands, a high affinity state that is bound to ligand or a distinct “inactive” state, which has a low affinity for ligands. Inside-out signaling is defined as modulation of the functional state of integrins *via* intracellular signaling processes. It is frequently mediated by interactions of intracellular components with the integrins’ cytoplasmic tails. Furthermore, signals from one integrin can regulate the activation state of another in a phenomenon referred to as integrin trans-regulation or cross-talk. Through trans-regulation, multiple integrins can be linked, *via* a combination of both outside-in and inside-out signaling (1).

Characteristics

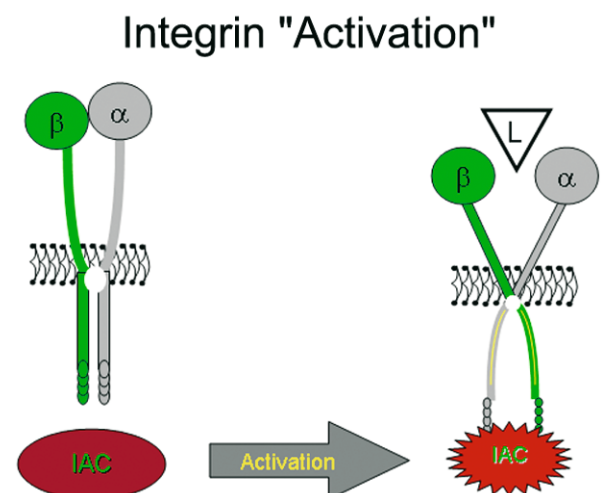
The switch from an inactive to an active state is the result of alterations in the conformation of the integrin extracellular domains. These conformational changes

can be triggered by protein-protein interactions at the integrin cytoplasmic face. These protein-protein interactions are regulated by different signaling pathways in different cells. However, in many contexts, activation of key intermediates such as the GTPase Rap1 or protein kinase C, promote increased integrin affinity. Conversely, activation of Raf-1 kinase often suppresses integrin activation. The activation states of individual integrin subtypes can be differentially regulated to affect integrin-specific cell adhesiveness. Therefore mechanisms of activation and suppression of individual integrins exist, in addition to common mechanisms affecting all integrins. Whatever the upstream events, it is likely that the final common events involve changes in interactions at the integrin cytoplasmic domain (Fig. 1).

Physiological Roles of Integrin Activation

Integrin Activation in Hematopoietic Cells

Integrin activation plays a crucial role in the biology of blood-borne cells. For example, the platelet integrin, $\alpha_{IIb}\beta_3$, is usually maintained in the inactive state, with very low affinity for its extracellular ligands, fibrinogen, fibronectin and von Willebrand factor. Upon platelet stimulation, the integrins are quickly activated, leading to fibrinogen binding and consequent platelet aggregation, an early step in hemostasis. Thus, loss of $\alpha_{IIb}\beta_3$ integrin activation impairs hemostasis. Conversely, platelet aggregation is central in arterial thrombosis, which is responsible for the majority of heart



Integrin Signaling. Figure 1 “Inside-out” signaling activates integrins. Binding of an integrin activating complex (IAC) to the cytoplasmic tails of an integrin heterodimer produces conformational changes in the extracellular head domains, leading to activation and high affinity binding to ligand (L).

attacks and strokes. Pharmacological blockade of the activation of integrin $\alpha_{IIb}\beta_3$ by agents such as aspirin or ticlopidine is therefore often used to treat patients at risk for these diseases. Similarly, rapid activation of dormant integrins on leukocytes allows these cells to attach firmly to the endothelial vessel wall and resist strong shear forces from the flowing blood. This step is essential for leukocyte migration into sites of inflammation. For this reason, loss of leukocyte integrin activation can lead to enhanced susceptibility to infections and its blockade may be useful in anti-inflammatory therapies.

Integrin Activation in Tissues

For adherent cells in tissues, such as ►fibroblasts or ►epithelial cells, stable cellular attachment to the ECM is necessary for maintaining tissue integrity. Many of the best-studied integrins in these cells, such as the α_5 , α_v and α_6 integrins, are indispensable components of cell-substrate attachment structures, ►focal adhesions, ►hemidesmosomes, and podosomes. In addition to attaching to the extracellular matrix, integrins participate in formation and remodeling of the ECM. Activation of tissue integrins can regulate the assembly and structure of the extracellular matrix and control the stability of integrin-matrix attachments. The migration of these cells in tissues is also regulated by the ability of integrins to become activated. Furthermore, integrin activation may be topographically localized across the basal cell surface. For example, activated $\alpha_v\beta_3$ integrins are preferentially localized to the leading edge of endothelial cells during cell migration.

Molecular Interactions

Tail-binding Proteins

Numerous cytoplasmic integrin-binding proteins have been described, and several have been reported to contribute to activation through binding to the integrin tails. The best-characterized example is the cytoskeletal linker protein ►talin, which promotes integrin activation by binding to a subset of integrin β subunit tails at a highly conserved ►NPxY motif. Mutational evidence suggests that integrin α - β subunit tail interactions inhibit activation. The ability of talin to activate the integrins has been attributed to competition with the α subunit for β tail binding. In addition to talin, other integrin-binding or talin-binding proteins may affect the activation states of integrins.

Membrane Proximal Residues

Deletion of the tail sequences adjacent to the putative transmembrane regions in either the α or the β tail can activate integrins, indicating that these residues are

essential for maintaining integrins in an inactive state. One model for activation entails a breaking of a salt bridge in this region that holds the tails together, leading to an “opening” of the tails *via* a scissor-like mechanism (Fig. 1). This leads to an allosteric rearrangement in the extracellular domain and a shift to the active state (1).

Extracellular Domain Conformational Changes

Precise conformational shifts at the ligand binding face of integrins result in high affinity binding to ligands. Integrin-ligand interactions require an essential Asp or Glu residue in the integrin recognition sequence in the ligand, as well as the presence of a divalent cation. Some of the integrin α subunits have an “inserted” or ►I/A domain near the N-terminus; these integrins contact ligand solely through the α subunit, although the β subunit plays an accessory role. The I/A domain manifests a metal ion-dependent adhesion site (►MIDAS). Key residues in the ligand cooperate with the MIDAS to coordinate the cation and allow integrin-ligand interaction. Conformational changes in the MIDAS during cation coordination and ligand binding subsequently result in secondary changes in the rest of the integrin, including a downward shift of a α -helix C-terminal to the I/A domain, thereby switching the integrin from the “closed” inactive state into the “open” active state. For integrins with a α subunit that does not contain an I/A domain, the I/A-like domain in the β subunit cooperates with the α subunit in a somewhat different fashion to mediate ligand interaction, but also through coordination of a divalent cation (2).

Connections to Intracellular Signaling Pathways

Although integrins themselves do not manifest any catalytic activity, they transmit signals through binding partners such as integrin-associated protein (IAP, CD47), integrin-linked kinase (ILK) and members of the tetraspanin transmembrane protein family. Integrin ligation activates members of the Ras family of GTPases and can also lead to activation of kinases such as JNK, PAK, and ERK/MAP kinase – the latter two are localized to integrin-rich adhesion sites. These enzymes can subsequently regulate the expression of genes involved in cell proliferation and differentiation. Furthermore, integrins play a role in the inhibition of detachment-induced apoptosis, called “anoikis”, through regulation of the PI3-kinase/Akt signaling pathway. Activation of these and other signaling pathways following integrin ligation must be precisely coordinated to affect productive cellular phenotypes. Thus, the integrin signaling proteome undergoes dynamic alterations depending on the cellular state.

Regulatory Mechanisms

“Inside-out” Signaling

Modulation of integrin function can occur by changes in ligand binding affinity or by changes in the clustering of these receptors. The former mechanism is often referred to as integrin “activation” and the latter as a change in avidity. These functional changes can be caused by signaling pathways initiated by agonist-receptor interactions at the plasma membrane. The appropriate regulation of integrin activation is required for the formation and maintenance of normal tissue architecture (3).

“Outside-in” Signaling

Occupancy of integrins leads to conformational changes that can propagate over long distances in the extracellular domain. It has been proposed that these allosteric rearrangements can be transmitted across the plasma membrane, changing the interactions of the α and β cytoplasmic domains with each other and thus heterodimer interactions with intracellular signaling molecules. Furthermore, occupancy leads to integrin clustering, which can also change the physical relationships of the integrin cytoplasmic domains with each other. This combination of occupancy and clustering is believed to initiate signals from integrins. These signals regulate the formation and strengthening of adhesion sites, the dynamics of cytoskeletal structure, cell shape

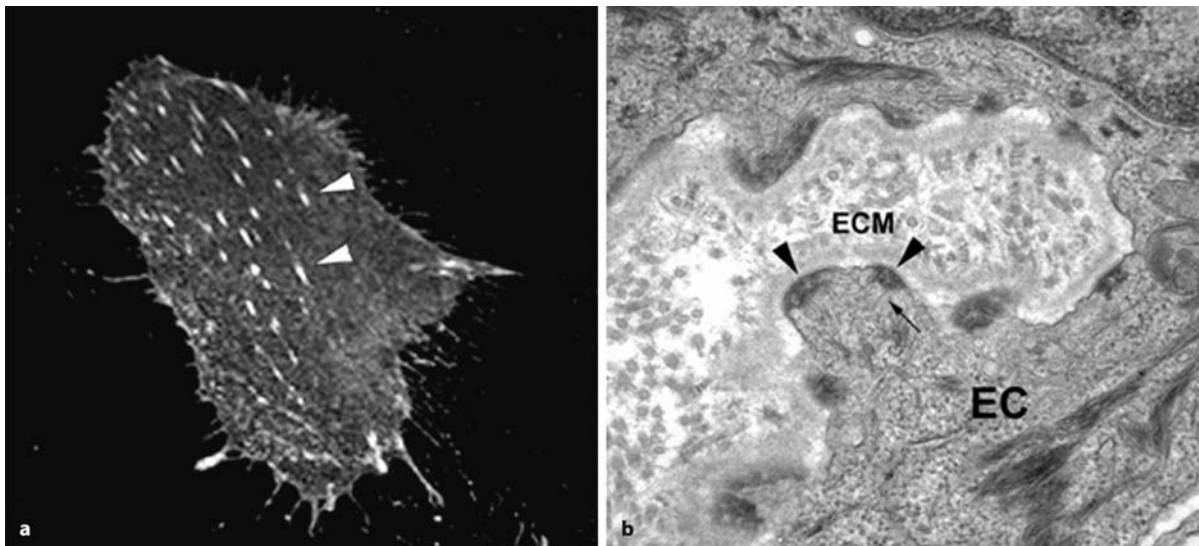
and size, cell polarity and cell migration. In addition to controlling morphological activities of cells, signaling through integrins also regulates gene expression, cell proliferation and apoptosis and many integrins are required for the development of particular cell lineages. Thus a more complete picture of the cellular responses to integrin ligation at the genomic level is necessary for a fuller understanding of the breadth of integrin signaling (3).

Avidity Modulation

In addition to affinity modulation through conformational changes in individual integrin molecules, cell adhesion can also be enhanced by increasing the density of ligand binding sites through integrin clustering, *via* an affinity-independent mechanism known as avidity modulation. Many [▶ integrin ligands](#) are multivalent; thus integrin clustering can promote adhesion through cooperative ligand binding. In many cases, a combination of affinity and avidity modulation both contribute to the regulation of cell adhesion (3).

Integrin Cross-talk

Ligand binding to one subtype of integrins can affect the activation state of another integrin subtype on the same cell, by modulating the ligand-binding affinity and/or avidity of that integrin. For example, outside-in signaling through ligand binding of $\alpha_4\beta_1$ integrin can



Integrin Signaling. Figure 2 Focal contacts and hemidesmosomes. Two types of integrin-ECM contacts are shown. (a) Focal contacts are visualized by labeling $\alpha_{11b}\beta_3$ integrin clusters in a fibroblast expressing $\alpha_{11b}\beta_3$ attached to fibrinogen. Arrowheads indicate two focal contacts. (b) Transmission electron micrograph of a cross section of bovine tongue. Hemidesmosomes (arrowheads) are seen as electron-dense structures at the basal surface of the tongue epithelial cells (EC), connecting the intermediate filaments (arrow) to the extracellular matrix (ECM). Image courtesy of Gregory W. deHart and Jonathan C.R. Jones.

affect $\alpha_L\beta_2$ -dependent lymphocyte adhesion and migration by regulating signaling through $\alpha_L\beta_2$. Furthermore, integrins are involved in cooperative signaling with other non-integrin receptors, such as the EGF receptor and the T cell receptor complex.

Contributions to Cellular Phenotypes

By connecting the ECM to the intracellular cytoskeleton, integrins also act as mechanosensors, transmitting information about the physical state of the extracellular matrix into the cell and altering cytoskeletal dynamics. Integrin ligation can activate key signaling components, such as the small GTPases Rho, Cdc42 and Rac, which promote cytoskeletal rearrangements leading to the formation of actin stress fibers, filopodia and lamellipodia respectively. Within focal adhesions, integrins can activate kinases important for downstream signaling, such as focal adhesion kinase (FAK) and Src tyrosine kinases. Activation of these enzymes results from signals initiated at the integrin cytoplasmic tails. Integrin tails also affect cytoskeletal rearrangements through other mechanisms, e.g. activation of Syk tyrosine kinase through its direct binding to the β_3 tail, which promotes Vav1-induced GTP loading of Rac and subsequent lamellipodial formation. Through such mechanisms, integrins regulate cell polarity, migration, shape and size (4).

Integrins in Adhesion Sites

Most integrins are physically linked to the actin cytoskeleton *via* interactions of the β tail with actin-binding proteins such as α -actinin, talin and filamin. By connecting the extracellular matrix with the actin cytoskeleton, clustered integrins form the essential core of complexes of structural and signaling molecules such as focal complexes and adhesions (contacts), and osteoclast structures called podosomes (which have many of the same protein components as focal adhesions). $\alpha_6\beta_4$ Integrins associate with the intermediate filament cytoskeleton through other linker proteins, BP180 (BPAG2), BP230 (BPAG1) and plectin in hemidesmosomes (5), which are specific to basal epithelial cells (Fig. 2).

Ligand binding and integrin clustering promote nucleation of nascent cell adhesion sites. Numerous cytoskeletal adapters and signaling molecules are subsequently recruited to the clustered integrin tails. These complexes serve as the initiation points for signaling by integrins to affect cellular phenotypes. Focal complexes “mature” to become focal adhesions (focal contacts) as more integrins and signaling molecules associate, thereby strengthening cellular attachments to the ECM. The intricate network of protein-protein interactions within focal adhesions contributes to signaling pathways regulating various cellular phenotypes (4).

Adhesion sites are dynamic structures, which undergo remodeling to fit the needs of the cell or of the region of the cell in contact with the extracellular matrix. Focal complexes form at the leading edges of lamellipodia and filopodia in migrating cells, where they mediate new attachments but do not mature to focal adhesions. Instead, these smaller complexes are subject to rapid turnover as the cell continues to extend protrusions in the direction of migration. Concomitantly, integrins in focal adhesions at the rear of the cell are recycled to the front *via* a treadmill-like mechanism. Focal adhesions, hemidesmosomes and podosomes break apart and reform by severing and re-establishing integrin-ECM contacts under conditions in which cells are motile, such as during development and in wound healing (4).

► Signal Transduction: Integrin-mediated Pathways

References

1. Schwartz MA, Schaller MD, Ginsberg MH (1995) Integrins: emerging paradigms of signal transduction. *Annu Rev Cell Biol* 11:549–599
2. Hynes RO (2002) Integrins: bidirectional, allosteric signaling molecules. *Cell* 110:673–687
3. Hughes PE, Pfaff M (1998) Integrin affinity modulation. *Trends Cell Biol* 8:359–364
4. Schoenwalder SM, Burridge K (1999) Bidirectional signaling between the cytoskeleton and integrins. *Curr Opin Cell Biol* 11:274–286
5. Bannan LJ, Goldfinger LE, Jones JCR et al (2001) Desmosomes and hemidesmosomes. In: Beckerle MC (ed) *Frontiers in Molecular Biology: Cell Adhesion*. Oxford University Press, New York p 339–351

Integrin-Associated Protein

► IAP

Integrins

Definition

Integrins are major receptors for cell-cell adhesion and cell binding to extracellular matrix. They are a family of transmembrane proteins composed of noncovalently linked heterodimers of an α and a β subunit. In mammals, 8 β subunits can dimerize with 18 α subunits to form 24 distinct integrins. These can be considered to belong to several subfamilies based on evolutionary relationships. All integrins containing an α_v or a β_1

subunit bind extracellular matrix proteins, which activates a signaling pathway, gene transcription and modifies cell behaviour. Inside the cell, integrins bind signalling molecules and other proteins connecting to actin, thus linking the extracellular matrix to the cytoskeleton.

- ▶ [Extracellular Matrix](#)
- ▶ [Gut Epithelium](#)
- ▶ [Hemidesmosomes](#)
- ▶ [Integrin Signaling](#)
- ▶ [Signal Transduction: Integrin-mediated Pathways](#)

Intein

Definition

Self splicing proteins are called inteins. Self-cleavage of the thioester bond can be induced by thiol reagents.

- ▶ [Protein Tags](#)

Intensified CCD Camera

Definition

An intensified CCD camera combines a ▶ [CCD camera](#) and an image intensifier with the intention to increase the sensitivity of the camera. The intensifier is a vacuum tube with a photocathode that emits electrons in response to the image focused onto it. The electrons are accelerated by an electrical field, and focused onto a phosphorescence plate at the exit of the device. The phosphor layer shows an image 10 to 1000× brighter than the original image. This image is then projected by a fibre bundle or lenses onto the CCD. ICCDs are gatable, and can function with very high frame rates, but the quantum efficiency of the primary photocathode is not as high as that of the best back-illuminated CCDs.

- ▶ [Fluorescence Microscopy: Single Particle Tracking](#)

ICCD Camera

- ▶ [Intensified CCD Camera](#)

Interaction Chromatography

- ▶ [Affinity Chromatography and In Vitro Binding \(Beads\)](#)

Interaction Discovery Mapping™

Definition

Interaction discovery mapping™ refers to a method to detect bio-molecular interactions such as protein-protein interactions, DNA-protein, antigen-antibody, receptor-ligand interactions with ▶ [SELDI](#) technology. One of the interacting molecules is immobilized on the ProteinChip Array and used as a bait to capture the interacting partner.

- ▶ [Mass Spectrometry: SELDI](#)

Interaction Map

Definition

Interaction map comprises of the network of protein-protein interactions in an organism, based on yeast two-hybrid or co-precipitation data. The complete map is referred to as the interactome.

- ▶ [C. Elegans as a Model Organism for Functional Genomics](#)

Interaction Trap

- ▶ [Two-Hybrid System](#)

Interactome

Definition

The collection of all protein-protein interactions present at any given moment in a cell or tissue, and its behaviour over time and cell states.

- ▶ [Interaction map](#)

Intercalated Disc

Definition

Intercalated disc defines an electron-dense junctional area between cardiomyocytes, which maintains the structural integrity of the heart and the end to end connection between the myocytes. The intercalated disc of adult cardiac muscle consists of three main junctional complexes; zonula adherens, desmosome, and gap junction, each with defined functions.

►Heart

Intercellular Adhesion Molecule

►ICAM

Intercellular Junctions

Definition

Intercellular junctions are plasma membrane-associated complexes that mediate adhesion or direct connection between adjacent cells, e.g. ►adherens junctions and ►desmosomes. They might also seal the membranes of the adjacent cells to prevent the diffusion of fluids or solutes between two cells, e.g. ►tight junctions. Others, for example, gap junctions, mediate the transport of small molecular weight substances between adjacent cells.

►Gap Junctional Communication

Interference

Definition

The amplification (constructive interference) and diminution (destructive interference) of a wave resulting from the combination of two or more waves with corresponding relative phases.

►X-Ray Crystallography–Basic Principles

Interference Filter

Definition

Interference filters are optical filters that operate by transmitting a selected wavelength region with high efficiency while rejecting, through reflection and destructive interference, all other wavelengths. There are different types, named band-pass, short and long-pass interference filters. They consist of glass substrates onto which thin layers of metal salts are deposited.

►Fluorescence Microscopy: Spectral Imaging vs. Filter-Based Imaging

Interferon

Definition

Interferon defines a large family of multifunctional glycoproteins that are produced mainly by cells of the immune system after exposure to a virus, bacterium, or parasite. Their most prominent functions are their antiviral properties. Interferons are used to treat cancer.

►Cardiac Signaling: Cellular, Molecular and Clinical Aspects

►Cytokines

Interferon-Regulatory Factor 1

Definition

IRF-1 belongs to the family of interferon regulatory factor (IRF) transcription factors. These transcription factors regulate interferons as well as interferon-inducible genes in response to infection by viruses.

►RNA Interference in Mammalian Cells

Interleukin–1 Receptor Associated Kinase

Definition

Interleukin–1 receptor associated kinase (IRAK) is a serine-threonine protein kinase that is involved in IL–1/

IL-18/ ▶Toll receptor signaling. Several IRAK proteins have been identified, including IRAK-1, IRAK-2, IRAK-M and IRAK-4.

▶Interleukins and Their Receptors, a Role in Regulating Hematopoietic Stem Cells

Interleukins

Definition

Interleukins comprise of a group of molecules (▶cytokines) involved in communication between cells of the immune system.

▶Cardiac Signaling: Cellular, Molecular and Clinical Aspects

▶Cytokines

▶Interleukins and Their Receptors, a Role in Regulating Hematopoietic Stem Cells

▶Rheumatoid Arthritis

Interleukins and Their Receptors, a Role in Regulating Hematopoietic Stem Cells

CLAUDIA ORELIO, ELAINE DZIERZAK
Department of Cell Biology and Genetics,
Erasmus University Medical Center, Rotterdam,
The Netherlands
c.orelio@erasmusmc.nl
e.dzierzak@erasmusmc.nl

Definition

▶Cytokines are low molecular weight proteins that are of great importance in the regulation of ▶hematopoiesis, ▶immunity, ▶inflammation and other processes. The earliest reports on the biological activity of cytokines in the 1960/1970s revealed that substances emanating from cells or present in conditioned medium could stimulate the growth and/or differentiation of hematopoietic colonies in semi-solid medium. Hence, these substances were referred to as ▶colony stimulating factors (CSFs). Several specific factors from different sources were identified that, for example, could stimulate macrophage colony formation, granulocyte colony formation or the formation of colonies containing both macrophage and granulocytes. The significance of the first hematopoietic colony forming assays facilitated the molecular isolation and

characterization of many colony stimulating factors in the late 1970s/early 1980s. Mainly for historical reasons, some CSFs acquired quite different names, e.g. erythropoietin, interleukins. Although they have little meaning with regard to classification, some of these names have been retained, particularly the interleukin designation.

The cytokine superfamily is usually divided into three subfamilies, the interleukins (ILs), the colony-stimulating factors and the tumor necrosis factors (TNFs). Cytokines are produced by many different cell types, both hematopoietic and non-hematopoietic, and can perform their function either in a local or systemic fashion by stimulating the target cells through interaction with specific cell surface receptors. Additionally, some cytokines act in an autocrine fashion, stimulating the cell by which they were produced. Cytokines have a wide range of biological activities, including the regulation of proliferation and/or differentiation, chemo-attraction and survival/apoptosis depending upon the target cell type (3).

Interleukins act as communication signals between leukocytic cells, hence the name meaning “between ▶leukocytes”. To date, more than 30 interleukins have been identified and this number is still increasing. The interleukin family consists of a structurally and functionally diverse group of proteins, but all interleukins are, in one way or another, regulators of the hematopoietic system (1, 3).

Despite the wide variety of cytokines, this review will focus on the interleukins and CSFs that have predominant effects on the pivotal cells of the hematopoietic system. In particular, we will focus on the cytokines that play an important role in the regulation of inflammatory reactions (i.e. IL-1, IL-6 and IL-18), in the regulation of the immature hematopoietic cells within the hematopoietic hierarchy, namely ▶hematopoietic stem and progenitor cells (i.e. IL-1, IL-3, IL-6 and IL-11 and GM-CSF), and those involved in the proliferation and/or differentiation of T and/or B ▶lymphocytes (i.e. IL-2, IL-4 and IL-7). Summaries of these cytokines and their receptors are presented in Tables 1 and 2, respectively. Included within Table 2 are the more recent nomenclature designations for molecules elaborated by cells in lineage differentiation hierarchies, i.e. clustering of differentiation or ▶CD numbers. These are provided for most of the cytokine receptors discussed in this review.

Characteristics

Interleukins: Expression and Biological Activity

Several subfamilies of interleukins are discussed in this section. The interleukins are grouped into subfamilies on the basis of their biological function within the hematopoietic system. Figure 1 shows the position and the cell type within the hematopoietic hierarchy upon

Interleukins and Their Receptors, a Role in Regulating Hematopoietic Stem Cells. Table 1 Characteristics of Interleukins (1, 3)

Interleukin	Protein size (kD)	Human chromosome	Mouse chromosome	Produced by	Biological activity & remarks
IL-1 α/β	17	2	2	<ul style="list-style-type: none"> • monocytes • macrophages • keratinocytes • endothelial cells • fibroblasts • lymphoid cells 	<ul style="list-style-type: none"> • pro-inflammatory cytokine • altering cell adhesion properties • radio-protective effects
IL-1Ra	17	2	2		<ul style="list-style-type: none"> • natural IL-1R antagonist, prevents IL-1 interaction with the IL-1R • heavily glycosylated
IL-2	15	4	3	<ul style="list-style-type: none"> • Activated T-cells 	<ul style="list-style-type: none"> • T-cell proliferation • activated B-cell proliferation
IL-3	15-17	5	17	<ul style="list-style-type: none"> • activated T-cells • keratinocytes • NK-cells • mast cells • endothelial cells • monocytes 	<ul style="list-style-type: none"> • promotes the proliferation and differentiation in the myeloid/erythroid lineage of hematopoietic progenitor cells • survival factor
IL-4	20	5	11	<ul style="list-style-type: none"> • activated Th2-cells • mast cells 	<ul style="list-style-type: none"> • B- and T-cell proliferation/ differentiation
IL-5	14	5	11	<ul style="list-style-type: none"> • T-cells 	<ul style="list-style-type: none"> • eosinophil differentiation • heavily glycosylated • forms homodimers
IL-6	22-28	7	5	<ul style="list-style-type: none"> • B and T cells • monocytes • fibroblasts • endothelial cells 	<ul style="list-style-type: none"> • affects antigen-specific immune responses • pro-inflammatory cytokine • acute phase reactions
IL-7	25	8	3	<ul style="list-style-type: none"> • BM stromal cells • thymic cells • keratinocytes 	<ul style="list-style-type: none"> • proliferation and differentiation of B- and T-cell precursors • supports megakaryocyte differentiation
IL-8	8	4		<ul style="list-style-type: none"> • monocytes • fibroblasts • endothelial cells • keratinocytes • hepatocytes 	<ul style="list-style-type: none"> • activation of neutrophilic granulocytes • chemotactic for migratory immune cells
IL-11	23	19	7	<ul style="list-style-type: none"> • BM stromal cells • mesenchymal cells 	<ul style="list-style-type: none"> • proliferation of HSCs and megakaryocyte progenitor cells • induces megakaryocyte maturation resulting in increased platelet production
IL-18	18	11	9	<ul style="list-style-type: none"> • monocytes • macrophages • T/B-cells • dendritic cells • keratinocytes • epithelial cells 	<ul style="list-style-type: none"> • growth and differentiation of Th cells • pro-inflammatory cytokine

which the specific interleukins act and Table 1 provides detailed information on each interleukin (1, 3).

Interleukin-1/Interleukin-18 (IL-1/IL-18)

IL-1 and IL-18 play an important role within the innate immune system and in the regulation of ►inflammation reactions. IL-1 and IL-18 are predominantly produced by macrophages and monocytes, but also by keratinocytes, endothelial cells, fibroblasts and lymphocytes. IL-1 and IL-18 mediate inflammatory reactions by recruitment and activation of neutrophils and macrophages. IL-1 regulates adhesion molecule expression on endothelial and hematopoietic cells thereby facilitating trans-endothelial transport. Interestingly, IL-1 is also a known regulator of immature hematopoietic cells. IL-1 stimulates (sometimes in synergy with other cytokines) the proliferation and/or differentiation of immature CD34⁺ hematopoietic cells and can further regulate hematopoiesis by inducing the expression of cytokines/growth factors by cells of the hematopoietic microenvironment. Similarly, IL-18 is a known regulator of lymphoid cells and plays a role in the proliferation and differentiation regulation of Th1 and B-cells. IL-18 is also a potent inducer of interferon-gamma (IFN γ) expression in T and B-cells.

Interleukin-2/Interleukin-4/Interleukin-7 (IL2/IL4/IL7)

The subfamily of interleukins that includes IL-2, IL-4 and IL-7 (and IL-9/IL-15) plays an important role in regulating lymphocytes within the hematopoietic system. IL-2 and IL-4 are produced predominantly by (sub)populations of T-cells and IL-7 is produced by BM stromal and thymic cells and keratinocytes. IL-2 and IL-4 regulate the proliferation and differentiation process of T-cells and B-cells respectively. IL-2 also plays an important role in the clonal expansion of antigen-activated T-cells in immune reactions. IL-7 is an important regulator of the immature ►common lymphoid progenitor (CLP), which gives rise to both B- and T-cells. The importance of these interleukins within the immune system was demonstrated by studies investigating immunodeficiency in humans and mice. Severe combined immunodeficiency results from defects in the gene encoding for the receptor gamma-common (γ c) subunit, which is commonly used by the receptors of the IL-2/IL-4/IL-7 subfamily. The γ c receptor gene is also known as the X-linked severe combined immunodeficiency (SCID-X) gene.

Interleukin-3/Interleukin-5/Granulocyte-macrophage Colony Stimulating Factor (IL-3/IL-5/GM-CSF)

IL-3, IL-5 and granulocyte-macrophage colony stimulating factor (GM-CSF) form a subfamily of interleukins that are important regulators of hematopoietic progenitor and myeloid cell development. Activated T-cells are the main physiological source of IL-3 and

IL-5, but stimulated mast and endothelial cells, fibroblasts and monocytes also produce these interleukins. IL-5, and to a lesser extent IL-3, is a regulator of eosinophil proliferation/differentiation and murine IL-5 plays a role in immature B-cell differentiation/proliferation. Additionally, IL-3 and GM-CSF are best known for their important roles in regulating HSC and hematopoietic progenitor cell proliferation and the differentiation of these cells towards the myeloid/erythroid lineage. Furthermore, IL-3 plays a pivotal role in regulating the survival of hematopoietic cells, *via* the up-regulation of the anti-apoptotic Bcl-2 gene and the down-modulation of the pro-apoptotic factors.

Interleukin-6/Interleukin-11 (IL-6/IL-11)

The IL-6 subfamily of cytokines includes IL-6 and IL-11, oncostatin M (OSM), leukemia inhibitory factor (LIF), ciliary neurotrophic factor (CNTF) and cardiotrophin-1 (CT-1). IL-6 and IL-11 are produced by several cell types, including lymphocytes, monocytes, epithelial cells and keratinocytes. IL-6 and IL-11 regulate the differentiation of megakaryocytes and macrophages and, to a lesser extent, the differentiation and proliferation of T-cells and B-cells respectively. However IL-6 and IL-11 are better known for their ability to induce expression of ►acute phase proteins, such as C-reactive protein (CRP), serum amyloid A (SAA) and fibrinogen, in the liver. Moreover, IL-6 and IL-11 can act in a synergistic manner to induce proliferation of multipotent HSCs and progenitor cells.

Interleukin Receptors

A wide variety of interleukin receptors have been identified. Some interleukins have very distinct receptors, while other interleukins utilize common receptor subunits. Here we discuss some of the biochemical features of the interleukin receptors. Figure 2 provides a schematic representation of several receptors and Table 2 provides additional details on their characteristics (1, 2, 3).

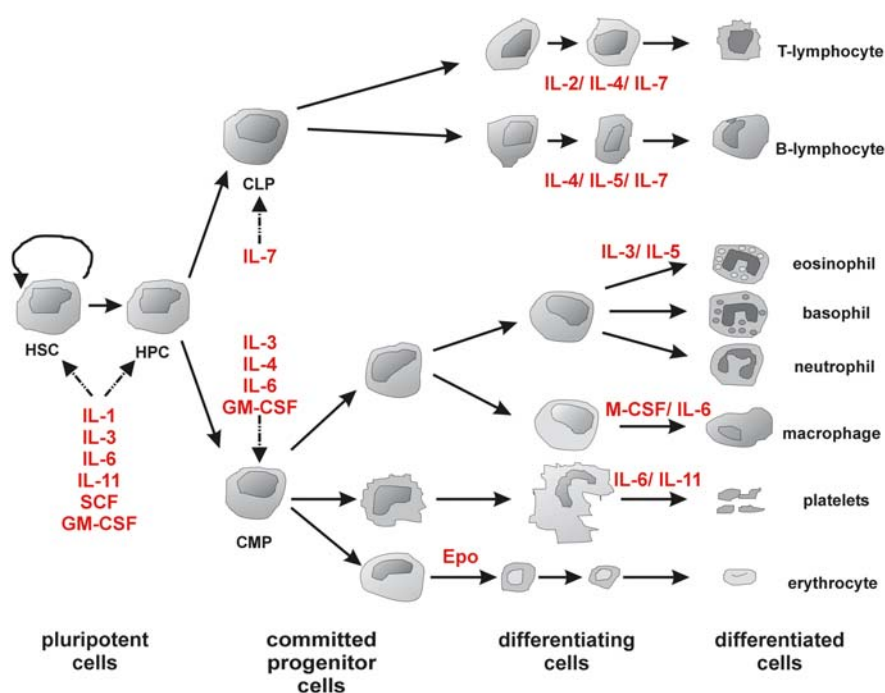
Within the interleukin receptor family, the IL-1 and IL-18 receptors are very distinct from other receptors. The IL-1 and IL-18 receptors contain several extracellular immunoglobulin (Ig)-like repeats and an intracellular ►Toll-IL1 receptor (TIR) domain involved in receptor proximal signal transduction events (Fig. 2a). The IL-1 receptor consists of ligand-binding subunit (receptor type 1; IL-1RI) and an accessory subunit (receptor associating protein; IL-1RAcP) that increases the affinity of the IL-1RI for its ligand. The type 2 receptor (IL-1RII) contains a short intracellular tail and lacks signal-propagating properties. In addition, soluble isoforms of the receptors have been identified and these receptors probably play a role in the availability of IL-1. Both the IL-1 and the IL-18 receptors activate similar ►signal transduction pathways, which could

Interleukins and Their Receptors, a Role in Regulating Hematopoietic Stem Cells. Table 2 Characteristics of Interleukin Receptors (1, 3)

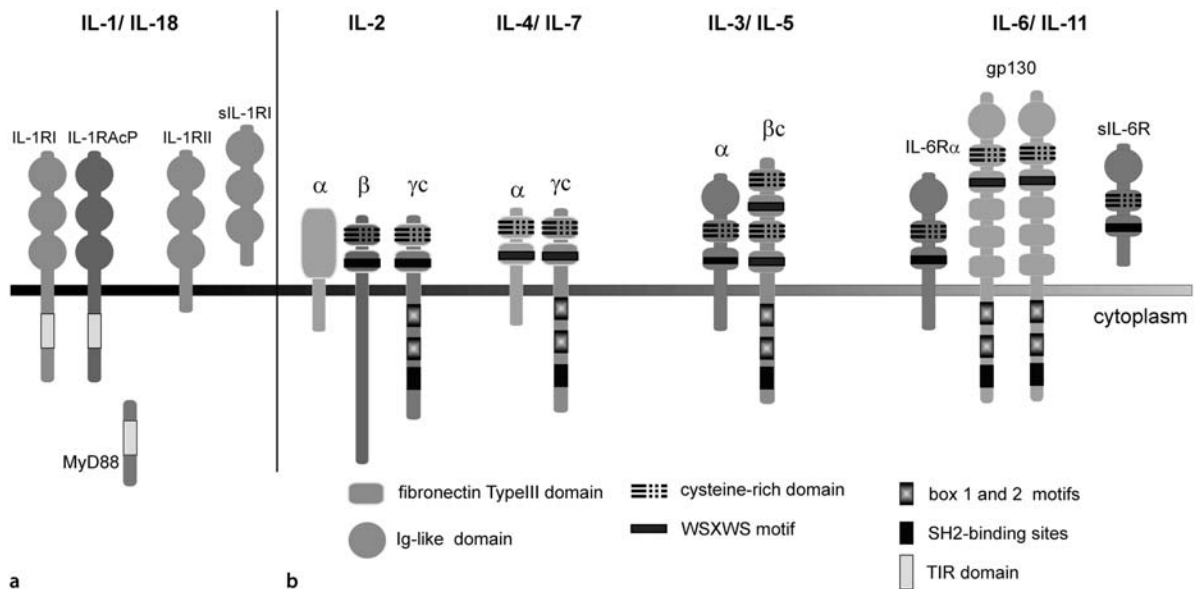
Interleukin	Protein size (kD)	Human chromosome	Mouse chromosome	CD number	Expression pattern	Remarks
IL-1RI	80	2	1	CD121a	T-cells mesenchymal cells	Signaling subunit associated with IL-1RAcP
IL-1RII	60	2	1	CD121b	B-cells granulocytes macrophages	Non-signaling subunit/ decoy receptor
IL-RAcP	66	3	16		see IL-IRI and RII	Accessory subunit
IL-2R α	55	10	2	CD25	(activated) T and NK cells	Ligand binding subunit
IL-2R β	75	22	15	CD122		Ligand binding subunit
IL-2R γ (yc chain)	64	X	X	CD132		Signaling subunit gamma-common chain
IL-3R α	150	X	14	CD123	Macrophages mast cells eosinophils/ basophils megakaryocytes HSCs/ HPCs	Ligand binding interacts with CSF2-R β
CSF2-R β (β c chain)	120	22	15	CD131		β common chain of IL-3R, IL-5R and CSF-R signaling subunit
IL-4R α	140	16	7	CD124	B-and T-cells NK cells HPCs	Ligand binding Extracellular domain is related to EPO-R, IL-6R and IL-2R β Associated with IL-2R γ chain
IL-5R α		3	6	CDw125	hematopoietic cells lymphoid cells	Ligand binding interacts with the β common chain of IL-3 receptor
IL-6R α	80	7	3	CD126	T-cells activated B-cells monocytes hepatocytes	Ligand binding subunit Resembles M-CSF, PDGF receptor interacts with gp130
IL-6 ST (gp130)	130	5	13	CD130		Signaling chain glycoprotein
IL-7R α	76	5	15	CD127	Activated T-cells T/ B-cell progenitors BM macrophages	Ligand binding associated with yc chain
IL-8R α (CXCR1)	59	2	-	CD128	neutrophils NK cells T-cells monocytes	Dimeric glycoprotein G-protein coupled receptor Important for chemotaxis related to stromal derived factor (SDF) and macrophage-inflammatory protein (MIP) receptors
IL-8R β (CXCR2)	67	2	1	CD128b		

Interleukins and Their Receptors, a Role in Regulating Hematopoietic Stem Cells. Table 2 Characteristics of Interleukin Receptors (1, 3) (Continued)

Interleukin	Protein size (kD)	Human chromosome	Mouse chromosome	CD number	Expression pattern	Remarks
IL-11R α	151	9	4	-	Megakaryocyte progenitor B- and T-cells HSCs/ HPCs	Ligand binding subunit associated with gp130 subunit
IL-18RI		2	1	-	Th1 cells Spleen /thymus cells Brain Various other tissues	Ligand binding subunit
IL-18RAcP		2	1			Accessory subunit



Interleukins and Their Receptors, a Role in Regulating Hematopoietic Stem Cells. Figure 1 Hematopoiesis is regulated by interleukins. A schematic representation of the hematopoietic hierarchy in the adult human and mouse providing an overview of the most prominent target cells of the interleukins discussed in this review. The hematopoietic hierarchy is founded by the hematopoietic stem cells (HSCs), which have the capacity to extensively self-renew and produce all the different mature hematopoietic cell types. HSCs, together with the hematopoietic progenitor cells (HPC), the common lymphoid progenitor (CLP) and the common myeloid progenitor (CMP), form the compartment of the immature hematopoietic cells. Cytokines that are involved in regulating HSCs and HPCs include interleukin (IL)-1, 3, 6 and 11, stem cell factor (SCF; also known as kit-ligand) and granulocyte-macrophage colony-stimulating factor (GM-CSF). For differentiation along the lymphoid lineages IL-7 plays an important role in regulating the CLP as well as T and B precursors. IL-2/4 and IL-4/5 play important roles in regulating respectively T-cell and B-cell proliferation and/or differentiation. IL-3/5/6 and GM-CSF play an important role in myeloid differentiation. IL-3 and 5 play a role in neutrophil/ eosinophil differentiation, while IL-6, GM-CSF and M-CSF regulate macrophage/ megakaryocyte differentiation. Although not addressed in this review, for clarity the central regulator of erythrocyte differentiation, erythropoietin-1 (Epo), is indicated.



Interleukins and Their Receptors, a Role in Regulating Hematopoietic Stem Cells. Figure 2 Interleukin Receptors. Schematic representation of several interleukin receptors and their structural and biochemical motifs. (a) The interleukin (IL)-1/18 receptor family has distinct features from the other interleukin receptors. The IL-1 receptor consists of two subunits, the type 1 receptor (IL-1RI) and the IL-1 receptor associating protein (IL-1RAcP). Both subunits contain several immunoglobulin (Ig)-like domains in the extracellular part of the receptor. Upon binding of IL-1 to the IL-1RI, the IL-1RAcP subunit is recruited to the receptor complex. IL-1RAcP increases the binding affinity of IL-1 for the IL-1RI and is involved in signaling. Upon ligand binding and receptor activation the signal transduction mediators interact with the IL-1RI and IL-1RAcP via the intracellular Toll/IL-1 receptor (TIR) domains. The MyD88 is an essential adapter protein for initiating signaling and this protein also contains a TIR domain to allow interactions with downstream signaling components. The IL-1 receptor type 2 (IL-1RII) and the soluble IL-1R (sIL-1R) receptor lack signal propagating properties and act as decoy receptors regulating the availability of the biological active ligands. (b) The IL-2/4/7, IL-3/5, IL-6/11 receptors all contain a ligand-interacting subunit, the α -chain and a signaling subunit. For the IL-2/4/7 subfamily the signaling subunit is the γ c chain. For the IL-3/5 and the IL-6 subfamilies these signaling subunits are the β c and the gp130 protein respectively. The extracellular part of the receptors consists of fibronectin type III and Ig-like structural domains with WSXWS motifs (forming a hinge structure) and the cysteine-rich motifs. The intracellular signal mediating domains include the box 1 and 2 motifs, to which the Janus kinases (JAKs) can bind. Moreover, several tyrosine residues are located in the intracellular tail of the receptors and upon phosphorylation these sites become docking sites for SH2-domain containing proteins.

explain the extensive overlap in the biological effects induced by these interleukins.

The receptors for IL-2/IL-4/IL-7, IL-3/IL-5/GM-CSF and the IL-6 subfamily share a number of characteristics (Fig. 2b). All the receptors consist of a specific ligand-binding α -chain, a common signal mediating receptor subunit and, in some cases, one or more auxiliary receptor subunits. For IL-2/IL-4/IL-7 the common subunit is named the gamma-common (γ c) chain, for the IL-3/IL-5 the signal mediating subunit is the beta-common (β c) subunit and the gp130 protein is the signaling subunit of all IL-6 subfamily receptors. These interleukin receptors share a number of structural (fibronectin type III), biochemical (WSXWS, cysteine-rich) and signaling (box motifs) domains (Fig. 2b). Since all interleukins within a subfamily signal *via* a common receptor subunit and the signaling domains in the intracellular part of the receptors are similar for all

mentioned subfamilies, it is not surprising that these receptors activate similar signaling pathways and overlap in biological function. In addition, it has been reported that soluble variants exist for several of these interleukin receptors. In most cases such soluble receptors are thought to play a role in regulating the bioavailability of the interleukins.

Molecular Interactions

Upon ligand binding, the intracellular domain of interleukin receptors becomes involved in signal transduction. However, interleukin receptors lack intrinsic kinase activity. Thus, other molecules, such as non-receptor protein kinases, are required to activate the signal transduction cascade. ► **Interleukin-1 receptor associated kinases (IRAKs)**, ► **Janus kinases (JAKs)** and members of the ► **Src kinase** family, such as Lck, Fyn and Lyn are located in close proximity to

interleukin receptors on the cytoplasmic side of the plasma membrane. These protein kinases are activated upon ligand binding and receptor activation. Subsequently, these kinases can directly phosphorylate and activate downstream signal transduction proteins. Alternatively, some kinases phosphorylate tyrosine residues on the intracellular part of the receptor to allow docking of SH2-domain signaling proteins that will in turn activate downstream targets (2, 4, 5).

IL-1/IL-18 Receptor Signaling

The IL-1/IL-18 receptors transduce a signal *via* a TIR domain that is a docking site for signaling proteins. Among the first proteins recruited to the activated receptor are the adapter protein MyD88 and IRAK family kinases, of which IRAK-4 seems to play a crucial role in IL-1 signaling. Once IRAKs are activated, ►TRAF (TNF receptor associated factors) adapter proteins and members of the ►MAPK family, such as ►TAK1 are recruited and activated. These signaling mediators will activate downstream signaling pathways, i.e. the nuclear factor kappa-B (NFκB) and the Jun kinase (JNK) pathways, which in turn will regulate the expression of IL-1 and IL-18 target genes.

IL-2, IL-3 and IL-6 Subfamily Receptor Signaling

Upon binding of these interleukins to their cognate receptors, one of the main signaling cascades activated is the Janus kinase/signal transducer and activator of transcription (►JAK/STAT) pathway. The ►Janus kinases are located in proximity to the intracellular tail of the interleukin receptors (near the box 1/2 motifs). They are activated upon receptor activation and phosphorylate their ligands, the STAT proteins. STAT proteins then translocate to the nucleus, bind to gene regulatory sites with the consensus sequence TTNCNNNAAA and subsequently activate their target genes. To date, four different mammalian JAKs (JAK 1/2/3 and Tyk2) and at least eight different STATs have been identified. A specific combination of JAK and STAT proteins is activated upon receptor activation by its cognate interleukin. This results in an activated signaling cascade and ultimately in a specific biological response. For example, the IL-2 receptor signals *via* JAK1/3 and STAT5, while the related IL-4 receptor signals *via* JAK1/2 and STAT6.

Besides the JAK/STAT pathway, Src-family kinases are also located in close proximity to interleukin receptors and are involved in the activation of downstream signaling mediators upon receptor activation. For example, the Src-kinases Lck, Lyn and Fyn play a role in IL-2 and IL-3 receptor signaling by phosphorylating tyrosine residues in the intracellular parts of these interleukin receptors. These phospho-tyrosine residues then become docking sites for SH2-domain signaling proteins (such as Shc and Grb2) that will activate

downstream targets (such as Ras-Raf). In addition, the Src-family kinases can directly phosphorylate downstream signal mediators.

Regulatory Mechanisms

As interleukins have a broad range of biological activities as well as numerous target cells, their expression requires tight regulation. Therefore, most interleukins are not expressed constitutively, but their expression is induced upon stimulation/activation of the producing cell with growth factors, cytokines or immune response inducing substances, such as pathogenic proteins. Moreover, some interleukins (including IL-1) regulate their own expression levels and that of their receptor, resulting in auto-regulatory feedback loops within the target cells. The expression of interleukins is regulated at several molecular levels, including the transcriptional and translational and at the protein level (glycosylation, proteolytic cleavage etc) (1, 3).

At the transcriptional level, several transcription factors were shown to regulate the gene expression of interleukins and their receptors. Among these are the ubiquitous gene expression regulators, such as members of the AP-1 family (jun, fos) and the CCAAT/enhancer binding protein (C/EBP). Transcription factors more restricted to the hematopoietic system are regulators of interleukin expression. For example, the ETS family member PU.1 plays an important role in the gene regulation of IL-1 and the β_c receptor. Another important hematopoietic transcription factor, Runx-1 (also known as AML-1) regulates IL-3 and GM-CSF expression. STAT and NFκB transcription factors play a role in the gene expression of multiple interleukins. Several interleukins are also regulated at the translational level. The mRNA for IL-1 contains stretches of AU rich sequences that prevent mRNA translation into functional IL-1 protein. However, upon activation of target cells with specific stimuli, this inhibitory effect can be overcome and functional protein can be produced. Similar mechanisms of regulation have been shown for other interleukins. Additionally, several interleukins require post-translational processing in order to become biologically active. For example, several interleukins contain signal peptides that direct the interleukins to the plasma membrane to allow secretion. Others are produced as precursor proteins and require proteolytic cleavage in order to be secreted or become biologically active. Proteolytic cleavage also regulates the subsequent degradation of the interleukins.

In addition to the regulation of gene expression, the affinity of interleukins receptors for their ligands can be regulated. Both high and low affinity receptors for a similar ligand have been described. Moreover, soluble receptors exist which in most cases compete with the

membrane-bound receptors for ligand binding and thereby prevent ligand-induced signaling. However, the soluble IL-6 receptor is an exceptional case. This soluble receptor can bind IL-6, interact with membrane-bound gp130 proteins and allow signaling in cell types that normally do not express the IL-6 receptor.

Summarizing Remarks

Prior to their molecular characterization, the biological activities of several interleukins and related cytokines and their effects on hematopoietic cells were already well known. Despite extensive research over the years, our knowledge of interleukins is incomplete and new interleukins and receptors are still being identified. Also, it is becoming more apparent that interleukins, besides their redundant biological functions, possess highly specific functions *in vivo*. Current interleukin research focuses on the elucidation of the interleukin signaling events and the regulation of interleukin target genes. The major challenge for the future will be to understand the interplay between interleukins and other cytokines and developmental factors within the hematopoietic system *in vivo*, especially within the compartment of the immature hematopoietic progenitor and stem cells.

References

1. Horst Ibelgauffs' Cytokines Online Pathfinder Encyclopaedia, Version 10.3 (July 2003), www.copewithcytokines.de
2. Helmreich EJ (2001) The Biochemistry of Cell Signaling. Oxford University Press, Oxford
3. Paul WE (2003) Fundamental Immunology. 5th edn. Williams & Wilkins Publishers, Philadelphia
4. Ihle JN, Kerr IM (1995) Jaks and Stats in signaling by the cytokine receptor superfamily. *Trends Genet* 11:69–74
5. Geijsen N, Koenderman L, Coffey PJ (2001) Specificity in cytokine signal transduction: lessons learned from the IL-3/IL-5/GM-CSF receptor family. *Cyt Growth Fact Rev* 12:19–25

Intermediate Filaments

E. BIRGITTE LANE

Division of Cell & Developmental Biology, University of Dundee School of Life Sciences, MSI/WTB Complex, Dundee, UK
e.b.lane@dundee.ac.uk

Definition

►Intermediate filaments are polymeric protein structures of the ►cytoskeleton and are found in most

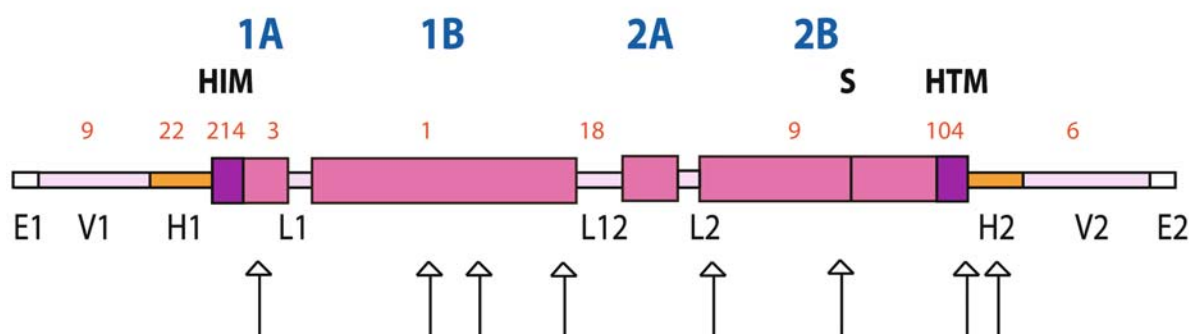
animal cells. Most types of intermediate filament proteins form a network of cytoplasmic filaments that are each about 10 nm in diameter and usually very long; ends are rarely seen. Their name comes from their intermediate size between actin and myosin filaments in muscle. They form one of the three fibrous systems of the cytoskeleton, the other two being microtubules (of tubulin) and actin microfilaments. Intermediate filament proteins differ from actin and tubulin systems in forming more stable filaments than either of them. They do not bind nucleotides and are not dependant on nucleotide hydrolysis to regulate their polymerisation and depolymerisation. The amount of intermediate filament protein in a cell varies widely, from less than 1% in hepatocytes of liver to around 80% of the total cell protein in differentiated ►keratinocytes of skin. Intermediate filaments are encoded by a large gene family and have distinct tissue expression patterns. They are found in nearly all cells of motile multicellular organisms and their common function is believed to be to provide physical resilience to cells in tissues. The best direct evidence for this comes from human skin fragility disorders caused by mis-sense mutations in the genes encoding ►keratin intermediate filament proteins. It is likely that different intermediate filament proteins possess different tissue-specific properties.

Characteristics

All intermediate filament proteins share the same basic secondary structure characteristics, shown in Fig. 1. Each has a central α -helical rod domain of around 310 residues, which coils to form an α -helix some 45 nm long. Only one group, the lamins, shows a significantly different rod length as these proteins have an additional 42 amino acid residues in helix 1B. The rod domains have a very consistent structure. They fall into 4 helical sections (1A, 1B, 2A, 2B) interspersed by 3 non-helical stretches (L1, L12, L2) and a fourth irregularity in 2B known as the "stutter". These non-helical linker domains are well conserved and are probably important for filament formation and flexibility. At either end of the rod domain are key conserved ►helix boundary motifs (Fig. 1) also needed for assembly; the helix termination motif has even been used to identify unknown intermediate filament genes.

The rod domain is flanked by non-helical amino-terminal head and carboxy-terminal tail domains, containing a region of quasi-repeats (V1 and V2) between charged end domains (E1, E2) and helix flanking regions that usually contain phosphorylation sites (H1, H2; see Fig. 1). The head and tail domains can vary hugely in size and are thought to determine the tissue-specific characteristics of different filament proteins.

The family of intermediate filament proteins is divided overall into six distinct types or classes,



Intermediate Filaments. Figure 1 Characteristics of intermediate filament proteins as illustrated with a type II keratin. All intermediate filament proteins have the basic structure as shown, with a central rod domain (pink: 1A, 1B, 2A, 2B), interspersed and flanked by non-helical regions (yellow, orange, pink). The conserved rod ends (dark pink) or helix boundary motifs are referred to as the helix initiation motif (HIM) and the helix termination motif (HTM) and are important for establishing the correct filament properties. Within the head and tail domains, E1 and E2 are the charged extreme ends; V1 and V2 are the variable quasi-repeat regions and H1 and H2 are the hypervariable regions that usually contain phosphorylation target sites for remodelling filaments by disassembly and reassembly. The non-helical linker domains are shown as L1, L12 and L2 and the stutter discontinuity in helix 2B is shown by the line labelled S. Numbers above the domains denote the number of published pathogenic mutations in keratins reported in that domain. Below is indicated the position of the intron-exon boundaries.

type I-type VI, based on similarities in their DNA coding sequences and the primary amino acid sequences of their proteins.

Expression

The spectrum of intermediate filament proteins and their expression in tissues was mapped using gel electrophoresis and antibodies (1, 2). Many well-characterised monoclonals are now available commercially as differentiation markers and tools for diagnostic pathology. There are over 60 intermediate filament proteins whose expression range in tissues is now known. The diversity of their expression is one of the most striking characteristics of the intermediate filament protein family and one of the most useful. Table 1 gives a brief overview of the diversity of known intermediate filament proteins and their expression in tissue differentiation, although a full account of tissue expression mapping of intermediate filament proteins is far beyond the scope of this article.

Type I and type II proteins are the keratins, or cytokeratins, the intermediate filament family expressed by the cells of **epithelia**. The classification of keratins as type I (smaller, acidic) and type II (larger, neutral to basic) proteins arises from original studies on wool proteins and this nomenclature has been expanded to cover all the intermediate filament proteins (3). There are thought to be 54 functional keratin sequences in the human genome and a significant proportion encode specialised hair keratins (trichocyte

keratins). The keratins are traditionally catalogued numerically by their charge and molecular weight. Hair keratins have a slightly different nomenclature. Keratins in tissues are coexpressed as pairs, from the largest type II keratin pairing with the largest type I keratin (K1 + K10, in differentiating epidermis), the next largest ones also pair, etc., down to the smallest pair (K8 + K18, in early embryos and simple epithelia). A potential revised nomenclature is currently under discussion (4).

Type III proteins are the vimentin-like group, expressed mostly in mesenchymal cells. Type IV are the neurofilament proteins in neurones. Type V are lamin proteins, expressed in the **nuclear lamina** of all cells and the only non-cytoplasmic group. The type VI group encompasses proteins that are clearly intermediate filament proteins but do not show obvious homology to the other groups; this group currently includes the divergent lens intermediate filament proteins, CP49 and filensin. Different members of each group are selectively expressed in different tissues and the intermediate filament expression profile of a cell is an indicator of the cell's differentiated state. This information is summarised in Table 1.

Intermediate filament genes are found widely throughout the animal kingdom, and proteins with related properties may occur elsewhere (e.g. some bacteria). They appear to be limited to motile multicellular organisms and cytoplasmic intermediate filaments are underrepresented in tissues protected by an exoskeleton.

Intermediate Filaments. Table 1 Intermediate filament proteins with known tissue expression

protein type	protein name (Mr)	Co-assembly	gene name + accession number	gene location	Main tissue expression
I (keratins)	K9 (64kD)	K1?	KRT9 NM_000226	17q21.2	Epithelia/ stratified / cornifying/ suprabasal/ palm and sole
	K10/11 (56kD)	K1	KRT10 J04029	17q21.2	Epithelia/ stratified / cornifying/ suprabasal
	K12 (55kD)	K3	KRT12 D78367	17q21.2	Epithelia/ stratified / cornea
	K13 (54kD)	K4	KRT13 X52426	17q21.2	Epithelia/ stratified / non-cornifying/ orogenital/ suprabasal
	K14 (50kD)	K5	KRT14 NM_000526	17q21.2	Epithelia/ stratified and complex/basal cells
	K15 (50kD)	K5?	KRT15 NM_002275	17q21.2	Epithelia/ stratified and complex/some basal cells
	K16 (48kD)	K6a	KRT16 NM_005557	17q21.2	Epithelia/ stratified / stress response and rapid turnover/ suprabasal
	K17 (47kD)	K6b	KRT17 NM_000422	17q21.2	Epithelia/ stratified / stress and fast turnover
	K18 (45kD)	K8, K7	KRT18 NM_000224	12q13.13	Epithelia/ Simple epithelia
	K19 (40kD)	K8	KRT19 NM_002276	17q21.2	Epithelia/ Simple epithelia; stratified/ some basal
	K20 (46kD)	K8, K7	KRT20 BC031559	17q21.2	Epithelia/ simple epithelia/ some gastrointestinal regions
	Ha1 (44kD, *47.2kD)	type II tricho- keratins	KRTHA1 X86570	17q21.2	Epithelia/ stratified/ hair follicle/ trichocytes
	Ha2 (42kD, *50.3kD)		KRTHA2 NM_002278	17q21.2	Epithelia/ stratified/ hair follicle/ trichocytes
	Ha3a (*45.9kD)		KRTHA3-I NM_004138	17q21.2	Epithelia/ stratified/ hair follicle/ trichocytes
	Ha3b (*46.2kD)		KRTHA3-II X82634	17q21.2	Epithelia/ stratified/ hair follicle/ trichocytes
	Ha4 (43kD, *44.7kD)		KRTHA4 NM_021013	17q21.2	Epithelia/ stratified/ hair follicle/ trichocytes
	Ha5 (*47.1kD)		KRTHA5 NM_002280	17q21.2	Epithelia/ stratified/ hair follicle/ trichocytes
	Ha6 (*52.2kD)		KRTHA6 NM_003771	17q21.2	Epithelia/ stratified/ hair follicle/ trichocytes
	Ha7 (*45.5kD)		KRTHA7 NM_003770	17q21.2	Epithelia/ stratified/ hair follicle/ trichocytes
	Ha8 (*46.3kD)		KRTHA8 NM_006771	17q21.2	Epithelia/ stratified/ hair follicle/ trichocytes

Intermediate Filaments. Table 1 Intermediate filament proteins with known tissue expression (Continued)

protein type	protein name (Mr)	Co-assembly	gene name + accession number	gene location	Main tissue expression
II (keratins)	K1 (68kD)	K10	KRT1 NM_006121	12q13.13	Epithelia/ stratified / cornifying/ suprabasal
	K2e (65.5kD)		KRT2A AF019084	12q13.13	Epithelia/ stratified / cornifying/ high suprabasal
	K2p (65kD)		HUMCYT2A M99063	12q13.13	Epithelia/ stratified / cornifying/ oral/ suprabasal
	K3 (63kD)	K12	KRT3 NM_057088	12q13.13	Epithelia/ stratified / cornea
	K4 (59kD)	K13	KRT4 NM_002272	12q13.13	Epithelia/ stratified / noncornifying/ oral/ suprabasal
	K5 (58kD)	K14 K15?	KRT5 NM_000424	12q13.13	Epithelia/ stratified and complex/basal cells
	K6a (56kD)	K16	KRT6A NM_005554	12q13.13	Epithelia/ stratified/stress and rapid turnover
	K6b (56kD)	K17	KRT6B NM_005555	12q13.13	Epithelia/ stratified/stress and rapid turnover
	K6hf	?	NM_004693	12q13.13	Epithelia/ stratified/ hair follicle
	K6irs1 (*57.3kD)	?	KRT6IRS AJ308599	12q13.13	Epithelia/ stratified/ hair follicle/ inner root sheath
	K6irs2 (*55.9kD)	?	K6IRS2 NM_080747	12q13.13	Epithelia/ stratified/ hair follicle/ inner root sheath
	K6irs3 (*59kD)	?	K6IRS3 AJ508776	12q13.13	Epithelia/ stratified/ hair follicle/ inner root sheath
	K6irs4 (*57.9kD)	?	K6IRS4 AJ508777	12q13.13	Epithelia/ stratified/ hair follicle/ inner root sheath
	K7 (52kD)	K18, K19	KRT7 NM_005556	12q13.13	Epithelia/ Simple epithelia
	K8 (52kD)	K18	KRT8 X74929	12q13.13	Epithelia/ Simple epithelia
	Hb1 (*54.8kD)	type I tricho-keratins	KRTHB1 NM_002281	12q13.13	Epithelia/ stratified/ hair follicle/ trichocytes
	Hb2 (*56.6kD)		KRTHB2 NM_033033	12q13.13	Epithelia/ stratified/ hair follicle/ trichocytes
	Hb3 (*54.1kD)		KRTHB3 NM_002282	12q13.13	Epithelia/ stratified/ hair follicle/ trichocytes
	Hb4 (*64.9kD)		KRTHB4 NM_033045	12q13.13	Epithelia/ stratified/ hair follicle/ trichocytes
	Hb5 (*55.7kD)		KRTHB5 NM_002283	12q13.13	Epithelia/ stratified/ hair follicle/ trichocytes
	Hb6 (*53.5kD)		KRTHB6 NM_002284	12q13.13	Epithelia/ stratified/ hair follicle/ trichocytes

Intermediate Filaments. Table 1 Intermediate filament proteins with known tissue expression (Continued)

protein type	protein name (M _r)	Co-assembly	gene name + accession number	gene location	Main tissue expression
III (vimentin-like)	Vimentin (55kD)	self	VIM NM_003380	10p13	Mesenchyme and other/ very widespread
	GFAP (glial fibrillary acidic protein) (58kD)	self	GFAP NM_002055	17q21	astroglial cells
	desmin (53kD)	self	DES NM_001927	2q35	Muscle (all types)
	Peripherin (58kD, 61kD)	self	PRPH NM_006262	12q12-q13	peripheral nervous system; some CNS
IV (neuro-filaments)	NF-H (200kD)	NF-L	NEFH NM_021076	22q12.2	Neurones
	NF-M (150kD)		NEFM	8p21	Neurones
	NF-L (68kD)	self	NEFL NM_006158	8p21	Neurones
	Nestin (220-240kD)	vimentin	NES NM_006617	1q23.1	Neuroepithelial stem cells, various
	α -internexin (66kD)	self	INA	10q24.33	
	Synemin (230kD)	desmin?	DMN	15q26.3	Muscle
	Syncoilin (*53.6kD)	desmin, nestin?	SYNC NM_030786	1p35.1	Muscle
V (lamins)	Lamin A (70kD)	lamin	LMNA NM_170707	1q21.2	Nuclei
	Lamin C (63kD)	lamin	LMNA NM_170707	1q21.2	Nuclei
	Lamin B1 (67kD)	lamin	LMNB1 NM_005573	5q23.2	Nuclei (less differentiated)
	Lamin B2	lamin	LMNB2 NM_032737	19p13.3	all nuclei (less differentiated)
	filensin (115kD)	CP49	BFSP1 NM_001195	20p12.1-p11.23	Eye lens
VI (other)	CP49/ phakinin (49kD)	filensin	BFSP2 NM_	3q21-25	Eye lens

*Calculated molecular mass; otherwise molecular weights are given as relative migration on gels. Eight further type I and five further type II gene sequences are present but their expression range is unclear

Genes

The genes are generally small and compact, with 8–9 exons in most genes separated by 7–8 introns. Within each type the gene organization is very similar. Keratin genes are located within two dense gene clusters but other intermediate filament genes are scattered

elsewhere throughout the genome (Table 1). Genes have now been identified for all known intermediate filament proteins. A systematic search of the human genome database in 2001 identified 65 sequences as intermediate filament genes (5), although not all of these had yet been identified in tissues. In the human

genome, type I keratins are on chromosome 17q and type II genes are all except one on chromosome 12q12-13. The gene for type I keratin 18 lies next to its co-expressed partner keratin K8 and is the only type I keratin on chromosome 12. The chromosomal location of the intermediate filaments genes is summarized in Table 1.

It now seems likely that all the intermediate filament genes have been identified. Genome analyses have confirmed that there are multiple genes for K1, K5, K6, and K2. Alternative splicing is found in the lamin A/C gene, LMNA, (which encodes both lamin A and lamin C), and in peripherin, nestin, syncoilin and synemin. Polymorphisms mostly involve the variable head and tail domains. Intronless, or processed, **pseudogenes** for the two embryonic keratins K8 and K18 are abundant and dispersed throughout the human genome and account for 90% of all the keratin sequences in the draft human genome sequence. A small number of pseudogenes have also been identified for K17, K19, K14 and K16.

It is thought that the mammalian intermediate filament genes probably evolved from a lamin-like precursor in a pre-vertebrate ancestor, as invertebrates have filament genes with a lamin-like extended helix. It is suggested that loss of the isoprenylation target sequence that leads to anchoring of lamins in the nucleus might have been sufficient to initiate the evolution of a cytoplasmic type of intermediate filament. Proteins with resemblance to type I/II keratins and type III intermediate filaments are certainly all present in early chordates.

Molecular Interactions

By electron microscopy, intermediate filaments in cells all look rather similar, forming long, sinuous filaments that traverse the cell cytoplasm, singly or in bundles, with visible free ends hardly ever seen in a normal cell. The filaments themselves are **apolar**, unlike actin filaments or microtubules and they do not have a directional “organising centre” in the cell. They do however interact with cell-cell junctions (**desmosomes**, through desmoplakin) and cell-substrate junctions (**hemidesmosomes**, through BP230 and plectin **plakins** or cytolinker proteins) at the cell membrane. Interactions with the cytolinker/plakin proteins also occur elsewhere in the cytoplasm.

Assembly

Studies with recombinant intermediate filament proteins confirm that they can polymerize spontaneously and rapidly without the need for additional cofactors or chaperones. Keratin assembly is especially fast. The α -helical rod domain forms the building block for assembly of the 10 nm-thick filaments. It is highly charged, with alternating positive and negative patches

on its surface, and has a strip of hydrophobic residues curving round its length. It is therefore unstable until it forms a dimer or higher order structure. Monomeric intermediate filament polypeptides start to assemble by dimerisation, 2 polypeptides forming a left-handed coiled-coil, parallel and in register, along their α -helical rods. This is through a conventional “**leucine zip**” mode of progressively interdigitating hydrophobic residues down the helical coil, probably initiated by key sequences at the ends of the rod domain. It is at this dimer formation stage that type I and type II keratins have to interact, as keratin homodimers are unstable and short-lived, and cannot assemble into filaments, thus keratins are obligate heteropolymers. Transfection experiments in cells have shown that if no complementary partner protein is available for a keratin to assemble with, the protein is rapidly degraded. The dimeric molecules associate into tetramers with a preferred antiparallel half-stagger configuration. Tetramers assemble laterally and longitudinally into filaments in a sequence of events that is not yet clear. From studies *in vitro*, Herrman and colleagues have proposed that this assembly passes through a stage of forming short fat “unit length filaments” which connect longitudinally as a string of beads, and finally undergo a compaction stage to form a 10 nm thick intermediate filament (6).

Certain other intermediate filament proteins also require mixtures of protein to form filaments efficiently. Neurofilaments NF-M and NF-H need NF-L, although NF-L can polymerise alone; eye lens proteins filensin and CP49 need each other, but with different stoichiometry from the keratins. Current thinking on the identity of preferred assembly partners for different intermediate filament proteins is indicated in Table 1.

Regulatory Mechanisms

Cellular and Molecular Regulation

Intermediate filaments are relatively stable structures in comparison to microtubules and actin filaments, yet there is clearly a lot of flux of subunits on and off the length of filaments in living cells. Filament assembly is controlled inside cells by phosphorylation (favouring disassembly) and dephosphorylation, mostly on serine and threonine residues located in the non-helical head and tail domains, close to the ends of the rod domain. Although many potential phosphorylation sites and kinases can be identified it is less clear which ones are important *in vivo*. Lamins are phosphorylated by p34 cdc2 at mitosis on sites in the head and tail domains for disassembly of the nuclear lamina at mitosis and during interphase they are targets for protein kinase A and C sites in the tail domain. The lamins are also subject to isoprenylation at their C terminus, which is necessary

for their correct intranuclear localization. Several serine phosphorylation sites in keratin heads and tails are now known [see review by Omary et al (7)] and again their main purpose seems to be remodelling during the cell-cycle. Phosphorylation may also regulate keratin interaction with chaperones. Neurofilament proteins can be heavily phosphorylated, especially NF-H, on repeated KSP sequence motifs in the tail domain. This leaves the tail domain highly charged so that it extends at right angles from the axis of the filament and this may be important for spacing filaments in the axon. At least some of the keratins are also subject to glycosylation, possibly affecting their solubility, and some intermediate filament proteins interact with chaperone molecules such as 14-3-3 or α B crystallin.

Disorders of Intermediate Filaments

Mutations in intermediate filament genes cause a wide range of genetic disorders. A web database has been set up to collate information on these mutations (at ►<http://www.interfil.org>), maintained at the University of Dundee. The first discoveries of intermediate filament mutations were made in a skin disorder, ►[epidermolysis bullosa simplex](#) (EBS). EBS is a skin blistering disease in which the keratinocytes of the epidermis are fragile due to mutations in the genes for either K5 or K14, the two keratins expressed in basal cells of the epidermis. Cells with these pathogenic mutations may lyse on scratching, rubbing or any traumatic contact, leading to intraepidermal fluid-filled blisters and sores on the skin. ►[Hyperkeratosis](#) also often develops on the hands and feet. These blisters will normally heal without scarring.

Pathogenic mutations are distributed non-randomly along the protein, as shown in Fig. 1. The most severe cases are caused by mutations arising in the helix boundary motifs that mark the ends of the rod domain, especially the type-specific helix initiation motif, and give rise to the Dowling-Meara form of EBS. Milder clinical phenotypes (Koebner and Weber-Cockayne forms) are associated with mutations in different hotspots in K5/K14, presumably reflecting regions where the detrimental effect on filament function is not quite so drastic.

Fragility of other epithelial cells is linked to mutations in different keratin genes (see www.interfil.org and references therein for accumulated information on keratin mutations). In the epidermis, mutations in the secondary (differentiation-specific) keratins K1 and K10 lead to bullous congenital ichthyosiform erythroderma (BCIE), so-called because of the formation of blisters and later the thickened ichthyotic (fish scale like) appearance of the skin, plus the reddening (erythroderma) that often accompanies this condition especially in children. K1/K10 disorders are also

sometimes called epidermolytic hyperkeratosis, referring to the major phenotypic characteristic of this disorder. Mutations in K2e cause ichthyosis bullosa of Siemens (IBS), whereas in K3 or K12 they cause Meesman corneal dystrophy. In K4 or K13, mutations cause oral and genital mucosal lesions of white sponge naevus (WSN). In K6a/K6b genes or K16 or K17, mutations are responsible for two forms of ►[pachyonychia congenita](#), PC-1 and PC-2, or the related and usually milder disorder of steatocystoma multiplex. Mutations in the palm- and sole-specific keratin K9 cause palmoplantar keratoderma. Nearly all of these are dominantly inherited and are caused by missense or nonsense mutations, but there are also several documented cases of recessive knockout mutations in K14 and some cases of compound heterozygotes. Some mutations in intron sequences have also been reported. The clustering of mutations, which tend to be the same regions in different keratin genes, must reflect the importance of conserving sequence to retain function. Therefore there must be other mutations in the rest of the keratin sequences that do not give rise to significant pathology.

Mutations have also been identified in the simple epithelial keratins K8 and K18, associated with disorders in liver, pancreas and intestinal epithelium (inflammatory bowel disease), but here the genetic link with disease is less clear. Inflammatory bowel disease is known to be of polygenic aetiology and moreover does not present clinically until the damage to the epithelium is already substantial, leaving less opportunity for morphological analysis. Finally, a number of mutations in the hair keratins that cause hair defects are also being uncovered. Disease-associated mutations remain to be uncovered in keratins K7, K15, K19 and K20. Hair disorders are also now beginning to be linked to keratin mutations. There are now altogether more than 400 published keratin mutations (www.interfil.org) and the field has recently been reviewed by several authors [e.g. Irvine & McLean (8)].

Outside the keratins the list of intermediate filament disease associations is also growing. Mutations in the type III desmin gene are associated with cardiomyopathies, disorders in which the stresses induced by constant contraction and relaxation of the muscles are thought to lead to cellular disruption analogous to the situation in skin keratinocytes. Type III mutations have also been seen in GFAP in patients with Alexander disease, a severe neurodegenerative condition that seems to be triggered and accelerated by incidences of brain trauma in the susceptible. Brain injury induces glial cell activation as part of the wound healing process and glial cells without intermediate filaments have been shown to behave aberrantly, which is

probably at the heart of the Alexander disease phenotype. Type V (lamin) genes are also implicated in disease. Mutations in the nuclear lamin A (A/C) gene are clearly responsible for a variety of pathologies ranging from familial partial lipodystrophy through Charcot-Marie-Tooth disease type 2B1 to Hutchinson-Gilford progeria: presumably loss of the reinforcement of the nuclear envelope in dividing cells leads to lethal cell damage in the presence of other undetermined stochastic factors.

Abnormalities in neurofilaments are seen in ALS and a number of other neurodegenerative disorders but the role of neurofilament mutations is mostly still obscure, as these are often late onset, variable severity disorders not amenable to genetic analysis. However mutations in NF-L have now been definitively linked to human Charcot-Marie-Tooth disease type 2E, a progressive muscle wasting disorder. In the eye lens, deletion of the CP49 gene in mice leads to loss of clarity of the lens, although no pathogenic mutations in the lens intermediate filament genes have yet been found in humans. Listing of mutations reported in intermediate filament genes can be found at ►<http://www.interfil.org>, and details on the individual disorders are recorded in ►OMIM (On-line Mendelian Inheritance in Man).

►Breast Cancer

►Mesenchymal Cells

References

1. Moll R, Franke WW, Schiller DL et al (1982) The catalog of human cytokeratins: patterns of expression in normal epithelia, tumors and cultured cells. *Cell* 31:11–24
2. Sun TT, Eichner R, Nelson WG et al (1983) Keratin classes: molecular markers for different types of epithelial differentiation. *J Invest Dermatol* 81:109s–115s
3. Geisler N, Weber K (1982) The amino acid sequence of chicken muscle desmin provides a common structural model for intermediate filament proteins. *EMBO J* 1:1649–1656
4. Hesse M, Zimek A, Weber K et al (2004) Comprehensive analysis of keratin gene clusters in humans and rodents. *Eur J Cell Biol* 83:19–26
5. Hesse M, Magin TM, Weber K (2001) Genes for intermediate filament proteins and the draft sequence of the human genome: novel keratin genes and a surprisingly high number of pseudogenes related to keratin genes 8 and 18. *J Cell Sci* 14:2569–2575
6. Herrmann H, Haner M, Brettel M et al (1999) Characterization of distinct early assembly units of different intermediate filament proteins. *J Mol Biol* 286:1403–1420
7. Omary MB, Ku NO, Liao J et al (1998) (eds) Keratin modifications and solubility properties in epithelial cells and *in vitro*. In: Intermediate filaments. Plenum Press, New York and London
8. Irvine AD, McLean WH (1999) Human keratin diseases: the increasing spectrum of disease and subtlety of the phenotype-genotype correlation. *Br J Dermatol* 140: 815–828

Intermediate Mesoderm

Definition

Intermediate mesoderm describes a specific region of embryonic mesoderm that forms the kidney and the sex organs.

►Kidney

Intermembrane Space

Definition

Intermembrane space is the region between the outer and the inner membrane of mitochondria.

►Mitochondria

Internal Ribosomal Entry Sites

Definition

Several eukaryotic cell and viral mRNAs initiate translation by an alternative mechanism, which involves internal initiation rather than cap-dependent translation initiation. These mRNAs contain complex nucleotide sequences, called internal ribosomal entry sites (IRES) where ribosomes bind in a cap-independent manner and start translation at the closest downstream AUG codon. As an IRES element can direct binding of ribosomes to the start codons of multiple genes on a single message, it can support coexpression of more than one protein under the control of a single promoter.

►CAP-independent Translational Control

►DNA-based Vaccination

►RNA Capping

►Translational Control in Eukaryotes

Interneuron

Definition

In general, the term ‘interneuron’ is used for neurons with a local axon plexus (projection approximately 0.5 mm) in the brain. Interneurons in the brain may show axonal projection both within a lamina and between laminae. In the neocortex, the majority of interneurons are inhibitory and use (-amino butyric acid as a

neurotransmitter. Exceptions to this rule are e.g. neocortical spiny stellate neurones and cerebellar granule cells, both of which are excitatory neurons.

► [Neurons](#)

Internode

Definition

Internode describes a myelinated axon segment between two nodes of Ranvier, up to 1 mm in length.

► [Glial Cells and Myelination](#)

Interphase

Definition

Interphase describes the long period of the cell cycle following mitosis, in which chromosomes are decondensed and general metabolism occurs. Interphase includes G1, S and G2 phases.

- [Cell Cycle – Overview](#)
- [Mitotic Recombination](#)
- [Nuclear Compartments](#)

Interpolar Microtubule

Definition

Interpolar microtubule refers to antiparallel microtubules that originate from opposite poles, and overlap at the midzone of mitotic spindles.

- [Cytoskeleton: Microtubules and Intermediate Filaments](#)
- [Mitotic Spindle](#)

Interspersed Repeats

Definition

Interspersed repeats are randomly inserted copies of transposable elements.

- [Repetitive DNA](#)
- [Transposons](#)

Intervertebral Disc

Definition

Intervertebral disc is a connective tissue that is located between the individual skeletal elements of the vertebrae.

- [Bone and Cartilage](#)

Intrabodies

Definition

Intrabodies are specifically designed antibodies which can be expressed intracellularly (e.g. in the cytoplasm).

- [Monoclonal Antibodies](#)

Intron

Definition

Intron refers to a DNA segment of a eukaryotic gene that is not represented in the mature mRNA transcript, because it is removed from the primary transcript in a process known as splicing. Introns often make up the bulk of sequence in the genes of higher eukaryotes.

- [Alternative Splicing](#)
- [Base Excision Repair](#)
- [Familial Hypercholesterolemia](#)
- [Full Length cDNA Sequencing](#)
- [Mutagenesis Approaches in Medaka](#)
- [Mutagenesis Approaches in the Zebrafish](#)
- [Recombinant Protein Production in Mammalian Cell Culture](#)
- [Splicing](#)

Intronic Mutation

Definition

Intronic mutation is a mutation that occurs within the

- [intron](#) of a gene.
- [Mutagenesis Approaches in Yeast](#)

Intronic Sequence Element

Definition

Intronic sequence element describes a short regulatory sequence that is located in an ► [intron](#).

► [Alternative Splicing](#)

Inversin

Definition

Inversin is a protein of unknown function in which a deletional mutation gives rise to the “inversion of embryo turning” (inv) mouse. Inversin localizes to the cilia, and the inv mutant mouse has a renal cystic phenotype.

► [Autosomal Dominant, Polycystic Kidney Disease](#)

Ion Channels/Excitable Membranes

HEINRICH TERLAU¹, FRANK KIRCHHOFF²

¹Institute of Experimental and Clinical Pharmacology and Toxicology, University Clinic of Schleswig-Holstein, Campus Lübeck, Lübeck, Germany

²Department of Neurogenetics, Max Planck Institute of Experimental Medicine, Göttingen, Germany
heinrich.terlau@uni-luebeck.de
kirchhoff@em.mpg.de

Definition

Investigation of electrical excitability in vertebrates has a long history. It started with the pioneering experiments of Alessandro Volta on frog muscles in the 18th century. Today electrical excitability of tissues or cells still fascinates life scientists, irrespective of whether they are physiologists, neurobiologists or biochemists. Thanks to numerous studies, we now have a rather detailed knowledge about the molecular and cellular mechanisms of electrical activity as well as their functional consequences (1).

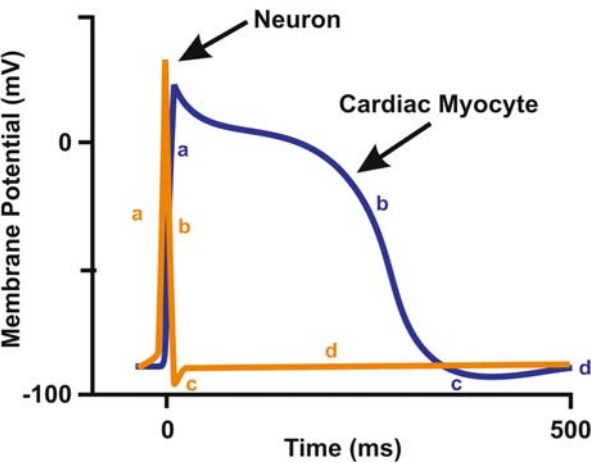
Membrane potential changes and secretion are the most prominent ways of information transfer in living organisms. Almost all cell types exhibit a voltage

difference, i.e. an electrical potential difference, across their membranes, which is negative at the cytoplasmic side compared to the extracellular space. This membrane potential (normally ranging from -90 to -50 mV) is mainly due to an uneven distribution of Na^+ , K^+ and Ca^{2+} ions at the extra- and intracellular sides of the plasma membrane, which harbors ion selective pores. Under resting conditions, the membrane potential is the equilibrium result of ion fluxes along chemical and electrical gradients. The negative voltage prevents small negative anions from entering the cytoplasm and thereby causing cell swelling. The electrochemical gradient is used by several transporters as an energy source for transporting metabolites or ions across the cell membrane. In electrically excitable cells (predominantly neurons and myocytes), transient potential changes and secretion of transmitters, hormones or growth factors convey cell-to-cell signaling, which represents the main mechanism for communication between these cells and their neighbors. The action potential, which defines the electrical excitability of a cell, is a fast redistribution of electrical charge across the membrane. In neurons and in myocytes, it can be divided into several phases which are characterized by different ion fluxes (Fig. 1): (a) sudden depolarization after reaching a certain voltage threshold, (b) repolarization, (c) hyperpolarization and (d) rest or equilibrium. In some cases, the repolarization phase is preceded by a plateau phase, which lasts a few ms in neurons and can last several 100 ms in cardiac myocytes.

Here we will focus on the molecular entities mediating these membrane potential changes underlying neuronal and muscle activity, i.e. voltage-gated Na^+ , K^+ and Ca^{2+} ion channels.

Characteristics

Neuronal action potentials can be initiated either at the afferent synapses or at the axon hillock of efferent ► [neurons](#), the most proximal part to the neuronal cell soma. The depolarizing upstroke of the action potential is mediated by Na^+ channels, the repolarization by K^+ channels. At synaptic terminals, action potentials activate voltage-gated Ca^{2+} channels. The subsequent influx of Ca^{2+} ions is a prerequisite for the vesicular release of transmitters, resulting in an activation of postsynaptic receptors. In non-myelinated axons, action potentials propagate as waves (diffusion of Na^+ ions) with a speed of about 1 m/s. In myelinated fibers, axons are electrically insulated by several neighboring lipid-rich myelin sheaths (a few hundred μm in length) intermitted by nodes of Ranvier. These nodes are characterized by a high-level expression of voltage-gated Na^+ and K^+ channels. Myelinated fibers are characterized by saltatory action potential propagation from one node to the other at ten- to hundred-fold higher



Ion Channels/Excitable Membranes. Figure 1 Time course of a neuronal and a cardiac action potential. Both types of action potentials (orange: neuron, blue: cardiac myocyte) are characterized by the same phases: (a) threshold-dependent depolarization, (b) repolarization (fast for a neuron, slow for the myocyte), (c) period of hyperpolarization and (d) resting state.

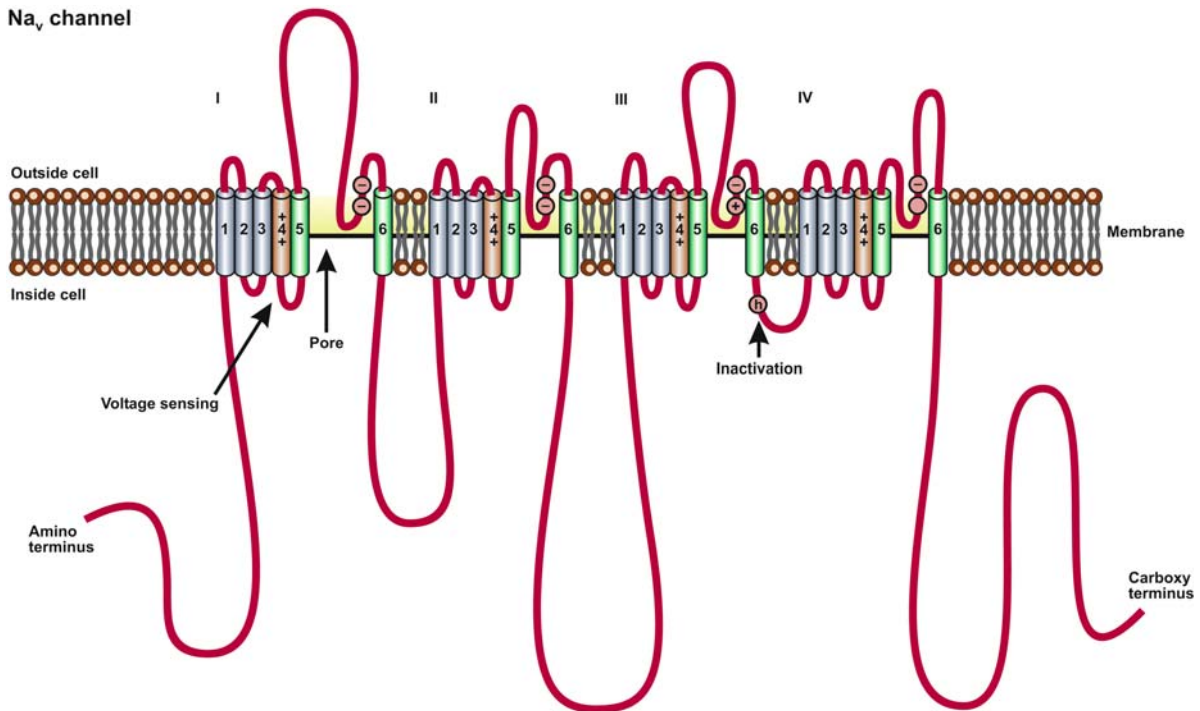
speeds compared to unmyelinated fibers. In cardiac myocytes, additional Ca^{2+} channels prolong the length of the depolarizing phase and mediate the substantial Ca^{2+} increase needed for muscle contraction.

Molecular Interactions
Voltage-gated Ion Channels

The superfamily of human voltage-gated and voltage-gated-like ion channels comprises a large set of 143 membrane-bound proteins that can be divided in 11 subfamilies according to sequence similarity (2), Na_v , Ca_v , K_v1-9 , K_v10-12 , K_{ir} , K_{Ca} , K_{2P} , CNG, HCN, TRP and TPC. They can all contribute to the membrane potential by their different selectivities for mono- and di-valent cations, depending on subcellular, cell- and age-selective expression, as well as by their modulation *via* accessory subunits and/or intracellular second messenger systems. According to their structure, these channels exhibit four variations (Fig. 2). (i) The Na_v and Ca_v channels consist of a principle, channel-forming α subunit composed of four homologous domains (Fig. 3). Each domain contains six membrane-spanning α helices (termed S1 – S6) and a membrane-reentrant loop (P loop) between the S5 and S6 segment, constituting the wall of the central ion pore. The S4 segment is an important part of the voltage sensor. (ii) K_v channels are composed of tetramers of subunits, which each resemble a single homologous domain of Na_v or Ca_v channels. (iii) K_{ir} channels display the simplest structure constituted by two transmembrane regions M1 and M2 separated by a P

Channel	Structure
Na_v Ca_v	
K_v1-9 K_v10-12	
K_{Ca} CNG HCN	
TRP	
TPC	
K_{ir}	
K_{2P}	

Ion Channels/Excitable Membranes. Figure 2 Structural building blocks of voltage-gated and voltage-gated-like ion channels. Na_v and Ca_v channels are formed by a single pore-forming α subunit only. Although evolutionarily separated, K_v as well as transient receptor potential (TRP) channels are composed of tetramers containing six transmembrane regions and the pore forming re-entrant loop. The voltage-sensing S4 segment is less conserved in TRP channels. Ca^{2+} -activated (K_{Ca}), cyclic nucleotide-gated (CNG) channels and hyperpolarization-activated, cyclic nucleotide-gated (HCN) channels are similarly constructed. Their α subunits carry a regulatory domain at the C-terminus. Recently discovered two-pore channel (TPC) proteins seem to be fusions of two TRP proteins. Inwardly rectifying K_{ir} channels represent the simplest ion channel family. Two-pore region (K_{2P}) channels form functional channels as dimers (adopted from 2).

Na_v channel

Ion Channels/Excitable Membranes. Figure 3 Detailed view of the Na_v channel α subunit. The long polypeptide chain of more than 260 kD is composed of four homologous domains each containing the pore-forming reentrant loop between α -helical segments S5 and 6. The voltage sensor is located in S4.

loop. (iv) Linking two of these structures together generates the architecture of K_{2P} channels. Only those channels that possess the S4 segment are voltage-gated in the strict sense. These are the Na_v, Ca_v and K_v channels.

Gating of Voltage-gated Ion Channels

During their activity, these proteins undergo conformational changes that are a prerequisite for ion permeation and thereby crucial for their physiological function. Under resting conditions, voltage-gated ion channels are usually in a closed conformation. Upon activation, the ion channel undergoes a conformational change ("gating") that opens a permeation pathway for cations (the pore) resulting in a selective permeation of Na⁺, K⁺ or Ca²⁺. From this so-called open state, these proteins can go into an inactivated state by an additional conformational change in which the protein is non-conductive. Upon repolarization to resting conditions, these proteins are deactivated by returning into the closed state.

The physiologically relevant ion channels are protein complexes that are known to interact with and be regulated by a whole variety of different proteins and other signaling molecules. Accessory subunits are

known to alter the current characteristics of a given channel but they are not usually part of the ion channel pore.

Na_v Channels

At resting potential, the probability of Na_v channels being in the open state is very low. Upon depolarization, the open channel probability increases, which under physiological conditions leads to the permeation of Na⁺ ions into a cell, which in turn leads to a depolarization of the membrane potential. Upon opening, Na_v channels inactivate fast. Accordingly, the proper kinetics of activation and inactivation of Na⁺ currents are crucial for the occurrence of the upstroke depolarization (Fig. 1a) and the shape of the action potential. To date, 10 different isoforms of α subunits with different developmental and tissue distributions have been identified and 9 of them fall into a single gene family.

K_v Channels

K_v channels are a heterogeneous group of ion channel proteins that are involved in many cellular properties. Accordingly, about 40 genes encoding different K_v channels have been identified and several families of

voltage-gated K_v channels ($K_{v1.x}$, $K_{v2.x...}$) are known. Under physiological conditions, the opening of K channels leads to an efflux of K^+ ions, which leads to a repolarization of the cell. Voltage-gated potassium channels are therefore crucial for the electrical excitability of cells, determining the repolarizing phase during an action potential (Fig. 1b, c).

Ca_v Channels

Ca_v channels mediate the Ca^{2+} influx in response to depolarizations. The physiologically relevant channel complex of this diverse group of proteins consists of three or four different subunits in addition to the pore-forming α subunit. Originally, Ca_v channels were grouped according to their different pharmacological properties and current characteristics into T-, L-, N-, P-, Q- and R-type channels. More recently, the genetically derived nomenclature for the different Ca_v channel families ($Ca_{v1.x}$, $Ca_{v2.x}$, etc.) is used. Since the intracellular Ca -concentration is much lower than the extracellular one, a transient rise in intracellular Ca^{2+} , which can be mediated by the activity of Ca_v channels, acts as a second messenger within intracellular signaling.

Regulatory Mechanisms

Pharmacology/Disease

A whole variety of different substances are known to interact with voltage-gated ion channels. Besides physiological substances such as ATP, Ca^{2+} , cyclic nucleotides, prostaglandins and biogenic amines, this also includes an enormous variety of different toxins from venomous organisms like snakes, scorpions, spiders and cone snails (3). In addition, ion channels are the prime target for numerous clinically relevant drugs like local anesthetics, antiarrhythmics or anti-convulsants. In recent years, several hereditary neurological, motor or cardiac diseases that are correlated with mutations in voltage-gated ion channels have been identified, for example special forms of epilepsy, deafness, migraine or heart arrhythmia. Certain malfunctions of ion channels have been identified for these so-called channelopathies. Ion channels, however, can also be affected by autoimmunity or acquired disorders of RNA processing.

The feature that mutations of different channels can cause very similar phenotypes such as cardiac arrhythmia in the case of some mutated Na_v and Ca_v channels is common to many channelopathies. In addition, different mutations of the same gene (e.g. $Na_v1.4$) may cause a broad bandwidth of myotonic disorders (4).

Interestingly, for diseases that are linked to mutations of Na_v channels, it can be shown that even subtle changes in the current kinetics can be responsible for

the clinical picture. This strongly demonstrates how well adjusted the “concert of activity” of the different ion channel proteins has to be for the proper electrical activity of a given cell.

While we already have a detailed knowledge of Na_v , Ca_v and K_v channel physiology, we are just at the beginning of understanding the diverse properties of the other channels (K_{2P} , TRP, TPC etc.) belonging to this exciting group of membrane proteins.

References

1. Hille B. (2001) Ionic channels of excitable membranes, 3rd edn. Sinauer, Sunderland, MA
2. Yu FH, Catterall WA (2004) The VGL-kanome: a protein superfamily specialized for electrical signaling and ionic homeostasis. *Sci STKE*2004:1–15
3. Terlau H, Olivera BM (2004) Conus venoms: a rich source of novel ion channel-targeted peptides. *Physiol Rev* 84:41–68
4. Kullmann DM, Hanna MG (2002) Neurological disorders caused by inherited ion-channel mutations. *Lancet Neurol* 1:157–166

Ion Trap

Definition

Ion trap describes an electrostatic lensing system (a three-dimensional quadrupole filter using a ring-electrode) that focuses ions created by electrospray (ESI), or matrix-assisted laser desorption (►MALDI), for analysis of their biophysical properties.

►Quadrupole Mass Analyser

►SNP Detection and Mass Spectrometry

Ion Trap Mass Spectrometer

Definition

The ion trap mass spectrometer uses three electrodes to trap ions in a small volume. The mass analyzer consists of a ring electrode separating two hemispherical electrodes. A mass spectrum is obtained by changing the electrode voltages to eject the ions from the trap. The advantages of the ion trap mass spectrometer include compact size, and the ability to trap and accumulate ions to increase the signal-to-noise ratio of a measurement.

►Mass Spectrometry: ESI

Ion-Exchange Chromatography

Definition

A chromatographic separation method based on the differences in electric charge of the components of the analyzed mixture. The column of ion-exchange chromatography consists of a synthetic resin containing fixed charged groups. In cation-exchange resins, negatively-charged groups are fixed to the resin, usually sulfonates, whereas anion-exchange resins bear positive-charged groups, usually alkyl-derivatives of ammonium. Ion-exchange separation is widely used to separate amino acid and protein mixtures.

► [Amino Acids](#), [Physicochemical Properties](#)

Ionic Interactions

Definition

Ionic interactions are electrostatic interactions between charged ions or partially charged atoms.

► [Protein/DNA Interaction](#)

Ionisation

Definition

Ionisation refers to a process of (electric) charge-transfer.

► [SNP Detection and Mass Spectrometry](#)

iProClass

Definition

The iProClass is an integrated resource that provides comprehensive family relationships and structural functional features of proteins.

► [Protein Databases](#)

IQ

Definition

The intelligence quotient (IQ) designates a score that is obtained in intelligence tests.

► [Fragile X Syndrome](#)

IRAK

► [Interleukin-1 Receptor Associated Kinase](#)

IRD

► [Infantile Refsum Disease](#)

IRES

► [Internal Ribosomal Entry Sites](#)

IRF-1

► [Interferon-Regulatory Factor 1](#)

Iron Storage Disease

► [Hemochromatosis](#)

I-Sce I Meganuclease

Definition

I-Sce I meganuclease is a super-rarely cutting endonuclease from yeast mitochondria that has an 18 bp recognition/cleavage site. The average cleavage frequency is once in 7×10^{10} bp.

► [Medaka as a Model Organism for Functional Genomics](#)

Isochromosome

Definition

An isochromosome is a symmetrical chromosome constituted of two identical arms.

► [Chromosome 21 Disorders](#)

Isoelectric Focusing

► [IEF](#)

Isoenzymes/Isozymes

Definition

Isoenzymes/Isozymes are forms of an enzyme with different amino acid sequences but identical substrate specificity. Isoenzymes may differ in activity level or regulation. They originate from different genes, but can also result from alternative splicing. Pseudo-isoenzymes are proteins that differ only in their post-translational modification.

► [Acute Intermittent Porphyria](#)

► [Affinity Chromatography and *In Vitro* Binding \(Beads\)](#)

Isoform

Definition

Isoform designates different forms of a protein.

► [Morpholino Oligonucleotides, Functional Genomics by Gene ‘Knock-down’](#)

Isolated Tumour Cells

► [Genomic Analysis of Single Disseminated Cancer Cells](#)

Isometric Contraction

Definition

Isometric contraction defines the force generation of a muscle at constant sarcomere length.

► [Muscle Contraction](#)

Isomorphic Replacement

Definition

Multiple Isomorphic Replacement (MIR) uses crystals of a macromolecule with additional atoms (mostly heavy metals) soaked in or co-crystallized to determine the phases. Only the normal scattering of the heavy atoms is used for phasing. The most striking problem is the isomorphism of the derivatized crystal with respect to the native crystal. Lattice parameters must not change by more than 1% to allow determination of the partial structure of the heavy atoms and the protein phases.

► [MAD Phasing](#)

Isopeptide Bond

Definition

Isopeptide bond describes an amide bond formed between a carboxyl group of one amino acid and an amino group of another amino acid with the exception of the conventional peptide bond, e.g. the C-terminus of ubiquitin is connected with the ϵ -amino group of a lysine-residue.

► [Sumoylation](#)

► [Ubiquitination](#)

Isoprenoid

Definition

A molecule containing isoprene units that are characterized by a pattern of 5 carbons with a methyl group and a double bond.

► [Protein Prenylation](#)

Isoprenylation

Definition

Isoprenylation defines the attachment of isoprenoids, which are branched, unsaturated carbon chains, to proteins containing a CAAX-box or the sequence CC or CXC at the C-terminus (see also CAAX-box). Isoprenoids are also intermediates in the biosynthesis of cholesterol.

► [Fatty Acid Acylation of Proteins](#)

Isothermal Titration Calorimetry

TIMM REINISCH, HANS-JURGEN HINZ
Institute for Physical Chemistry, WWU Münster,
Münster, Germany
hinz@uni-muenster.de

Definition

Isothermal Titration Calorimetry is one of the most powerful analytical techniques in modern drug design. It exploits the method of thermal titrations to obtain the number of bound ligands per mole of protein (or biopolymer in general), i.e. the stoichiometry, as well as the relevant thermodynamic parameters such as standard reaction enthalpy ΔH^0 , standard reaction entropy ΔS^0 and the standard Gibbs energy change ΔG^0 which relates to the equilibrium constant. Isothermal measurements can be performed in the temperature range from close to 0°C up to 80°C. This permits to obtain the important information on the variation with temperature of all parameters.

Description

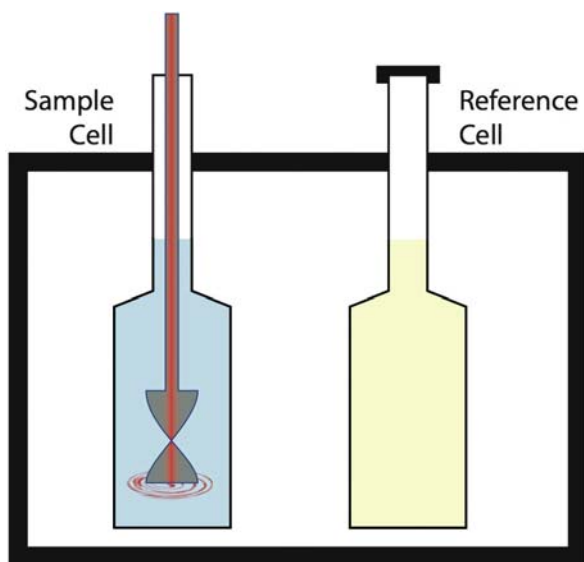
Binding phenomena are the essence of biological regulation. The quantitative understanding of these

interactions in molecular terms forms therefore the basis of all pharmaceutical and medicinal activities to evolve strategies for coping with malfunctions of organisms. The indispensable vital parameters for the characterization of binding reactions are the number and molar ratio of interacting molecules (termed the stoichiometry, n), the binding constant K_B , (or dissociation constant $K_D = 1/K_B$); the standard reaction enthalpy, ΔH^0 , the standard reaction entropy ΔS^0 , the standard Gibbs energy change ΔG^0 and the variation with temperature of ΔH^0 , the so called heat capacity change ΔC_p^0 . An elegant, simple and fast method for the determination of these parameters is provided by high sensitivity isothermal titration calorimetry (ITC). As implied by the name a solution is titrated at constant temperature with another solution containing the potential ligand. The heat resulting from the interaction is used as an indicator of the degree of complex formation. As practically all reactions are associated with heat release (exothermic reaction) or heat absorption (endothermic reaction) the field of application of ITC is nearly unlimited. Even reactions that are apparently athermal at the temperature chosen for the experiment are likely to involve non-zero heat changes at a higher or lower temperature.

This feature renders ITC a very useful technique for the determination of physical properties of binding reactions. They include interactions between macromolecules such as DNAs, RNAs and proteins or between these molecules and small ligands, drugs and coenzymes or inhibitors as well as binding reactions to membranes and receptors, to name a few.

Figure 1 shows the schematic view of an ITC instrument. It generally consists of two identical reaction cells, one serving as reference the other as sample cell. At specified time intervals a small volume of the ligand solution (red) [typically 5–20 µL] is injected into the sample solution (blue) [0.5–2 mL] under stirring to provide rapid mixing of the solutions. The ITC instrument registers the compensatory power (J/s) required to maintain a “zero temperature difference” between the reference and sample cell as a function of time. Depending on the reaction being exothermic or endothermic the compensatory power of the instrument will be negative or positive relative to the zero reaction level. Figure 2 shows the simulation of an ITC titration based on the parameters given in the figure legend.

Integration of the individual power peaks over the time interval defined by the time elapsed between the start of the injection and the return of the power signal to the original baseline provides the heat change per injection, δq . The magnitude of the heat change per injection is largely determined by the change in the number of moles of ligand bound, δn_B , and the molar binding



Isothermal Titration Calorimetry. Figure 1 Schematic view of an ITC instrument.

enthalpy ΔH^0 , however, it also contains heat contributions from stirring and dilution of both ligand and sample. $\Delta q = \Delta H^0 \cdot \Delta n_B + (q_{\text{stirring}} + q_{\text{dilution}})$ The latter heat contributions shown in brackets must be corrected for by separate experiments in which a) buffer is injected into the sample solution and b) ligand solution is injected into the sample free buffer. Usually the correction resulting from the small dilution of the sample solution is negligible, while the heat associated with the dilution of the ligand can be significant. Division of the corrected heat change, δq , by the number of moles of ligand bound, δn_B , results in the value for the binding enthalpy per mole of ligand ΔH^0 . Optimal conditions for the determination of the binding enthalpy are therefore a high binding constant and an excess of sample over ligand concentration in the cell, i.e. conditions that guarantee complete binding of the injected ligand. This situation often prevails during the first injections at the beginning of a titration and is apparent by the identical magnitude of the initial power peaks as seen in Fig. 2. The enthalpy change ΔH^0 calculated in this manner can be further broken down into an intrinsic enthalpy value ΔH_{int} and an enthalpy contribution resulting from protonation or deprotonation effects, ΔH_{H^+} , if a change in the degree of protonation originates from ligand binding. The relevant equation is

$$\begin{aligned} \Delta H^0 &= \Delta H_{\text{int}}^0 + \Delta H_{\text{H}^+} \\ &= \Delta H_{\text{int}}^0 + n_{\text{H}^+} \cdot (\Delta H_{\text{ion, sample}} + \Delta H_{\text{ion, buffer}}) \end{aligned} \quad [1]$$

ΔH_{H^+} is given by the number n_{H^+} of protons transferred, times the sum of the ionization enthalpies

(with the proper sign) of the groups of the sample and that of the buffer used.

Performing the titration under otherwise identical conditions in a series of buffers with different heats of ionization the number of protons involved in the binding reaction can be obtained. If in addition conditions can be found where ligand binding is pH independent ($n_{\text{H}^+} = 0$) the observed enthalpy is equal to the intrinsic binding enthalpy.

Favourable conditions for the determination of binding enthalpies and the number of protons involved are unfavourable for the determination of binding constants. Extraction of binding constants from titration curves requires a significant change in the degree of ligation of the sample in the course of the titration which is equivalent to having a sizeable amount of unbound ligand in solution. A good estimate of the proper sample concentration that will yield analyzable titration curves can be obtained from the equation

$$a = K_B \cdot [\text{bs}]_{\text{total}} \quad [2]$$

where K_B is the site binding constant and $[\text{bs}]_{\text{total}}$ is the total concentration of binding sites. In the range of a values between 1 and 1000 the binding isotherms can be deconvolved to obtain binding constants; best accuracy is achieved with a values in the range of 5 to 500.

There is an elegant mathematical procedure to analyze both ITC and DSC experiments. We shall consider here the part relevant to ITC measurements. We give the derivation for proteins but the formalism applies of course also to all other macromolecules or binding partners. The enthalpy change of a reaction relative to a [reference state](#), for example the unliganded protein, can be obtained from the derivative with respect to temperature of the logarithm of the relative partition function Z as shown in Eq. 3:

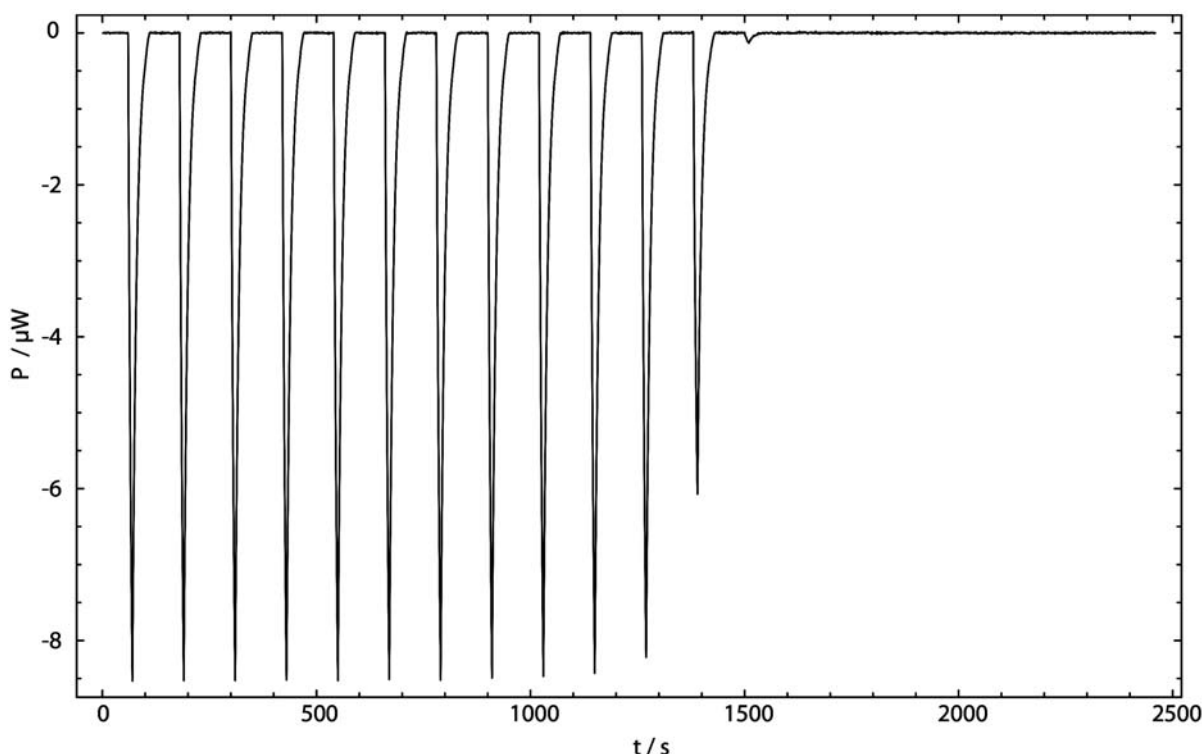
$$H - H_N = \Delta H = -\frac{\partial}{\partial \beta} (\ln Z) = R \cdot T^2 \frac{\partial \ln Z}{\partial T} = \frac{RT^2}{Z} \frac{\partial Z}{\partial T} \quad [3]$$

β is defined as $1/RT$, $R = 8.3145 \text{ JK}^{-1}\text{mol}^{-1}$ is the Avogadro constant and T is absolute temperature.

Generally the relative partition function depends on the free ligand concentration x_{free} , which can be calculated from the mass conservation equation

$$x_{\text{total}} = x_{\text{free}} + \bar{x}_{\text{bound}} \cdot c_{\text{prot}} \quad [4]$$

x_{total} refers to the total and x_{free} to the free and ligand concentration in the calorimetric cell, \bar{x}_{bound} is the average number of ligands bound per mole of protein and c_{prot} is the total protein concentration. x_{free} can be calculated once \bar{x}_{bound} is known, as c_{prot} and x_{total} are



Isothermal Titration Calorimetry. Figure 2 Simulation of an ITC-Experiment: $K = 10^7$, $c_{\text{cell}} = 0,380 \text{ mM}$, $c_{\text{injector}} = 4.2 \text{ mM}$, $\Delta H = 10 \text{ kJ/mol}$, time between 10 μl injections: 120 s.

the accessible analytical concentrations of protein and ligand, respectively. For simplicity we write in the following x instead of x_{free} .

The average number of ligands bound per mole of protein is easily calculated from the relative partition function Z according to the equation

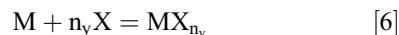
$$\bar{x}_{\text{bound}} = \frac{\partial \ln Z}{\partial \ln x} = \frac{x}{Z} \frac{\partial Z}{\partial x} \quad [5]$$

Thus once the relative partition function Z can be formulated for the various binding equilibria both the binding enthalpy ΔH^0 and the binding constant K_B can be obtained from non-linear curve fitting algorithms.

Binding to Independent Sites

A widely used theoretical model for the analysis of binding phenomena in biology is based on the assumption of multiple sets of independent binding sites. For generality we assume k classes of binding sites, each of which is characterized by its own binding constant K_y . This is indicated by the index y varying from 1 to k . Each of the k classes can have an arbitrary number of independent binding sites m_y ($y = 1$ to k) and each of the k binding sites of one class accommodates n_y ligands ($y = 1$ to k). Though in many cases n_y will be equal to 1 in biological systems, we

allow here for values different from the sake of generality.



The partition function Z_y for the single site reaction is described by the equation:

$$Z_y = (1 + K_y x^{n_y}) \quad [7]$$

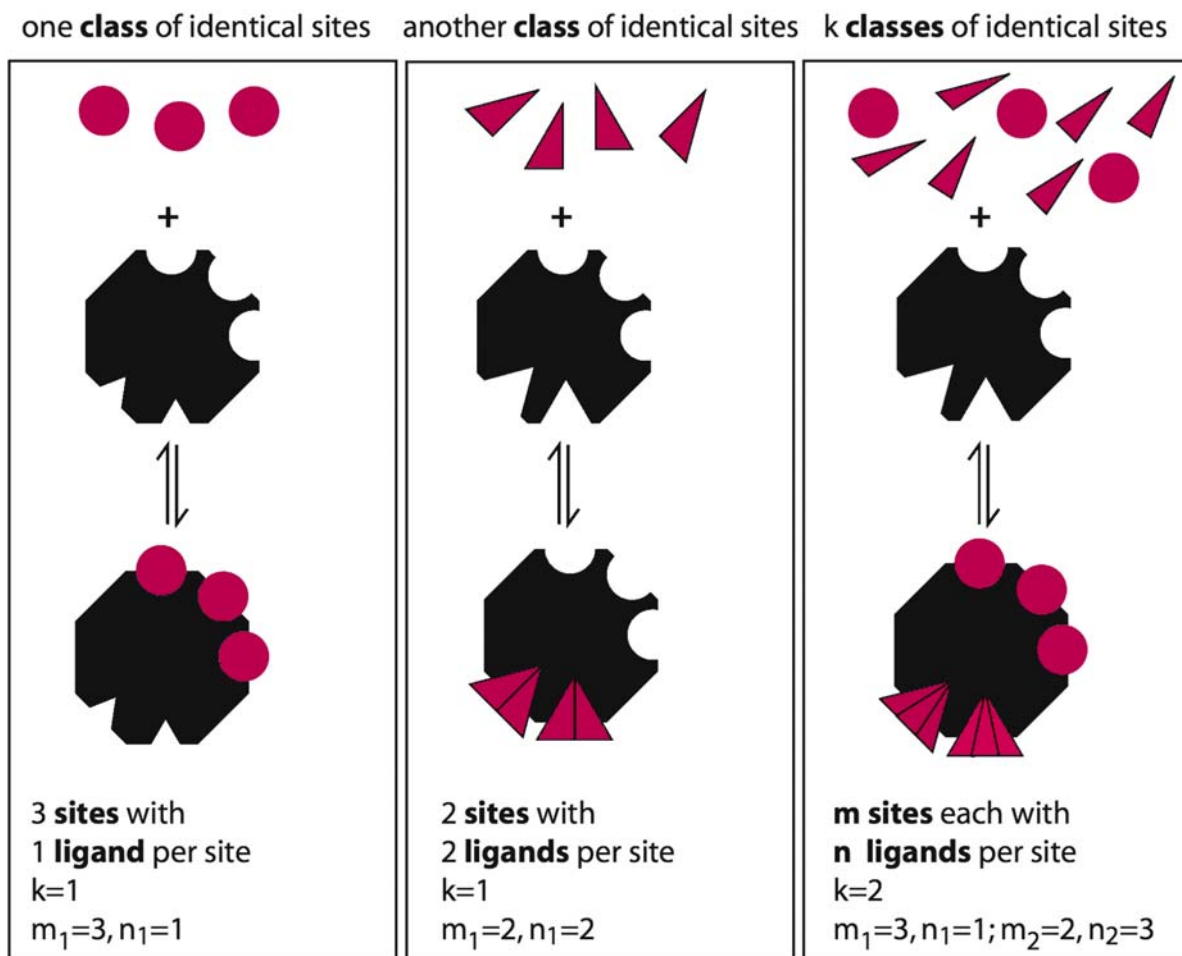
x is the free ligand concentration and K_y is the binding constant of the class.

For m_y identical independent sites with the same binding constant K_y the corresponding class partition function Z_y^{class} is the product of the individual partition functions.

$$Z_y^{\text{class}} = (Z_y)^{m_y} = (1 + K_y x^{n_y})^{m_y} \quad [8]$$

Finally, if there are m_y independent classes, the overall partition function Z is the product of the class partition functions

$$\begin{aligned} Z &= (1 + K_1 x^{n_1})^{m_1} (1 + K_2 x^{n_2})^{m_2} \dots (1 + K_k x^{n_k})^{m_k} \\ &= \prod_{y=1}^k (1 + K_y x^{n_y})^{m_y} = \prod_{y=1}^k Z_y^{\text{class}} \end{aligned} \quad [9]$$



Isothermal Titration Calorimetry. Figure 3 Illustration of the terms used in the independent and equal sites model.

Knowledge of the overall partition function for the general “multiple sets of independent binding sites model” permits now the calculation of the binding enthalpy using Eq. 10:

$$\Delta H^0 = \frac{RT^2}{Z} \frac{dZ}{dT} \quad [10]$$

With Z from Eq. 9 one obtains:

$$\begin{aligned} \Delta H^0 &= RT^2 \sum_{y=1}^k \frac{1}{Z_y^{\text{class}}} \frac{dZ_y^{\text{class}}}{dT} \\ &= RT^2 \sum_{y=1}^k \frac{1}{Z_y^{\text{class}}} m_y q_y^{m_y-1} \cdot x^{n_y} \cdot \frac{K_y \Delta H_y^0}{RT^2} \quad [11] \\ &= \sum_{y=1}^k m_y \Delta H_y^0 \frac{K_y x^{n_y}}{1 + K_y x^{n_y}} \end{aligned}$$

The average number of bound ligands \bar{x}_{bound} is determined from Eq. 5

$$\begin{aligned} \bar{x}_{\text{bound}} &= \frac{\partial \ln Z}{\partial \ln x} = x \sum_{y=1}^k \frac{1}{Z_y^{\text{class}}} \frac{\partial Z_y^{\text{class}}}{\partial x} \\ &= \sum_{y=1}^k \frac{1}{Z_y^{\text{class}}} m_y Z_y^{m_y-1} \cdot n_y K_y \cdot x^{n_y} \quad [12] \\ &= \sum_{y=1}^k n_y q_y^{-1} \cdot m_y K_y x^{n_y} \\ &= \sum_{y=1}^k m_y \cdot n_y \cdot \frac{K_y x^{n_y}}{1 + K_y x^{n_y}} \end{aligned}$$

With the use of the expression for \bar{x}_{bound} the free ligand concentration x_{free} can be obtained for each total ligand concentration x_{total} employing the mass conservation constraint

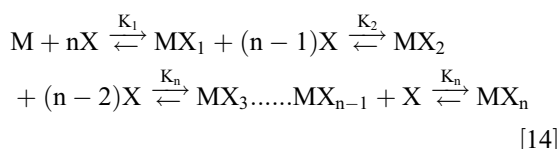
$$0 = x_{\text{free}} + \left(\sum_{y=1}^k m_y \cdot n_y \cdot \frac{K_y x^{n_y}}{1 + K_y x^{n_y}} \right) \cdot c_{\text{prot}} - x_{\text{total}} \quad [13]$$

As this equation has no analytical, closed form solution for $k > 3$, iterative numerical procedures must be used to solve the equation for x_{free} .

On the basis of this general algorithm the measured total enthalpy change as well as the individual enthalpy change per injection can be expressed as a function of the known parameters: total ligand and total protein (sample in general) concentration in the calorimetric cell.

Sequential Binding

Obligatory sequential binding can be modelled by the following binding scheme:



The relative partition function representative of this equilibrium is

$$Z = 1 + K_1x + K_1K_2x^2 + K_1K_2K_3x^3 + \sum_{i=4}^n x^i \prod_{y=1}^i K_y \quad [15]$$

The K_i values refer to the microscopic binding constants characteristic for each binding equilibrium. The treatment of the partition function is facilitated, if instead of these microscopic constants K_i the stoichiometric constants \tilde{K}_i are introduced. They are defined by the relation

$$\begin{aligned} \tilde{K}_i &\equiv \prod_{y=1}^i K_y \\ (\tilde{K}_1 &= K_1; \tilde{K}_2 = K_1K_2; \tilde{K}_3 = K_1K_2K_3 \text{ etc.}) \end{aligned} \quad [16]$$

Introducing the stoichiometric binding constants transforms the partition function into

$$Z = 1 + \tilde{K}_1x + \tilde{K}_2x^2 + \tilde{K}_3x^3 + \sum_{i=4}^n x^i \prod_{y=1}^i K_y \quad [17]$$

Differentiation of $\ln Z$ with respect to the logarithm of free ligand concentration, yields the average number of ligands bound \bar{x}_{bound} :

$$\bar{x}_{\text{bound}} = \frac{\partial \ln Z}{\partial \ln x} = \frac{\sum_{y=0}^n i \tilde{K}_y x^y}{1 + \sum_{y=1}^n \tilde{K}_y x^y} \quad [18]$$

Introduction of this expression into the mass conservation equation results in the conditional equation for the free ligand concentration

$$0 = x_{\text{free}} + \left(\frac{\sum_{y=0}^n i \tilde{K}_y x^{ny}}{1 + \sum_{y=1}^n \tilde{K}_y x^{ny}} \right) \cdot c_{\text{prot}} - x_{\text{total}} \quad [19]$$

There is no analytical solution to this equation for $n > 2$ but numerical methods are available that allow for the extraction of the equilibrium constant, the binding enthalpy and the stoichiometry.

Thermodynamic Relationships

The change in Gibbs energy, ΔG , determines the direction of a spontaneous reaction at constant pressure and temperature.

$$\Delta = G - G^0 + RT \ln \frac{\prod_i [c_i]_{\text{products}}}{\prod_i [c_i]_{\text{educts}}} \quad [20]$$

$R = 8.3145 \text{ J K}^{-1} \text{ mol}^{-1}$ is the gas constant, T is the absolute temperature, c_i is the molar concentration of educts and products, respectively. ΔG^0 is the standard state Gibbs energy change, if all c_i are kept at 1 M. At that condition $\Delta G = \Delta G^0$. In case the c_i 's assume arbitrary values and experimental conditions are such that these analytical concentrations are kept constant by removal or replenishment of material, ΔG is the driving force of the reaction at the specified concentrations.

In a titration experiment the binding reaction tends to approach equilibrium. As no work can be performed at equilibrium ΔG equals zero at equilibrium and one obtains from Eq. 20 the important relationship between the equilibrium constant K_{eq} and the standard Gibbs energy change, ΔG^0 .

$$\Delta G^0 = -RT \ln \frac{\prod_i [c_i]_{\text{products,eq.}}}{\prod_i [c_i]_{\text{educts,eq.}}} \equiv -RT \ln K_{\text{eq}} \quad [21]$$

The Gibbs energy change is related to the enthalpy (ΔH) and entropy (ΔS) change of the reaction by the eq.22

$$\Delta G = \Delta H - T\Delta S \quad [22]$$

This is a significant relationship, since interactions with similar affinities can have a largely different energy and entropy signature. Generally both ΔH and ΔS vary with temperature and for biological reactions a linear dependence of δH satisfies usually the observations.

$$\begin{aligned}
 \Delta H(T) &= \Delta H(T_R) + \Delta C_p \int_{T_R}^T dT \\
 &= \Delta H(T_R) + \Delta C_p (T - T_R) \\
 \Delta S(T) &= \Delta S(T_R) + \Delta C_p \int_{T_R}^T \frac{dT}{T} \\
 &= \Delta S(T_R) + \Delta C_p \ln\left(\frac{T}{T_R}\right)
 \end{aligned} \quad [23]$$

T_R is a convenient reference temperature, for example, 25°C (298.16 K). Allowing for the temperature dependence of ΔH and ΔS the equation for ΔG assumes the form

$$\begin{aligned}
 \Delta G(T) &= \Delta H(T_R) - T\Delta S(T_R) \\
 &+ \Delta C_p \left(T - T_R - T \ln \frac{T}{T_R} \right)
 \end{aligned} \quad [24]$$

In analogy the same relation written in terms of standard properties can be used as fit equation to determine the dependence on temperature of the equilibrium constant.

$$\begin{aligned}
 -RT \ln K_{eq} &= \Delta G^0 = \Delta H^0(T_R) - T\Delta S^0(T_R) \\
 &+ \Delta C_p \left(T - T_R - T \ln \frac{T}{T_R} \right)
 \end{aligned} \quad [25]$$

The heat capacity change on binding, ΔC_p , can be obtained conveniently by performing ITC measurements at different temperatures. Calculation of the slope of the ΔH^0 versus T plot yields

$$\Delta C_p = \left(\frac{\Delta H^0(T_2) - \Delta H^0(T_1)}{T_2 - T_1} \right) \quad [26]$$

Depending on whether ΔH^0 increases or decreases with temperature, ΔC_p can be positive or negative. It has been observed frequently that ligand binding reactions to proteins are characterized by negative values of ΔC_p in the range of a few $\text{kJ mol}^{-1} \text{K}^{-1}$.

Application of ITC in the Life Sciences

Enthalpy changes are a characteristic and universal property of chemical and biological reactions and can therefore be utilized to follow the degree of complex formation without the requirement of modification of the samples under study. A major field of application is in drug discovery, target validation and assay development. As the chemical nature of the compounds plays virtually no role, both samples and ligands can consist of small molecules or polymers. Thus protein-protein, protein-DNA, protein-polysaccharide interactions are equally accessible as the reactions of these biopolymers

with small ligands such as ions, drugs, enzymes, allosteric effectors, protons and water. Another wide field of application is the characterization of vesicular micellar systems, the analysis of ligand-induced protein folding mechanisms, and the identification of **co-operativity** of ligand binding.

Determination of Kinetic Parameters of Enzyme Reactions Using ITC

Enzyme kinetics can be determined by ITC because the thermal power generated as the reaction proceeds is a direct probe of the reaction velocity. The total heat involved in the conversion of n moles of substrate into product P is

$$Q = n\Delta H^0 = [P]_T \cdot V_{\text{cell}} \Delta H^0 \quad [27]$$

$[P]_T$ is the total molar concentration of product in the calorimetric cell having the volume V_{cell} . ΔH^0 refers to the enthalpy change per mole of product generated.

Assuming Michaelis-Menten kinetics the reaction rate r is determined by the equation

$$r = -\frac{d[S]_{\text{inst}}}{dt} = \frac{k_{\text{cat}} \cdot [E]_T \cdot [S]_{\text{inst}}}{K_M + [S]_{\text{inst}}} \quad [28]$$

k_{cat} is the catalytic rate constant, $[E]_T$ is the total enzyme concentration in the calorimetric cell, and $[S]_{\text{inst}}$ is the instantaneous substrate concentration. The rate of disappearance of substrate $-\frac{d[S]_{\text{inst}}}{dt}$ is equal to the rate of product formation $\frac{d[P]}{dt}$. The power $[J s^{-1}]$ generated by the reaction is given by the equation

$$\frac{dQ}{dt} = \frac{d[P]}{dt} \cdot V_{\text{cell}} \cdot \Delta H^0 \quad [29]$$

Rearrangement of the equation provides the dependence of the reaction rate on the measured power signal dQ/dt , the cell volume V_{cell} and the enthalpy change ΔH^0 per mole of substrate transformed into product. The latter can be determined in a separate experiment, in which the heat for a known, completely converted amount of substrate is measured.

$$r = \frac{d[P]}{dt} = \frac{\frac{dQ}{dt}}{V_{\text{cell}} \cdot \Delta H^0} \quad [30]$$

Calculation of k_{cat} and K_M on the basis of eq.(28) becomes possible by fitting routines if $[S]_{\text{inst}}$ is known. Under continuous assay conditions $[S]_{\text{inst}}$ can be obtained using the relation

$$[S]_{\text{inst}} = [S]_{t=0} - \frac{\int_0^t \left(\frac{dQ}{dt} \right) dt}{V_{\text{cell}} \cdot \Delta H^0} \quad [31]$$

$[S]_{t=0}$ is the analytical substrate concentration at the beginning of the experiment, the integral is proportional to the area under the power-time curve (the output signal) t seconds after the start.

The resulting instantaneous substrate concentrations together with the measured reaction rates, r , can be employed to extract k_{cat} and K_M values from the Michaelis-Menten rate equation using non-linear fitting routines. Activation parameters can be obtained from performing these measurements at different temperatures. Different buffer systems can be used for signal amplification in reactions where proton evolution is coupled to product formation. (Eq. 1)

Summary

Isothermal Titration Calorimetry is an extremely useful, universally applicable, non-invasive technique to quantify interactions by providing all necessary thermodynamic parameters. These include reaction enthalpies, ΔH^0 , their temperature dependence, ΔC_p , and the binding constant K_B . The second rapidly growing field of application of ITC is the determination of kinetic parameters of enzyme reactions. There are no assay requirements other than the fact that the reaction be associated with enthalpy changes. This condition is practically universally met or can be arranged by a proper choice of the experimental temperature.

References

1. Wiseman T, Williston S, Brandts JF et al (1989) Rapid measurement of binding constants and heats of binding using a new titration calorimeter. *Anal Biochem* 62:30–36
2. Freire E, Mayorga OL, Straume M (1990) Isothermal Titration Calorimetry. *Anal Chem* 62:950A–959A
3. Todd MJ, Gomez J (2001) Enzyme kinetics determined using calorimetry: A general assay for enzyme activity? *Anal Biochem* 296:179–187
4. Cliff MJ, Gutierrez A, Ladbury JE (2004) A survey of the year 2003 literature on applications of isothermal titration calorimetry. *J Mol Recognit* 17:513–523
5. O'Brien R, Haq J (2004) In: Ladbury JE, Doyle M (eds) *Biocalorimetry 2. Applications of Biocalorimetry: Binding, Stability and Enzyme Kinetics*. John Wiley and Sons, Ltd

Isotonic Contraction

Definition

Isotonic contraction defines the sarcomere shortening caused by relative movement of thin and thick filaments at constant load.

► [Muscle Contraction](#)

Isotope Labeling

Definition

Isotope labeling is the selective incorporation of an isotope (^{13}C , ^{15}N , or ^2H) into a molecule. Isotope labeling, for instance, allows the tracing of the labeled molecule in pharmacokinetic studies. ► [NMR-based screening](#) utilizes the magnetic properties of the isotope. Isotope-edited NMR spectra facilitate the NMR resonance assignments of complex biomolecules, and allow the selective observation of an isotope-labeled compound in a composite mixture. (Synonym: radioactive labeling). In mass spectrometry, isotope labeling enables mass spectrometric quantification of labeled proteins or peptides within a mixture.

► [NMR-Based Screening](#)

► [PET](#)

► [Protein-Ligand-Interaction by NMR](#)

► [Two-Dimensional Gel Electrophoresis](#)

Isozymes

► [Isoenzymes/Isozymes](#)

ITC

► [Isothermal Titration Calorimetry](#)

ITIM

Definition

ITIM stands for immunoreceptor tyrosine-based inhibition motif. It characterizes an amino acid sequence in the cytoplasmic tail of receptor complex proteins which are modified during inhibition of cell function.

► [Adhesion Molecules](#)

Jag1 and 2

► Jagged (Jag) 1 and 2

Tyk2. They phosphorylate members of the STAT family, which translocate to the nucleus and activate promoters of specific genes.

► Cytokines

► Interleukins and Their Receptors, a Role in Regulating Hematopoietic Stem Cells

Jagged 1 and 2

Definition

Jagged (Jag) 1 and 2 are the two mouse orthologues of the *Drosophila* Serrate gene.

► Notch Pathway

Janus Kinase

► JAK

JAK

Definition

Janus Kinase (JAK) is a protein tyrosine kinase that activates ► STAT mediated transcription.

► Autosomal Dominant, Polycystic Kidney Disease

► JAK/STAT

Jet Lag

Definition

Jet lag is the result of long distance travel towards east/west crossing time zones at a rapid rate. Symptoms such as sleep disturbance, loss of appetite, reduced psychomotor efficiency and general malaise may occur.

► Circadian Clocks

JAK/STAT

Definition

Janus kinases (JAK)/signal transducers and activators of transcription (STAT) mediate important biologic responses to a number of cytokines (e.g. IL-6, interferons). JAKs are protein tyrosine kinases that directly associate with the intracellular domains of the respective receptors. Four mammalian JAK members have been identified; JAK-1, JAK-2, JAK-3 and

JNK

Definition

c-Jun N-terminal kinase (JNK), (also known as SAPK) responds to activation by environmental stress and pro-inflammatory cytokines by phosphorylating a number of transcription factors, primarily components of activating protein-1 (AP-1) such as c-Jun and ATF2, and thus regulates AP-1 transcriptional activity.

► Jun/Fos

► Signal Transduction: Integrin-Mediated Pathways

Joule Heating

Definition

In the presence of an electric field, charges gain in kinetic energy, which may be lost as heat due to inelastic collisions.

► **Proteomics in Microfluidic Systems**

Jun/Fos

PETER ANGEL, MARINA SCHORPP-KISTNER
Deutsches Krebsforschungszentrum, Division of
Signal Transduction and Growth Control, Heidelberg,
Germany
p.angel@dkfz-heidelberg.de/angel@dkfz.de
marina.schorpp@dkfz.de

Definition

The term “Fos/Jun” is used for a number of structurally and functionally related members of the Fos and Jun protein families of transcription factors, collectively described as activating protein-1 (► **AP-1**). AP-1 mediates gene regulation in response to cytokines, growth factors and stress signals during cell proliferation, differentiation and apoptosis or transformation and tumorigenesis.

Characteristics

Much of our present knowledge about transcription factors comes from the discovery and study of the activating protein-1 (AP-1) family. AP-1 has served to detect the first decisive DNA binding motifs required for gene regulation by a variety of extracellular signals including growth factors, cytokines, tumour promoters, such as the phorbol ester TPA (12-O-tetradecanoyl-phorbol-13-acetate) and carcinogens, for example UV irradiation and other DNA damaging agents. One AP-1 subunit, the heterodimer Fos-Jun was found in the mid 80s to contain the viral oncogene product Fos. AP-1 was defined as an activity that controls both basal and inducible transcription of several genes containing AP-1 sites (5'-TGAG/CTCA-3'), also known as TPA-responsive elements (► **TRE**).

Within one year it became evident that the AP-1 containing polypeptides correspond to the products of members of the *jun* and *fos* gene families and that the first member of the Jun family, c-Jun, represents

the cellular homologue of the transforming oncogene (v-Jun) of the chicken retrovirus ASV-17.

At present, the Jun protein family consists of c-Jun, JunB and JunD; the Fos protein family comprises c-Fos, FosB, Fra-1 and Fra-2. During the past decade additional proteins, such as members of the ATF family (ATFa, ATF-2, ATF-3) have been identified (mostly by yeast-two-hybrid screening) that share structural homologies and form heterodimeric complexes predominantly with Jun proteins (see below) to bind to TRE-like sequences.

General Structure of the AP-1 Subunits

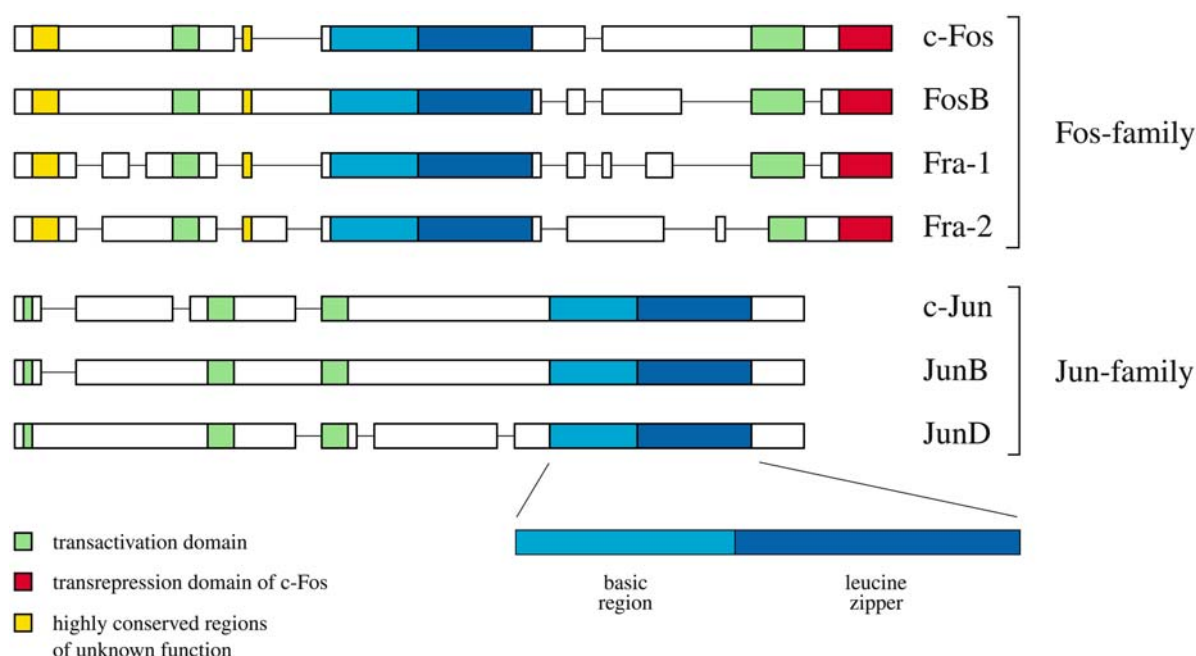
For its function in controlling gene expression, the prototype of a transcription factor has to include at least two structures; one region of the protein that is responsible for binding to a specific DNA recognition sequence (DNA binding domain) and a second region that produces transcriptional activation (transactivation domain) once the protein is bound to DNA.

DNA Binding Domain

The DNA binding domain is evolutionarily conserved between the Jun, Fos and CREB/ATF proteins, thus defining the protein family called “► **bZip**” proteins. bZip stands for the amino acid sequences of the two independently acting sub-regions of the DNA binding domain, the “basic domain” rich in basic amino acids which are responsible for contacting the DNA and the “leucine-zipper” region characterised by heptad repeats of leucine, part of the well known “4–3 repeats” forming a coiled-coil structure, which is responsible for dimerisation and is a pre-requisite for DNA binding (Fig. 1). In addition to the leucines, other hydrophobic and charged amino acid residues within the leucine zipper region are responsible for the specificity and stability of homo- or hetero-dimer formation between the various Jun, Fos or CREB/ATF proteins. The Fos proteins do not form stable homodimers but heterodimerise efficiently with the Jun proteins. The Jun proteins can form homodimers, although with reduced stability compared to Jun/Fos or Jun/ATF. Jun-Jun and Jun-Fos dimers preferentially bind to the 7bp motif 5-TGAG/CTCA-3' whereas Jun-ATF dimers or ATF homodimers prefer to bind to a related, 8 bp consensus sequence 5'-TTACCTCA-3'. Therefore, individual AP-1 dimers are expected to regulate specific subsets of AP-1 target genes depending on the characteristics of the AP-1 site in their promoter.

Transactivation Domain

In contrast to the well-defined DNA binding domain, the structural properties of the domains in the AP-1 proteins mediating transcriptional activation of target genes (transactivation domain, TAD) are less well



Jun/Fos. Figure 1 Structural organisation of the Fos and Jun proteins.

understood. The activity of the TAD can be transferred to heterologous DNA binding domains, such as the yeast transcription factor GAL4. By employing such chimeric proteins (which, in contrast to the wild type proteins do not depend on a dimerisation partner) critical amino acids in the TADs were identified. Moreover, it became clear that the various Jun, Fos and ATF proteins significantly differ in their transactivation potential. For example, c-Fos, FosB and c-Jun are strong transactivators, whereas JunB, JunD, Fra-1 and Fra-2 exhibit only weak transactivation potential. Under specific circumstances, the latter may even act as repressors of AP-1 activity by competitive binding to AP-1 sites or by forming inactive heterodimers with c-Fos, FosB or c-Jun. Most importantly, transactivation studies using fusion proteins led to the identification of protein kinases that bind to and phosphorylate AP-1 proteins in the TAD in response to extracellular signals, thereby controlling expression of AP-1 target genes (see below).

Molecular Interactions

In addition to the “classical” AP-1 members (Jun, Fos, ATFs), several new bZip proteins have been defined on the basis of DNA sequence specificity and heterodimer formation with Jun and Fos proteins. These include Maf and Maf-related proteins, Nrl, Smads and Jun-dimerising partners (JDs). The function of these proteins in AP-1-regulated processes remains to be determined. Binding of AP-1 to DNA may also support binding of other transcription factors to adjacent or

overlapping binding sites (composite elements) to allow the formation of quaternary complexes. The interaction of NFAT and Ets proteins with DNA on the IL-2 and collagenase promoters respectively, may serve as paradigms for this type of protein/protein interaction. The mutual interference between AP-1 and steroid hormone receptors, particularly the glucocorticoid receptor (GR) represents another extensively analysed example of protein-protein interaction-based cross-talk. There is evidence that the anti-inflammatory and immunosuppressive activities of glucocorticoids are mediated, at least in part, by GR-mediated repression of AP-1 activity. In addition to GR, transcription factors NFκB, MyoD, the LIM-only protein FHL2 and YY1, as well as other types of cellular proteins, such as DExD/H-box RNA helicase RHH/Gu, BAF60a subunit of the chromatin remodelling complex SWI/SNF, CBP and p300 have been found to modulate AP-1 activity. In most cases, the exact mechanism of interaction between AP-1 and these proteins remains to be determined. The N-terminus of JunD is the site of interaction with the tumour suppressor menin, which represses JunD-activated transcription through interference of ERK-dependent phosphorylation and/or association with an mSinA-histone deacetylase complex.

Regulatory Mechanisms

Transcriptional Regulation of *fos* and *jun* Genes

The *jun* and *fos* genes are members of a class of cellular genes, termed early response or “immediate-early”

genes which are characterised by a rapid and transient activation of transcription in response to changes in environmental conditions, such as growth factors, cytokines, tumour promoters, carcinogens and expression of certain oncogenes. Since this type of regulation of promoter activity is also observed in the absence of ongoing protein synthesis, it is generally accepted that pre-existing factors, whose activity gets altered by changes in post-translational modification (described in detail in the subsequent section), are responsible for the regulation of promoter activity.

Most of our current knowledge on transcriptional activation of ►immediate early genes is derived from studies on deletion and point mutations in the *c-fos* and *c-jun* promoters, combined with *in vitro* and *in vivo* footprinting analyses. The serum response element (SRE) is required for the majority of extracellular stimuli, including growth factors and phorbol esters. This element is bound in a ternary complex containing the transcription factor p67-SRF and p62-TCF (which stands for a class of related proteins described as Elk, SAP). Changes in the phosphorylation pattern of SRF and, predominantly, TCF regulate *c-fos* promoter activity by these stimuli. Other elements include the c-AMP response element (CRE) and the Sis-inducible enhancer (SIE), which is recognised by the STAT group of transcription factors, which are at the receiving end of the Jak/Stat signalling pathway initiated by specific classes of cytokines. In contrast, the element responsible for negative autoregulation of the *c-fos* promoter has not yet been identified conclusively.

Analysis of deletion mutants of the *c-jun* promoter identified two AP-1-like binding sites (Jun1, Jun2), which are recognised by Jun/ATF heterodimers or ATF homodimers and which are involved in transcriptional regulation in response to the majority of extracellular stimuli affecting *c-jun* transcription. In response to G-protein coupled receptor, EGF and other growth factors, the AP-1 sites and an additional element in the *c-jun* promoter recognised by MEF2 proteins cooperate in transcriptional control of the *c-jun* gene. Like the factors binding to the *c-fos* promoter, the activity of factors binding to the *c-jun* promoter is regulated by their phosphorylation status.

Post-Translational Control of AP-1 Activity

The most important members of the class of protein kinases regulating the activity of AP-1 in response to extracellular stimuli are mitogen-activated protein kinases (MAPKs). Depending on the type of stimuli these proline-directed kinases can be dissected into three subgroups. The extracellular signal-regulated kinases (Erk-1, Erk-2) are robustly activated by growth factors and phorbol esters, but only weakly activated by cytokines and cellular stress-inducing stimuli (UV irradiation, chemical carcinogens). In contrast

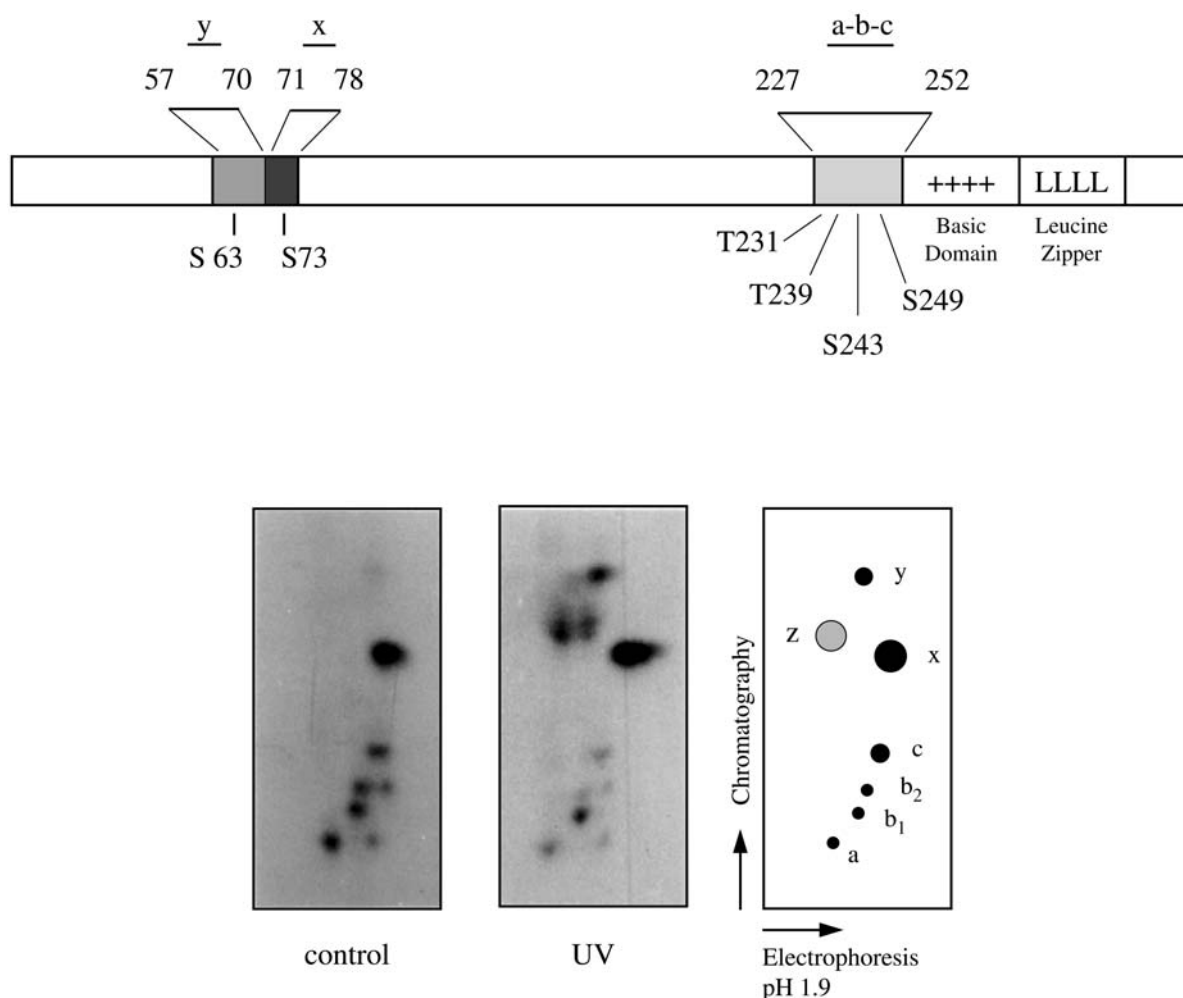
Jun-N-terminal kinases (►JNK-1, -2, -3), also known as stress-activated kinases (SAPK) and a structurally related class, p38 MAP kinases (p38 α , - β , - γ), are strongly activated by cytokines and environmental stress, but are poorly activated by growth factors and phorbol ester.

These kinases themselves are under the strict control of upstream kinases and phosphatases, which are part of individual signalling pathways initiated by specific classes of extra- and intra-cellular stimuli (growth factors, DNA damaging agents, oncoproteins). This network, which exhibits a high degree of evolutionary conservation between yeast, *Drosophila* and mammals is, however, far too complex to be discussed in greater detail in this review (for in-depth information on this subject see 7, 10, 11, 12).

Erk1 and Erk2 carry out mitogen-stimulated phosphorylation of Elk/SAP proteins. The sites phosphorylated by Erks reside in the TAD of TCF proteins and have a positive regulatory role in transactivation. The JNK/SAPKs were originally identified by their ability to specifically phosphorylate c-Jun at two positive regulatory sites (Ser-63, Ser-73) residing within the transactivation domain (Fig. 2). ►Hyperphosphorylation of both sites, which was originally identified by 2D-phospho-amino acid-peptide mapping (peptides x, y; Fig. 2), as well as additional sites at Thr-89 and -91 is observed in response to stress stimuli and oncoproteins and is responsible for transcriptional activation of c-Jun target genes. The JNKs can also phosphorylate and stimulate the transcriptional activity of ATF-2 and, to a lesser extent, TCF proteins. The same positive sites on ATF2 also serve as phospho-acceptor sites for p38, while Ser-63 and -73 of c-Jun are not affected by p38. Most probably, hyperphosphorylation of Jun, ATF and TCF proteins results in a conformational change in the TAD, allowing more efficient interaction with co-factors, such as CBP, which facilitate and stabilise the connection with the RNA polymerase II/ initiation complex to enhance transcription of target genes. Interestingly, in addition to enhanced transactivation, hyperphosphorylation of the TAD of c-Jun also regulates the stability of c-Jun by reducing ubiquitin-dependent degradation of c-Jun. Similarly, phosphorylation-dependent changes in the half-life of c-Fos have been observed.

In unstimulated cells, the DNA binding domain of c-Jun becomes phosphorylated at multiple sites (Fig. 2) by GSK-3 and/or casein kinase II (CK-II) resulting in reduced DNA binding. In response to extracellular stimuli, such as UV, phosphorylation is reduced, leading to enhanced DNA binding. The mechanism (reduced activity of the kinase or enhanced activity of a phosphatase) has not yet been defined conclusively.

In addition to phosphorylation, other mechanisms have been identified which regulate AP-1 activity. These



Jun/Fos. Figure 2 Top: schematic diagram of the human c-Jun protein. Amino acids are numbered. The numbers on top refer to the trypsin cleavage sites that lead to the appearance of phosphopeptides after *in vivo* labelling of cells with ^{32}P -orthophosphate. The location of the tryptic peptides “a-c” in the DNA binding domain and peptides “x” and “y” in the transactivation domain are indicated. Bottom: Autoradiogram of *in vivo* labelled c-Jun protein, isolated by immunoprecipitation from untreated and UV-treated cells, digested with trypsin and separated by gel electrophoresis into two dimensions. On the right the positions of the tryptic peptides are schematically illustrated. Peptide “z” most probably represents a peptide containing residual phosphorylation at threonine-89 and/or threonine-91 of c-Jun.

include redox-dependent DNA binding and regulation of nuclear localisation. Moreover, positive and negative interferences between AP-1 and other cellular proteins (in addition to the protein kinases and coactivators described above) have been identified (see “Molecular Interactions”).

In Vivo Functions of Fos and Jun Proteins in Physiology and Pathology

As described above, AP-1 activity is enhanced when cells are stimulated by mitogens. Moreover, oncogenic versions of c-Jun and c-Fos have been isolated from retroviruses and various membrane-associated or

cytoplasmic oncogenes (Ras, Src, Raf) permanently up-regulate AP-1 abundance as part of their transforming capacity, suggesting that AP-1 members play an important role in cell proliferation and transformation. Initial evidence for this assumption has been obtained by over-expression studies in various cell lines, by blocking AP-1 activity either through expression of a transdominant-negative c-Jun mutant, by expression of antisense sequences or by microinjection of Jun- and Fos-specific antibodies. Under these conditions, cell-cycle progression was disturbed and the efficiency of oncoprotein-mediated cell transformation was reduced. *In vivo* in transgenic mice, expressing a

Jun/Fos. Table 1 Summary of the phenotypes of mice containing loss-of-function mutations (introduced by gene targeting)

AP-1 member	Knockout/Knockin mice	References
c-Jun	Complete loss: embryonic lethal E12.5, hepatic failure, heart defect Liver-specific loss (albumin-cre): viable, impaired postnatal hepatocyte proliferation and liver regeneration Ubiquitous mosaic loss (Bal1-cre): malformations of axial skeleton Notochord and sclerotome-specific loss (collagen2a1-cre): increased apoptosis of notochordal cells, fusion of vertebral bodies, scoliosis of axial skeleton Epidermis-specific loss (K5-cre, K14-cre): eye lid closure defect	Reviewed by (1, 4, 8)
c-Jun ^{AA/AA}	Loss of N-terminal phosphorylation of c-Jun: viable, resistant to epileptic seizures and neuronal apoptosis induced by the excitatory amino acid kainate	
<i>junB</i> in <i>c-jun</i> locus	Rescue of <i>c-jun</i> ^{-/-} phenotype by JunB (cardiac and liver defects); rescue of c-Jun/c-Fos-, but not c-Jun/ATF- dependent gene expression	
JunB	Complete loss: embryonic lethal E8.5-E10.0, placentation defect Myeloid cell lineage: myeloproliferative disease T cells: altered Th2 cytokine expression, impaired allergen-induced airway inflammation Compromised expression in epiphyseal growth plate chondrocytes and bone lining osteoblasts: defective endochondral ossification Epiblast-specific loss (MORE-cre): osteopenia, myeloproliferative disease Macrophage/osteoclast-specific loss (lysozyme-cre): osteopetrosis	Reviewed by (2, 3, 4, 5, 8)
JunD	Viable, reduced postnatal growth, age-dependent defects in male reproductive function dysregulated T helper cell cytokines	(4, 6)
c-Fos	Viable, defective bone remodelling, osteopetrosis, light-induced apoptosis of photoreceptors CNS-specific loss: impaired long-term memory, NR2A-type NMDA receptor-dependent synaptic plasticity, resistant to chemically induced skin carcinogenesis Trp53/c-fos double knock out: development of rhabdomyosarcomas	(1)
<i>fra-1</i> in <i>c-fos</i> locus	Dose-dependent rescue of the <i>c-fos</i> ^{-/-} phenotype (bone remodeling, photoreceptor apoptosis)	
FosB	Viable, nurturing defect	(4, 8)
Fra-1	Embryonic lethal E10, placenta defect	
ATF-2	Hypomorph allele: decreased postnatal viability and growth, defects in endochondral ossification, ataxic gait, hyperactivity, decreased hearing, decreased number of Purkinje cells, decreased immediate inflammatory gene induction Complete loss: Neonatal lethality, Meconium aspiration syndrome, decreased cytotrophoblasts in placenta	

transdominant-negative mutant of c-Jun (which compromises total AP-1 activity) in keratinocytes interferes with chemically induced skin carcinogenesis. Already in these systems, different lines of evidence, including variations in the expression pattern and phosphorylation status of AP-1 members during the cell-cycle, had suggested that the members of the Jun and Fos families play specific roles during these processes or may even antagonise each other. The generation of mice

harbouring genetic disruption and/or transgenic over-expression (Tables 1, 2), as well as the availability of genetically defined mutant cells isolated from these animals represent a major breakthrough in our understanding of the regulatory functions of AP-1 subunits. The distinct phenotypes of the individual **knock out** mice (Table 1) induced by defects in cells or tissues in which the subunit was particularly important or where its absence became rate-limiting, support the notion that

Jun/Fos. Table 2 The consequences of over-expression of Fos and Jun proteins in mice (gain of function) are listed

AP-1 member	Over-expression in transgenic mice	References
c-Jun	H2K-c-jun: None H2K-v-jun: None, but upon wounding development of fibrosarcomas In cooperation with c-Fos: contribution to formation of skeletal osteosarcomas	Reviewed by (1, 4, 8)
JunB	UbC-junB: None CD4-junB: increased Th2 cytokine expression in developing Th1 cells	(8)
JunD	UbC-junD: reduced numbers of peripheral T and B cells, impaired activation of T cells due to reduced levels of IL-4, CD25 and CD69	(6)
c-Fos	H2Kb-c-Fos: osteosarcoma	(1, 4)
FosB	H2Kb-FosB: None Tcrb-δFosB: impaired T cell differentiation	(1)
Fra-1	H2Kb-Fra-1: osteosclerosis	(1)
Fra-2	CMV-Fra-2: ocular malformations	(1)

AP-1 dimers exhibit specific and independent functions *in vivo*. As a general rule derived from all studies, the AP-1 subunits must be present in a complementary and co-ordinated manner in order to ensure proper development or physiology of the organism.

Embryonic development of c-Fos, FosB or ATF-2 null mice is normal. Adult *fos*^{B^{-/-}} females, however, nurture insufficiently. ATF-2-deficient mice die shortly after birth and suffer from a disease resembling a severe type of human meconium aspiration syndrome. Loss of c-Jun, JunB or Fra-1 results in embryonic lethality. While c-Jun null embryos die at midgestation (E12.5 to 13.5) due to failure in liver and heart development, lethality of JunB (E8.0 to E10) and Fra-1 (E10.5) deficient embryos is caused by placentation failure due to multiple defects in the extraembryonic tissues. These data suggest that JunB and Fra-1, possibly as heterodimers, address common target genes responsible for the generation of a functional placental labyrinth. Despite the general association of AP-1 with proliferation, these developmental phenotypes are caused by differentiation failures rather than aberrant cell proliferation. Tissue- and cell-type specific ablation confirmed initially seen phenotypes but also revealed non-overlapping as well as common functions within preferred cellular programs. While c-Jun plays a unique role in liver regeneration in response to partial hepatectomy, loss of JunB results in a transplantable myeloproliferative disease resembling human chronic myeloid leukaemia. Viable JunD null mice develop age-dependent defects in reproduction, hormone imbalance and impaired spermatogenesis. Specific functions in skeletal and bone morphogenesis can be

assigned to a majority of the AP-1 subunits; presence of c-Jun in notochord and sclerotome protects from scoliosis of the axial skeleton, while compromised expression of JunB as well as of ATF-2 results in defective endochondral ossification. Epiblast-specific and macrophage/▶osteoclast-specific loss of JunB caused osteopenia or osteopetrosis respectively, due to failure in ▶osteoblast and osteoclast differentiation and physiology. c-Fos null mice are viable and also show defective bone remodelling, suggesting common target genes or even a genetic program affecting bone homeostasis regulated by c-Fos, JunB and ATF-2. Indeed, c-Fos is a major regulator of osteoclastogenesis, while deregulated transcriptional regulation of cyclin A in the absence of either ATF-2 or JunB has been associated with defective chondrogenesis. T cells represent another cell type addressed by various AP-1 members. JunB and JunD are required for the regulation of T helper cell specific cytokine expression and ATF-2 is needed for immediate inflammatory gene induction.

AP-1 activity is greatly enhanced upon treatment of cells with genotoxic agents, implying that AP-1 target genes are involved in the cellular ▶stress response, such as ▶DNA repair, induction of survival functions or initiation of the apoptotic program. While JunD participates in an anti-apoptotic pathway, JunB appears to be part of a pro-apoptotic pathway through negative regulation of anti-apoptotic genes, at least in myeloid cells. However, it is important to note that AP-1 subunits, depending on the cell type and stimuli, may be involved in both anti-apoptotic and pro-apoptotic responses. When N-terminal phosphorylation of c-Jun

is lost upon mutation of ser-63 and -73 to alanine *via* ►knock-in, mutant mice become resistant to epileptic seizures and neuronal apoptosis induced by the excitatory amino acid kainate. Moreover, JunAA fibroblasts from these mice show proliferation defects as well as apoptotic defects upon stress induction. By contrast, T-cell proliferation and differentiation appears to be independent of c-Jun N-terminal phosphorylation whereas efficient T-cell receptor-induced thymocyte apoptosis is affected. Several pro- and anti-apoptotic functions have also been described for c-Fos depending on the cell type and nature of the inducing signal. Light-induced apoptosis of retinal photoreceptors requires c-Fos, yet in fibroblasts loss of c-Fos results in hypersensitivity to UV-induced DNA damage. Reduced apoptosis in response to genotoxic agents was also observed in mice lacking members of the JNK/SAPK family of protein kinases, suggesting that c-Jun and ATF proteins are the major substrates of JNK/SAPKs in mediating the cellular stress response. The findings that the AP-1 members can exert ambiguous functions as pro- or anti-apoptotic proteins or as transcriptional activators or repressors depending on the target gene, the cellular context and the inducing extracellular signal open an avenue for understanding the high level of complexity surrounding AP-1.

Interestingly, knock-in approaches revealed that some members are able to substitute at least for some functions of a near relative, when expressed at sufficient levels. Knock in mice having the *c-jun* allele replaced by *junB* undergo normal embryogenesis and develop a normal liver and heart. Analysis of indicative marker genes showed that expression of genes regulated by Jun/Fos, but not those regulated by Jun/ATF dimers, are restored, thereby rescuing c-Jun-dependent defects *in vivo* as well as in primary fibroblasts and foetal hepatoblasts *in vitro*. Fra-1-deficient embryos can be rescued from embryonic death by JunB over-expression. The c-Fos-dependent functions in bone cells can be substituted by Fra-1, when the *fos* locus is deleted and replaced by the *fosl1* gene encoding Fra-1. *Fosl1* is a transcriptional target of c-Fos during osteoclast differentiation and can fully complement for the lack of c-Fos in bone development in a gene-dosage-dependent manner. However, Fra-1 is not able to induce expression of c-Fos target genes in fibroblasts derived from the knock in mice, suggesting the need for additional tissue-specific factors.

Genetically altered mice represent an excellent system for the isolation of mutant cells. Primary and immortalised fibroblasts were isolated from almost all mice lacking individual AP-1 members. Analysis of these cells revealed that c-Jun acts as a positive regulator of the cell-cycle by suppressing ►p53 and, indirectly, the p53 target gene p21. Moreover, loss of

c-Jun results in reduced ►cyclin D1 activity. JunB, on the other hand, serves as both a negative and a positive regulator of cell-cycle progression by induction of the cyclin-CDK inhibitor p16, down-regulation of c-Jun and cyclin D1 expression and transcriptional activation of cyclin A. Data from fibroblasts lacking both c-Fos and FosB established a critical role for these AP-1 subunits in cyclin D expression. Moreover, fibroblasts lacking either *c-jun* or *c-fos* cannot be transformed by oncogenes, such as Ras and Src. JunD-deficient fibroblasts exhibit specific alterations in cell proliferation, depending on p53 and p19-ARF expression. In addition to these cell-autonomous effects, critical and antagonistic functions of c-Jun and JunB in cell proliferation and differentiation *in trans* were observed. c-Jun and JunB-dependent expression of critical cytokines (KGF, GM-CSF) could be demonstrated in an *in vitro* skin equivalent model system composed of primary keratinocytes and mouse embryonic fibroblasts of wild type, *c-jun*^{-/-} or *junB*^{-/-} genotypes. This system mimics many characteristics of cutaneous wound healing.

Gain-of function studies have been performed with various AP-1 members using promoters that mediate ubiquitous gene expression, such as the H2kB, the human ubiquitin C and the CMV promoter as summarised in Table 2. In principle, for some subunits these studies were in accordance with the knockout data, showing the opposite phenotype. Tissue-specific over-expression of JunB in T lymphocytes induced an enhanced T helper cell maturation. C-Fos, Fra-1 and δ FosB over-expression resulted again in bone phenotypes causing osteosarcoma in c-Fos and osteosclerosis in either Fra-1 or δ FosB transgenic mice.

Despite the fact that AP-1 was identified more than a decade ago, it still maintains much of its mystery. Further research on tissue-specific inactivation of AP-1 members and the identification of subunit-specific target genes may yield an even more complex picture of the function and regulation of AP-1.

Acknowledgements

The work in the laboratory of the authors is supported by grants from the Deutsche Forschungsgemeinschaft, European Community and the German-Israeli Foundation for Scientific Research and Development.

References

- 1) Eferl R, Wagner EF (2003) AP-1: a double-edged sword in tumorigenesis. *Nat Rev Cancer* 3:859–68
- 2) Hartenstein B, Teurich S, Hess J et al (2002) Th2 cell-specific cytokine expression and allergen-induced airway inflammation depend on JunB. *EMBO J* 21:6321–6329

- 3) Hess J, Hartenstein B, Teurich S et al (2003) Defective endochondral ossification in mice with strongly compromised expression of JunB. *J Cell Sci* 116:4587–4596
- 4) Jochum W, Passegue E, Wagner EF (2001) AP-1 in mouse development and tumorigenesis. *Oncogene* 20:2401–2412
- 5) Kenner L, Hoebertz A, Beil T et al (2004) Mice lacking JunB are osteopenic due to cell-autonomous osteoblast and osteoclast defects. *J Cell Biol* 164:613–623
- 6) Meixner A, Karreth F, Kenner L et al (2004) JunD regulates lymphocyte proliferation and T helper cell cytokine expression. *EMBO J* 23:1325–1335
- 7) Minden A, Karin M (1997) Regulation and function of the JNK subgroup of MAP kinases. *Biochem Biophys Acta* 1333: 85–104
- 8) Hess J, Angel P, Schorpp-Kistner M (2004) AP-1 subunits: quarrel and harmony among siblings. *J Cell Sci* 117:5965–5973
- 9) Szabowski A, Maas-Szabowski N, Andrecht S et al (2000) c-Jun and JunB antagonistically control cytokine-regulated mesenchymal-epidermal interaction in skin. *Cell* 103:745–755
- 10) Weston CR, Davis RJ (2002) The JNK signal transduction pathway. *Curr Opin Gen Dev* 12:14–21
- 11) Wilkinson M, Millar J (2000) Control of the eukaryotic cell cycle by MAP kinase signalling pathways. *FASEB J* 14:2147–2157
- 12) Yang SH, Sharrocks AD, Whitmarsh AJ (2003) Transcriptional regulation by the MAP kinase signalling cascades. *Gene* 320:3–21

Juvenile Diabetes

► [Diabetes Mellitus, Genetics](#)

Juvenile Polyposis

Definition

Juvenile polyposis is an autosomal-dominant inherited gastrointestinal polyposis syndrome that is associated with mutations in either SMAD4, PTEN, or BMP receptor 1A.

► [Colorectal Cancer](#)

Juvenile Spinal Muscular Atrophy

► [Spinal Muscular Atrophy](#)

kanMX4

Definition

A gene conferring kanamycin resistance.

► [Mutagenesis Approaches in Yeast](#)

Karyopherins

Definition

Karyopherins (or importins/exportins) are nuclear transport receptors (Ran-binding transport receptors), which are responsible for the translocation of their specific cargo in or out of the cell nucleus through the nuclear pores. They act as an “adaptor” to bind or target a signal bearing cargo to the nuclear pore complex. Karyopherins bind to either nuclear localization signals (NLS) or nuclear export signals (NES) on cargo proteins, respectively.

► [Nuclear Import and Export](#)

► [RNA Export](#)

Karyotype

Definition

Karyotype defines a full set of chromosomes of a cell that is arranged with respect to size, shape and number. In a stem cell it is a must to have a normal and stable karyotype, i.e. all chromosomes in the nucleus even after unlimited cell divisions.

► [Stem Cells: an Overview](#)

► [Tumor Suppressor Genes](#)

Karyotypic Abnormalities

Definition

Karyotypic abnormalities are chromosome characteristics (size, shape number) revealed by analysis of metaphase chromosomes that differ from the usual state.

► [Hyper- and Hypoparathyroidism](#)

kb

Definition

Kilobases, i.e. thousands of bases in sequence within DNA. A gene or a DNA segment is measured in number of bases.

KD

► [Kennedy Disease](#)

Kennedy Disease

Definition

Kennedy disease (KD), also known as spinobulbar muscular atrophy, is an X-chromosomal recessively transmitted, adult-onset, neurodegenerative disorder caused by death of the spinal and bulbar motor neurons and dorsal root ganglia, due to mutations in the

► [androgen receptor gene](#) (sometimes combined with signs of partial ► [androgen insensitivity](#)).

- ▶ Polyglutamine Disease, the Emerging Role of Transcription Interference
- ▶ Repeat Expansion Diseases
- ▶ Spinobulbar Muscular Atrophy

Keratin

Definition

Keratins are a large family of intermediate filament proteins encoded by more than 50 genes in humans. About half of these are the epithelial keratins that are found in soft epithelial tissues of the human body. The rest are high-sulphur, hard keratins of which hair and nail are composed. Keratins are subdivided into Type I and Type II proteins based on their charge, and form heteropolymers consisting of a specific pair of Type I and Type II keratins in well-defined tissuespecific and differentiation-specific patterns. Keratins share a central alpha-helical rod domain, which is important for dimerization through coiled-coil interaction between a Type I and a Type II protein. These dimers form tetramers which undergo further stages of assembly to produce 10nm keratin intermediate filaments.

- ▶ Cytoskeleton
- ▶ Heritable Skin Disorder
- ▶ Intermediate Filaments
- ▶ Skin and Hair

Keratin End Domains (Head and Tail)

Definition

Keratin end domains are non-helical domains flanking the central rod domain of keratin polypeptides of variable length and sequence.

- ▶ Heritable Skin Disorders

Keratin Homologous Domain

Definition

A keratin homologous domain is a short, conserved segment of keratin polypeptides, immediately

upstream of the beginning of the rod domain in Type II keratins.

- ▶ Heritable Skin Disorders

Keratin Linker Domains/Regions

Definition

Keratin linker domains are three short, non-helical segments of the keratin polypeptide that interrupt the alpha-helical rod. The gene segments encoding the keratin linker regions represent hotspots for mutations.

- ▶ Heritable Skin Disorders
- ▶ Intermediate Filaments

Keratin Rod Domain

Definition

The keratin rod domain is the central alpha-helical domain of keratin polypeptides responsible for dimerization and higher order polymerization.

- ▶ Heritable Skin Disorders

Keratinocyte

Definition

Keratinocyte refers to the epithelial cell type which forms the major cellular unit of stratified squamous epithelia, such as the ▶epidermis of skin or the buccal mucosal epithelium. A certain percentage of keratinocytes in the basal layer of squamous epithelia undergo cell division, while keratinocytes of the more superficial layers undergo programmed cell death, thereby ensuring a permanent self renewal of skin/mucousal epithelia. Keratinocytes express keratin ▶intermediate filaments -K5 and K14 in the undifferentiated state and a variety of alternative secondary pairs upon differentiation, the keratin pair depending on the body site location. They are connected to each other and to the underlying extracellular matrix by desmosome and hemidesmosome anchorage junctions.

- ▶ Cytoskeleton
- ▶ Desmosomes
- ▶ Heritable Skin Disorders
- ▶ Intermediate Filaments
- ▶ Wound Healing

Kidney

SEPPO J. VAINIO

Biocenter Oulu and Department of Biochemistry,
Linnanmaa, Faculties of Science and Medicine,
University of Oulu, Oulu, Finland
seppo.vainio@oulu.fi

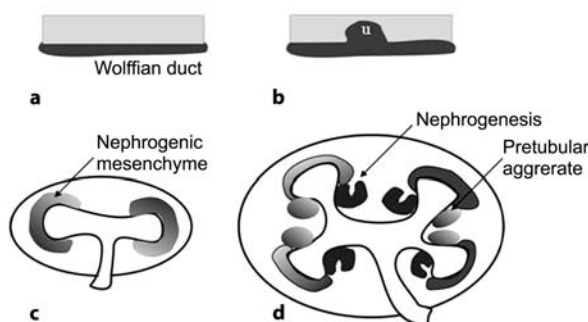
Definition

The kidney is one of the main excretory and homeostatic organs of the body. It performs two essential functions; it excretes most of the end-products of metabolism, removing them from the blood to produce urine and it controls the concentration of certain constituents of the body's fluids, such as salt and uric acid, by controlling their excretion and reabsorption. The kidney has served as a model organ for developmental studies for over 40 years. Like many other organs kidney develops as a result of sequential and reciprocal tissue interactions (1, 2). Many basic developmental events are associated with kidney organogenesis. These include epithelial branching morphogenesis, inductive tissue interactions, differentiation, cell polarization, mesenchymal-to-epithelial transformation and pattern formation. One of the aims in the field is to identify the signals that determine inductive tissue interactions, trigger nephrogenesis and later pattern the [nephron](#). Generation of mutant animals has been useful in addressing molecular mechanisms of kidney assembly and such animals have proved to be useful models of inherited human kidney disorders (1, 2).

Characteristics

Morphogenesis of the Kidney

The adult mammalian kidney, the metanephros, develops from the Wolffian duct and the metanephric mesenchyme. Both of these derive from the [intermediate mesoderm](#). The metanephric kidney begins to develop when the [Wolffian duct](#) has produced an outgrowth called the ureteric bud (Fig. 1). The ureteric bud is an epithelial tissue that invades the [metanephric blastema](#), which occurs around E10.5–11 in the mouse and E35 to E37 in humans. On invading this mesenchymal tissue, the ureteric bud induces the mesenchymal cells that surround it to condense to form a cap of closely associated cells in the [nephrogenic mesenchyme](#) (Fig. 1). The [stroma](#) develops initially at the periphery of this mesenchyme and later also in between the ureteric branches. The condensed mesenchymal cells induce the ureteric bud to branch to form two new ureteric tips and to begin to



Kidney. Figure 1 The morphological stages and essential genes of early kidney development. (a) The mouse kidney starts to develop when the ureter bud forms at the caudal end of the Wolffian duct at ~E10.5 in mice. (b) The ureter bud grows into the metanephric mesenchyme and induces the mesenchyme adjacent to the tips of the ureter bud to condense and to form the stomal mesenchymal zone. (c) The condensed mesenchyme (dark area) induces the ureter bud to branch from E11.5 onwards. (d) In association with ureter branching morphogenesis, mesenchymal cells are induced in each epithelial tip region to undergo a mesenchymal to epithelial transformation and to form the nephron. U, ureteric bud.

form pretubular aggregates that undergo a mesenchyme-to-epithelial transition to form epithelial tubules *via* S-shaped and comma-shaped bodies. [Endothelial cells](#), which invade the kidney at the initiation of its development and also develop from the kidney mesenchyme, contribute to the formation of the functional nephron (Fig. 2) (1, 2).

Molecular Interactions

Induction of Kidney Development by Mesenchymal Signals

The metanephric mesenchyme is currently thought to initiate organogenesis by inducing the formation of the ureteric bud that forms the [collecting duct](#). The initiation signals are still largely unknown.

The Wilm's tumour suppressor gene, *Wt1*, encodes a transcription factor that is an important initiation factor. *Wt1* is first expressed in the condensing mesenchymal cells and, in the *Wt1* knockout mouse, the metanephric mesenchyme forms but the ureteric bud fails to grow out of the Wolffian duct. Due to lack of ureteric growth, the metanephric blastema will degenerate *via* apoptosis. Amphiregulin, a member of the epidermal growth factor (EGF) family of secreted signals is a direct target of *Wt1* at the initiation of kidney development and mirrors its expression. The details of the mechanism of *Wt1*'s action at the initiation of kidney development remain unclear.

Wt1 regulates another gene, the paired homeobox gene *Pax2*. *In vivo*, *Pax2* expression is reduced in kidney mesenchyme in *Wt1* deficient embryos, suggesting regulatory interactions. *Pax2* controls the initiation of kidney development as it controls the expression of the gene that encodes glial cell line-derived neurotrophic factor (*Gdnf*), a member of the transforming growth factor (TGF) β family of signalling molecules. *Gdnf* also has an important role in ureteric bud induction. The absence of *Pax2* causes *Gdnf* expression to be lost from the metanephric mesenchyme and *Pax2* also regulates *Gdnf* transcription in cell culture conditions (3).

Gdnf is expressed in the kidney mesenchyme prior to ureteric bud induction and lack of *Gdnf* function leads to the failure of ureteric bud formation. As the local application of *Gdnf* to the side of the Wolffian duct *in vitro* induces supernumerary ureteric buds, *Gdnf* is evidently sufficient for the induction of ureteric bud formation.

Eya1 is critical in controlling *Gdnf* expression in kidney mesenchyme. *Eya1* is a homologue of the *Drosophila* eyes absent (*Eya*) gene, which encodes another transcription factor. *EYA1* mutations in humans are associated with the branchio-oto-renal (BOR) syndrome. As with the *Wt1*, *Pax2* and *Gdnf* deficient embryos, in *Eya1* null mouse embryos the ureteric bud fails to invade the kidney mesenchyme. As the expression of *Gdnf* is lost but *Pax2* expression persists in the mesenchyme of *Eya1* mutant embryos. *Eya1* may function in kidney mesenchyme upstream of *Gdnf* but downstream of *Pax2* to initiate kidney development.

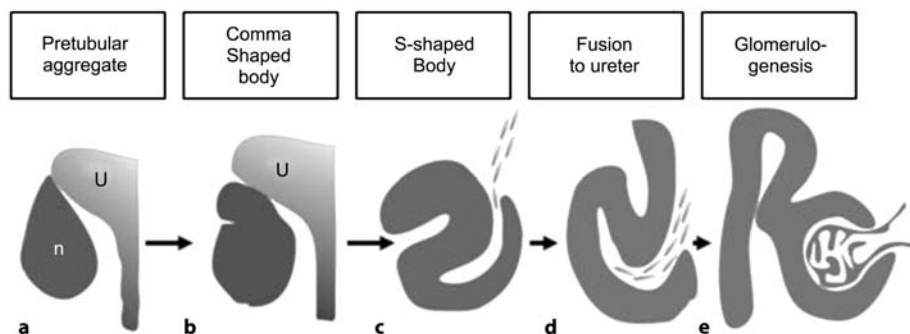
Sal1 is a new class of molecules implicated in early kidney development. *Sal1* is a mammalian homologue of the *Drosophila* region-specific homeotic gene *Spalt*. Heterozygous *SALL1* mutations in humans induce a

Townes-Brocks syndrome characterized by kidney anomalies. *Sal1* is expressed in the kidney mesenchyme and *Sal1* deficiency leads to incomplete ureteric bud growth and to a failure in nephrogenesis. Interestingly, *Sal1* deficient mesenchyme can form tubules when induced by heterologous inducer tissue. This suggests that the deficiency is transmitted to the ureteric bud. *Gdnf*, *Eya1* and *Wt1* remain expressed in *Sal1* mutant mesenchyme, which suggest that *Sal1* function lies downstream of these signals or possibly in a separate pathway.

Foxc1 is also expressed in the kidney mesenchyme and might control the location where the ureteric bud forms. *Foxc1* encodes a transcription factor of the forkhead/winged-helix family on the Wolffian duct. This conclusion is based on the fact that double ureter buds and kidneys can form in *Foxc1* knockout embryos. *Foxc1* is thought to regulate ureteric bud formation by repressing *Gdnf* and *Eya1* in the anterior region of the nephrogenic mesenchyme.

Initiation of Ureteric Bud Formation Induction of Ureteric Bud

The tyrosine kinase receptor *Ret* expressed in the ureteric bud has been implicated in the control of ureteric bud growth (2, 4). As *Gdnf* co-precipitates with *Ret*, *Gdnf* induces ureter budding in a *Ret*-dependent manner and defects in *Ret* and *Gdnf* signalling *in vivo* generate very similar kidney phenotypes suggesting that *Gdnf* contributes to ureteric bud induction via *Ret* signalling. *Gfra1*, a gene that encodes a co-receptor for *Ret*, is also expressed in the ureteric bud, and its deficiency in mice causes a phenotype related to that of *Ret* and *Gdnf* knockout mice. This *in vivo* data support the role for *Gfra1* in *Gdnf* signalling. The *Ret*/*Gdnf*/*Gfra1* system may initiate ureteric budding by promoting the growth of the ureteric bud towards the



Kidney. Figure 2 Schematic stages of nephrogenesis. (a) Ureteric bud signalling leads to formation of mesenchymal cell condensates. (b) The condensate starts to undergo mesenchymal-epithelial transformation to generate an epithelial tubule. The epithelialised tubule undergoes simple morphogenetic steps via comma and S-shape stages (c) to assemble the nephron. During nephrogenesis, endothelial cells are induced to migrate into the developing nephron that will fuse to the ureteric branch (d). The interactions between the endothelial cells and the nephron control glomerulogenesis.

mesenchymal source of Gdnf *via* a chemo attractive mechanism.

Branching of Ureteric Bud

The Ret/Gdnf/Gfra₁ system coordinates ureteric bud growth with other secreted factors, possibly pleiotrophin (Ptn, also known as HB-GAM) and Bmp4. Like *Wt1*, *Eya1*, *Sall1* and *Foxc1*, *Ptn* is expressed in the kidney mesenchyme and can stimulate the branching of isolated ureteric buds in culture together with Gdnf. *Bmp4*, a member of the ▶bone morphogenetic factor family is in turn expressed in the mesenchymal cells that surround the Wolffian duct and ureteric stalk. *Bmp4* is thus in the right place to regulate ureteric bud formation. *Bmp4*^{+/-} mouse embryos on a C57BL/6 background have kidney defects. These are thought to be due to the misregulated development of the ureteric bud. The defect mimics human congenital anomalies of the kidney and urinary tract (CAKUT).

In addition to the secreted signalling molecules, ▶extracellular matrix (ECM) proteins, such as the ▶proteoglycans (PGs), are also involved in ureteric bud development. PGs have many functions, being involved in growth factor signalling and ECM interactions. PGs consist of a core protein and sulphated ▶glycosaminoglycan (GAG) side chains. The glypican 3 (*Gpc3*) gene is mutated in humans with Simpson-Golabi-Behmel syndrome (SGBS) who have renal dysplasia. *Gpc3* contains HS GAG chains attached to the core protein, which is in turn anchored to the plasma membrane by the lipid, glycosyl-phosphatidylinositol. *Gpc3* is expressed in the ureteric bud and kidney mesenchyme. *Gpc3* inactivation leads initially to increased ureteric bud branching, while later the proliferation of the cells in the proximal collecting duct is inhibited. Hence *Gpc3* might control the cellular responses of the collecting duct to inducing or inhibitory growth factors.

Gpc3 is also a low affinity receptor for endostatin, a carboxyl-terminal proteolytic cleavage product of the extracellular matrix component type XVIII collagen. Type XVIII collagen and its fragment endostatin, are expressed in the ureteric bud and exogenously applied endostatin inhibits migration of renal epithelial cells towards hepatocyte growth factor (HGF). HGF in turn is another growth factor that, together with its oncogenic receptor c-met, is thought to regulate ureteric epithelial branching *in vitro*. Interestingly, blocking endostatin function with antibodies enhances the outgrowth and branching of isolated ureteric buds *in vitro*. This is an effect that is similar to that seen in *Gpc3* deficient mice. It is therefore possible that endostatin might normally inhibit ureteric bud branching and use *Gpc3* in the process.

Induction of Kidney Tubulogenesis

One of the central aims of the field of kidney research has been to identify the inducers of kidney tubules. Tubule formation in the mesenchyme is induced *in vivo* by the ureteric bud and *in vitro* by various ▶heterologous inducers, such as the spinal cord. The classic view has been that the mesenchymal cells that come into contact with the ureteric bud and its signal go on to form the kidney tubules.

The inactivation of *Emx2* in mice provided genetic support for the model according to which the ureteric bud secretes an initial inductive signal to trigger tubule formation *in vivo*. *Emx2* is a homeobox-containing transcription factor that is expressed initially in the ureteric bud. In *Emx2* knockout mice, the ureteric bud invades the metanephric mesenchyme normally, but no tubules are induced and kidney development fails. Although *Wt1*, *Gdnf* and *Ret* are expressed, the expression of *Wnt4* in mesenchymal cells is lost. As discussed below, *Wnt4* is believed to be an important initiating signal for tubulogenesis. *Emx2* might therefore regulate the expression of one or more unidentified ureteric bud-derived inducers that mediate epithelial signalling to the kidney mesenchyme to trigger tubulogenesis.

Wnt Signals as Inducers of Kidney Tubulogenesis

Wnts compose a large family of secreted signals that regulate key morphogenetic steps during embryogenesis. Since the discovery in 1994 that *Wnt1* can replace the spinal cord in kidney organ culture assays as an inducer of tubulogenesis, Wnts have been implicated as tubule inductive signals. However, *Wnt1* is not normally expressed in the embryonic kidney.

Of the Wnts, *Wnt4* is expressed in the condensed mesenchyme and pretubular aggregates. In *Wnt4*^{-/-} kidneys, the mesenchyme initially condenses on the ureteric bud outgrowth and expresses several early markers of mesenchymal induction, such as *Wt1* and *Pax2*, but then the mesenchymal-to-epithelial transformation fails and tubules do not form. *Wnt4* therefore appears to be a mesenchymal signal that is involved in the transition of mesenchyme to epithelium. The function of *Wnt4* signalling in triggering tubulogenesis was tested by generating NIH3T3 mouse cells to express *Wnt4*. Interestingly, both *Wnt4*-expressing cells and those expressing a panel of other Wnts were sufficient to induce tubulogenesis in the kidney mesenchyme *in vitro* (5). These studies point to an important tubule-inducing role for *Wnt4*. Interestingly, another *Wnt*, *Wnt6*, is also expressed in the developing kidney at the tips of the ureteric bud. Indeed, NIH 3T3 cell lines that express *Wnt6* can induce *Wnt4* gene expression and kidney tubules *in vitro*. *Wnt6* may

therefore function as an epithelial Wnt that triggers tubulogenesis by activating the Wnt4 signal-transduction pathway. Alternatively, Wnt6 signalling may be redundant with respect to Wnt9b and act as an inducer for this reason.

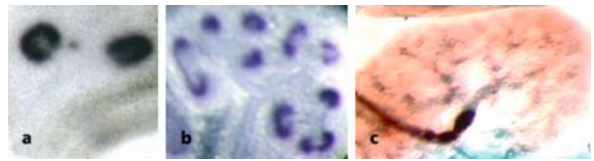
The Stroma Is Critical for Kidney Development

As discussed above, the kidney mesenchyme gives rise to cell types other than those that contribute to the nephron. One such cell type is the stromal cell, which does not differentiate into nephrons or into the collecting duct system but may be active in signalling and controlling kidney development.

The winged-helix transcription factor *Foxd1* gene (formerly called the *BF2* gene) is expressed in the stromal cell compartment. However its inactivation leads to defects in the collecting duct system and the nephrons. Hence, *Foxd1* may contribute to kidney development by regulating the expression of a stromal cell-derived signal that promotes tubulogenesis and ureteric bud development.

Vitamin A has been shown to be involved in stromal signalling as well (6). The vitamin A signal is transduced by the nuclear retinoic acid receptors (Rars). *Rara* is expressed at low levels throughout the embryonic kidney, whereas *Rarβ2* expression is restricted to stromal cells; these receptors co-localise in stromal cells. In double null *Rar* and *Rar2* mutants, ureteric bud growth is reduced, indicating a redundant function for these genes in the stroma. Consistent with this ureteric phenotype, *Ret* expression is down-regulated in the ureteric bud. When this hypothesis was tested by expressing *Ret* in the ureteric bud of double mutant mice, ectopic *Ret* expression rescued kidney development. These observations provided *in vivo* evidence that stromal cell-derived signalling maintains *Ret* expression in the ureteric bud, which in turn triggers signalling that regulates the differentiation of mesenchymal cells. These data indicate that a signalling loop might exist between the stroma, ureteric bud and kidney mesenchyme that coordinates kidney development. Indeed recent studies showed that Wnt11 (Fig. 3) and Sprouty signalling operate in a Ret/Gdnf signalling loop to contribute to ureteric branching, while Sonic hedgehog acts as an ureteric bud signal to regulate part of the smooth muscle cells' fates in kidney mesenchyme. Wnt7b is expressed in presumptive collecting duct and may regulate its development (Fig. 3).

Although the nature of the signals secreted by the early stromal cells remains elusive, there is some evidence to suggest that Fgf and Wnt signals may be involved. *Fgf7* is expressed in the stromal cells that surround the ureteric bud and developing collecting duct, whereas



Kidney. Figure 3 Expression of some important genes in the embryonic kidney. (a) At E11.5, *Wnt-11* is expressed in the ureteric bud that has initiated branching morphogenesis. (b) During subsequent ureteric bud branching, *Wnt11* is maintained at the tips of the ureteric bud tree (E13.5). (c) *Wnt7b* expression is excluded from the very tips of ureteric buds while its expression is localised mainly to presumptive collecting duct.

the *Fgf7* receptor is expressed in the ureteric bud itself. In *Fgf7*-deficient mice, the ureteric bud has branched less and ~30% fewer nephrons are formed. When FGF7 is added to organ culture *in vitro*, it stimulates the growth of isolated mutant ureteric buds. This indicates that FGF7 serves as a growth factor for ureteric bud development. *Wnt2b* is expressed in the early stromal mesenchyme of the kidney, suggesting that it might function in this kidney compartment. Indeed, kidney organogenesis can be reconstituted *in vitro* by treating isolated ureteric buds with cells that express *Wnt2b*, prior to co-culturing the ureteric buds with mesenchyme. Wnt2b may therefore function as a reciprocal mesenchymal signal that regulates early ureteric branching; its inactivation in mice may provide proof of this possibility.

Regulatory Mechanisms

Even though knowledge of the regulatory mechanisms of kidney development is still rather limited, we can draw some hypothetical models of the molecular mediators involved. Initiation of kidney development involves several transcription factors such as WT-1, Pax2, Foxc1 and Sall1. Gdnf is a critical factor together with pleiotrophin in regulating ureteric branching, while Emx2 controls ureteric growth and signalling, with modifiers of proteoglycans such as glypican 3 involved in the control of ureteric branching. Wnt11 regulates ureteric patterning at the initiation of trifurcation of the ureteric bud while Sonic [hedgehog](#) regulates smooth muscle cell fates as a ureteric bud derived signal. Wnt4 acts as a critical tubule inductive signal. The possibility of applying the technology of conditional mutagenesis will permit the next generation of experiments to target gene functions in organogenesis of the kidney to be done. A selection of the genes implicated in early kidney development is presented in Table 1.

Kidney. Table 1 Some genes that are implicated in early kidney development

Type/Name	Tissue type expressing	Phenotype
Transcription factors		
<i>Emx2</i>	U, MM	Kidneys, U, genital tract completely missing
<i>Eya1</i>	MM	Absence of U growth and failure of induction of MM
<i>Foxc1</i>	MM	Two kidneys and double U's
<i>Foxd1</i>	S	Mutant kidneys are small, few nephrons, fused longitudinally
<i>Pax2</i>	U,MM	Deficient U outgrowth, MM uninduced
<i>RARα & RARβ2</i>	U, S,MM	Hypoplasia/agenesis
<i>Sall1</i>	MM	Incomplete U outgrowth, failure
<i>Wt1</i>	MM	MM undergoes unrescuable apoptosis tubule formation, apoptosis
Growth factors		
<i>Bmp4</i>	MM	+/- Hypo/dysplastic kidneys, hydroU, ectopic uterovesical junction, double collecting duct
<i>Bmp7</i>	U, MM	Severe hypoplasia with few nephrons and collecting ducts
<i>Fgf7</i>	S	Small kidneys, fewer U branches and nephrons
<i>Gdnf</i>	MM	No kidney as U bud fails to grow
<i>Wnt4</i>	MM	Failure of kidney tubule formation
<i>Wnt11</i>	U	Defect in ureteric patterning
Sonic hedgehog	U	Defect in smooth muscle cell differentiation
Growth factor/receptor		
<i>Gfra1</i>	U, MM	Agenesis of the kidney as in <i>Ret</i> and <i>Gdnf</i>
<i>Notch2</i>	MM*	Glomerular defects
<i>Ret</i>	U	Failure of U growth
Proteoglycans and their biosynthetic enzymes		
<i>Hs2st</i>	U, MM	Renal agenesis due to lack of U branching and mesenchymal condensation
<i>Glypican3</i>	U, MM	Selective degeneration of medullary collecting duct

U, ureteric bud; MM, metanephric mesenchyme; S, stromal cells. Expression at later steps on nephrogenesis not indicated. * In maturing nephron and glomerulae

Acknowledgement

MSc. Herkko Sikkilä is acknowledged for critically reading the manuscript. The work was financially support by the Academy of Finland, Sigrid Jusélius Foundation and EU program. Due to space limitations many important contributions to the field could not be cited.

References

1. Saxén L (1987) Organogenesis of the kidney. Cambridge, University Press, pp 1–173
2. Vize P, Woolf AS, Bard JBL (2003) The Kidney. From congenital development to congenital disease. Academic Press, pp 1–519

3. Brophy PD, Ostrom L, Lang KM et al (2001) Regulation of ureteric bud outgrowth by Pax2-dependent activation of the glial derived neurotrophic factor gene. *Development* 128:4747–4756
4. Schuchardt A, D'Agati V, Larsson-Blomberg L et al (1994) Defects in the kidney and enteric nervous system of mice lacking the tyrosine kinase receptor Ret. *Nature* 367:380–383
5. Kispert A, Vainio S, McMahon AP (1998) Wnt-4 is a mesenchymal signal for epithelial transformation of metanephric mesenchyme in the developing kidney. *Development* 125:4225–4234
6. Batourina E et al (2001) Vitamin A controls epithelial/mesenchymal interactions through Ret expression. *Nat Genet* 27:74–78

Kinase

Definition

Kinase/protein kinases are cytosolic regulatory enzymes that modify the functional properties of certain cytosolic proteins (substrates), by catalysing the transfer of phosphate groups from a high-energy phosphate-containing molecule (as ATP or ADP) to the side chains of serine, threonine or tyrosine residues. This large group of enzymes not only comprises those transferring phosphate, but also diphosphate, nucleotidyl residues, and others. Kinases have also been subdivided according to the acceptor group.

- ▶ Bone and Cartilage
- ▶ Cell Cycle – Overview
- ▶ Cell Polarity
- ▶ Peptide Chips
- ▶ Peutz-Jeghers Syndrome
- ▶ Protein Microarrays as Tool for Protein Profiling and Identification of Novel Protein-Protein Interactions
- ▶ Receptor Serine/Threonine Kinase
- ▶ Tyrosine Kinases

Kinetochores

Definition

Kinetochores refer to a proteinaceous structure at the condensed chromosomes, which serves as the attachment site for microtubules during mitosis. Microtubules attach at the kinetochore and segregate the two sets of chromosomes.

- ▶ Cell Cycle – Overview
- ▶ Centromeres

- ▶ Cytoskeleton
- ▶ Meiosis and Meiotic Recombination
- ▶ Mitotic Spindle

Klinefelter Syndrome

Definition

Klinefelter syndrome is the most common chromosomal disorder defined by the 47,XXY chromosome constitution. Phenotypic features include tall stature, hypogonadism, infertility, breast enlargement, learning and behavioural problems.

- ▶ Marfan Syndrome

Knockdown

Definition

Knockdown describes a treatment that results in the loss or reduction of function of a targeted gene/protein. This may be the result of a dominant inhibitory mutant, antibody, ▶ Morpholino Oligonucleotides, Functional Genomics by Gene ‘Knock-down’, siRNA, etc. In the classical genetic nomenclature, this type of mutation is analogous to a hypomorphic or amorphic allele, where the gene expression level is reduced but not knocked out completely.

- ▶ *Drosophila* as a Model Organism for Functional Genomics
- ▶ Large-Scale Homologous Recombination Approaches in Mice
- ▶ *Xenopus* as a Model Organism for Functional Genomics

Knock-In (Mutation)

Definition

Knock-in (mutation) refers to a mutation that is created by replacing the selectable marker of a knock-out mutation by the desired artificially generated allele of the gene.

- ▶ Jun/Fos
- ▶ Mutagenesis Approaches in Yeast
- ▶ Transgenic and Knock-out Animals

Knock-Out

Definition

Knock-out describes the targeted disruption of a specific gene by homologous recombination, leading to the complete absence of the gene product.

- ▶ Common Diseases, Genetics
- ▶ Heritable Skin Disorders
- ▶ Jun/Fos
- ▶ Transgenic and Knockout Animals

Knock-Out Animals

- ▶ Transgenic and Knock-out Animals

Knock-Out Mice (Null Mice)

Definition

The term knock-out or knock-out mutation characterises the inactivation of specific genes in order to study their function. Knock-out mice (also called Null mice) are derived from embryonic stem cell clones, in which a particular gene was disrupted (knocked-out) by targeted integration of a defective gene construct. If the mice differ from their normal litter mates, it is assumed that the phenotype is a consequence of the gene disruption. Knock-out mouse models allow the study of disease predisposing genes identified in humans.

- ▶ Cre/Lox P Strategies
- ▶ Defective Protein Folding Disorders
- ▶ DNA Recombination
- ▶ Glial Cells and Myelination
- ▶ Large-Scale ENU Mutagenesis in Mice

- ▶ Large-Scale Homologous Recombination Approaches in Mice
- ▶ Mutagenesis Approaches in Yeast
- ▶ Transgenic and Knock-out Animals
- ▶ Wound Healing

Kremen

Definition

Kremen refers to a family of Type I transmembrane proteins which are high affinity receptors for ▶ dickkopf-1 (Dkk).

- ▶ Wnt/Beta-Catenin Signaling Pathway

Kugelberg-Welander Disease

- ▶ Spinal Muscular Atrophy

Kupffer Cells

Definition

Kupffer cells are the resident macrophages of the liver located in the walls of the liver sinusoids. These cells are intensely phagocytic and – together with the liver sinusoidal endothelial cells – constitute the most powerful scavenger system in the body, removing waste macromolecules or bacteria from the blood of the portal vein. Kupffer cells are also able to secrete several types of mediators such as cytokines, oxygen and nitrogen species, and prostaglandins, mostly after stimulation through infection.

- ▶ Hemochromatosis

Labeling

Labeling is a method of adding a chemical (e.g. radioactive isotope; ►[isotope labeling](#)) or biological compound to a system in order to identify, visualize or trace proteins or specially designed tags.

- [Electron Tomography](#)
- [PET](#)

Laboratory Informatics Management System

Definition

Laboratory informatics management system describes the specialized application of information technology to track laboratory operations. This comprises the full path of data flow, from data acquisition (instrument interfacing) to data storage analysis and provision.

- [Automated High-Throughput Functional Characterization of Human Proteins](#)

LacZ

Definition

The *LacZ* gene codes for the enzyme beta-galactosidase, and is often used as a reporter system in mammalian cells and tissues. In knockouts, and when the *LacZ* has been incorporated appropriately into a knockout vector, it allows the *in vitro* visualization of expression from the targeted gene.

- [Large-Scale Homologous Recombination Approaches in Mice](#)

Lagging Strand

Definition

Lagging strand refers to a DNA strand that is replicated discontinuously in direction, opposite to that of the fork movement (direction of synthesis and replicative fork moving are opposite).

- [Replication Fork](#)
- [Replication Origins](#)

Lamellipodium

Definition

Lamellipodium is a flattened projection from the anterior region of a cell. It is an actin rich zone that is formed in response to receptor mediated signal transduction. It propulses the forwarding movement of a migrating cell.

- [Cell Migration](#)
- [Focal Complexes/Focal Contacts](#)
- [Rho, Rac, Cdc42](#)

Laminar Flow

Definition

Laminar flow is defined as the state of fluid (or gas) flow wherein particles move along parallel, ordered paths.

- [Proteomics in Microfluidic Systems](#)

Laminin

Definition

Laminin designates a family of glycoproteins that form an integral part of basement membranes. All laminins are heterotrimers assembled from α , β , and γ -chains. From the known 5 α , 3 β , and 3 γ -chains in vertebrates, at least 15 different laminin trimers can be formed, which are assembled into ternary protein complexes, via coiled coil interactions. The complete trimers have been named by arabic numerals (Laminins 1 to 15), roughly in order of their discovery. They are considered important for organ development.

- ▶ [Extracellular Matrix](#)
- ▶ [Hemidesmosomes](#)

Lamm Equation

Definition

Lamm equation refers to the partial differential equation predicting the changes with time of the macromolecular concentration distribution as a result of diffusion, sedimentation, and chemical reaction fluxes in a sector-shaped solution column in the centrifugal field.

- ▶ [Analytical Ultracentrifugation](#)

Langer-Giedion Syndrome

Definition

Langer-Giedion syndrome describes a rare contiguous genetic disease (1:100,000 newborns) caused by loss of the TRPS1 and EXT1 genes on one homologue of human chromosome 8q24.1. Infants show short stature, facial anomalies, redundant skin, hypotonia, and frequent infections; later signs include multiple exostoses, rheumatoid deformation of fingers and toes, and mental retardation.

- ▶ [Microdeletion Syndromes](#)

LAP

Definition

LAP stands for Latency Associated Peptide. It consists of the N-terminal part of the TGF- α propeptide, which

is proteolytically cleaved from the mature C-terminal peptide, and remains non-covalently bound to mature TGF- α and disulfide-linked to other proteins of the latent TGF- α complex, such as latent TGF- β binding proteins (▶ [LTBPs](#)). The function of the propeptide is to confer inactivity to the mature C-terminal peptide.

- ▶ [Receptor Serine/Threonine Kinase](#)

Large Scale ENU Mutagenesis In Mice

MARTIN HRABÉ DE ANGELIS, JOHANNES BECKERS, SIBYLLE WAGNER, DIAN SOEWARTO
GSF Research Center for Environment and Health, Neuherberg, Germany
hrabe@gsf.de

Definition

Chemical mutagenesis by ▶ [ENU](#) (ethylnitrosourea) is a method for efficient production of mouse ▶ [mutants](#). ENU-mutagenesis is used in the field of forward genetics to create mouse models with specific ▶ [phenotypes](#) (e.g. disease models) and unknown ▶ [genotypes](#). The latter are discovered by genetic-mapping and cloning strategies. ENU mutagenesis can also be used as a reverse genetics approach, creating mouse models with a known ▶ [mutation](#) and yet unknown phenotype. It can be applied on a large-scale basis and/or in combination with other mutagenesis strategies to genetically dissect phenotypes or molecular pathways. ENU is the most powerful ▶ [mutagen](#) known in mammals.

Characteristics

ENU is a highly effective mutagen in mammalian germ cells with mutations rates higher than any form of radiation mutagenesis tested to date (9). ENU was originally used to study the influences of a number of variables on germ cell mutagenicity such as gender, sensitivity of germ cell stages and dose fractionation. In the mouse, ENU is known to have its main effect on pre-meiotic stages of spermatogonial stem cells. These stages are characterized by biological parameters such as cell division (mitosis, meiosis), DNA replication and DNA lesion repair, which presumably affect the sensitivity to mutation induction by ENU mutagenesis. ENU depletes multiple stem cells, such as hematopoietic stem cells, which causes temporary sensitivity to pathogens. Therefore, ENU treated male mice not only undergo a period of sterility, but are also more susceptible to infections than untreated males and have a shorter life span.

ENU transfers its ethyl group to nitrogen or oxygen in the DNA and can lead to mispairing and base pair substitution after DNA replication if DNA repair mechanisms do not revert the mutation. ENU mainly induces point mutations, namely base pair substitutions such as A/T→T/A transversions and A/T to G/C transitions, covering about 82% of all induced mutations or small deletions. Base pair substitutions in coding regions result in missense, nonsense or mRNA

▶**splicing** defects.

When used in its optimal dose, ENU induces approximately one mutation in every 700 gametes per locus. There is a dose response correlation for ENU induced specific locus mutations in mouse spermatogonia, which have been assessed in the ▶**specific locus test** (SLT). Several studies showed that specific locus mutations were not induced when a lower threshold dose up to approximately 38 mg ENU/kg body weight was applied. This threshold is presumably due to the presence of DNA repair mechanisms that, below a certain dose, are not saturated and able to repair the ENU induced adducts efficiently. Above this threshold, a linear correlation could be demonstrated between the ENU concentration up to a dose of 400 mg ENU/kg body weight (fractionated dose of 4×100 mg/kg) and the mutation frequency.

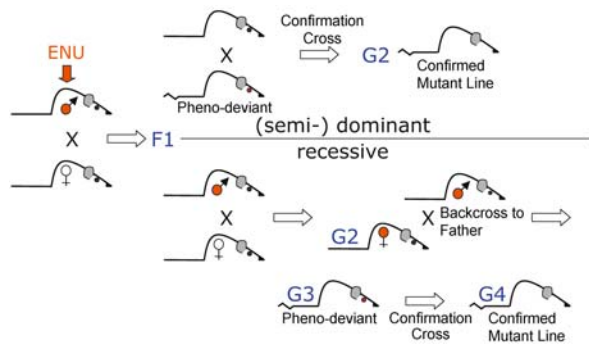
To date, mutation frequencies for ENU have been determined mainly in specific locus experiments but can be different for other loci depending on the phenotyping assay applied.

New molecular technologies in mutation detection at the DNA level and ▶**sequencing** have revealed mutation frequencies of ENU induced mutations in mouse embryonic stem (ES-cells) of one mutation in a gene in every 150 ES cells (2).

The efficiency of an ENU mutagenesis project is dependent on different issues such as the choice of appropriate mouse strains or the optimal dose of ENU (5). It is known that certain inbred strains exhibit different sensitivities to chemical mutagens, presumably due to different characteristics of DNA repair systems in their germ cells. In most instances, inbred strains are used for mutagenesis experiments in order to avoid the difficulties of a polymorphic genetic background in chromosomal mapping or for phenotypic analysis and characterization. Another issue is the application of the optimal dose of ENU which has already been determined for some of the most common inbred strains considering a) the period of sterility, b) the percentage of male mice that regain fertility, c) the mortality rate before or after regaining fertility and d) the mutation frequency in the F1 or G3 generation.

Forward Genetics

Within the analysis of gene functions, mutant phenotypes resulting from spontaneous or induced mutations



Large Scale ENU Mutagenesis In Mice.

Figure 1 Generalized scheme of a dominant (or semi-dominant) and recessive ENU mutagenesis screen. In both cases male mice are treated with ENU to mutagenize spermatogonia and crossed to a wild type female (generally from the same genetic background). In the dominant screen, mice of the resulting F1 generation are examined for significant phenotypic variation from the wild type phenotype. The genetic (non-congenital) nature of such pheno-deviant mice needs to be established by a confirmation cross, generally to a mouse of the same inbred strain. Only if the phenotypic change is confirmed in the resulting (G2) generation (by dominant or semi-dominant inheritance) may the pheno-deviant of the F1 generation be designated as the founder of a new mutant line. The scheme shown for the recessive ENU mutagenesis screen allows the efficient identification of potential recessive alleles in a rather small number of G3 animals. Male mice from the F1 generation are crossed back with wild type females to generate females heterozygous for the mutant allele. To generate homozygous G3 animals a limited number of females from the G2 are crossed back to their father. Pheno-deviants may be identified in the third generation (G3). To establish a mutant line, the genetic inheritance of the allele needs to be demonstrated by crossing the G3 pheno-deviant to either its G2 mother or F1 grandfather (depending on the gender of the pheno-deviant). The mutated gene may be identified by inter- or back-cross; the mouse line may be analyzed for gene function, distributed to the scientific community and archived by cryopreservation of germ cells.

have always had an important role. This strategy is also called forward genetics or the phenotype-driven approach (Fig. 1). The ability of ENU to induce different types of ▶**alleles** including loss- and gain-of-function, hypo- and hyper-morphs as well as dominant negative alleles allows the creation of allelic series of single genes. Such allelic series are a valuable tool in the analysis of gene function by revealing different and novel gene functions. The main interest in such phenotype driven approaches is the implementation of appropriate procedures to assess the mutant

phenotypes of interest. Systematic ENU mutagenesis screens have successfully been performed in several non-mammalian organisms such as fruit fly or zebra fish from which most of the genes and pathways that are involved in early development have been identified. The first ENU mutagenesis screens that were carried out in the mouse focused on the specific chromosomal region of the T locus. One of the largest genome-wide screens for specific phenotypes identified dominant mutations causing cataracts in mice out of which 35 different alleles were isolated. Although several thousand different mouse mutations are available, this represents only a small fraction of the estimated total number of about 30,000 genes in mammals. To close this “phenotype gap”, several large-scale and genome-wide ENU mutagenesis programmes have been launched all over the world in recent years.

In order to identify the causative mutation of an abnormal phenotype, it is first necessary to map the chromosomal localization of the mutation by ► [linkage analysis](#) using an outcross/backcross or an outcross/intercross mapping strategy in different inbred strains. An optimal condition for linkage analysis is the use of two mouse-inbred strains having high polymorphism. A backcross strategy is preferred and more efficient:

- if the mutant allele is fully penetrant or
- if the mutant allele is recessive and homozygotes are fertile.

An intercross strategy is preferred and more efficient:

- if the homozygous individuals are not fertile and
- if the mutation is recessive and affected homozygous individuals are not fertile.

Linkage analysis is performed using a genome-wide panel of ► [microsatellite](#) markers (www.informatics.jax.org) or using single-nucleotide polymorphism (► [SNP](#)) maps (www.genome.wi.mit.edu/SNP/mouse/). Computer programs such as Map manager, GeneLink or Mapmaker are available for the analysis of mouse linkage data. The analysis of for example 100 meioses results in the determination of a chromosomal region of about 10 cM (► [centi Morgan](#)).

The mapped location of the chromosome is then applied to determine the mutation in a new or already known gene by a ► [candidate gene approach](#) or ► [positional cloning](#).

Modifier/Sensitized Screen

Modifier or sensitized screens are also phenotype-driven strategies to study specific genetic interaction in particular biological pathways. The aim of such screens is to detect new alleles that enhance or suppress the phenotype of the original mutation or transgenic model (in an allelic or non-allelic, non-complementary manner). In these sensitized screens, ENU treated

males may be mated, e.g. to female mice carrying mutations in the pathways of interest. In the G1 generation, mutant phenotypes may be recovered 1) that show a completely new dominant phenotype owing to the presence of a dominant mutation that might be dependent or independent from the pathway of interest 2) that show a less (suppressor) or more severe (enhancer) appearance compared to the original mutant phenotype due to a dominant mutation, which directly affects and modifies the pathway of interest and 3) that show a mutant phenotype caused by new alleles at the locus of the original recessive mutation.

Reverse Genetics

In contrast to the phenotype-driven approaches described above, the gene-driven approach starts with a gene of interest to analyze the effect of the mutation on the phenotype. Nevertheless, most of the induced mutations by gene targeting resulted in null alleles.

With the combination of new techniques in cryobiology and the development of DNA based mutation detection methods on a high-throughput level (1), ENU mutagenesis has become important in gene-driven approaches, also called reverse genetics. The cryopreservation of spermatozoa of male mice offers the possibility of establishing a sperm and tissue bank from a large colony of male F1 or G3 offspring whose parents have been treated with ENU. In order to get mice with a mutation in a gene of interest, it is now possible to screen large numbers of DNA samples in an efficient way. The re-derivation of interesting mutants can be carried out by *in vitro* fertilization and re-implantation of two cell embryos or blastocysts into pseudo-pregnant female mice.

Another gene-driven strategy is the treatment of embryonic stem cells with ENU. The aim is the establishment of a cryopreserved bank of ENU treated ES cells that can be screened for mutations in any gene of interest. Once the mutation is identified, the phenotypic consequences are analyzed in animals derived from the germ line contribution of the mutated ES cells.

A high mutation rate coupled with high throughput mutation detection technology renders these approaches applicable to generating mutations in any gene of interest and thereby creating an allelic series of mutations.

Clinical Relevance

The relevance of ENU mutagenesis for human diseases can be assessed on at least two levels, the relation of phenotypic screens in the mouse to human diseases and the correspondence of mutated gene and phenotype in man and mice. Some examples of both aspects are given below. In addition, it will be shown how the application of ENU mutagenesis in reverse genetic

approaches can support the isolation of new alleles of known genes with importance for human diseases and how sensitized screens allow the identification of novel genes that may be associated in particular with ►polygenic diseases.

Relevance of “Model-Phenotypes” to Disease

The majority of mouse models that have been generated in ENU mutagenesis screens are of high clinical relevance (4, 8). This is primarily due to the fact that many phenotype driven screens were and are specifically designed to examine parameters that are also affected in human diseases. The general concept of phenotype driven approaches is to identify mutant lines based on, for example, morphological, physiological or behavioral parameters that are distinct from established baselines of the corresponding wild type strain. The frequency with which disease models can be identified in genome wide mutagenesis screens may be substantially increased when a large variety of parameters are assessed for each mutant line. For these reasons new non-invasive phenotypic screens are continuously being adapted for mice from human clinical examinations.

External and non-invasive screens for dysmorphologies may reveal phenotypic abnormalities, for example of sense organs, the limb or axial skeleton, the skull, pigmentation of hair or skin or other anatomical structures, which may be relevant for the identification of genes affected in human congenital malformations. X-ray and computer tomography have been adapted to small rodents, so that they can reveal subtle dysmorphologies of the skeleton and soft tissues comparable to the precision achieved in human examinations. Simple behavioral tests including, for example, click-box-tests (for hearing), have successfully been used to identify new mouse models for congenital malformations and progressive degeneration of the sensory patches or vestibular structures of the inner ear. Examination of the eye lens using the split lamp or electro-retinograms to measure action potentials in the retina are examples of efficient tools to identify models for diseases affecting the eye. Clinical chemical and biochemical metabolite assays may assess, for example, blood plasma enzyme activities and plasma concentrations of electrolytes and metabolites using blood auto-analyzers that are comparable to those used in clinics. These parameters are of particular importance for human disorders. For example, abnormalities in glucose metabolism may be indicative of diabetes, serum urea levels may be related to gout and so on. In addition, alterations in the immune system that are characteristic of, for example, autoimmune disease, allergic reactions or pathological changes in immune cell populations in man may also be assessed in the mouse by applying fluorescence associated cell sorting

(FACS) or enzyme-linked immunosorbent assays (ELISA). Atherosclerosis, hypertension, and other cardiovascular diseases are major causes of mortality. The screening of cardiovascular disorders in mouse models involves the examination of cardiac functions, blood pressure regulation and measurement of cardiovascular hormones and lipoprotein metabolites. Primary, non-invasive screens in the mouse may include, for example, echocardiography or plethysmographic measurement of blood pressure. Exercise capacity examinations are also possible in the mouse model. Such cardiovascular screens are directly adapted from human examinations.

Other relevant screening parameters to identify mouse models for human diseases include, for example, lung function tests, metabolic screens, measurement of metabolized energy, standardized behavioral tests with relevance for human psychology and genome wide gene expression analyses to name just a few. The phenotypic screens that are being used to find novel mouse models for human diseases in ENU mutagenesis (and other mutagenesis and targeting) approaches continue to be refined and standardized. Novel screens are being adapted while the mouse model system becomes increasingly important for the study of the molecular and genetic bases of human diseases.

Relevance of “Model-Mutations” to Diseases

After having generated several thousands of mouse models in large-scale ENU mutagenesis screens, attention is beginning to shift from the generation of new disease models to the identification of the underlying mutation and the affected gene(s). Two examples will be given in which the identification of the mutated gene in the mouse was accompanied by the identification of mutations in the homologous gene in man resulting in related disease phenotypes. These findings demonstrated that the mouse provides not only a “phenotypic model” but also a direct “genetic model” for the human disease, at least for some gene/phenotype relations in both species.

The first example is an ENU induced mouse mutant line that was discovered as a congenital zonular cataract with progressive opacity (*Po*) and a semidominant mode of inheritance. The induced T to A substitution was found to affect the fourth Greek key structural motif in the β A1- and β A3-crystallins, both encoded by the *Cryba1* gene (3). A mutation in the human homologue, *CRYBA1*, is phenotypically very similar to the *po* allele. Thus, this particular mouse mutation provides an excellent model for a human congenital zonular cataract with suture opacities, caused by a mutation in the homologous gene.

The second example is the semidominant *Beethoven* (*Bth*) missense mutation in the *Transmembrane Cochlear-Expressed* gene (*Tmc1*). This mutant line

is phenotypically characterized by progressive hearing loss and degeneration of cochlear hair cells (10). At least one ►**dominant allele** (DFNA36) has been identified in the human homologue *TMC1* that is associated with progressive hearing loss (6). This makes *Bth* an invaluable model to study postlingual deafness and could also provide new insights into the factors needed for long-term survival of hair cells. The examples above stand for many other mutant lines that have demonstrated the clinical relevance of ENU induced mutations.

Application of Reverse-Mutagenesis-Genetics to Identify Novel Disease-Alleles

With the availability of high-throughput sequencing infrastructures and the hundreds of thousands of mutagenized genomes that have been archived, it has now become possible to profit from ENU mutagenesis in reverse genetic approaches also. In this strategy, selected genes of clinical relevance may be sequenced in thousands of ENU mutagenized mouse genomes to identify potentially interesting mutations in these genes. The corresponding mouse mutant lines may then be re-derived from frozen sperm or oocyte archives and the phenotype can be assessed.

Application of Sensitized Mutagenesis to Identify Polygenic Disease Genes

In sensitized screens, a known mutation in a single gene is used to simplify the identification of genes or pathways involved in polygenic diseases. In such an analysis one mouse mutant line carries a mutation that is associated with a particular disease or ►**trait**. The second line is mutagenized (but may also carry the sensitizing mutation – depending on whether modifiers are sought in the heterozygous or homozygous background). The phenotypic screen in the filial generation of both lines may identify, for example, dominant modifiers (enhancers or repressors) of a quantitative or qualitative trait (recessive modifiers may be identified in subsequent generations). The relevance for human diseases in sensitized polygenic trait analyses is through the selection of the sensitizing mutation (involved in a disease) and the (disease-relevant) phenotypic screen (7).

References

1. Beier DR (2000) Sequence-based analysis of mutagenized mice. *Mamm Genome* 11:594–597
2. Chen Y, Schimenti J, Magnuson T (2000) Toward the yeastification of mouse genetics: chemical mutagenesis of embryonic stem cells. *Mamm Genome* 11:598–602
3. Graw J, Jung M, Loster J et al (1999) Mutation in the betaA3/A1-crystallin encoding gene *Cryba1* causes a dominant cataract in the mouse. *Genomics* 62:67–73
4. Hrabe de Angelis MH, Flaswinkel H, Fuchs H et al (2000) Genome-wide, large-scale production of mutant mice by ENU mutagenesis. *Nat Genet* 25:444–447
5. Justice MJ, Carpenter DA, Favor J et al (2000) Effects of ENU dosage on mouse strains. *Mamm Genome* 11:484–488
6. Kurima K, Peters LM, Yang Y et al (2002) Dominant and recessive deafness caused by mutations of a novel gene, *TMC1*, required for cochlear hair-cell function. *Nat Genet* 30:277–284
7. Matin A, Nadeau JH (2001) Sensitized polygenic trait analysis. *TIG* 17:727–31
8. Nolan PM, Peters J, Strivens M et al (2000) A systematic, genome-wide, phenotype-driven mutagenesis programme for gene function studies in the mouse. *Nat Genet* 25:440–443
9. Russell WL, Kelly PR, Hunsicker PR (1979) Specific-locus test shows ethylnitrosurea to be the most potent mutagen in the mouse. *Proc Natl Acad Sci USA* 76:5918–5922
10. Vreugde S, Erven A, Kros CJ et al (2002) Beethoven, a mouse model for dominant, progressive hearing loss DFNA36. *Nat Genet* 30:257–258

Large Scale Protein Production

Definition

Large scale protein production denotes a cost-efficient method for protein production, usually in large fermenters.

►**Protein Tags**

Large-Scale Gene Trap Approaches in Mice

PATRICIA RUIZ

Center for Cardiovascular Research, Charité University Medicine Berlin and Max Planck Institute for Molecular Genetics, Berlin, Germany
ruiz@molgen.mpg.de

Definition

Since mouse mutants are valuable resources for the study of mammalian gene function in the context of the whole organism, large-scale mouse ►**mutagenesis** programs have recently gained increasing attention. Several mutant collections have been established with this aim using either a phenotype- (1) or a gene-driven

approach (2). The development of high-throughput mutagenesis screens in the mouse using ethylnitrosourea (ENU) and gene trap vectors has enabled an efficient large-scale mutagenesis approach. The main advantages of these systems are that (i) many more genes can be mutated at once and (ii) no initial knowledge of gene structure or sequence is necessary to mutate a given gene.

Within the concept of tackling gene function with a ►gene-driven approach, there are basically two methodological ways of approaching the problem:

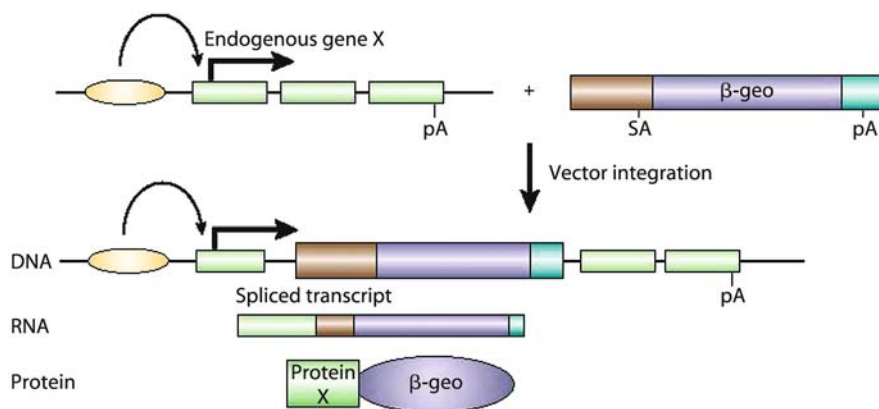
1. the generation of defined mutations in the gene(s) of interest (e.g. over-expression, ectopic expression or inactivation of a gene and knock-in of a modified gene product isoform, conditional mutations using cre-loxP).
2. the non-directed, systematic generation of mutants without any previous knowledge of the gene to be mutated (e.g. gene traps, mutagen-treated ES-cells).

Homologous recombination in ES-cells has led to the development of hundreds of mouse mutants that have aided in the annotation and functional characterization of many genes (3). Although very successful, this method is still very work- and time-consuming and has been prone to a high degree of redundancy; over the last 15–20 years ~7,000 genes have been inactivated. An overall estimation would therefore predict 70–100 years before a given mouse mutant would be available for all mouse genes (assuming the current calculation of 25–35,000 genes/transcripts). Gene trap technology was developed to contribute to the goal of saturating the mouse genome with mutations. Since it combines gene trap vector and ES-cell technologies, it offers the possibility of applying high-throughput methods to this ambitious goal.

Characteristics

The non-directed insertion of different selector cassettes with or without minimal promoters had led to the identification of many new genes and enhancer elements in bacteria, *Drosophila* and *C. elegans*. The enhancer traps were adapted to the mouse system 15 years ago, but had to be largely modified to identify and at the same time mutate genes of a much more complex organism such as the mouse. Therefore, splice-acceptor (SA) sites were inserted 5' to a reporter/selector gene cassette (2). Whenever such a vector integrates into an intron of an active gene, the cassette becomes posttranscriptionally fused to the 5' upstream exon. A fusion transcript and a fusion protein are obtained and neomycin (G418) resistant ES-cell clones can be selected (►intron trap vectors) (Fig. 1).

Gene trap vectors without an additional splice acceptor (SA) allow the insertion into exonic regions of transcriptionally active genes (►exon trap vectors). Again, a fusion protein allows the selection of neomycin-resistant ES-cell clones with the corresponding gene trap insertion and mutation. One limitation of these vectors is that only genes transcriptionally active in undifferentiated ES-cells can be trapped. In order to trap genes transcriptionally inactive in ES-cells, a new generation of vectors has recently been developed, the ►poly-A trap vectors, carrying a promoter-driven neomycin cassette but no poly-A signal. In this case, a functional neomycin-resistance gene cannot be encoded by the vector itself. Instead, a splice donor (SD) site was added 3' to the neomycin cassette. By these means an integration into an intron leads to splicing to the next 3' exon, leading to the polyadenylation and therefore to the translation of the neomycin cassette. Since expression of the neomycin cassette does not depend on the endogenous promoter,



Large-Scale Gene Trap Approaches in Mice. Figure 1 The insertion of the β geo vector generates a fusion transcript and a fusion protein from the β geo reporter gene and the upstream exon of gene X, providing that gene X is transcriptionally active in undifferentiated embryonic stem (ES) cells (adapted from Stanford et al. 2001).

genes transcriptionally inactive in ES-cells can be trapped, identified and eventually functionally characterized.

At the same time retroviral vectors were developed. These vectors carrying the SA recognition site of the adenovirus and a LacZ/neomycin fusion cassette are transferred into ES-cells *via* infection. Without regard to the vector type used, ES-cell clones can be further selected based on the expression of the reporter gene, e.g. ►LacZ. This feature will be essential later on when analyzing gene expression in the organism, e.g. at different stages of embryonic development or in different cell types. The crucial and common attributes of all the different vectors are (i) that the expression of the reporter/selector cassette is under the regulatory control of the trapped gene and (ii) that the trapped gene is mutated due to the insertion.

Many mouse mutant lines have been established meanwhile with all the different types of vectors. It has thereby been shown that the LacZ reporter gene does indeed reflect the endogenous gene expression of the trapped gene *in vitro* as well as *in vivo* (Fig. 2).

In addition, the analysis of gene trap mouse mutants has shown that the mutagenicity of the vectors seems comparable to that obtained using homologous recombination in ES-cells. The trapped genes in homozygous mutant animals were usually inactivated and in ~60% of the homozygous mutant mice an obvious ►phenotype could be detected.

The characterization of the trapped genes is indispensable for the functional analysis of the genes and for

the phenotype analysis (genotype-phenotype correlations) of mouse mutants. The trapped genes can be identified using PCR-based strategies (5' RACE, rapid amplification of cDNA ends) and in combination with ES-cell technology (e.g. injection of ES-cell clones into host blastocysts or aggregation chimeras) a mouse mutant line can be generated. In the first instance, RNA from neomycin-resistant ES-cell clones is isolated, cDNA synthesized using either random or poly-d(T) primers and primed with vector-specific as well as with synthetic primers in a nested-PCR approach. The obtained PCR products contain at their 5' ends sequences corresponding to the exons of the endogenous gene. By direct sequencing of these PCR products and sequence comparison analysis using publicly available databases the trapped genes are identified. Currently available databases (e.g. Ensembl, GeneCards, SwissProt, OMIM) enable the characterization of many of the trapped genes at genomic, transcript and protein levels, offering a wealth of information and a basis for the identification of genes relevant for e.g. human disease in general and developmental processes in particular.

Gene trap vectors have been developed and modified more or less extensively according to many prerequisites (4). For instance, an induction-trap vector was designed to trap, identify and mutate genes relevant to certain signal transduction pathways. Different soluble factors (e.g. retinoic acid, growth factors) were added to the ES-cell cultures and cells selected for the expression of a given reporter gene. By inducing or repressing the expression of such a reporter gene, genes can be trapped that are under the control of the given factor. Furthermore, many secreted and/or receptor molecules represent interesting targets for pharmacological approaches and putative pharmaceutical studies. Since such targets seem to be underrepresented in screens employing classical vectors, the design of alternative vectors has been enforced (5, 6). A prominent example is the U3Ceo vector, a retroviral gene trap vector containing a human CD2 cell surface antigen/neomycin-phosphotransferase fusion gene in the U3 region of its LTR. U3Ceo was used to screen the mammalian genome for genes encoding secreted and/or transmembrane proteins (6). Molecular analysis of a random sample of neomycin-resistant clones revealed that the U3Ceo gene trap vector indeed preferentially disrupted genes coding for secreted and transmembrane proteins.



Large-Scale Gene Trap Approaches in Mice.

Figure 2 Whole mount LacZ staining of a 14.5 dpc embryo with a gene trap insertion into the neurochondrin gene. Note that LacZ expression is detected by an enzymatic reaction, in blue and reflects neurochondrin gene expression (with courtesy of T. Floss, unpublished).

The German Gene Trap Consortium

In 1998 the German Gene Trap Consortium (GGTC) was established within the framework of the German Human Genome Project (DHGP) and carries out large-scale gene trap mutagenesis in ES cells. Its goal is to contribute to the saturation mutagenesis of the mouse



Large-Scale Gene Trap Approaches in Mice. Figure 3 The German Gene Trap Consortium's database (a) offers access to more than 18,000 mutated and characterized ES-cell clones. In addition, it enables searching of the data using keywords (b), DNA sequences and accession numbers.

genome and to generate a mouse model for each gene in cooperation with the International Mouse Mutant Consortium (IMMC). Towards this goal, the GGTC has generated 16,000 mutant ES cell lines and has identified the gene trap integration sites. Of the generated gene trap sequence tags (GTSTs), 8,689 informative sequences were obtained of which 6,077 corresponded to known mouse genes, 1,288 to known homologous genes, 483 to ESTs and 295 to putative novel genes (7). Data obtained from individual clones, such as GTSTs, expression patterns and phenotypes, are deposited in the GGTC's database, which is publicly accessible via <http://genetrap.de> (Fig. 3). For the first time, the insertion performance of four different gene trap vectors, pT1 β geo, pT1ATG β geo, U3 β geo and Rosa β geo was compared (7). The integration sites of all vectors were found randomly and evenly distributed over the entire mouse genome, although several hot spots were identified for the different vectors. These turned out to be common to all the different vectors as well as vector-specific, although gene size independent. In addition integrations into all the different classes of gene products were identified. Signal-sequence containing proteins were underrepresented.

Clinical Relevance

In attempting to engineer a mouse model for a human disease, it is important to know what kind of mutation causes the disease and if possible introduce the same kind of mutation into the mouse gene. Mutants corresponding to genes associated with human diseases are of particular value and interest. An ideal mutant collection would therefore contain mutants for every human gene

and for every phenotype or disease, to the limit of species differences between humans and the mouse.

According to the Online Mendelian Inheritance in Man database (OMIM), out of all the gene trap integrations into previously characterized genes 159 insertions characterized by the German Gene Trap Consortium occurred in genes involved in human disease (e.g. Emery-Dreyfuss muscular dystrophy, epidermolysis bullosa dystrophica, Huntington disease, xeroderma pigmentosum). Furthermore, over 100 germline chimeras were generated by this consortium and 60% of the resulting homozygous mutants exhibited an obvious phenotype either at embryonic stages or during adulthood. Indeed, the suitability of these mouse mutants as models for human disease was also shown; gene trap mouse mutants with an insertion into the *Nephrin* gene perfectly mimic the human congenital nephrotic syndrome CNF (8). However, to circumvent embryonic lethality and enable the analysis of gene function in the adult mouse, novel gene trap vectors that induce conditional mutations in ES cells are currently being developed and implemented (9).

One can conclude that gene trap mutagenesis is an efficient approach for annotating and dissecting the function of mammalian genes. Its large-scale implementation has already enabled the worldwide establishment of several databases containing GTSTs from thousands of mouse genes (e.g. <http://genetrap.de>, <http://baygenomics.ucsf.edu>). Collectively, these databases provide an unprecedented resource for the scientific community in the postgenomic era, as clones from the corresponding ES cell libraries can be used immediately to cost-effectively generate mouse models

of human disease. Clearly, the goal of understanding the function of every gene in the genome could be attained more quickly with the establishment of ES cell libraries with mutations in every single gene. Since each gene trap vector appears to have its own set of genomic integration targets, the most effective generation of an ES cell library saturated with mutations must involve a collection of different gene trap vectors. The ongoing collaboration within the International Mouse Mutagenesis Consortium can be expected to achieve the complete saturation of the mouse genome within the next few years.

References

1. Angelis MH de, Fuchs H, Rathkolb B et al (2000) Genome-wide, large-scale production of mutant mice by ENU mutagenesis. *Nat Genet* 25:444–447
2. Skarnes WC, Auerbach BA, Joyner AL (1992) A gene trap approach in mouse embryonic stem cells: the lacZ reporter is activated by splicing, reflects endogenous gene expression, and is mutagenic in mice. *Genes Dev* 6:903–918
3. Hicks GG, Shi EG, Li XM et al (1997) Functional genomics in mice by tagged sequence mutagenesis. *Nat Genet* 16:338–344
4. Stanford WL, Cohn JB, Cordes SP (2001) Gene-trap mutagenesis: past, present and beyond. *Nat Rev Genet* 2:756–768
5. Skarnes W, Moss J, Hurtley S et al (1995) Capturing genes encoding membrane and secreted proteins important for mouse development. *Proc Natl Acad Sci USA* 92:6592–6596
6. Gebauer M, Melchner H von, Beckers T (2001) Genomewide trapping of genes that encode secreted and transmembrane proteins repressed by oncogenic signaling. *Genome Res* 11:1871–1877
7. Hansen J, Floss T, Sloun P van et al (2003) A large scale, gene-driven mutagenesis approach for the functional analysis of the mouse genome. *Proc Natl Acad Sci USA* 100:9918–9922
8. Rantanen M, Palmen T, Patari A et al (2002) Nephron TRAP mice lack slit diaphragms and show fibrotic glomeruli and cystic tubular lesions. *J Am Soc Nephrol* 13:1586–1594
9. Schnutgen F, De-Zolt S, Van Sloun P et al (2005) Genomewide production of multipurpose alleles for the functional analysis of the mouse genome. *Proc Natl Acad Sci USA* 102: 7221–7226

Large-Scale Homologous Recombination Approaches in Mouse Embryonic Stem Cells

MICHAEL V. WILES¹, GABRIELE PROETZEL²

¹The Jackson Laboratory, Maine, USA

²Scil Proteins GmbH, Halle, Germany

mvwiles@jax.org

g.proetzel@scil.com

Definition

Homologous recombination (HR) is the precise strand exchange of a section of host DNA sequence for an externally derived sequence. Homologous recombination is dependent upon shared sequence homology between the host target DNA (e.g. the embryonic stem cell genome) and the externally derived sequence (i.e. a targeting construct or vector) and aims to genetically modify the host target region permanently (1, 2, 3).

Mouse embryonic stem (ES) cells are immortal, non-transformed cell lines derived from the inner cell mass of a ►blastocyst (~3.5 days post fertilization). Mouse ES cell lines have been derived from multiple inbred mouse strains, including 129/Sv, C57BL/6, BALB/c, etc and from F1 hybrids (4, 5, 6). When reintroduced into the embryo, ES cells can colonize the developing embryo, giving rise to chimeras containing ES cell contributions to all cell lineages including the germline. HR in mouse ES cells is the principal tool for creating specifically genetically modified mice. The process is known as a targeted mutation or gene ►knockout (KO) and produces “knockout mice” (generally designed to generate a ►null phenotype, although other modifications are possible), ►knock-in mice, where the modification is a functional replacement with an alternative sequence and knockdown mice, where the gene expression level is reduced.

Large scale HR in ES cells represents a shift from the research scale, where a few genes or host sequences are modified per year per lab, to a scale where many hundreds of genes per year are modified at a single facility. This shift involves a fundamental change in laboratory philosophy from a single gene being the principal driving parameter to the overall robustness and final efficiency of hundreds of gene modifications.

Characteristics

It is estimated using comparative data from the mouse and the human genome that there are ~35,000 genes in the mammalian genome. These genes operate interactively with the environment defining the organism, be it woman, man or mouse.

HR in ES cells with subsequent germ line transmission is a methodology to help us understand how a gene or other DNA sequence functions, by modifying it and observing the effect (the phenotype) in the resulting organism. In its crudest form, this involves destroying the gene under study by the precise removal or interruption of part or all of the genomic sequence under study (null mutation) with the resulting (homozygous) animals having a null phenotype, although varying degrees of partial loss or modified function (►hypomorph, ►hypermorph, etc) can occur at times serendipitously.

In more advanced incarnations, HR is used to create more subtle modifications of a gene or other region of

the genome. These modifications can range from a single base change mimicking, for example a sequence polymorphism, to major changes reconstructing a human disease ►[phenotype](#) in the mouse. Such modifications, called “humanization” are being used to assist pharmacogenetic research, including disease control and toxicology studies e.g., allowing the monitoring of a drug-substrate interaction with a human protein in a non-human animal test system (7). HR gene modification is also used to engineer changes that are under spatial and/or temporal control. Such genetically engineered mutations are referred to as conditional KO’s as the gene modification effect only occurs in cells upon the fulfilling of defined preset conditions. There are many rapidly evolving methods to achieve conditional gene modifications. One approach, for example, uses a vector containing ►[LoxP](#) sites flanking the region to be modified. This leaves the region unchanged until a recombination event is initiated using the LoxP sites. These sites act as a recombination substrate for ►[Cre recombinase](#). When the recombinase is present in the cell, sequences between the two LoxP sites are deleted. This genetically modifies the targeted locus irreversibly, but only in cells where Cre recombinase is expressed (8, 9). Using animals that contain the Cre recombinase under specific (i.e. temporal and/or spatial) control allows target loci to be genetically modified upon command, giving exquisite control of when and where the targeted gene/region is modified. Cre recombinase expression can be made inducible (e.g. by tetracycline), again allowing exact control of a genetic modification event.

Large-Scale Processes – Outline of Operating Concepts

When planning a production line it is essential to view the process as a whole. This global view of a process is one of the principal philosophy shifts when a research tool or process is industrialized. This viewpoint allows the definition of operational principals and current working methodologies and begins to help develop policies. Below is a non-exhaustive listing of considerations to help to visualize such a pipeline.

Detailed laboratory protocols for various steps of a HR pipeline are available in publications such as (10, 11, 12, 13).

Operating considerations include the following:

- a) Definition of the desired outcome, including expected output animals, goals, including costs and timelines, i.e. what is the goal of this pipeline.
- b) Which genes – i.e. selection criteria for targeted gene families and/or gene types.
- c) Desired gene modifications, e.g. KO, knock-in, conditional KO, ►[knockdown](#), etc.
- d) Derivation, number and method of construction of DNA vectors, e.g. isogenic DNA, use of

- [bacterial artificial chromosome](#) (BAC) with recombinogenic engineering (14), use of lambda phage DNA or DNA generated by ►[long range PCR](#) (LR-PCR) using available sequence information followed by ligation construction (15), etc.
- e) Vector verification ►[quality control](#) (QC) methods and depth, e.g. ►[PCR](#), restriction digests, sequencing etc.
- f) Use of cell expression lineage markers incorporated into the vector design, e.g. ►[LacZ](#), green fluorescent protein (GFP), ►[alkaline phosphatase](#) reporter (16) etc and reporter localization signals, e.g. nuclear, membrane etc.
- g) System/s for the selection for HR events in ES cells during culture, e.g. G418, positive-negative selection etc.
- h) ES cell line germline competence and ►[genetic background](#), e.g. inbred strains such as 129S1/SvImJ, C57BL/6J or hybrid F1 strains etc. This also influences the need for feeders and the overall robustness of the cell culture stage.
- i) Selection and isolation methods after ES cell ►[electroporation](#), as well as the number of ES cell clones to be screened for HR events and the screening method used, e.g. ►[Southern blot](#), LR-PCR, ►[FISH](#), ►[quantitative PCR](#) etc, including independent verification of HR events.
- j) Number of independent ES cell clones required for each construct and each gene.
- k) Chimera production, e.g. *via* ►[aggregation](#) or ►[blastocyst injection](#), definition of host embryo strain, details to include the acceptable chimera level and subsequent breeding strategies.
- l) Number and quality of chimeras produced per ES cell clone.
- m) Methods for mouse genotyping and verification of HR transmission of chimera offspring, e.g. coat color followed by PCR analysis or Southern blot.
- n) Establishing goals for each stage of the output as well as reiteration limits, i.e. when to give up on a particular gene, construct, electroporation, ES cell screening, number of chimeras, number of breeding rounds for ►[germline transmission](#) etc (i.e. define step reiterations for each stage). This should also include what actions needs to be taken to deal with recalcitrant genes, e.g. repeat, change construct, drop etc.
- o) An effective pipeline communication and information storage/control system is an absolute requirement, i.e. a controlling database allowing efficient pipeline process control and interrogation.

The first step in designing a large-scale process is to obtain a functional pilot pipeline. This is used to define feasibility and limits for each part of the process, allowing adaptation etc where needed. It is crucial to

Large-Scale Homologous Recombination Approaches in Mouse Embryonic Stem Cells. Table 1 For the process of Large Scale HR in ES cells suggested operational units could be as follows

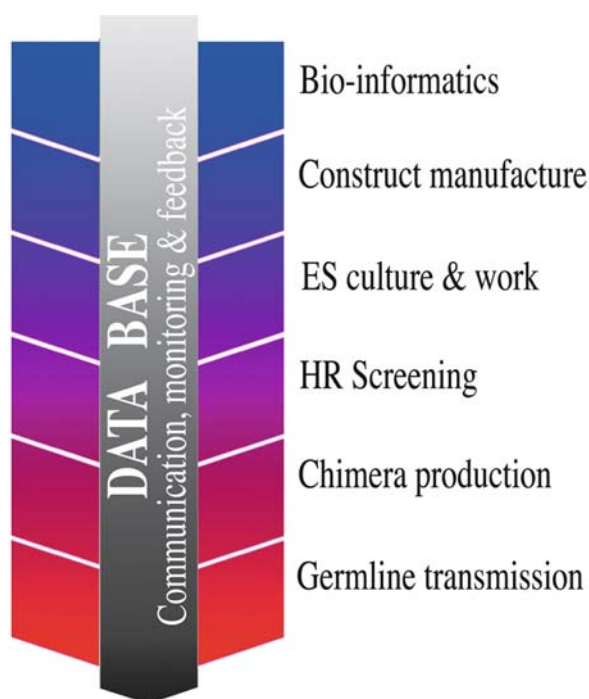
Units		Responsibilities
1	Bio-informatics	Target identification, theoretical construct design, ►QC for pseudogenes and published data
2	Construct manufacture	Vector building and QC, plus designing analytical probes for screening HR events and subsequent genotyping of animals
3	ES culture	General ES cell maintenance and QC, vector introduction, clone isolation, cryopreservation and clone DNA for screening
4	HR Screening	Screening for HR events and QC of HR events
5	►Chimera production	Supply of host embryos, ►pseudopregnant females and production of chimeras
6	►Germline transmission	Initial screening for germline transmission, breeding/maintenance of genetically modified animals

define each operational methodology before charging ahead. Essentially, if a process does not work (well) on a small scale, it won't work when scaled up. On the other hand, no matter how exhaustive the pilot studies are, they can never be perfect in predicting how the scale up will perform. It is a balance between seeking perfection in the pilot study and the desire to move on and scale up rapidly. As each of the processes is defined, scale up can begin – carefully. These ideas are obvious but they are too often ignored, possibly leading to the scale up becoming an expensive quagmire of technical problems and delays.

As the pipeline scales up, communication between the various units becomes the essential tool in maintaining process control. The best communication tool is a user-friendly, flexible database allowing complete access and interchange of information between the various units over the whole process. The control database should be designed to provide complete process tracking information and, importantly, serve as an early warning system. If designed well, it will assist in spotting problems *before* they impact the entire process.

Outline Units

It is readily apparent that a process pipeline for homologous recombination in ES cells is a complex operation involving many quite different technologies. It is therefore convenient that the process be broken down into operational units or teams (Table 1), each unit having defined goals, process QC methods and trouble shooting routines. This facilitates control and continual optimization as each unit's input and output can be isolated and monitored. Further, by placing people in teams responsible for a unit, the management of the whole process is simplified (Fig. 1).



Large-Scale Homologous Recombination Approaches in Mouse Embryonic Stem Cells.

Figure 1 Process flow pilot studies will help define a series of units. Although the subdivisions may appear artificial in some cases, the idea is that they be viewed as "gates" allowing the isolation of stages, each with defined limits, inputs, outputs, QC and specific trouble shooting routines. If a unit's goals are not reached, this can be easily seen and the situation analyzed in isolation.

It is essential for any industrial process that ►**standardized operating procedures** (SOPs) and relevant QC protocols be established for each procedure and a unit's output. Further, it must be stressed that these SOPs and QC protocols be followed, updated when necessary and not allowed to fall into disuse. Abandonment of SOPs is a recurring problem giving rise to "►**protocol mutagenesis**", i.e. loss of control of the methodologies used. If unchecked this will lead to degradation of the pipeline and a loss of the ability to trouble shoot individual steps efficiently. At the same time, innovation must be allowed to enter the process. New ideas, methods or technologies are continually arising; these should be looked at and if worthwhile implemented with care, together with the updating of all SOPs and QC protocols.

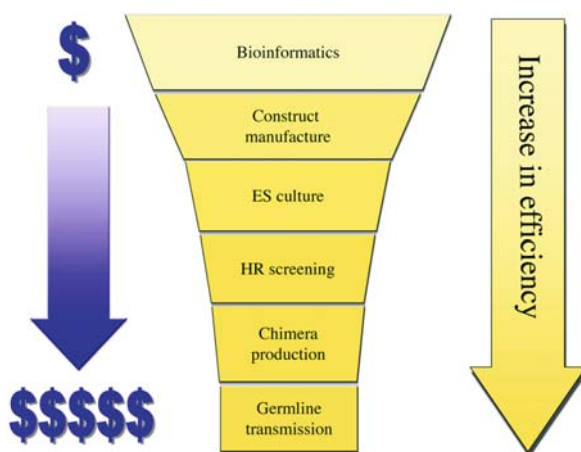
In establishing large scale HR in ES cells, it is the final speed and efficiency that are the key measures of success. For any multiple stage process the final efficiency is 100% minus the cumulative failure of each of the process units (Fig. 2). For example, in a pipeline consisting of ten units, if each unit has an internal failure rate of 5% the pipeline will produce ~60% of its original input (e.g. if 1000 genes are selected as genetic targets, ~600 will make it though the complete process). Accordingly input has to be adjusted. The use of units allows the allocation of resources in the most efficient manner. However, when dealing with limited resources, not all units are equal. In general it is acceptable that units at the initial stages

can have a higher failure rate than later ones. This is due to cumulative costs – i.e. as an individual project travels through the process pipeline its actual value, as calculated by the resources spent on it, increases. Hence errors need to be removed before they have traveled far in the pipeline, otherwise they will have a higher negative financial impact. A QC process should be established which reflect this, for example, it is better to stop a 'dubious' DNA vector construct when the investment is still low, before the whole process of making chimeras is reached.

General Discussion

The process described, especially if it is to be developed on a truly large scale, is complex and expensive. It is therefore worth investing effort in the forward planning, as once errors are built into a process they are costly financially and in time lost. Pilot studies will help in building the system and will assist in developing a whole series of QC measures to prevent expensive errors. However, the main elements in such a labor-intensive endeavor are the people involved. It is vital that the teams and especially the team leaders fully understand not only the processes involved in their section but also how their work is an integral part of the whole process. Personnel are also one of the most expensive elements. The use of robotic systems can increase efficiency and reduce human errors in what can be boring, but high precision, work. As a pipeline is piloted, some steps will obviously be open to using robotic approaches. Robots can also reduce repetitive strain injuries, an intrinsic problem with this type of work. If used imaginatively, robots can bring in completely novel methods that, for example, would not be practical for people. However, such machines must be viewed with caution. For example, will a robot actually do the work more efficiently than a person equipped with electro-mechanical pipette aids? Further, robotic systems generally involve a large capital investment. Once installed, they can inhibit subsequent innovative process enhancements, i.e. they lack adaptability. However, in general, once a process has matured and the methodology is stable, robotics brings considerable increases in efficiency and accuracy in processing, when applied to labor-intensive sections and process bottlenecks.

In any industrial process and especially in the rapidly developing field of molecular biology, it is important to avoid ossification. Whatever pipeline is constructed, it will be a combination of innovative ideas and expedient compromises. As such, the process pipeline should be under constant review for efficiency improvement, ranging from the individual units to reassessing the reasoning and method of the entire process, i.e. keep asking, "Why are we doing this" and "Is this the best way to do it?" In the case of HR in ES cells, the general need is to produce animals that have specific permanent



Large-Scale Homologous Recombination Approaches in Mouse Embryonic Stem Cells.

Figure 2 For a multistage process the final efficiency is 100% minus the cumulative failure of each process step. When setting up large-scale HR screens with the aim of generating genetically modified mice, one has to keep in mind that there is an increase in the money invested in the later stages of the process. The QC processes should reflect this.

genetic modifications. In this field new technologies are continually arising, eventually they will have an impact on any current approach used. For example, possible new methods that may give us similar results include reverse genetic methods, transgenics expressing RNAi or perhaps even HR directly in oocytes (17). Such novel methodologies are the substance of opportunity, not threats. The final goal must be kept in mind, not the protection of antiquated processes.

Clinical Relevance

The use of homologous recombination in ES cells allows the development of genetically modified animals that offer almost unlimited possibilities for developing models that advance the understanding of gene function. Gene targeting allows the design of tractable model systems, assisting target discovery and validation. The data obtained allow the dissection of disease pathways and establish new disease models for drug development including safety assessment (18, 19, 20). Additionally, knock-ins using human genes have the potential to revolutionize drug development.

Gene targeting technology is now at the level where it can be performed on a scale similar to the human genome-sequencing project (21). The data obtained are already building a bridge between ►genomics and ►proteomics – from sequence to function.

References

1. Thomas KR, Folger KR, Capecchi MR (1986) High frequency targeting of genes to specific sites in the mammalian genome. *Cell* 44:419–28
2. Doetschman T, Gregg RG, Maeda N et al (1987) Targeted correction of a mutant HPRT gene in mouse embryonic stem cells. *Nature* 330:576–8
3. Smithies O (2001) Forty years with homologous recombination. *Nat Med* 7:1083–6
4. Kaufman MH, Robertson EJ, Handyside AH et al (1983) Establishment of pluripotential cell lines from haploid mouse embryos. *J Embryol Exp Morphol* 73:249–261
5. Ledermann B, Burki K (1991) Establishment of a germline competent C57BL/6 embryonic stem cell line. *Exp Cell Res* 197:254–8
6. Schoonjans L, Kreemers V, Danloy S et al (2003) Improved generation of germline-competent embryonic stem cell lines from inbred mouse strains. *Stem Cells* 21: 90–7
7. Nebert D, Dalton T, Stuart G (2000) “Gene-swap knock-in” cassette in mice to study allelic differences in human genes. *Ann NY Acad Sci* 919:148–70
8. Fukushige S, Sauer B (1992) Genomic targeting with a positive-selection lox integration vector allows highly reproducible gene expression in mammalian cells. *Proc Natl Acad Sci USA* 89:7905–9
9. Smith AJH, Desousa MA, Kwabiaddo B et al (1995) A site-directed chromosomal translocation induced in embryonic stem cells by cre-loxp recombination. *Nat Genet* 9:376–385
10. Wassarman PM, DePamphilis ML (eds) (1993) *Guide to Techniques in Mouse Development*. 3rd edn *Methods in Enzymology* Vol 225 Academic Press, San Diego 1021
11. Turksen K (ed) (2001) *Embryonic Stem Cells: Methods and Protocols*. 1st edn *Methods Mol Biol* Vol 185 Humana Press
12. Nagy A, Gersenstein M, Vintersten K (eds) (2002) *Manipulating the mouse embryo: a laboratory manual*. Cold Spring Harbor, Cold Spring Harbor Laboratory 800
13. Joyner AL (2000) Gene Targeting: a practical approach. In: Hames BD (ed). *The practical approach series*. 2nd edn Oxford University Press, Oxford, New York, London
14. Valenzuela D, Murphy A, Frendewey D et al (2003) High-throughput engineering of the mouse genome coupled with high-resolution expression analysis. *Nat Biotechnol* 21:652–9
15. Kuwayama H, Obara S, Morio T et al (2002) PCR-mediated generation of a gene disruption construct without the use of DNA ligase and plasmid vectors. *Nucleic Acids Res* 30:E2
16. Le Mouellic H, Lallemand Y, Brûlet P (1990) Targeted replacement of the homeobox gene Hox-3.1 by the Escherichia coli lacZ in mouse chimeric embryos. *Proc Natl Acad Sci USA* 87:4712–4716
17. Bibikova M, Beumer K, Trautman J et al (2003) Enhancing gene targeting with designed zinc finger nucleases. *Science* 300:764
18. Kirkland DJ (1998) Use of transgenic animals in toxicology. *Pharmaceutical Science & Technology Today* 1:62–68
19. West DB, Iakougova O, Olsson C et al (2000) Mouse genetics/genomics: an effective approach for drug target discovery and validation. *Med Res Rev* 20:216–30
20. Zambrowicz BP, Sands AT (2003) Knockouts model the 100 best-selling drugs – will they model the next 100? *Nat Rev Drug Discov* 2:38–51
21. Austin CP, Battey JF, Bradley A et al (2004) The knockout mouse project. *Nature Genetics* 36:921–924

Laryngospasm

Definition

Laryngospasm describes a spasmodic closure of the glottic aperture.

►Hyper-and Hypoparathyroidism

Laser Capture Microdissection

Definition

Laser capture microdissection (LCM) refers to a method for isolating a homogeneous population of a

single cell type from any tissue section. LCM uses infrared laser to adhere a single cell or region of cells to a specially coated cap, which can then be lifted from the tissue.

► [Genomic Information and Cancer](#)

LCA

► [Leber's Congenital Amaurosis](#)

Laser Interferometer

Definition

Laser interferometer is an optical system that uses the interference from two coherent beams to assess the optical path length differences, resulting from differential refractive index or distance changes along the light paths.

► [Analytical Ultracentrifugation](#)

LCM

► [Laser Capture Microdissection](#)

LCR

► [Locus Control Region](#)

Lateral Element

Definition

When homologous chromosomes synapse, the axial elements form the lateral elements of the tripartite synaptonemal complex.

► [Meiosis and Meiotic Recombination](#)

LCRs

► [Low Copy Repeats](#)

Lateral Inhibition

Definition

Lateral inhibition is a well studied mechanism in *Drosophila*, and describes how Notch signalling leads to the selection of a single neural precursor cell from a cluster of equivalent cells that are marked by the expression of proneural genes.

The term is also used to describe how activity in one area of the visual processing system will tend to produce a dampening effect on other areas of the same system. For example, activity in a given photoreceptor of the retina will inhibit activation in nearby photoreceptors via inhibitory synapses.

► [Neural Development](#)

LD

► [Linkage Disequilibrium](#)

LDL

► [Low Density Lipoprotein](#)

LDL Receptor

Definition

LDL (low density lipoprotein) receptor is a membrane protein of 839 amino acids, which is responsible for cholesterol uptake into cells via receptor-mediated endocytosis of cholesterol-rich particles secreted by the liver. LDL receptors are receptors on the cell surface that remove LDL and some other forms of lipoproteins from the plasma into the cell.

The LDL receptor gene is located on chromosome 19 (p13.1-p13.3). It spans 45 kb and comprises of 18 exons and 17 introns.

► [Genetic Screening in Populations](#)

compartments or for secretion. Often the leader signal is cleaved off during the export of the peptide.

► [Nuclear Import and Export](#)

► [Recombinant Protein Expression in Yeast](#)

LDL-A Domain

Definition

LDL-A domain resembles the class A domain of the ► [LDL receptor](#), and probably mediates protein-protein interactions.

► [Autosomal Dominant \(Inherited Disorder\)](#)

► [Polycystic Kidney Disease, Autosomal Dominant](#)

Leading Strand

Definition

Leading strand designates a DNA strand that is replicated continuously in the same direction of the unwinding duplex DNA (direction of synthesis and replicative fork moving are the same) lagging strand.

► [Replication Fork](#)

LDLR-/-Mice

Definition

LDLR-/-mice are genetically modified mice that are deficient in the low-density lipoprotein ► [\(LDL\) receptor](#) gene.

► [Tangier Disease](#)

Learning Disabilities

Definition

This term refers to a variety of disorders that affect the acquisition, retention, understanding, organisation or use of verbal and/or non-verbal information.

► [Fragile X Syndrome](#)

Leader Signal

Definition

Leader signal denotes a specific amino acid sequence at the N-terminus of a protein targeting it to specific cell

Leber's Congenital Amaurosis

Definition

Leber's congenital amaurosis (LCA) denotes a heterogeneous group of inherited retinal dystrophies, causing severe visual loss with nonrecordable or a largely reduced ERG response, either congenitally or in very early childhood. The fundus of the eye is marked by an usually flecked appearance with loss of retinal pigment epithelium and/or pigment deposit. Many cases of Leber's congenital amaurosis are caused by dysfunction of an isomerase important for the retinoid cycle induced by a mutation in the ► [RPE 65 gene](#).

► [Retinitis Pigmentosa](#)

Leber's Hereditary Optic Neuropathy

Definition

Leber's hereditary optic neuropathy comprises of a group of disorders most commonly affecting males, and is inherited through the maternal line affecting vision and resulting from pathological mutations of mitochondrial DNA.

► [Genetic Predisposition to Multiple Sclerosis](#)

LEF/TCF Family

Definition

LEF-1 (lymphoid enhancer factor-1) and TCF (T cell factor) are members of a family of DNA binding proteins (HMG box transcription factors) that affect the expression of specific genes, and are downstream components of the Wnt signaling pathway. All proteins interact with β -catenin at an N-terminal domain, and with DNA via the HMG (high mobility group) box.

► [Colorectal Cancer](#)

► [Desmosomes](#)

Legless(Lgs)/BCL19

Definition

Legless (Lgs)/BCL19 is a conserved nuclear protein recently discovered in *Drosophila*. It functions as an adaptor between Pygopus and Armadillo/ β -catenin and permits transcriptional activation.

► [Wnt/Beta-Catenin Signaling Pathway](#)

Lens Connexins

Definition

Lens connexins are protein components, mainly connexins 46 and 50, which form with other channels, e.g. aquaporins, the microcirculation system aimed at supplying deeper fiber cells of the lens with nutrients and clearing away waste products.

► [Intermediate Filaments](#)

Lentiviruses

Definition

Lentiviruses describes the genus of complex retroviruses that include HIV.

► [Retroviruses](#)

Leptin

Definition

Leptin is a hormone of ~16 kDa which is encoded by the obese (ob) gene. Leptin is an important regulator of food intake, body weight and energy metabolism. Leptin is mainly produced and secreted from adipocytes (fat cells). It appears that as adipocytes increase in size due to accumulation of triglyceride, they increase their rate of leptin secretion. Thus, there is a strong positive correlation of serum leptin concentrations with percentage of body fat. Via leptin receptors in areas of the hypothalamus the hormone signals to the brain the size of the peripheral adipose tissue deposits. Leptin levels are considerably decreased in narcolepsy.

► [Narcolepsy](#)

► [Obesity](#)

Leptomycin B

Definition

Leptomycin B is a secondary metabolite (an unsaturated branched fatty acid chain) produced by *Streptomyces sp.* with known antifungal and antitumor effects. It is a potent inhibitor of nucleocytoplasmic export and exhibits significant antiproliferative activity.

► [Actin Cytoskeleton](#)

Leptotene

Definition

Leptotene describes the earliest stage of prophase 1 of the meiosis, when the despiralized chromosomes first

become visible as fine threads and have not yet associated in pairs. (Gk. leptos, thin; taenia, ribbon.)

► [Meiosis and Mitotic Recombination](#)

Lethal Mutations

Definition

Lethal mutations are mutations in genes that result in the death of the organism. Such mutations can be genetically visualised using genetic balancers covering the region of the mutation.

► [C. elegans Genome, Comparative Sequencing](#)

Leucine Rich Repeat

Definition

Leucine rich repeat describes a protein domain the function of which is unknown. It is likely to be involved in protein-protein interactions.

► [Autosomal Dominant \(Inherited Disorder\)](#)

► [Polycystic Kidney Disease, Autosomal Dominant](#)

Leucine Zipper

Definition

The leucine zip or zipper is a motif that mediates protein-protein interaction. It describes the coiled-coil dimerization of two α -helical-like polypeptides, by interdigitation of periodic leucines or other hydrophobic or bulky amino acids, which lie along the cylinder of the monomeric α -helix. This has the effect of sequestering the hydrophobic residues away from the aqueous environment of the cytoplasm and so forms a very stable dimer. It is an association seen in a large number of protein dimers from transcription factors (usually short stretches) to intermediate filaments (where they can be very long).

► [Intermediate Filaments](#)

► [NF \$\kappa\$ B Pathway](#)

► [Transcription Factors and Regulation of Gene Expression](#)

Leucine Zipper Transcription Factors: bZIP Proteins

ANDREAS G. BADER, PETER K. VOGT

The Scripps Research Institute, La Jolla, CA, USA
pkvogt@scripps.edu

Definition

bZIP proteins are ► [transcription factors](#) that consist of three modular functional regions mediating dimerization, DNA binding and transcriptional regulation. The hallmark of these proteins is the bZIP (basic region, leucine zipper) domain, a well-defined motif in eukaryotic proteins (1, 2). The basic region of the bZIP domain, rich in lysines and arginines, represents the DNA contact surface, interacting with the major groove of the DNA double helix. The zipper region is an ► [amphipathic](#) helix of 30–40 residues with every seventh residue a leucine. These leucines, together with the intervening residues, form a zipper-like structure that dimerizes with the corresponding region of a partner bZIP protein.

Characteristics

Proteins with a bZIP domain are grouped into one super-family. They are distinct from other proteins that contain similar structural elements like the numerous proteins with only leucine zippers or the basic helix-loop-helix leucine zipper proteins in which basic region and leucine zipper are functionally equivalent to those of bZIP proteins but are separated by an intervening helix-loop-helix region (2, 3). There are seven bZIP protein families, Jun, Fos, ATF, CREB, C/EBP, Maf and PAR (Table 1). This classification is based on ► [protein homology](#) and DNA-binding specificity (4, 5). To become biologically functional, a bZIP protein must dimerize with another bZIP protein. Dimers can form between members of the same bZIP family or between members of different families. Some bZIP proteins form exclusively homodimers; others are only capable of heterodimerization. A third category can homo- as well as hetero-dimerize, with one combination energetically favored over all other possible unions. Dimerization specificity is determined by the non-leucine residues in the zipper region (5, 6).

Cellular and Molecular Regulation

One of the best-characterized bZIP proteins is Jun. Jun was originally recognized as the ► [oncoprotein](#) of the avian ► [retrovirus](#) ASV17. Sequence homology and DNA binding properties led to the identification of the cellular Jun as a component of the ► [AP-1](#) transcription

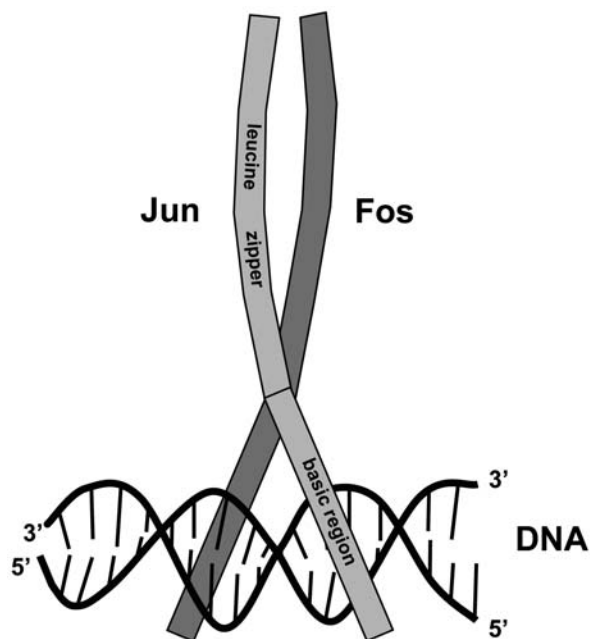
Leucine Zipper Transcription Factors: bZIP Proteins. Table 1 Human bZip protein families and their members

Protein family	Name	Protein members
Jun	v-jun avian sarcoma virus 17 oncogene homolog	c-Jun, JunD, JunB,
Fos	FBJ murine osteosarcoma viral oncogene homolog	c-Fos, FosB, Fra-1, Fra-2, JDP1, JDP2
ATF	activating transcription factor	ATF 1-7
CREB	cAMP responsive element binding protein	CREB1, CREB3, CREM, CREB-H, CREB-L1
C/EBP	CCAAT/enhancer binding protein	CEBPa, CEBPb, CEBPd, CEBPe, CEPBg
Maf	avian musculoaponeurotic fibrosarcoma virus homolog	c-Maf, MafB, MafG, MafF, MafK, NRL
PAR	proline and acidic rich	DBP, TEF, HLF

factor complex (7, 8). The classic AP-1 dimer consists of Jun and Fos. Other combinations include the Jun relatives JunB and JunD; Fos-related proteins or ATFs are also found as components of AP-1. Jun can form weak homodimers; heterodimers with Fos are significantly more stable. Fos family proteins cannot homodimerize. *In vivo*, the composition of bZIP dimers is determined by partner concentrations and subcellular distribution (9).

The three-dimensional structure of the bZIP domain has been determined by X-ray crystallography of the Jun homolog GCN4 from yeast and of the Jun-Fos heterodimer (10, 11, 12). The bZIP dimer consists of a pair of continuous α helices with the zipper region forming a parallel **coiled coil** that diverges in the amino terminally located basic region. This creates a Y-shaped structure in which the α helices of the basic region reach into the major groove of the DNA double helix. The coiled coil of the leucine zipper extends away from the longitudinal axis of the DNA at almost 90° (Fig. 1).

In the absence of DNA binding, the basic region is disordered; it assumes helical structure only upon contact with DNA. There is a limited degree of flexibility in the joint region that connects the zipper and basic domains. Basic residues close to the zipper domain interact with the DNA backbone; more amino terminally positioned alanines and one asparagine as well as other residues make specific contacts with DNA base pairs (Fig. 2). The sequence of the basic region and the composition of the dimer specify the recognition site on the double helix. Because of the dimer structure of bZIP proteins the corresponding DNA target sequences usually have dyad symmetry. Upon binding to DNA, bZIP proteins may introduce a bend in the target molecule. The degree of bending is specified by dimer composition and DNA sequence. The coiled coil of leucine zippers determines the stability of the bZIP dimer. The α helices in these coiled



Leucine Zipper Transcription Factors: bZIP Proteins. Figure 1 Illustration of the Jun-Fos heterodimer bound to DNA (redrawn from Ref. 11). The leucine zippers of Jun and Fos facilitate dimerization; the basic regions of the protein dimer reach into the major groove of the DNA double helix.

coils have a periodicity of 3.5 residues per helical turn and hence every seventh amino acid is in the same structural environment. The hallmark of the zipper structure is the seven-residue periodicity of leucines or alternative hydrophobic residues. The leucine zipper consists of a series of heptads where the position of each residue is designated by a lower case letter *gabcdef* (helix 1) and *g'a'b'c'd'e'f'* (helix 2). Leucines occur in the *d* and *d'* positions; additional hydrophobic residues are found in the *a* and *a'* positions. Residues in

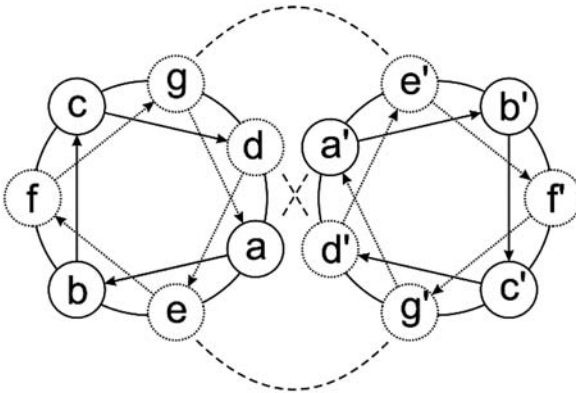
protein	bZIP domain												aa of bZIP domain	total aa of protein
	basic region						leucine zipper							
GCN4	SSDPAALKRARNT	EAARRSR	AKLQRMKQ	LED	KVEE	LS	KNYH	EN	EVAR	LKK	LVGER		224-281	281
c-Jun	ERIKAEKRKMRN	RIAASKR	KRKKLERIAR	LEE	KVKTKA		QNSELAS		TANMLRE	QVAQLKQ	KVMNHVN		251-317	331
CREB1	AARKREVRLMKN	REAAAREC	RRKKKEYVKC	LEN	RVAVLEN		CNKTLIE	ELKAKD	LYCHKSD				268-327	327
CEBPa	NSNEYRVRR	RNNIAVR	KSRDKAKQRNVE	TOQ	KVLELTS		DNDRLEK	RVEQLSR	ELDTURG	IFRQ	LPE		281-347	358
ATF2	PDEKRRKFLERN	RAAASRCR	QKRKRVWVQS	LEK	KAEDLSS		LNGQLQS	EVTLLRN	EVAQLRQ	LLLAKHD			333-399	487
c-Fos	EEEKRRIRRE	RNKMAAAK	CRNRRRLTDT	LQA	ETDQLED		EKSALQT	EIANLLK	EKEKDF	ILAAHRP			136-202	380
c-Maf	IRLKQKRRTL	LKNRGYAQ	SCRFRKVQQRHV	LES	EKNQLQ		QVDHLKQ	EISRLVR	ERDAYKE	KYEK	VS		287-353	403
heptad														
heptad pos.														
consensus	-----BB-BN--AA-B-R-BB-----	L--	----L--	-N--L--	----L--	----L--	----L--	----L--	----L--	----L--	----L--			

Leucine Zipper Transcription Factors: bZIP Proteins. Figure 2 Alignment of the basic regions and the leucine zipper domains of various human bZip-proteins and yeast GCN4. Conserved leucines within the zipper region are shadowed black, conserved charged residues are shadowed gray. The consensus sequence is shown at the bottom of the figure. The sequences were taken from Pubmed according to their accession numbers as follows: GCN4, NP_0109107; c-Jun, NP_002219; CREB1, NP_004370; CEBPa, NP_004355; ATF2, P15336; c-Fos, NP_005243; c-Maf, AAC27038. aa, amino acid; B, basic residue arginine or lysine.

the *a* and *d* positions form the hydrophobic core of the coiled coil (Figs. 2, 3). The side chains of the leucines and of hydrophobic residues in the *a* positions form “knobs” that fit into “holes” generated by the spaces between the side chains of the partner helix. The hydrophobic residues of the dimerizing helices are not in exactly opposite positions but interact from side to side as in a handshake (Fig. 3). In homodimerizing bZIP proteins, the *a* position of the second heptad is occupied by an asparagine which favors interaction with another asparagine in the same position of the partner helix and appears to be critical for homodimerization. The *g* and *e* positions typically contain charged amino acids that affect partner selection and dimer stability. They undergo electrostatic *g-e'* and *g'-e* interactions where the *e* position belongs to the adjacent heptad, five residues C-terminal. The electrostatic interactions can be attractive or repulsive and regulate homo- and hetero-dimerization. A leucine zipper can have up to five heptad repeats, allowing for a great diversity of electrostatic *g-e'* and *g'-e* interactions. The distinct dimerization patterns determined by the leucine zippers of bZIP proteins can be used to define protein families (5).

Clinical Relevance

Several bZIP proteins have been identified as oncogenic components of retroviruses establishing a fundamental role in transcription in cancer. Transcriptional regulation has an inherent oncogenic potential, and deregulated transcription is a principal mechanism of oncogenic transformation (7). *maf* and *mafB* gene rearrangements are found in about 5–10% of multiple myelomas presumably resulting in a gain of function; c-Maf is over-expressed in 50% of multiple myelomas



Leucine Zipper Transcription Factors: bZIP Proteins. Figure 3 Helical wheel diagram of the leucine zipper. The view is parallel to the α helical axis from the amino terminus starting with position *a*. Heptad positions are labeled *a-g*. Residues in the first helical turn are in solid circles, residues in the second turn are in dotted circles. Hydrophobic *a-a'* and *d-d'* connections, as well as electrostatic interactions between *g-e* pairs are indicated by dashed lines. Note that the residues at position *g* interact with the residues at position *e* from the following heptad.

contributing to enhanced growth of these tumors (13). A dominant negative Jun protein interferes with the growth of human tumor cells (for review see 8).

Acknowledgements

This is TSRI manuscript number 16587-MEM. This work was supported by grants from the National Cancer Institute, National Institutes of Health.

References

1. Landschulz WH, Johnson PF, McKnight SL (1988) The leucine zipper: a hypothetical structure common to a new class of DNA binding proteins. *Science* 240:1759–1764
2. Pabo CO, Sauer RT (1992) Transcription factors: structural families and principles of DNA recognition. *Annu Rev Biochem* 61:1053–1095
3. MacFarlane SN, Levitan IB (2001) Unzipping ion channels. *Sci STKE* 98:PE1
4. Hurst HC (1995) Transcription factors 1: bZIP proteins. *Protein Profile* 2:101–168
5. Vinson C, Myakishev M, Acharya A et al (2002) Classification of human B-ZIP proteins based on dimerization properties. *Mol Cell Biol* 22:6321–6335
6. Vinson CR, Hai T, Boyd SM (1993) Dimerization specificity of the leucine zipper-containing bZIP motif on DNA binding: prediction and rational design. *Genes Dev* 7:1047–1058
7. Vogt PK (2002) Fortuitous convergences: the beginnings of JUN. *Nat Rev Cancer* 2:465–469
8. Vogt PK (2001) Jun, the oncoprotein. *Oncogene* 20:2365–2377
9. Hu CD, Kerppola TK (2003) Simultaneous visualization of multiple protein interactions in living cells using multicolor fluorescence complementation analysis. *Nat Biotechnol* 21:539–545
10. Ellenberger TE, Brandl CJ, Struhl K et al (1992) The GCN4 basic region leucine zipper binds DNA as a dimer of uninterrupted alpha helices: crystal structure of the protein-DNA complex. *Cell* 71:1223–1237
11. O'Shea EK, Klemm JD, Kim PS et al (1991) X-ray structure of the GCN4 leucine zipper, a two-stranded, parallel coiled coil. *Science* 254:539–544
12. Glover JN, Harrison SC (1995) Crystal structure of the heterodimeric bZIP transcription factor c-Fos-c-Jun bound to DNA. *Nature* 373:257–261
13. Hurt EM, Wiestner A, Rosenwald A et al (2004) Overexpression of c-maf is a frequent oncogenic event in multiple myeloma that promotes proliferation and pathological interactions with bone marrow stroma. *Cancer Cell* 5:191–199

Leukemia

RALF BARGOU, WOLF-DIETER LUDWIG
Department of Hematology, Oncology and Tumor Immunology, Robert Rössle Clinic at the HELIOS Clinical Center Berlin-Buch, Charité, Berlin, Germany
ludwig@rrk.charite-buch.de

Definition

Acute Leukemias

Acute leukemias represent highly heterogeneous groups of malignancies that, nowadays, require an extensive routine diagnostic workup of their phenotypic

and genetic characteristics for the most appropriate classification and tailoring of therapy. The recently introduced World Health Organization (WHO) classification takes into account morphologic, immunophenotypic, genetic and clinical features of acute leukemias to define entities that are biologically homogeneous and that have clinical relevance (1).

Through the use of conventional and molecular cytogenetic analyses, acute leukemias have been recognized as genetic diseases reflecting deranged and disordered haematopoiesis, that results from a series of acquired or, rarely, inherited mutations in the structure of certain genes. These mutations are passed from the original transformed stem cell or progenitor cell to its clonal descendants. The diversity of molecular abnormalities in acute leukemias is, to a large extent, explained by the complexity of the genetic circuits that regulate the cell cycle, proliferation, differentiation, survival and cell death (►apoptosis) of the different haematopoietic progenitor cells that provide the biological targets for leukemogenesis. Although general mechanisms underlying the induction of acute myeloid (►AML) and acute lymphoblastic leukemias (►ALL) are often similar (e.g. the conversion of a transcription factor from an activator to a repressor of genes), there are significant differences in the genetic alterations implicated in AML and ALL pathogenesis.

Acute Myeloid Leukemia

The term AML collectively refers to a mixture of distinct disorders that differ with regard to their pathogenetic evolution, genetic abnormalities, clinical features, response to therapy and prognosis. AML arise in primitive, myeloid-lymphoid stem cells or from the accumulation of lineage-restricted myeloid progenitor cells arrested at early stages of differentiation. The different subtypes correspond roughly to maturation stages of granulocytic, monocytic, erythroid and megakaryocytic development (1). Age-specific incidence of AML, the most common acute leukemia in adults, rises linearly after age 40 with a median age of approximately 65 years. Most cases are sporadic, but congenital disorders (e.g. Fanconi's, ►Down's syndrome and Diamond-Blackfan syndromes) can increase the relative risk of developing AML. According to karyotypic analysis, three different prognostic groups with favourable, intermediate and unfavourable cytogenetic features have been defined. As treatment outcome differs substantially between these groups, with survival rates at five years ranging from 60–70% in the favourable to less than 20% in the unfavourable prognostic subgroup, the presence or absence of recurrent cytogenetic aberrations is currently used to identify the appropriate therapy.

Acute Lymphoblastic Leukemia

The distinct subtypes of ALL are characterized by leukemic lymphoblasts that are arrested in their differentiation at a precursor stage of B- or T-cell development. Although the general mechanisms underlying the induction of ALL are similar, the level of leukemic transformation, the frequency of particular genetic subtypes and cure rates differ markedly in children and adults. Lineage-committed stem cells are active in early development and provide potential targets for transformation in childhood ALL, whereas in adult ALL stem cells predominantly reside at the multipotential, lymphoid-myeloid cell level (e.g. B-cell precursor ALL with *BCR-ABL* fusion gene), and hence are preferentially at risk (4). ALL is the most common malignancy in children and has a bimodal age distribution with an early peak between 2 and 6 years of age and a second peak in the fifth decade. With the use of modern treatment strategies, about 80% of children and 35–40% of adults can be cured (5).

Chronic Leukemias

In contrast to acute leukemias, which show an aggressive clinical course, chronic leukemias have a more indolent clinical course often requiring less aggressive therapy. Chronic leukemias represent a heterogeneous group of different clonal disorders and belong, according to the WHO classification of tumours of the haematopoietic and lymphoid tissues, either to the mature B- or T-cell neoplasms or to the chronic myeloproliferative disorders. The most common chronic leukemias are chronic B-cell lymphocytic leukemia (B-CLL) and chronic myelogenous leukemia (CML). Whereas ▶B-cell is characterized by a marked genetic heterogeneity, which allows division of this malignant disorder into different prognostic subgroups, ▶CML is typically characterized by one specific genetic abnormality, which is the central event in the pathogenesis of this disease.

Chronic Lymphocytic Leukemia

B-CLL is the most common leukemia of adults in western countries and might originate either from immature pregerminal B-cells or from mature post-germinal centre B-cells. Although objective remissions can be achieved by various chemotherapy regimens, patients will relapse and ultimately succumb to their disease. Thus, CLL cannot be cured by currently available treatment options.

Chronic Myelogenous Leukemia

CML is a chronic myeloproliferative disease that originates in an abnormal pluripotent haematopoietic stem cell and is consistently associated with the Philadelphia chromosome (Ph) or its molecular equivalent, the *BCR-ABL* fusion gene.

Characteristics

Acute Myeloid Leukemia

AML involves a large variety of genetic alterations, including mainly chromosomal translocations that are selectively associated with distinct subtypes of the disease, point mutations and deletions (2). The chromosomal translocations can be clustered into groups based on similar structure and function of the expressed genes (Table 1). Some of the chromosomal translocations that occur in disparate loci in the human genome are associated with clinically indistinguishable phenotypes. Chromosomal translocations in AML most often result in loss of function of transcription factors or transcriptional coactivators that impair normal haematopoietic differentiation. There is now abundant evidence, however, that many of the fusion oncogenes resulting from balanced chromosomal translocations are not sufficient by themselves to cause AML. The observation that chimeric fusion genes generated by chromosomal translocations commonly coexist with a defined spectrum of mutations affecting tyrosine kinases has led to the notion of cooperating gene lesions, which couple a block in differentiation with enhanced cellular proliferation (Fig. 1). Such cooperating mutations have been divided into Class I and Class II mutations. Class I mutations (e.g. constitutively activated tyrosine kinases or their downstream effectors) confer a proliferative and/or survival advantage to haematopoietic progenitors, but do not have any effect on haematopoietic differentiation. Class II mutations (e.g. chromosomal translocations targeting core binding factor, ▶CBF, and the PML/RAR α oncogene) impair haematopoietic differentiation and might provide a survival advantage by interfering with terminal differentiation and apoptosis. Together, Class I and Class II mutations result in the AML phenotype characterized by proliferation and impaired haematopoietic differentiation of haematopoietic progenitor cells. While this model for a multistep pathogenesis, also referred to as the “two-hit” model of oncogenes in AML, might be an oversimplification, there is convincing experimental support for this model in mouse models of leukemia and it has a number of therapeutic implications (Fig. 1).

Acute Lymphoblastic Leukemia

Molecular analysis of the common genetic alterations in ALL has contributed greatly to the understanding of the pathogenesis of ALL and has been used successfully as a diagnostic and prognostic tool that gives important clues to risk-adapted therapeutic interventions. These genetic alterations contribute to the leukaemic transformation of haematopoietic stem cells or their committed progenitors by altering key regulatory processes that maintain or enhance an unlimited capacity for self-renewal, subvert the

Leukemia. Table 1 Main biological subtypes, recurrent chromosomal changes, and genes affected by chromosomal breakpoints in acute myeloid leukemias and acute lymphoblastic leukemias

Subtype	Cell type involved (Immunophenotype)	Chromosome abnormality	Molecular lesion	Function
AML	Acute myeloid leukemia with maturation	t(8;21)(q22;q22)	AML1-ETO fusion	Chimeric TF
AML	Acute myeloid leukemia with abnormal bone marrow eosinophils	inv(16)(p13q22) t(16;16)(p13;q22)	CBF β -SMMHC fusion	Chimeric TF
AML	Acute promyelocytic leukemia	t(15;17)(q22;q12-21)	PML-RAR α fusion	Chimeric TF
AML	Acute myeloid leukemia	t(9;11)(p22;q23)	MLL-AF9 fusion	Modified TF, dysregulated expression of HOX genes
AML	B-cell precursor (pro-B)	t(4;11)(q21;23)	MLL-AF4 fusion	Modified TF, dysregulated expression of HOX genes
ALL	B-cell precursor (common, pre-B)	t(12;21)(p13;q22)	TEL-AML1 fusion	Chimeric TF
ALL	B-cell precursor (pre-B)	t(1;19)(q23;p13)	E2A-PBX1 fusion	Chimeric TF
ALL	B-cell precursor (common, pre-B)	t(9;22)(q34;q11)	BCR-ABL fusion	Activated kinase
ALL	T-cell precursor	1p33 deletion (CD3- or + $\alpha\beta$)	TAL1-SCL fusion	dysregulated TF
ALL	T-cell precursor	t(5;14)(q35;q32)	HOX11L2 rearrangement	HOX11L2 over-expression
ALL	T-cell precursor	t(10;14)(q24;q11)	HOX11 rearrangement	HOX11 over-expression

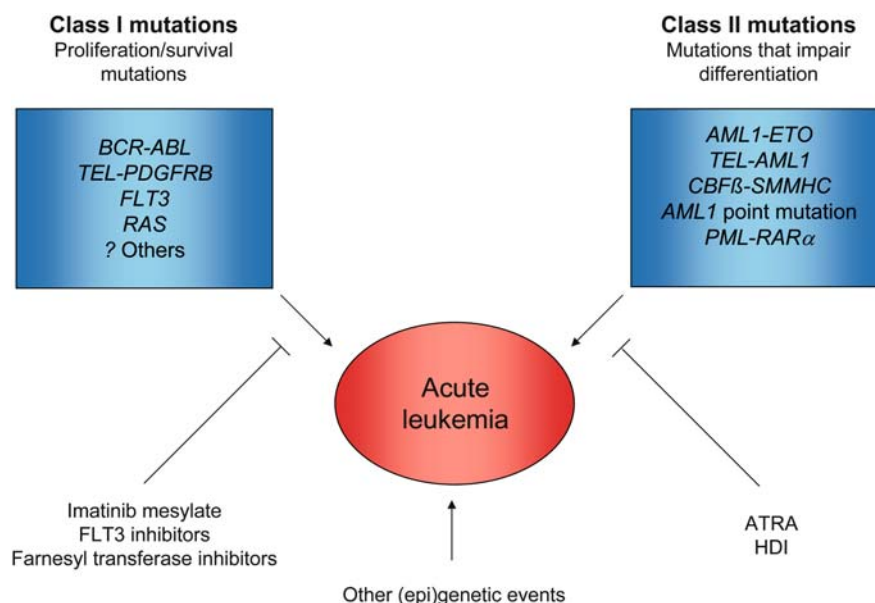
AML, acute myeloid leukemia; ALL, acute lymphoblastic leukemia; TF, transcription factor

controls of normal proliferation, block differentiation and promote resistance to apoptosis. Like AML, the conversion by chromosomal translocation of a transcription factor from an activator to a repressor of genes (e.g. the t(12;21) translocation and translocations involving the *MLL* gene, see below) or the generation of fusion genes leading to constitutive activation of a tyrosine kinase (e.g. the t(9;22) translocation, also referred to as the Philadelphia (Ph) chromosome) are recurrent pathogenetic mechanisms in ALL (Table 1). These oncogenic events triggered by chromosomal rearrangements probably cooperate with a second class of mutations that alter the proliferation and survival of haematopoietic progenitors, including over-expression of **▶FLT3** or alterations in the retinoblastoma protein (RB) and **▶p53** pathways that have a central role in both tumour suppression and the response of tumour cells to chemotherapy. Alternatively, transcription factor genes in ALL are dysregulated by their

juxtaposition with transcriptionally active T-cell receptor (*TCR*) or immunoglobulin (*Ig*) genes. In addition, numerical chromosomal changes, especially hyperdiploidy involving more than 50 chromosomes, occur frequently in childhood ALL and are associated with more favourable presenting features and higher cure rates than other major prognostic subgroups.

Chronic Lymphocytic Leukemia

Although many patients with B-CLL have an indolent course with a median survival of 7–10 years, the disease is characterized by a marked clinical heterogeneity; some patients are refractory to initial chemotherapy and die from the disease within a few months, whereas others live for 20 years or longer. The clinical heterogeneity is reflected in a considerable genetic heterogeneity and certain genomic aberrations and mutations have turned out to be associated with either a more aggressive or a more indolent clinical course (6).



Leukemia. Figure 1 Multistep pathogenesis of AML and targeted treatment strategies. Model for cooperation between gene rearrangements and mutations that confer a proliferative and/or survival advantage to leukaemic blasts and those that impair differentiation (modified from reference 2). Targeted therapies have been developed and have proven to be effective (e.g., small molecule inhibitors, such as imatinib mesylate for BCR-ABL or TEL/PDGFR β , and all-trans-retinoic acid (ATRA) or arsenic trioxide for the treatment of APL). Other molecularly targeted drugs, including histone deacetylase, farnesyl transferase, and FLT3 inhibitors, are being tested in AML. There may be therapeutic synergy achieved by targeting each class of mutation, such as a combination of FLT3 inhibitors and ATRA in treatment of APL.

Chronic Myelogenous Leukemia

CML is characterized by the expansion of all maturation stages of the myeloid lineage, but other haematopoietic lineages may regularly be affected as well. The clinical course of the disease is triphasic: 1) an initial chronic phase with an indolent clinical course (median duration 5–6 years), 2) an accelerated phase (median duration 6–9 months) and 3) a blast crisis which resembles acute leukemia with a poor prognosis (median survival 3–6 months). Treatment of patients in chronic phase with interferon- α and/or cytotoxic drugs induces stable remissions and can lead to disease control for several years. However, despite treatment the disease is progressive and ultimately fatal. Currently, the only curative therapy is allogeneic bone marrow or peripheral blood stem cell transplantation.

Molecular Interactions Acute Myeloid Leukemia

Transcription factors that are targeted by multiple chromosomal translocations in AML include CBF, retinoic acid receptor alpha (RAR α), and members of the \blacktriangleright *HOX* family of transcription factors (Table 1). As chromosomal translocations involving CBFs are a frequent target of gene rearrangements and mutations in acute leukemias, molecular interactions impairing

haematopoietic differentiation will be exemplified by CBF gene rearrangements (2). CBF is a heterodimeric transcription factor composed of a DNA-binding component, *AML1* (also known as *PEBPA2B* and *CBFA2*; recently renamed by the HUGO as Runt-related transcription factor 1 – *RUNX1*) and the CBF β subunit that does not contact DNA but increases transcriptional activity of *AML1*. CBF transactivates expression of a broad spectrum of genes that are critical for normal haematopoietic differentiation. The three most common translocations involving CBF are t(8;21) and inv(16), accounting for approximately 25% of adult AML, and t(12;21), occurring in 20–25% of childhood ALL. The molecular consequence of these chromosomal translocations is the production of AML1-ETO (acute myeloid leukemia 1/eight twenty one), CBF β -SMMHC (smooth muscle myosin heavy chain), and TEL (translocation ETS leukemia)-AML1 fusion proteins, respectively. *In vitro* data and experiments with knock-in mice have provided convincing evidence that the main physiological function of these fusion proteins is dominant inhibition of CBF-mediated transcription, resulting in severely impaired haematopoietic differentiation, a phenotypic hallmark of AML. Transcriptional repression of CBF target genes by CBF-related fusion proteins is mediated in part through

aberrant recruitment of the nuclear corepressor/histone deacetylase complex. Some of these mechanisms of transcription repression are similar to those reported for the PML (promyelocytic leukemia)-RAR α and PLZF-RAR α fusion proteins that are associated with acute promyelocytic leukemias (APL), and that also recruit the nuclear corepressor complex.

Genetic alterations involving CBFs probably cooperate with a second class of mutations that alter the proliferation and survival of haematopoietic progenitors. Recently, several such mutations have been identified in AML patients with CBF leukemias, including internal tandem duplications in the juxta-membrane domain of FLT3 as well as activating loop mutations of FLT3 or \blacktriangleright c-KIT, resulting in constitutive activation of the Class III receptor tyrosine kinases, FLT3 or c-KIT and activation of growth-related signalling pathways (3). In addition, activating mutations in *RAS* (rat sarcoma) family members, primarily *N-Ras* and *Ki-Ras*, have been identified in a subset of AML patients.

Acute Lymphoblastic Leukemia

ALL-specific translocations can oncogenically activate transcription factors by at least two mechanisms (4, 5). First, especially in B-cell precursor ALL, translocations fuse discrete portions of two different genes to create chimeric transcription factors with oncogenic properties. This mechanism can be exemplified by translocations involving the mixed-lineage leukemia (*MLL*) gene, also called myeloid lymphoid leukemia gene. Structural genetic alterations of the *MLL* locus (11q23) occur in more than 80% of infants with ALL, 5% of AML cases and in most therapy-induced leukemias caused by topoisomerase II inhibitors. Translocations involving *MLL* result in chimeric proteins consisting of the N-terminal portion of *MLL* fused to the C-terminal portion of one of more than 40 partners. The most common translocations represent the t(4;11), t(9;11) and t(11;19) which fuse *MLL* to a family of genes (*AF-4*, *AF-9* and *ENL*) that may contribute similar functional domains to the chimeric proteins. The *MLL* fusion proteins have a dominant gain-of-function effect than enhances their transcriptional activity. These alterations disrupt the normal pattern of expression of *HOX* genes, resulting in a change in the self-renewal and growth of haematopoietic stem cells and progenitor cells.

Alternatively, B- and T-cell precursor ALL are characterized by the juxtaposition of strong enhancer elements from *TCR* or *Ig* genes with various transcription factors leading to protein over-expression and ultimately to leukaemic transformation. For example, recurring translocations in T-cell precursor ALL often involve transcriptionally active sites of the

beta (7q34) or the delta locus (14q11) of the TCR, leading to dysregulated expression of transcription factors. These rearrangements may result from mistakes in the normal recombination process involved in the generation of functional antigen receptors. Developmentally important transcription factor genes altered in T-cell precursor ALL include members of the basic helix-loop-helix (bHLH), LIM and homeodomain families.

Chronic Lymphocytic Leukemia

Genetic analysis of the variable region of the immunoglobulin heavy chain gene (\blacktriangleright IgV_H) revealed that CLL can be divided into two subgroups, tumours with unmutated IgV_H genes are associated with rapid clinical progression and unfavourable clinical course, whereas tumours with hypermutated IgV_H genes have a better prognosis requiring late or no treatment. Physiologically, an unmutated IgV_H locus reflects an immature pregerminal B-cell maturation stage, whereas B-cells with hypermutated IgV_H have passed through the germinal centre and represent a positively selected mature differentiation stage (postgerminal centre B-cells). Therefore, it has been proposed that B-CLL may originate from two different B-cell stages, immature (unmutated) pregerminal B-cells or mature (hypermutated) postgerminal B-cells.

Although the link between IgV_H mutations and clinical outcome is still unclear, one crucial factor probably relates to the different chromosomal abnormalities associated with each of the two subgroups. The malignant cells from patients with unmutated IgV_H genes frequently have cytogenetic changes that are associated with a poor clinical outcome such as an 11q or a 17p deletion, whereas cells from patients with hypermutated IgV_H more frequently have chromosomal aberrations with a benign clinical course (e.g. 13q deletions). Although certain chromosomal aberrations allow division of B-CLL into different prognostic subgroups, relatively little is known about the molecular correlates and consequently about the molecular mechanisms underlying the pathogenesis of this disorder (Table 2).

The deleted regions of the affected chromosomes 17 and 11 are known to contain tumour suppressor genes that regulate proliferation, apoptosis and cellular response to DNA damage. Deletions of the short arm of chromosome 17 (17p) affect the tumour suppressor gene *p53* and deletions of the long arm of chromosome 11 (11q) can affect the \blacktriangleright *ATM* (ataxia telangiectasia mutated) gene (Table 2). Accordingly, it could be demonstrated that B-CLL tumours that have inactive ATM or p53 proteins show defects in p53 DNA damage response and in ionizing radiation (IR)-induced apoptosis.

Leukemia. Table 2 The most common chromosomal abnormalities found in patients with CLL. Little is known about the molecular correlates of these chromosomal abnormalities. The tumour suppressor gene *p53* is affected by 17p deletions, the *ATM* gene is affected by 11q deletions

Aberration	Affected Genes
13q deletion	unknown
11q deletion	<i>ATM</i>
12q trisomy	unknown
17p deletion	<i>p53</i>
6q deletion	unknown
8q trisomy	unknown

Chronic Myelogenous Leukemia

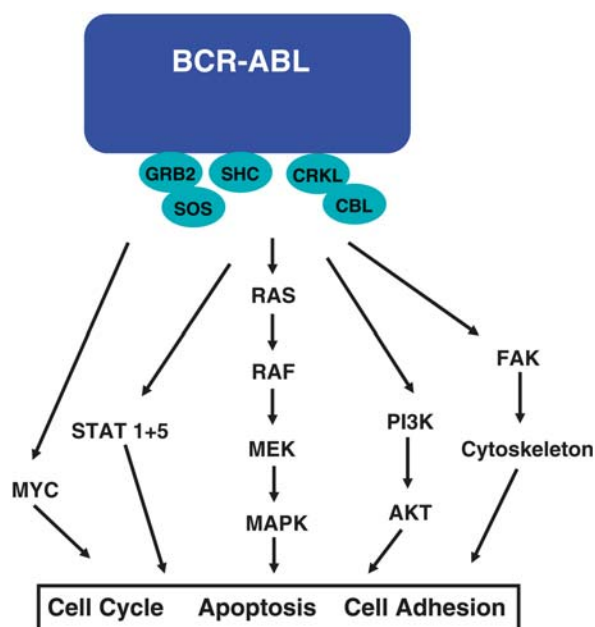
At diagnosis, 90–95% of CML cases have a reciprocal chromosomal translocation involving the long arm of chromosomes 9 and 22, resulting in the Ph chromosome. This translocation fuses the *BCR* gene (break-point cluster region) on chromosome 22 with regions within the *ABL* gene (Abelson mouse leukemia proto-oncogene) on chromosome 9. Physiologically, ABL is a non-receptor tyrosine kinase that activates growth regulating signalling pathways upon growth factor stimulation. The *BCR-ABL* fusion gene leads to the expression of a kinase with constitutive activity. The reason for the constitutive activity of the **BCR-ABL** fusion kinase is still unclear. The aberrant BCR-ABL fusion kinase leads to the constitutive activation of signalling pathways which mediate cell survival, cell-cycle progression, and cell adhesion (Fig. 2). Thus, BCR-ABL has been described to activate the RAS/MAPK, STAT 1 + 5 and PI3K/AKT signalling pathways, which are known to be involved in the regulation of proliferation, differentiation and apoptosis. The *BCR-ABL* fusion gene is the hallmark of CML, defines the disease and is most likely the major causative event in disease development (7).

Thus, at least in the chronic phase, the expansion of the malignant clone is largely dependent on this specific genetic defect. During disease progression to accelerated phase and blast crisis, the malignant clone acquires additional genetic defects.

Regulatory Mechanisms

Acute Myeloid Leukemia

The better understanding of the molecular pathogenesis of AML, especially CBF leukemias, and the growing number of tyrosine kinases now known to be activated by mutations in leukemic blasts from AML patients,

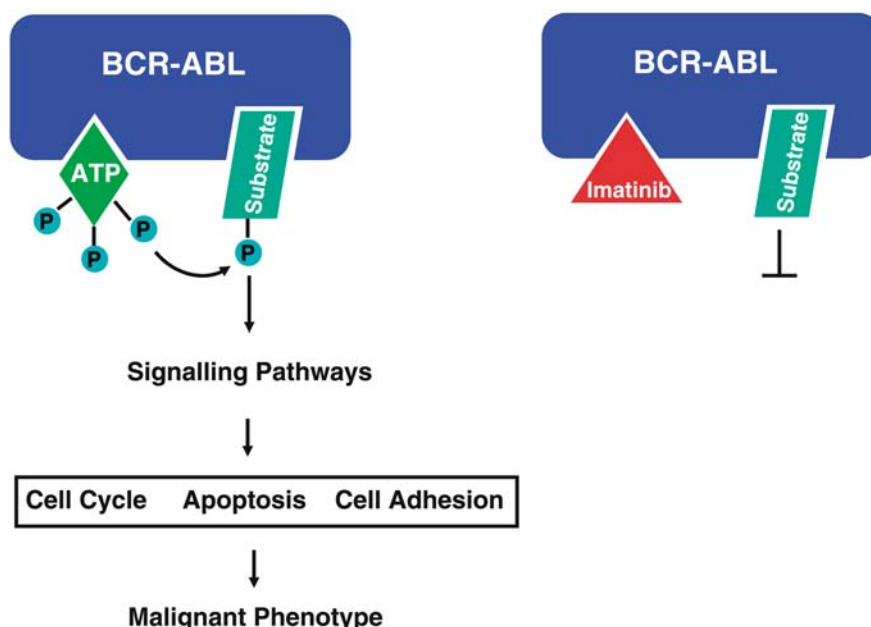


Leukemia. Figure 2 BCR-ABL triggered signalling pathways in CML. The Philadelphia chromosome produces a constitutively activated BCR-ABL tyrosine kinase, which leads to the constitutive activation of different signalling pathways (e.g., RAS/RAF/MEK/MAPK, STAT 1 + 5, PI3K/AKT, FAK) and oncogenes (e.g., MYC). The activation of these pathways protects the cell from apoptosis, promotes cell-cycle progression and alters proteins of the cytoskeleton and cell adhesion properties, thereby critically contributing to the process of malignant transformation. The activation of RAS, PI3K, STAT and FAK is mediated by a series of adapter proteins (SHC, SOS, GRB2, CRKL, CBL).

provide a foundation for the development of new molecularly targeted therapies. These include histone deacetylase inhibitors (HDI), specific inhibitors of tyrosine kinases such as FLT3 and c-KIT, farnesyl transferase inhibitors and the identification of small molecules that specifically inhibit fusion proteins without affecting the function of native CBF components or specifically disrupt association of the nuclear corepressor complex with leukaemogenic fusion proteins. Several of these new molecularly targeted therapies, including HDI and FLT3 inhibitors, have entered clinical trials. A combination of basic, clinical and translational studies will ultimately determine the clinical utility of these agents and their mechanism of action.

Acute Lymphoblastic Leukemia

The different clinical outcomes associated with the various subtypes of B- and T-cell precursor ALL can be



Leukemia. Figure 3 Mechanism of action of imatinib mesylate. The BCR-ABL fusion protein is a constitutively activated non-receptor tyrosine kinase that is causally linked to malignant transformation in CML. Imatinib mesylate selectively blocks the binding site for ATP in the ABL kinase, thereby preventing phosphorylation of tyrosine residues on substrate proteins. This leads to the disruption of the downstream signalling cascades and might revert the malignant phenotype.

attributed mainly to drug sensitivity or resistance of leukaemic lymphoblasts harbouring specific genetic abnormalities. The unusual sensitivity of leukaemic lymphoblasts with a hyperdiploid karyotype to chemotherapy, especially in childhood ALL, is an example. Although the precise mechanism of leukaemic transformation in hyperdiploid ALL is unknown, leukaemic cells from patients with this numerical abnormality have an exquisite sensitivity to chemotherapy. This *in vivo* response correlates with the propensity of leukaemic cells to undergo spontaneous apoptosis *in vitro* and to have higher than average intracellular concentrations of methotrexate and its active polyglutamate metabolites. The latter probably results from three or four copies of chromosome 21, which harbours a gene encoding the transporter of methotrexate into cells, in more than 97% of hyperdiploid blasts. Recent emphasis has been placed on the study of genetic polymorphisms in drug-metabolizing enzymes, drug transporters and targets of drug action. The information should be useful for optimising drug doses and drug combinations. As with AML, identification of specific oncoproteins and better understanding of the molecular processes regulating treatment response have paved the way for therapy directed to pivotal molecular targets (e.g. ►[imatinib](#)

[mesylate](#) for the treatment of *BCR-ABL*-positive ALL). Progress in the molecular classification of ALL, e.g. by using DNA microarrays, coupled with methods to assess the functional significance of newly discovered genes or through proteomic techniques will hopefully lead to identification of new targets for specific treatments.

Chronic Lymphocytic Leukemia

ATM functionally interacts with the p53 protein. The p53 protein preserves genome integrity by regulating cell-cycle arrest and apoptosis. In response to DNA damage (e.g. DNA double strand breaks induced by IR), ATM stabilizes and activates p53 through phosphorylation. As a transcription factor, activated p53 represses or induces genes that are involved in the regulation of cell-cycle arrest, apoptosis and DNA repair. Thus, mutations or chromosomal deletions of the *ATM* or the *p53* gene might provide some explanation as to why patients with these genetic aberrations often respond poorly to cytotoxic chemotherapy and have an unfavourable clinical course. Activated p53 acts as a transcription factor and drives the expression of proapoptotic genes such as *Bax*, cell-cycle inhibitors such as *p21* and DNA-repair genes. Therefore, besides their role in DNA repair, mutations

in *p53* and *ATM* may contribute to the pathogenesis of a certain unfavourable subgroup of B-CLL.

Chronic Myelogenous Leukemia

Since the BCR-ABL fusion protein has been thought to be causative for the expansion of the malignant clone, attempts have been made to develop selective inhibitors. Recently, an inhibitor specific for a small family of tyrosine kinases, including PDGF-R, c-KIT and BCR-ABL, was generated. This small compound, designated imatinib mesylate, binds as a competitive inhibitor to the ATP binding site of these kinases, thereby blocking their enzymatic activity (8). Consequently, the activation of apoptosis and proliferation regulation signalling pathways is blocked as well (Fig. 3). In first clinical trials, imatinib mesylate has induced complete haematological and cytogenetic remissions in patients with CML. In particular in the chronic phase, high remission rates could be achieved and in a phase III trial imatinib mesylate appeared to be superior to the conventional standard therapy consisting of interferon- α and cytosine arabinoside. However, despite these promising clinical results, resistance against this novel drug might develop preferentially by the selection of malignant clones which have point mutations affecting the imatinib mesylate binding site of *BCR-ABL*. Furthermore, despite the induction of high remission rates, imatinib mesylate alone probably cannot cure the disease and how long this selective tyrosine kinase inhibitor can keep CML patients in stable clinical remission is still unclear. Nevertheless, treatment of CML with imatinib mesylate is a *par excellence* example of molecularly targeted therapy that has been developed based on the understanding of the molecular mechanisms of malignant transformation.

References

1. Jaffe ES, Harris NL, Stein H (2001) World Health Organization Classification of Tumours. Pathology and genetics of tumours of haematopoietic and lymphoid tissues. IARC Press, Lyon
2. Kelly LM, Gilliland DG (2002) Genetics of myeloid leukemias. *Annu Rev Genomics Hum Genet* 3:179–98
3. Scheijen BGJ (2002) Tyrosine kinase oncogenes in normal hematopoiesis and hematological disease. *Oncogene* 21:3314–3333
4. Greaves MF, Wiemels J (2003) Origins of chromosome translocations in childhood leukemia. *Nat Rev Cancer* 3:639–649
5. Pui C-H RM, Downing JR (2004) Mechanisms of disease – Acute lymphoblastic leukemia. *N Engl J Med* 350:1535–1548
6. Döhner H, Stilgenbauer S, Benner A et al (2000) Genomic aberrations and survival in chronic lymphocytic leukemia. *N Engl J Med* 343:1910–1916
7. Pane F, Intrieri M, Quintarelli C et al (2002) BCR/ABL genes and leukemic phenotype: from molecular mechanisms to clinical correlations. *Oncogene* 21:8652–8667
8. Schindler T, Bornmann W, Pellicena P et al (2000) Structural mechanism for STI-571 inhibition of abelson tyrosine kinase. *Science* 289:1938–1942

Leukocytes

Definition

Leukocytes (or white blood cells) are differentiated hematopoietic cells expressing the surface marker CD45. They include cells of the myeloid and lymphoid lineage (i.e. monocytes, macrophages and lymphocytes), but exclude the erythroid lineage. Leukocytes help the body fight infections and other diseases.

► [Interleukins and Their Receptors, a Role in Regulating Hematopoietic Stem Cells](#)

► [Neutrophils](#)

► [Proteomics in Cardiovascular Disease](#)

Leukodystrophies

Definition

Leukodystrophies comprise of a heterogenous group of diseases affecting synthesis or function of CNS myelin due to hereditary enzymopathies (e.g. storage of lysosomal lipids).

► [Glial Cells and Myelination](#)

Lewy Body

Definition

Lewy body refers to intracellular eosinophilic proteinaceous inclusion, which is found on neuropathological evaluation in some neurons of ► [Parkinson's Disease: Insights from Genetic Causes](#) patients and in some other disorders (such as Diffuse Lewy Body disease). It stains strongly with antibodies to ubiquitin, and even more strongly with antibodies to α -synuclein.

► [Parkinson's Disease: Insights from Genetic Cause](#)

Leydig Cells

Definition

Leydig cells are cells of the testis that regulate ►[Sertoli cell](#) function by secreting testosterone in response to luteinizing hormone (LH).

►[Mammalian Fertilization](#)

LFS

►[Li-Fraumeni Syndrome](#)

Lgs/BCL19

►[Legless\(Lgs\)/BCL19](#)

LH

Definition

The luteinizing hormone is a glycoprotein (28 kD) consisting of 2 polypeptide chains (α and β), which is generated in the pituitary, and stimulates the growth of the ►[Leydig cells](#) in males and the synthesis of androgens; in females it initiates the maturation of follicles, ovulation and the synthesis of estrogen and progesterone. The release of LH is controlled by LHRH (Luteinizing hormone releasing hormone).

LH Domain

Definition

The LH (Lipoxygenase homology) domain, also known as the Polycystin-1, Lipoxygenase, and Alpha-Toxin domain (PLAT), is thought to mediate protein-protein interaction and interaction of proteins with membrane lipid.

►[Autosomal Dominant \(Inherited Disorder\)](#)

►[Polycystic Kidney Disease, Autosomal Dominant](#)

Library

Definition

In the context of molecular biology, a library is a collection of cloned DNA fragments ('clones'), normally all derived from the same source DNA and in the same vector and host. Unlike real-life libraries, no index or inventory is required for a library of clones to merit the name.

►[Protein Interaction Analysis: Suppressor Hunting](#)

►[YAC and PAC Maps](#)

Library Amplification

Definition

Certain cDNA libraries are often a precious resource and may become limited. To conserve the complexity of a library, the amplification of the cDNA library may be required. This can be achieved by growing the entire library or aliquots of it as a mixed culture. Both chance (in the early part of growth) and varying growth rates mean that this is associated with loss of some clones and high abundance of others, possibly ones with a selective advantage caused by rearrangements such as insert deletion.

►[YAC and PAC Maps](#)

Liddle's Syndrome

Definition

Liddle's Syndrome is an autosomal dominant disorder that is caused by persistent hyperactivity of the epithelial Na-channel (ENaC). Its symptoms mimic aldosterone excess, but plasma aldosterone levels are actually reduced (pseudoaldosteronism). The disease is characterized by early onset arterial hypertension hypokalemia and metabolic alkalosis. Disease causing mutations are found in the cytoplasmic regulatory region of the β - and γ subunits of the epithelial sodium channel genes.

►[Hypertension](#)

►[Mendelian Forms of Human Hypertension and Mechanisms of Disease](#)

►[Mendelian Hypertension](#)

LIE

Definition

The linear interaction energy (LIE) method estimates binding free energies from molecular dynamics simulation averages of the electrostatic and van der Waals interaction energies in the bound and the free state of the binding partners, using an empirically determined scaling factor.

► [Molecular Dynamics Simulation in Drug Design](#)

Li-Fraumeni Syndrome

Definition

Li-Fraumeni Syndrome (LFS) is a rare autosomal dominantly inherited disorder, characterized by the diagnosis of bone- or soft-tissue sarcoma at an early age, in an individual who has one first-degree relative with early-onset cancer or sarcoma diagnosed at any age. Li-fraumeni syndrome is caused by mutation in the tumor protein p53 (TP53), checkpoint kinase 2 (CHK2) or other genes.

► [Tumor Suppressor Genes](#)

Ligand

Definition

A ligand is the signal protein or protein complex in a signalling pathway that binds to and activates the corresponding receptor protein. Ligands can be macromolecules themselves, like proteins or DNA, or they might define a small molecule, as for example peptides, hormones or compounds obtained by synthetic organic chemistry.

► [Notch Pathway](#)

► [Nuclear Hormone Receptors: Regulators of Gene Transcription and Cellular Signaling](#)

► [Protein-Ligand-Interaction by NMR](#)

Ligation

Definition

Ligation is a method in recombinant DNA technology, by which the enzyme T4-DNA ligase is used to

covalently join the ends of DNA fragments previously cut by restriction enzymes.

► [DNA Ligases](#)

► [Recombinant Protein Expression in Bacteria](#)

LIM Gene

Definition

LIM genes constitute a large gene family of homeobox genes, whose gene products carry the LIM domain, a unique cysteine-rich zinc-binding domain. The acronym stands for *lin-11*, *islet-1* (*isl-1*) and *mec-3*. At least 40 members of this family have been identified in vertebrates and invertebrates, and are distributed into 4 groups according to the number of LIM domains and to the presence of domains such as homeodomains and kinase domains.

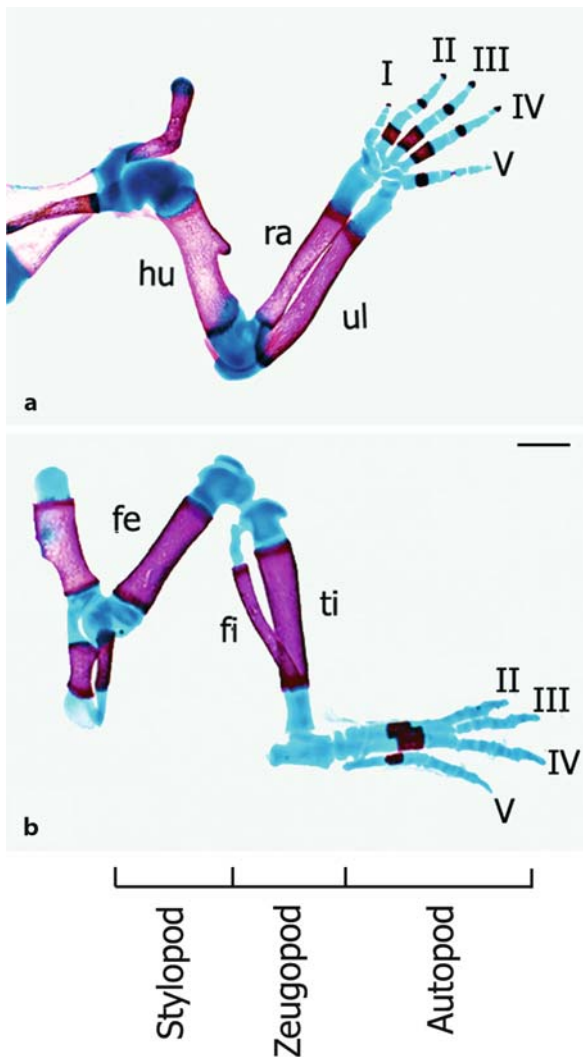
► [Hypothalamic and Pituitary Diseases genetics](#)

Limb Development

NATALIA SOSHNIKOVA, WALTER BIRCHMEIER
Max-Delbrueck-Center for Molecular Medicine,
Berlin, Germany
wbirch@mdc-berlin.de

Definition

The developing limb is a classical model system for studying cytological and genetic aspects of pattern formation and signal transduction. The vertebrate limb is an outgrowth of the embryonic body wall at a specific position along the anterior-posterior axis of the embryo. The forelimbs always arise at the cervical to thoracic transition and the hind limbs at the lumbosacral transition. The limb consists of mesenchymal tissues covered by a layer of ectoderm. Cartilage and bone elements develop from the limb ► [mesenchyme](#). Other tissues such as muscles, nerves and blood vessels evolve from cells that migrate into the limb bud. Three discrete compartments form in vertebrate limbs, proximal (stylopod), middle (zeugopod) and distal (autopod) (Fig. 1).

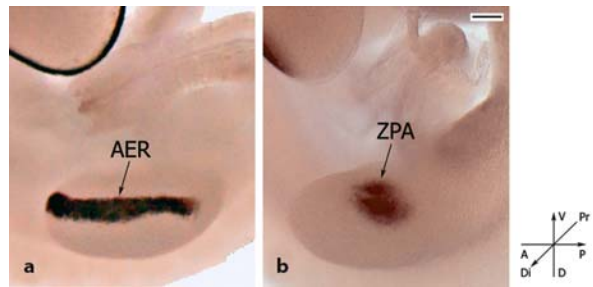


Limb Development. Figure 1 Skeletal pattern of a newborn mouse forelimb (a) and hindlimb (b). The stylopod contains humerus (hu) in the forelimb and femur (fe) in the hindlimb. The zeugopod consists of radius (ra) and ulna (ul) in the forelimb, and tibia (ti) and fibula (fi) in the hindlimb, whereas the autopod comprises wrist and five fingers in the forelimb, and ankle and five toes in the hindlimb. Alcian blue and alizarin red staining produces coloration of cartilage elements in blue and bone elements in red. Bar: 2 mm.

Characteristics

The outgrowth and patterning of the limb depends on the establishment and maintenance of three distinct signaling centers, the ►**apical ectodermal ridge** (AER), the ►**zone of polarizing activity** (ZPA) and the non-ridge ectoderm of the limb bud (1, 2).

- The AER is a specialized epithelial structure positioned at the border between dorsal and ventral



Limb Development. Figure 2 Signaling centers in the limb bud. A posterior view of mouse limb buds at the 11th day of embryonic development. (a) The AER is positioned at the border between dorsal and ventral ectoderm as specified by the expression of *Fgf8*. (b) The ZPA is located in the posterior mesenchyme as marked by the expression of *Shh*. The expression of genes is visualized by whole-mount RNA *in situ* hybridization. Pr, proximal; Di, distal; A, anterior; P, posterior; D, dorsal; V, ventral. Bar: 150 µm.

ectoderm (Fig. 2a). It is responsible for the continuous proliferation of the underlying mesenchymal cells that leads to the outgrowth of the limbs along the proximal-distal axis (shoulder to finger tip). When the AER is surgically excised, proliferation of mesenchymal cells is affected and the limb is truncated. Truncations are more severe when the AER is removed early in development, indicating that there is a differential temporal necessity for the AER. The AER produces fibroblast ►**growth factors** (FGFs) that are the key signals for stimulation of cell division in the subjacent mesenchyme.

- The ZPA, a group of mesenchymal cells, is located directly below the posterior margin of the AER (Fig. 2b). Transplantation of posterior mesenchymal cells into the anterior side of the limb results in mirror duplication of digits, indicating that the ZPA organizes structures within the limb along the anterior-posterior axis (thumb to little finger). The ZPA cells secrete a ►**morphogen**, Sonic Hedgehog (Shh). Cells near to the ZPA receive a high concentration of Shh and adopt posterior fates, while cells further from the ZPA receive a lower Shh concentration and adopt anterior fates.
- The non-ridge limb ectoderm regulates the patterning of the limb along the dorsal-ventral axis (back of the hand to palm). Inversion of limb ectoderm along the dorsal-ventral axis leads to inversion of muscle, skeleton and skin appendages. The earlier rotation is performed, the more complete is the reversal of the dorsal-ventral limb pattern. Ventral identity of the limb is specified by ►**bone morphogenetic protein** (BMP) signals. Wnt-7a signals from dorsal ectoderm are required for limb cells to adopt a dorsal character.

Importantly, there are regulatory interactions among the different signaling centers that coordinate limb development along the three axes. For example, signals from the AER and the dorsal ectoderm cooperate to maintain ZPA activity. In a reciprocal manner, signals from the ZPA are necessary for the maintenance of the AER. Moreover, dorsal-ventral polarity and AER formation are processes initiated by a common mechanism.

Molecular Interactions

Significant progress has been achieved in defining signals that control patterning of each of the cardinal axes of the developing limb (1, 2, 3). These signals do not act in isolation but rather interactions between them direct the formation of various limb elements.

The Wnt/ β -catenin-FGF Regulatory Loop (Fig. 3a)

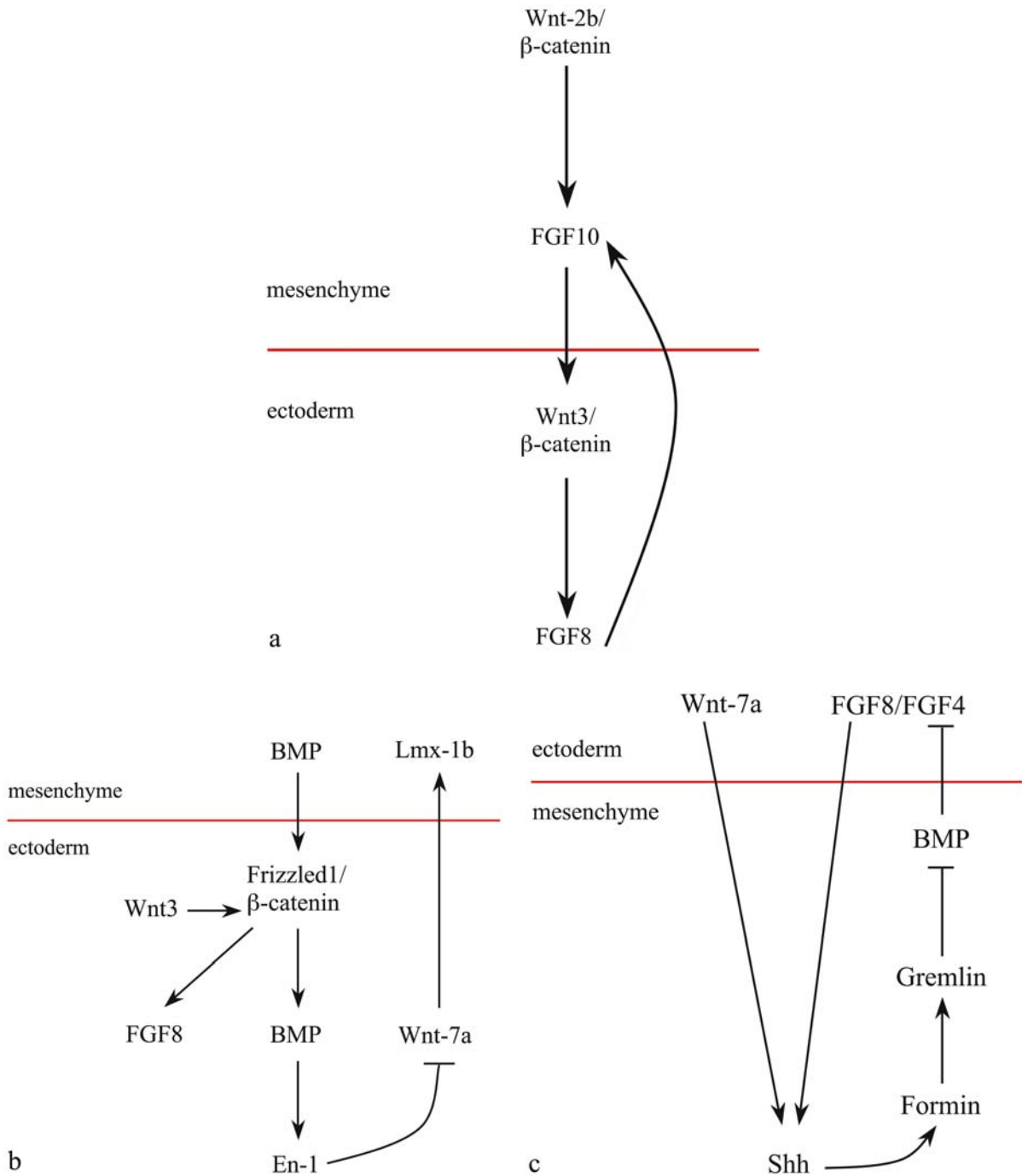
The crucial event in vertebrate limb development is the interaction of the presumptive limb mesenchyme with the surface ectoderm that results in the formation of the AER. Among the earliest genes expressed in the presumptive fore- and hind-limb mesenchyme are *Wnt-2b* and *Wnt-8c*, respectively. Wnt signals are mediated by the Frizzled receptors, which act to stabilize β -catenin in the cytoplasm (4). Upon nuclear translocation, β -catenin interacts with LEF/TCF transcription factors to regulate expression of target genes. Mutations in β -catenin, which increase stability of the protein, cause constitutive transduction of the signal to the nucleus, even in the absence of Wnt ligands. Ectopic expression of *Wnt-2b*, *Wnt-8c* or β -catenin results in induction of ectopic limbs in the chick embryonic flank (3). It has been shown that both *Wnt-2b* and *Wnt-8c*, via β -catenin, activate *Fgf10* expression in the limb mesenchyme. Genetic experiments in mouse and embryological manipulations in chick have demonstrated that FGF10 is a key factor for limb outgrowth. FGF10 acts specifically on the epithelium and induces the expression of *Wnt-3a* (in chick) or *Wnt3* (in mouse) in the limb ectoderm. Recently, it has been shown that *Wnt3*/ β -catenin signaling plays a crucial role in AER formation. Ablation of β -catenin in the limb ectoderm using the Cre/loxP system results in a block in AER induction and a loss of *Fgf8* and *Fgf4* expression in the presumptive AER cells (5). In contrast, activation of β -catenin-mediated signaling leads to ectopic AER formation and *Fgf8* expression in the limb ectoderm of both mouse and chick embryos. The expression of *Fgf8* and *Fgf4* in the limb ectoderm is required for the maintenance of *Fgf10* expression and for the proliferation of the underlying mesenchymal cells. Thus, three Wnts signaling through β -catenin mediate the FGF10/FGF8 loop during formation of the AER. Inactivation of any gene that encodes

components of this regulatory loop results in deficits of limb outgrowth along the proximal-distal axis.

The BMP-Wnt/ β -catenin Regulatory Network (Fig. 3b)

Conditional *Wnt3* and β -catenin mutant mice display a complete absence of *Bmp2* and *Bmp4* expression in the limb ectoderm (5). It has been demonstrated that BMP signaling plays an essential role in AER induction and proximal-distal outgrowth of the limbs (2). Mice carrying a *Bmp receptor IA* loss-of-function mutation in the limb ectoderm fail to form the AER and express *Fgf8*, whereas ectopic expression of constitutively active *BMP receptor IA* results in induction of *Fgf8* expression in the limb ectoderm of chick embryos. Moreover, inactivation of the *p63* gene, which is regulated by both Wnt/ β -catenin and BMP signals, also results in a loss of AER formation and in severe truncation of the limbs. To verify the epistatic relationship between Wnt/ β -catenin and BMP signaling pathways, compound mutant mice that carry β -catenin gain-of-function and *Bmp receptor IA* loss-of-function mutations in the limb ectoderm have been generated (5). Analysis has revealed that β -catenin-mediated signaling acts downstream of the BMP receptor IA during AER formation. Specifically, BMP signals from the ventral mesenchyme are required for expression in the ventral limb ectoderm of the *Frizzled1* gene, which acts as a permissive signal to establish competence of the limb ectoderm to respond to *Wnt3* signals. β -Catenin does, however, control *Bmp2* and *Bmp4* expression in the ectoderm and is responsible for the initiation of a positive feedback loop.

The formation of the AER and of the dorsal-ventral axis is tightly linked during early limb development (1, 2, 5). For example, conditional β -catenin or *Bmp receptor IA* loss-of-function mutants demonstrate absence of ventral structures of the limb such as dermal pads and duplication of dorsal structures such as nails. These findings indicate that both β -catenin and BMP receptor IA-mediated signaling are required not only for AER formation but also for the correct dorsal-ventral patterning of limbs. Interestingly, analysis of the compound mutant mice that carry β -catenin gain-of-function and *Bmp receptor IA* loss-of-function mutations in the limb ectoderm has demonstrated that β -catenin acts upstream of or in parallel to the BMP receptor IA for the establishment of the dorsal-ventral polarity of the limb. Gain-of-function experiments in the chick have shown that BMP signaling can induce the expression of *En-1* in the ventral limb ectoderm. *En-1*, a homeobox-containing transcription factor, is essential for the acquisition of the ventral pattern in limbs. In *En-1* mutant mice, dorsal structures such as nails and hair follicles also develop on the ventral side



Limb Development. Figure 3 Signaling pathways regulating limb morphogenesis. (a) The Wnt/β-catenin-FGF regulatory loop. (b) The BMP-Wnt/β-catenin regulatory network. (c) The FGF-Shh regulatory loop.

of the limb. Remarkably, *Wnt-7a* is ectopically expressed in the ventral limb ectoderm of *En-1* mutants. It has been suggested that *En-1* functions to restrict *Wnt-7a* expression to dorsal limb ectoderm. *Wnt-7a* activates the *Lmx-1b* expression in the dorsal

mesenchyme that is required for dorsal fate specification in limbs. Ectopic expression of *Wnt-7a* or *Lmx-1b* causes dorsalization of ventral mesenchyme, leading to the loss of dermal pads and to the formation of nails and hair follicles on the ventral side of limbs. In contrast,

loss of *Wnt-7a* or *Lmx-1b* results in formation of dermal pads on the dorsal side of the limb and in ventralization of dorsal muscles and tendons. Taken together, the intricate interplay between BMP and Wnt signaling pathways provides the molecular basis for the coordination of the development of the proximal-distal and dorsal-ventral axes in the limb.

The FGF-Shh Regulatory Loop (Fig. 3c)

Specific interactions between the AER and the ZPA establish growth and patterning of the limb along the proximal-distal and the anterior-posterior axes (1, 2, 3). It has been demonstrated that removal of the AER or the dorsal ectoderm results in rapid loss of *Shh* expression in the posterior mesenchyme of chick limbs. *Shh* plays a crucial role in specifying the identity of the limb along the anterior-posterior axis. Genetic studies in mice have shown that FGF8 and FGF4 from the AER and *Wnt-7a* from dorsal ectoderm are required for the induction of *Shh* expression in posterior mesenchyme. Moreover, regulatory interactions between *Shh* and FGF4 are reciprocal. Application of *Shh* to the anterior side of the chick limb causes ectopic expression of *Fgf4* in the overlying anterior ectoderm. Consistently, *Shh* mutant mice display a loss of *Fgf8* and *Fgf4* expression, the precocious regression of the AER and absence of digits 2–5. Recently, the molecular mechanism for *Shh*-dependent limb outgrowth has been revealed. *Shh* acts *via* formin to maintain the expression of *Gremlin*, an antagonist of BMP signaling, in limb mesenchyme. The inhibition of BMP signaling is required for the maintenance of *Fgf4* expression in the posterior AER. Loss of *formin* leads to a failure of *Fgf4* expression in the AER and subsequently to a decrease in the expression of *Shh* in limb mesenchyme. Thus, a positive feedback loop between FGFs and *Shh* controls outgrowth and patterning of the limb along the proximal-distal and anterior-posterior axes.

Limb Defects in Humans

Mechanisms controlling limb outgrowth and patterning in chick and mouse also function in a similar manner during human limb development. Identification of mutations in human genes that cause limb malformations has confirmed this (3, 6).

Mutations in genes that are responsible for early steps in limb development such as AER formation result in a complete loss of limbs (amelia) or in severe truncations of limbs. Tetra-amelia is a rare genetic disorder characterized by complete absence of all four limbs. Recently, it has been demonstrated that a [▶nonsense mutation](#) in the *Wnt3* gene causes the disorder. The Q83X mutation in the NH2-terminus of *Wnt3* creates a stop codon resulting in a truncated protein. Hence, the

loss-of-function in both copies of *Wnt3* is the mechanism for the disease.

The Holt-Oram syndrome is an autosomal dominant disorder characterized by forelimb abnormalities and congenital heart defects. Mutations in the *Tbx5* gene have been shown to be responsible for this syndrome. *Tbx5* encodes a T-box containing transcription factor that is expressed in embryonic forelimbs prior to induction of the AER. Frameshift, splicing and nonsense mutations that produce truncated TBX5, or an in-frame deletion of 27 nucleotides in the T-box that reduces the DNA binding ability of TBX5 have been reported in humans. Moreover, mutated TBX5 protein has been observed predominantly in the cytoplasm, indicating that impaired nuclear localization and dysfunction in DNA binding and transcriptional activity are the mechanisms for the loss-of-function mutation in these patients. Consistently, genetic studies in mice have demonstrated that a deletion of *Tbx5* results in a complete absence of limb bud outgrowth. It has been shown that TBX5 initiates limb bud outgrowth by activation of *Fgf10* in the early forelimb mesenchyme.

Nail patella syndrome (NPS) is an autosomal dominant disorder, characterized by a small or absent patella, malformations of the radial head, iliac horns and an absence or hypoplasia of nails on the fingers and toes. Many [▶missense mutations](#) in the homeodomain of LMX1b that abolish DNA-binding activity of the protein or in the LIM-domain of the protein that affect essential co-factor interactions have been identified in individuals with NPS.

Mutations in genes, which are responsible for patterning of the autopod cause abnormal development of digits, polydactyly (extra digits), syndactyly (fusion of digits) and brachydactyly (shortening of digits). Mutations in *Fgfr2* are associated with human disorders of limb patterning, such as Pfeiffer syndrome and Apert syndrome. Mutations of the splice sites surrounding the alternatively spliced exon IIIc of *Fgfr2* cause Pfeiffer syndrome, which is characterized by the broadening of thumbs and halluces, and radio-humeral or -ulnar synostosis. FGFR2c functions to transmit FGF8 signals from the AER to limb mesenchyme. Exon IIIc splicing mutations lead to ectopic expression of FGFR2b, a receptor isoform for FGF10, in limb mesenchyme. Moreover, mutations in the IgIIIa or IgIIIc domains of FGFR2c either substitute one of the paired cysteine residues (at C278 and C342) or create extra cysteines (W290C and S351C). The production of an odd number of cysteine residues enables mutant FGFR2 molecules to form covalently cross-linked dimers, resulting in constitutive activation. In Apert syndrome, there is a complex bony syndactyly involving the central three digits, radio-humeral or

-ulnar synoptosis. Mutation P253R in the extracellular domain of the FGFR2c leads to loss of ligand selectivity. The P253R mutant aberrantly binds FGF10 resulting in ectopic FGF10 signaling in limb mesenchyme.

Missense mutations in the p63 gene account for most cases of ectrodactyly-ectodermal dysplasia-clefting (EEC) syndrome. Arginine codons were frequently found to be mutated in the DNA binding domain of p63. Since the deletion of p63 is not associated with any of the signs of EEC syndrome, it has been suggested that a gain-of-function in p63 causes the limb defects.

Greig cephalopolysyndactyly (GCPS) is a rare autosomal dominant disorder that affects limb and craniofacial development. In the hands both preaxial and postaxial polydactyly and in the feet duplication of the hallux combined with syndactyly of toes 1–3 can occur. Haploinsufficiency of the Gli3 gene is the molecular mechanism that underlies GCPS. GLI3 is a dual-function transcription factor, which is crucial for the formation of the anterior-posterior axis of the limb.

Synpolydactyly (SPD) is a dominantly inherited disorder characterized by 3/4 finger and 4/5 toe syndactyly and urogenital tract abnormalities. The disease is caused by an expansion of CGC trinucleotide repeats (greater than 15 repeats) in the first exon of the Hoxa13 or Hoxd13 genes. The CGC repeats are translated into polyalanine sequence in the NH2-terminus of HOX proteins. The functions of the domain containing the polyalanine tract are unclear, although it may be involved in interactions with other proteins. Target inactivation of Hoxd13 in mice results in the loss of several skeletal elements of autopods. However, Hoxd13 loss-of-function mutation produces milder phenotypes than does synpolydactyly homolog (Hoxd13spdh) mutation, suggesting that mutant HOXD13 interferes with the activity of other proteins.

References

1. Tickle C (2003) Patterning systems-from one end of the limb to the other. *Dev Cell* 4:449–458
2. Niswander L (2003) Pattern formation: old models out on a limb. *Nat Rev Genet* 4:133–143
3. Capdevila J, Izpisua Belmonte JC (2001) Patterning mechanisms controlling vertebrate limb development. *Annu Rev Cell Dev Biol* 17:87–132
4. Huelsken J, Birchmeier W (2001) New aspects of Wnt signaling pathways in higher vertebrates. *Curr Opin Genet Dev* 11:547–553
5. Soshnikova N, Zechner D, Huelsken J et al (2003) Genetic interaction between Wnt/beta-catenin and BMP receptor signaling during formation of the AER and the dorsal-ventral axis in the limb. *Genes Dev* 17:1963–1968
6. Tickle C (2002) Molecular basis of vertebrate limb patterning. *Am J Med Genet* 112:250–255

Limb Girdle Muscular Dystrophies

ISABELLE RICHARD

Généthon CNRS, Evry, France

richard@genethon.fr

Definition

Limb girdle muscular dystrophies (LGMD) are inherited degenerative disorders of the skeletal muscle (1, 2, 3, 4). As their name indicates, they are characterized by progressive atrophy and weakness of pelvic and scapular girdle muscles. Other clinical features include an elevation of the serum creatine kinase level and a dystrophic pattern on muscle biopsies. At the histological level, this pattern typically includes heterogeneity in fiber size, areas of muscle necrosis and regeneration, fiber splitting and ultimately increased amounts of fat and connective tissue.

Despite their common name, there is a marked clinical and genetic heterogeneity within this group of muscle diseases. Following the genetic definitions generated by the application of molecular strategies, a classification was established in which dominant and recessive forms are designated LGMD1 and LGMD2, respectively. The different subtypes are distinguished by addition of capital letters in the order of identification. To date, 7 different subtypes have been recognized for the dominant forms (LGMD1A to 1G) and ten for the recessive forms (LGMD2A to 2J). The genes responsible for 3 dominant and all recessive forms have been identified. The corresponding proteins belong to different groups of proteins with respect to their subcellular localization (nucleus, ►cytoplasm, ►cytoskeleton and ►sarcolemma) and to their functions (structural, enzymatic, etc.).

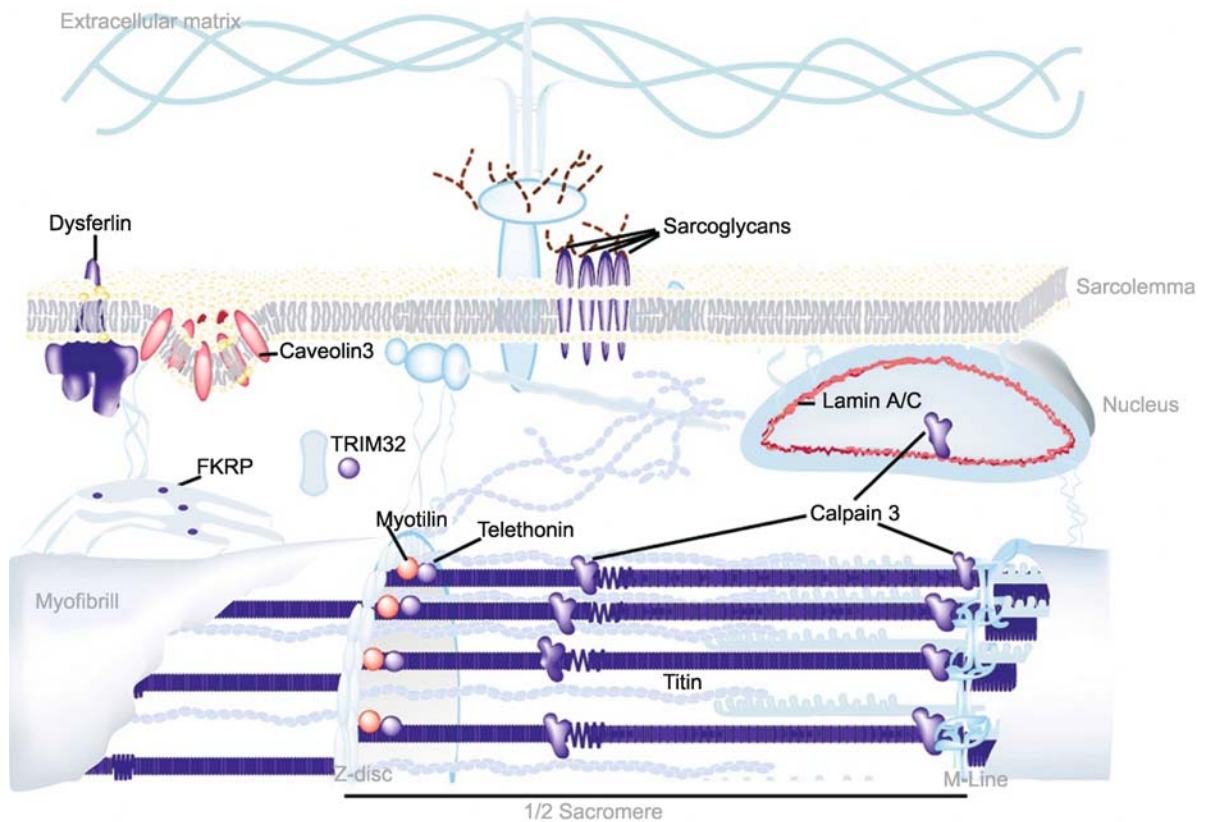
Characteristics

►Autosomal Dominant (Inherited Disorder) LGMD

The dominant forms usually show a mild course and are often associated with other clinical features. Only a few families have been described for each form.

LGMD1A: Genetic Localization: 5q31; Causative Protein: Myotilin (TTID)

LGMD1A is characterized by an adult-onset proximal weakness, later progressing to distal extremities and is often associated with a dysarthric pattern of speech. Myotilin is a structural protein located in the ►Z-disc of the ►sarcomere in cardiac and skeletal muscles. By its association with other proteins that include α -actinin, γ -filamin and F-actin, it may participate in Z-disc organization. Histologically, the mutations



Limb Girdle Muscular Dystrophies. Figure 1 Schematic representation of the muscle cell and localization of the proteins involved in LGMDs. The proteins involved in dominant and recessive LGMD are in red and purple, respectively. The main proteins of the muscle are depicted in blue.

induce a large number of autophagic vesicles and disruption of the Z-disc.

**LGMD1B: Genetic Localization: 1q11-21;
Causative Protein: Lamin A/C (LMNA)**

LGMD1B is characterized by muscle weakness starting in the proximal lower-limb muscles and preceding the appearance of life threatening cardiac dysrhythmias. Lamins A/C are components of the nuclear lamina, a network of fibrous proteins located in the internal face of the internal nuclear membrane. A role for lamin has been suggested in DNA replication, chromatin organization, spatial organization of nuclear pore complexes and anchorage of nuclear envelope proteins. It is interesting to note that mutations in LMNA can lead to a number of other tissue-specific disease phenotypes.

LGMD1C: Genetic Localization: 3p25; Causative Protein: Caveolin 3 (CAV3)

LGMD1C presents as a mild to moderate proximal muscle weakness and calf hypertrophy, starting in childhood around 5 years of age. Caveolin 3 is a muscle-specific member of the caveolin family and is

the principal structural protein of caveolae in striated muscles. Caveolae are vesicular invaginations of the plasma membrane participating in vesicular trafficking events and signal transduction processes. It has been shown that LGMD1C mutants of caveolin 3 form unstable high molecular aggregates that are retained within the ►Golgi complex instead of being targeted to the plasma membrane.

Other LGMD1:

Four other genetic loci have been identified for dominant LGMD although the genes involved have not yet been identified. LGMD1D was mapped to chromosome 7q in two families. It is characterized by proximal weakness starting in adulthood in the lower limbs. LGMD1E was mapped to chromosome 6q23 and is associated with dilated cardiomyopathy with conduction defect (CMD1F, OMIM602067). LGMD2F was mapped to chromosome 7q32.1-q32.2 in a Spanish pedigree. Occurrence of genetic anticipation was proposed by the authors as severity appears to worsen between successive generations. LGMD1G was mapped to chromosome 4p21 in a Brazilian family.

It has a late-onset and is associated with progressive fingers and toe flexion limitation.

►Autosomal recessive LGMDS

Autosomal recessive limb-girdle muscular dystrophies have a wide spectrum of clinical severity. In the milder forms of LGMD, onset may be in the first, second or even third decade; loss of ambulation usually occurs after the third decade. In the most severe forms, onset usually occurs between 3 and 5 years and progression is very rapid with loss of ambulation usually before the third decade. There is significant variability of clinical features in unrelated patients and within families

LGMD2A: Genetic Localization: 15q15.1-15.3; Causative Protein: Calpain 3 (CAPN3)

LGMD2A is also referred as calpainopathy. It is characterized by a symmetric involvement of the proximal muscles, especially those of the posterior compartment of the lower limbs. LGMD2A has a worldwide distribution and represents a large proportion of the LGMD2s. Calpain 3 is the muscle specific member of the family of calpains, ►cysteine proteases whose activities are calcium-dependent. In humans, the calpain 3 protein can be localized either in the cytoplasm where it binds to titin, a giant elastic protein of the sarcomere, or in the nucleus. The presence of mutations seems to impair mainly the proteolytic activity and is associated with a perturbation of the apoptotic-regulating pathway of NF- κ B/I κ B α .

LGMD2B: Genetic Localization: 2p13; Causative Protein: Dysferlin (DYSF)

LGMD2B is also referred as dysferlinopathy. It is characterized by a very high serum creatine kinase level and prominent inflammatory changes. It is allelic to a distal phenotype, Miyoshi myopathy, and, like LGMD2A, is a frequent LGMD2. Dysferlin is a 230 kD protein localizing to the muscle fiber membrane and belonging to the ferlin family. The dysferlin mutations induce its intracellular accumulation, in the form of nuclear staining or abundant cytoplasmic vesicles, associated with a reduction in its sarcolemmal expression. It was shown that dysferlin is involved in membrane repair.

LGMD2C, D, E, F: Genetic Localization: 3q12, 17q21-q21.33, 4q12, 5q33-34; Causative Proteins: γ -, α -, β -, δ -Sarcoglycans (SGCG, SGCA, SGCB and SGCD)

This group of diseases is referred as sarcoglycanopathies and includes the most severe forms of LGMD2s. Alpha-sarcoglycanopathy is the only sarcoglycanopathy with a universal distribution. The sarcoglycans form a subcomplex in the ►dystrophin glycoprotein complex (DGC) linking the subsarcolemmal cytoskeletal proteins to the

extracellular matrix (ECM) via β -dystroglycan and dystrophin. In each of the sarcoglycanopathies, most of the mutations destabilize the whole sarcoglycan complex at the plasma membrane.

LGMD2G: Genetic Localization: 17q11-q12; Causative Protein: Telethonin (T-CAP)

LGMD2G is a mild myopathy with some involvement of the distal muscles. Rimmed vacuoles were observed in the muscle biopsies. Telethonin is a sarcomeric protein of 19 kD found exclusively in striated and cardiac muscle and localized to the Z-disk. Telethonin is phosphorylated by the serine kinase domain of titin in early differentiating myocytes. This event seems to be involved in the reorganization of the cytoskeleton during ►myofibrillogenesis. Through binding, telethonin inhibits the secretion of myostatin, an inhibitor of skeletal muscle growth.

LGMD2H: Genetic Localization: 9q31-33; Causative Protein: Tripartite-motif-containing Gene-32 (TRIM32)

This mild muscular dystrophy was only found in Hutterite population in North America. TRIM32 is a member of the growing family of tripartite motif proteins. Members of this family may form, through homodimerization, high molecular weight complexes that define subcellular compartments. The tripartite motif includes a RING-finger domain that mediates transfer of ►ubiquitin to substrates, suggesting that TRIM 32 is a ►E3 ubiquitin ligase.

LGMD2I: Genetic Localization: 19q13.3; Causative Protein: Fukutin-related protein (FKRP)

LGMD2I is a mild to severe LGMD2 presenting with calf hypertrophy. FKRP is ubiquitously expressed with the highest level in skeletal muscle and heart. It was identified through a homology search to fukutin, the protein involved in ►Fukuyama myopathy. These two proteins are Golgi-resident ►glycosyl transferases that seem to cooperate in the transfer of sugars to α -dystroglycan. Alteration of ►glycosylation of α -dystroglycan reduces its ability to bind to ECM ligands such as laminin- α 2 and agrin, disrupting linkage of the internal cytoskeleton to the ECM.

LGMD2J: Genetic Localization: 2q24.3; Causative Protein: Titin (TTN)

Homozygous mutations in Mex6, the last exon of titin, are responsible for LGMD2J. The same mutation in a heterozygous state causes tibial muscular dystrophy, a late-onset autosomal myopathy affecting mostly the anterior compartment of the lower limb. Titin is a giant elastic protein spanning half a sarcomere. LGMD2J is associated with a secondary deficiency in calpain 3 that has a binding site in the vicinity of the mutation.

Limb Girdle Muscular Dystrophies. Table 1 Chromosomal localization and causative genes of the different LGMDs

	Chromosomal localization	Causative gene	Locus	Size protein (kD)	Site of main expression	Subcellular localization
LGMD1A	5q22.3-q31.3	Myotilin	TTID	57	Sk. and cardiac muscles	Z-disc
LGMD1B	1q11-q21	Lamin A/C	LMNA	80 and 65	Ubiquitous	Nuclear lamina
LGMD1C	3p257q	Caveolin-3	CAV3	17-20	Sk. and cardiac muscles	Sarcolemma
LGMD1D	7q	-	-	-	-	-
LGMD1E	6q23	-	-	-	-	-
LGMD1F	7q32.1-q32.2	-	-	-	-	-
LGMD1G	4p21	-	-	-	-	-
LGMD2A	15q15.1-q15.3	Calpain 3	CAPN3	94	Sk muscles	Sarcomere, nucleus
LGMD2B	2p13-p16	Dysferlin	DYSF	230	Sk. and cardiac muscles, placenta	Sarcolemma
LGMD2C	13q12	γ -Sarcoglycan	SGCG	35	Striated muscles	Sarcolemma
LGMD2D	17q21-q21.33	α -Sarcoglycan	SGCA	50	Striated muscles	Sarcolemma
LGMD2E	4q12	β -Sarcoglycan	SGCB	43	Sk., cardiac and smooth muscles	Sarcolemma
LGMD2F	5q33-34	δ -Sarcoglycan	SGCD	35	Ubiquitous	Sarcolemma
LGMD2G	17q11-q12	Telethonin	T-CAP	19	Sk. and cardiac muscles	Z-disc
LGMD2H	9q31-q33	Tripartite-motif protein 32	TRIM32	45-50	Sk and cardiac muscles, testis	Cytoplasm
LGMD2I	19q13.3	Fukutin-related protein	FKRP	55	Ubiquitous	Golgi apparatus
LGMD2J	2q24.3	Titin	TTN	3 000	Sk. and cardiac muscles	Sarcomere

Sk: skeletal

Cellular and Molecular Regulation

Muscular dystrophies could be seen as a consequence of an imbalance between muscle cell death and survival. Considering the nature of the LGMD proteins, it is clear that such an imbalance can be the consequence of a diversity of pathological mechanisms, highlighting some aspects that are of utmost importance for skeletal muscle ►[homeostasis](#).

Skeletal muscle endures important mechanical constraints during muscular contractions. The structural characteristics of its complex cytoskeleton and the linkage of the DGC with the ECM confer on the muscle the necessary load-bearing properties to support these stresses. In addition to providing structural integrity,

there is increasing evidence that many of the cytoskeletal proteins also support cellular signaling processes important for cell survival. The LGMDs, that can be put in the category where these functions are impaired, are: LGMD2I where dysglycosylation of α -dystroglycan disrupts the sarcolemma-ECM linkage, the sarcoglycanopathies where the absence of sarcoglycans lead to alterations in membrane permeability and possibly of ►[signal transduction](#), LGMD1A where myotilin mutations induce a disruption of the Z-disc and LGMD1B in which deficiency in laminA/C makes the nuclear envelope fragile. Another mechanism of protection, in which dysferlin plays a role and that is therefore impaired in LGMD2B, is the capacity to

reseal membrane disruptions generated during mechanical injury.

Another characteristic of the skeletal muscle is its important plasticity. By constant alterations in synthesis, degradation and assembly of metabolic, structural and regulatory proteins, it is capable of responding to changes in cellular physiology such as modification of contractile activity, nutrient supply or environmental factors. Many of the LGMDs may be linked to defects in those processes, especially for those that are supported by cytoskeletal networks. In LGMD1C, it can be hypothesized that signaling pathways supported by the caveolae are impaired. There is also a possible perturbation in the regulation of ►glycolysis, as phosphofructokinase, an important enzyme of this pathway, is one of the major caveolin binding proteins. Calpain 3 may participate in linking sarcomeric to nuclear function and may prevent cell death by controlling the ►apoptosis regulatory pathway of NF- κ B/I κ B α in response to mechanical stress. Telethonin deficiency may be related to an impairment of muscle growth with respect to its inhibitory role in myostatin secretion. Mutations in TRIM 32 may result in defects in protein turnover by preventing the targeting of one or a group of proteins to the ►proteasome. Titin mutations in LGMD2J are not associated with disruption of the sarcomere and may more probably be associated with a signaling pathway perturbation, possibly in relationship with calpain 3.

Independently of the pathological mechanism, mutations in the increasingly diverse group of the LGMD proteins lead to an overall similar outcome, which is the predominance of proximal muscle weakness. Much work remains to be done to fully understand the biochemical basis of susceptibility to cell death, especially with respect to the unsolved issue of the muscle specificity of these disorders.

Clinical Relevance

Even though the molecular discoveries led to a better definition of each disorder, each genetically distinct entity may present heterogeneity of clinical signs resulting in an overlap in presentation and making them often difficult to distinguish clinically. Therefore, differential diagnosis should rely on molecular tools and especially on protein analysis by western blotting or immunohistochemistry with specific antibodies (5). However, it is complicated by the fact that a genetic alteration could lead to a secondary defect in another protein.

None of the LGMDs are effectively treatable to date. However, there is hope that advances in knowledge of disease mechanisms and in solving the issues involved by ►gene-based therapies will lead to applications in treating these muscular dystrophies. A number of viral and non-viral vector systems have been evaluated for

muscle-directed gene transfer in several dystrophic mouse models bringing proof-of-principle for feasibility and efficacy (6). Currently, ►adeno-associated virus vectors may be the best choice with regard to efficiency and longevity of gene transfer in muscle tissues. However, the gene transfer approach for muscular dystrophy presents significant challenges, especially regarding gene delivery into such large amounts of tissue as muscles, which represent 40 % of the body mass. In parallel to gene-based therapy, a better understanding of causative protein functions and of downstream pathological consequences of the genetic defect could facilitate the identification of novel drug targets. In addition, systematic drug screening could help to identify compounds that would improve the condition of the patients.

References

1. Brown SC, Muntoni F, Sewry CA (2001) Non-sarcolemmal muscular dystrophies. *Brain Pathol* 11:193–205
2. Ozawa E, Nishino I, Nonaka I (2001) Sarcolemmopathy: muscular dystrophies with cell membrane defects. *Brain Pathol* 11:218–30
3. Piccolo F, Moore SA, Mathews KD et al (2002) Limb-girdle muscular dystrophies. *Adv Neurol* 88:273–91
4. Mathews KD, Moore SA (2003) Limb-girdle muscular dystrophy. *Curr Neurol Neurosci Rep* 3:78–85
5. Bushby KM (1999) The limb-girdle muscular dystrophies: diagnostic guidelines. *Eur J Paediatr Neurol* 3:53–58
6. Gregorevic P, Chamberlain JS (2003) Gene therapy for muscular dystrophy – a review of promising progress. *Expert Opin Biol Ther* 3:803–14

Limbic System

Definition

The limbic system is a heterogeneous neuronal system that is involved in the generation of feelings, emotions and motivations, and is intimately involved in learning and memory consolidation related to drug effects. Functional imaging techniques in humans demonstrate that craving for drugs of abuse involves areas (e.g. limbic system) predicted from animal experiments.

►Addiction, Molecular Biology

LIMS

Definition

LIMS (Laboratory Information Management System) refer to the output from high throughput analyses of the proteome with clinical information, protein structure,

sample storage and stock, and the status of various items for data production.

► [Protein Databases](#)

Lin

Definition

C. Elegans mutant in the cell lineage

► [C. Elegans as a Model Organism for Functional Genomics](#)

LINE

► [Long Interspersed Repeat](#)

Lineage Determination

Definition

Lineage determination designates the restriction of developmental options of cells to one or few differentiation pathways.

► [Muscle Development](#)

Linear Unmixing

Definition

If several fluorochromes are used simultaneously to record images in fluorescence microscopy, the resulting image is the sum of the contributions of the single fluorochromes and a separation of those contributions is desirable. Linear unmixing refers to a mathematical algorithm used to separate the signals from different fluorochromes at each pixel within an image. It is based on a matrix where the relative contributions of each of the fluorochromes into each detection channel are first recorded as reference spectra, and subsequently applied to the spectral image. This operation deconvolves the weighting coefficients of the component spectra from the measured sum spectra, and thus reveals each individual contribution.

► [Fluorescence Microscopy: Spectral Imaging vs. Filter-Based Imaging](#)

Linkage

Definition

Linkage means the tendency of (series of) polymorphisms at specific loci to be inherited together. If two genes (genetic loci) are in close proximity on the same chromosome they are said to be linked. Genetic linkage between a marker and a disease gene implies that alleles at the marker locus co-segregate with the disease allele within families, but that different marker alleles may co-segregate with the disease allele in different families. Under linkage, the overall frequencies of the marker alleles calculated from population-based samples need not vary between affected and control groups.

► [Chromosome 21, Disorders](#)

► [Cleft Lip Palate](#)

► [COPD and Asthma, Genetics](#)

► [Diabetes Mellitus, Genetics](#)

► [Genetic Epidemiology](#)

► [Genetic Predisposition to Multiple Sclerosis](#)

► [Mendelian Forms of Human Hypertension and Mechanisms of Disease](#)

Linkage Analysis

Definition

Linkage analysis tests, by use of statistical methods, for cosegregation of a marker and disease phenotype within a pedigree, to determine whether a genetic marker and a disease-predisposing gene are physically linked, that is, in close physical proximity to each other (see ► [Linkage Map](#)).

► [Cleft Lip Palate](#)

► [Common \(Multifactorial\) Diseases](#)

► [COPD and Asthma Genetics](#)

Linkage Disequilibrium

Definition

Linkage disequilibrium (LD) is the result of non-random association of alleles at the population level of variants at the marker and disease genes. This usually implies close physical linkage of the marker and disease gene, reflecting the historical origin of a mutation on a specific chromosome with a characteristic set of variations. Evidence for linkage disequilibrium can be

helpful in mapping disease genes, since it suggests that the two may be very close to one another.

- ▶ Atopy Genetics
- ▶ Common (Multifactorial) Diseases
- ▶ Genetic Epidemiology
- ▶ Genetic Predisposition to Multiple Sclerosis
- ▶ Manic Depression
- ▶ Mendelian Forms of Human Hypertension and Mechanisms of Disease

Linkage Map

Definition

Linkage map defines the relative positions of genetic loci on a chromosome, obtained by ▶ [linkage analysis](#). This procedure, based on meiotic recombination frequencies, determines how often the loci are inherited together. Distance is measured in centimorgans (cM).

- ▶ Human Genome Project Information
- ▶ Large-Scale ENU Mutagenesis in Mice
- ▶ Mendelian Forms of Human Hypertension and Mechanisms of Disease
- ▶ Mutagenesis Approaches in Medaka
- ▶ SNP Detection and Mass Spectrometry

Linker DNA

Definition

Linker DNA denotes the DNA tract that connects adjacent nucleosomes in the chromatin fiber.

- ▶ Nucleosomes

Lipid Asymmetry

Definition

Lipid asymmetry defines the asymmetric distribution of lipids between the two layers of a biological membrane, which causes differences in the physicochemical and functional properties between both layers.

- ▶ Biological Membranes

Lipid Peroxidation

Definition

Lipid peroxidation is a term commonly used for free radical-driven oxidative destruction of unsaturated fatty acid residues in complex lipids that ultimately leads to biomembrane disruption. Lipid peroxidation can also be initiated by lipoxygenases. Such enzymatic lipid peroxidation by 15-lipoxygenase, has been implicated in, e.g. the elimination of mitochondria during erythrocyte maturation. If catalyzed by 5-lipoxygenase it represents the initial step of leukotriene biosynthesis.

- ▶ Free Radicals

Lipid Rafts

Definition

Lipid rafts are membrane microdomains enriched in lipids such as sphingolipids and cholesterol. These “rafts” or microdomains have been implicated in cell signaling and in apical transport of plasma membrane proteins in polarized cells.

- ▶ Cell Polarity
- ▶ Epithelial Cells

Lipidophilicity

Definition

Lipidophilicity describes the affinity of a molecule or a moiety for a lipophilic environment, commonly measured by its distribution behaviour in a biphasic system.

- ▶ Hydrophobicity
- ▶ QSAR

Lipopolysaccharide

- ▶ LPS

Lipoprotein

Definition

Lipoprotein refers to a complex of lipid and protein. It is the way lipids travel in the blood.

► [Familial Hypercholesterolemia](#)

Liposomes

Definition

Liposomes are spherical particles that consist of an aqueous compartment enclosed by a lipid bilayer.

► [Surface Plasmon Resonance](#)

Liquid Chromatography

Definition

Chromatography basically involves the separation due to differences in the equilibrium distribution of sample components between two different phases. One of these phases is a mobile phase and the other is a stationary phase. In *liquid chromatography* a sample mixture is passed through a column packed with solid particles which may or may not be coated with another liquid. With the proper solvents, packing conditions, some components in the sample will travel the column more slowly than others resulting in the desired separation.

► [Proteomics in Cancer](#)

► [Protein-Protein Interaction](#)

► [Two Hybrid System](#)

Liquid Phase Photo-Polymerization

Definition

In Liquid Phase Photo-Polymerization (LP3), 3D structures are created from a mixture (liquid state) of monomer, cross-linker and photo-initiator, using light to initiate polymerization.

► [Proteomics in Microfluidic Systems](#)

LOAD

Late-onset Alzheimer disease (AD).

► [Alzheimer's Disease](#)

Loader

Definition

Loader refers to an enzyme (e.g. replication factor C) that can load another protein (e.g. proliferating cell nuclear antigen) onto the place of action (DNA). The loader itself is an ATPase machine and does not stay at the place of action, where the loaded protein was brought.

► [Replication Fork](#)

Localization Precision

Definition

The particle localization precision is defined as that distance from the found particle position, in which the true particle position is located with a probability of 0.68. The localization precision depends strongly on the signal-to-noise ratio of the particle in question.

► [Fluorescence Microscopy: Single Particle Tracking](#)

Locus

Definition

A unique chromosomal location defining the position of an individual gene or DNA sequence on a chromosome.

► [Common \(Multifactorial\) Diseases](#)

► [Diabetes Mellitus, Genetics](#)

► [Mouse Genomics](#)

Locus Control Region

Definition

The Locus Control Region (LCR) is a regulatory DNA segment, originally identified as an extended cluster of DNase I hypersensitive (HS) regions in the beta globin locus. An LCR is typically extended over several kilobases and contains several enhancer-active subsegments of 100-300 bp each. The LCR may be considered an umbrella term that encompasses all enhancers required for the correct developmental expression pattern of one or more genes under its influence. It also may direct the position-independent and copy number-dependent expression of transgenes.

- ▶ [Enhancer](#)
- ▶ [Transgene Silencing](#)

Locus Heterogeneity

Definition

Locus heterogeneity refers to mutations in different genes at different chromosomal loci that cause an identical phenotype.

- ▶ [Chromosomal Instability Syndromes](#)

Lod Score

Definition

Lod score stands for (decadic) logarithm of the odds ratio. In genetic linkage analysis, the lod score quantifies the empirical evidence in favour of linkage between two loci of interest.

- ▶ [Genetic Epidemiology](#)

LOH

- ▶ [Loss of Heterozygosity](#)

Long (Range) PCR

Definition

Long-Range PCR refers to a polymerase chain reaction (PCR) optimized to facilitate the amplification, detection

and synthesis of long (~2 to ~35 kb) DNA fragments. The DNA produced is used to construct vectors.

- ▶ [Large-Scale Homologous Recombination Approaches in Mice](#)
- ▶ [Mitochondrial Myopathies](#)

Long Interspersed Repeat

Definition

Long interspersed repeat (LINE) refers to any non-LTR (long terminal repeat) retrotransposon that encodes reverse transcriptase and endonuclease, which catalyzes nuclear reverse transcription and integration of cDNA in the host genome.

- ▶ [Repetitive DNA](#)

Long Term Depression

Definition

Long term depression (LTD) is a synaptic plasticity phenomenon that corresponds to the decrease in the synaptic strength (decrease in the post-synaptic response observed for the same stimulation of the presynaptic terminals) observed after a specific stimulation of the afferent (input) pathway. This decrease is still observed hours after its induction.

Long Term Potentiation

Definition

Long term potentiation (LTP) is a synaptic plasticity phenomenon that corresponds to the increase of synaptic strength (increase in the post-synaptic response observed for the same stimulation of the presynaptic terminals), observed after a high-frequency stimulation (tetanus) of nerve fibers. The increased response in synaptic sensitivity is long lasting, and is still observed hours or even days after the tetanus. The phenomenon is often observed at glutamatergic synapses and involves, in most cases, the activation of the N-methyl-D-aspartate (NMDA) subtype of ionotropic glutamate receptors.

- ▶ [Fragile X Syndrome](#)
- ▶ [Ion Channels](#)
- ▶ [Neutrophilic Factors](#)

Long Terminal Repeats

Definition

Long terminal repeats (LTRs) define, in retroviruses, the DNA structure at the ends of the provirus created by copying the U3, R, and U5 regions of the RNA genome twice.

► [Retroviruses](#)

The so-called loss of heterozygosity assay is used to detect the loss of a polymorphism at a specific gene locus, such as occurs when a second hit inactivates the remaining normal gene in a cell already carrying a germline mutation for that gene.

► [Autosomal Dominant \(Inherited Disorder\)](#)

► [Double-Strand Break Repair](#)

► [Genomic Information and Cancer;](#)

► [Peutz-Jeghers Syndrome](#)

► [Polycystic Kidney Disease, Autosomal Dominant](#)

► [Repetitive DNA](#)

Loss of Function Mutations

Definition

In 'loss of function' mutations, the mutation results in a protein change that abolishes the function of the translation product.

► [Huntington's Disease](#)

► [Mouse Genomics](#)

► [Repeat Expansion Diseases](#)

Loss-of-Function Screen

Definition

Mutational screen based on the generation of mutations that either reduce or eliminate gene functions. Loss-of-function mutations may be induced by chemical mutagenesis, resulting in point-mutations or small deletions in a gene, or by integration of P elements.

► [Drosophila as a Model Organism for Functional Genomics](#)

Loss of Heterogeneity

Definition

Loss of heterogeneity is a molecular finding in which a patient's (tumor) sample retains only one allele of a genetic marker, while the germline cells of the patient have two alleles. This is a common mechanism for the inactivation of a tumor suppressor gene, thus favouring the development of cancer.

► [Loss of Heterozygosity](#)

► [Tuberous Sclerosis](#)

Loss-of-Function Studies

Definition

Loss-of-function studies are aimed at achieving insight into protein function by acting at the genomic or transcriptional level. Loss-of-function mutations may be induced by chemical mutagenesis resulting in point-mutations, or small deletions in a gene, or by integration of ► [P-elements](#). Analysis of the resulting phenotype affecting the organism provides evidence for the function of the gene product of interest.

► [Drosophila as a Model Organism for Functional Genomics](#)

► [High-Throughput Approaches to the Analysis of Gene Function in Mammalian Cells](#)

Loss of Heterozygosity

Definition

Loss of heterozygosity (LOH) designates the loss of one allele in somatic cells of an individual, which is heterozygous for this genetic locus in the germ line. Multiple mechanisms can lead to homozygosity of the defective allele. The two-hit-hypothesis of Knudson explains the formation of tumors by an inactivating germ line mutation in one allele, combined with the acquired loss of the functional second allele in somatic cell populations.

LOV Domain

Definition

LOV domain is the acronym for light, oxygen, voltage protein domains which constitute a family of blue light

receptors. It is a subset of the Per-ARNT-Sim (PAS) superfamily.

► [Photoreceptors](#)

Lov-1

Definition

Lov-1 is the closest ► *C. elegans* homologue of polycystic kidney disease gene 1 (PKD1) in humans.

► [Autosomal Dominant \(Inherited Disorder\)](#)

► [Polycystic Kidney Disease, Autosomal Dominant](#)

Low Copy Repeats

Definition

Low copy repeats (LCRs) are highly similar, large blocks of duplicated ► [pseudogenes](#) or repetitive DNA located within a specific chromosomal region. LCRs can give rise to chromosomal rearrangements (micro-deletions, microinversions).

► [Microdeletion Syndromes](#)

Low Density Lipoprotein

Definition

LDL is a lipoprotein (comprised of a lipid and a protein part) that acts as a carrier for cholesterol and fats in the blood. High blood levels of LDL increase the risk of arteriosclerosis.

► [Proteomics in Cardiovascular Disease](#)

Low Density Lipoprotein Receptor Related Proteins 5 and 6

Definition

Low density lipoprotein receptor related proteins 5 and 6 (LRP5/6) define large Type I transmembrane receptors which serve as co-receptors for Wnt. Mutation of the *Drosophila* LRP5/6 homologue, Arrow, resembles a Wnt null.

► [Wnt/Beta-Catenin Signaling Pathway](#)

Lower Motor Neurons

Definition

Lower motor neurons is a clinical term referring to motor neurons whose cell bodies are located in the anterior horn of the spinal cord, and whose axons project to make connections with muscle at the neuromuscular junction.

► [Hereditary Spastic Paraplegias](#)

LoxP and Cre Recombinase System

► [Cre/loxP strategies](#)

► [Large-Scale Homologous Recombination Approaches in Mice](#)

► [LOXP site](#)

LoxP Site

Definition

LoxP site is a 34 bp DNA element consisting of two 13 bp inverted repeats, separated by a spacer sequence of 8 bp. Two *loxP* sites are required by Cre to recombine DNA. The DNA flanked by *loxP* sites is termed 'floxed' DNA.

► [Cre/loxP Strategies](#)

► [Mouse Genomics](#)

LP3

► [Liquid Phase Photo-Polymerization](#)

LPS

Definition

The outer leaflet of the outer membrane of Gram-negative bacteria is composed primarily of lipopolysaccharides

(LPS). These molecules all contain an endotoxic lipid A, with which the LPS is anchored in the bacterial membrane, and a polysaccharide unit presented at the cell surface that is specific to the bacterial species. LPS are essential for the survival of Gram-negative bacteria, providing a permeation barrier for harmful substances. LPS induces an inflammatory response.

► [Inflammatory Response](#)

LPS Recognition Complex

Definition

LPS (lipopolysaccharide) is recognized by many cell surface molecules (CD55, Hsp70, Hsp90, GDFS, CD14, LBP, MD2, TLR4, CD11/18, CD16/32/64, CXCR4, CD36, CD81) on a single cell that forms a recognition complex to induce inflammation upon contact with this bacterial compound.

► [Inflammatory Response](#)

LRP5/6

► [Low Density Lipoprotein Receptor Related Proteins 5 and 6](#)

LR-PCR

► [Long \(Range\) PCR](#)

LTBP

Definition

► [Transforming-growth-factor \$\beta\$ \(TGF- \$\beta\$ \)](#) is secreted from cells as a latent complex. The complex is formed by LTBP (Latent TGF- β binding protein), which is disulfide-linked to the latency-associated propeptide (► [LAP](#)) of TGF- β . LTBP contributes to either latency

of the complex or anchoring of TGF- β in proper extracellular compartments.

► [Receptor Serine/Threonine Kinases](#)

LTD

► [Long Term Depression](#)

LTP

► [Long Term Potentiation](#)

LTRs

Definition

LTRs are long sequences of double stranded DNA, a sequence directly repeated at both ends of retroviral DNA. LTRs originate from reverse transcription of retroviruses and contain strong promoters as well as transcription termination signals.

► [Recombinant Protein Expression in Yeast](#)

► [Retroviruses](#)

L-Type Calcium Channel

Definition

L-type calcium channels are voltage-operated calcium channels with high electrical thresholds found in both excitable and nonexcitable tissue. It is essential for most of the early plateau current of the action potential, and thus for normal myocardial and vascular smooth muscle contractility. The channels are formed by five subunits (alpha-1, alpha-2, beta, gamma, and delta).

► [Cardiac Signaling: Cellular, Molecular and Clinical Aspects](#)

Lung

SP DE LANGHE, FG SALA, AA MAILLEUX,
S BELLUSCI

Developmental Biology Program, Department of
Surgery, USC Keck School of Medicine and the Saban
Research Institute of Childrens Hospital Los Angeles,
Los Angeles, CA, USA
sbellusci@chla.usc.edu

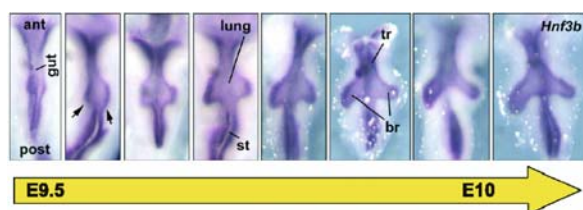
Definition

The mouse embryonic lung is an excellent example of a branched organ. The branching pattern in the lung is extremely reproducible from one embryo to another suggesting a tight temporal-spatial genetic control of the branching process. It is only during the last decade that the molecular interactions at the base of the branching process in this organ have begun to become elucidated. This review will focus on the current knowledge of the ►growth factors, produced either by the ►epithelium or the ►mesenchyme, playing a key role in controlling branch formation. For more information about lung development, we will address the reader to excellent reviews published recently (1, 2) and some key papers.

Early Lung Development and Branching

►Morphogenesis

The mouse lung arises at E9.5 (Fig. 1), from the ventral foregut, just anterior of the developing stomach. The primordia of the trachea and the two lung buds consist of an epithelial layer of ►endoderm surrounded by ►splanchnic (lateral plate) ►mesoderm. The trachea is separated from the esophagus by means of a longitudinal septation of the foregut, while the rest of the lung develops from two ventral buds that form at the posterior end of the trachea. Initially the primary buds



Lung. Figure 1 Lung branching morphogenesis illustrated by *Hnf3 β* expression. Embryonic lung development in mouse. Lung development starts at E9.5 with the emergence of two primary buds from the ventral foregut (only one is shown in the picture). At E10.5, the lung is composed of a trachea (tr) and two bronchi (br).

undergo repetitive outgrowth and branch laterally at invariant positions. They develop asymmetrically, so that in the mouse the right bud gives rise to four lobes, whereas the left bud gives rise to only one. The number of secondary buds on the left and right sides varies amongst species and is regulated by the pathways that control left-right asymmetry. As lung morphogenesis continues, dichotomous branching is observed at the tip of each duct, driven by reciprocal interactions between the distal mesenchyme and endoderm. The early lung branching pattern is stereotypic and invariant, indicating that it is probably genetically programmed.

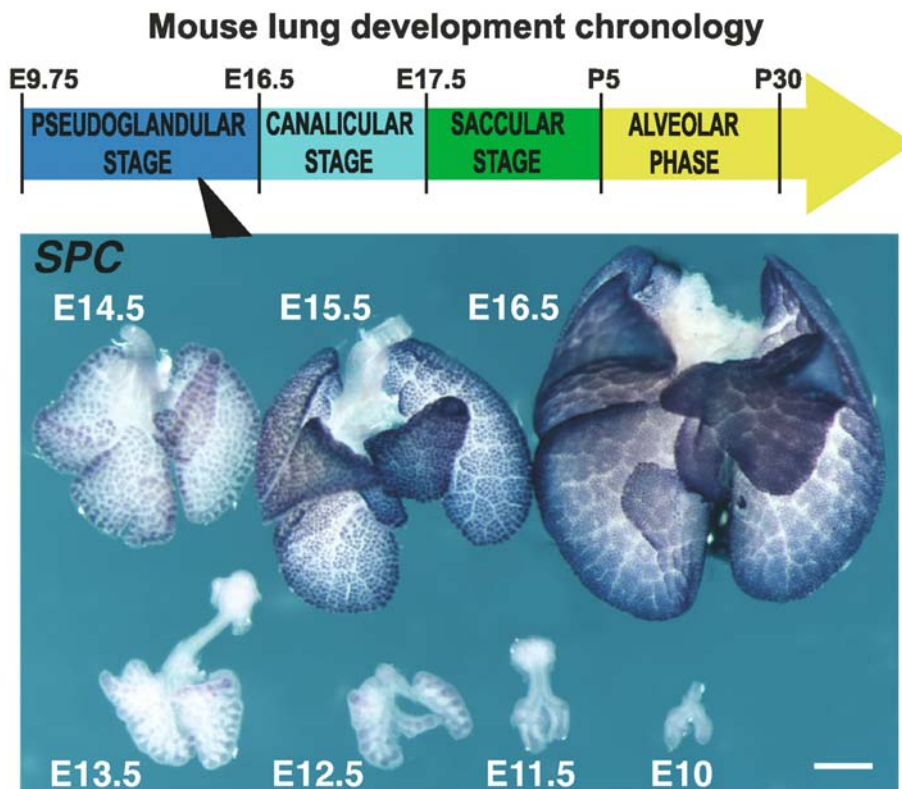
Characteristics

Histologically, lung development has been divided into four stages in mouse (Fig. 2): 1) pseudoglandular stage (E9.5–16.5), characterized by the development of the bronchial and respiratory tree and a relatively undifferentiated distal endoderm; 2) canalicular stage (E16.5–17.5), when the distal endoderm begins to form terminal sacs and vascularization begins; 3) sacular stage (E17.5–P5), characterized by thinning of the mesenchyme, an increase in the number of terminal sacs, vascularization and ►differentiation of the endoderm into type I and type II cells; 4) alveolar stage (P5–P30), when the terminal sacs develop into mature alveolar ducts and alveoli. Unlike in the mouse, the alveolarization of the human lung begins before birth. This review will focus mostly on the pseudoglandular stage.

Molecular Interactions

Fibroblast Growth Factor 10 (*Fgf10*) Expression is Associated with Directional Growth of the Epithelial Buds

Fgf10 (Fig. 3a), which is expressed focally in the embryonic lung distal mesoderm adjacent to stereotypically determined branching sites, has been shown to trigger chemotaxis and proliferation on the adjacent epithelium. This results in directional growth of the primary lung buds. Mutant embryos defective in signaling for either *Fgf10* or *Fgfr2IIIB* (encoding the principal and highest affinity FGF10 receptor) die at birth as a consequence of lung defects; they develop a trachea and primary bronchi, but no respiratory tree. *FGFR2IIIB* is expressed in the lung epithelium with a higher expression distally. *Fgf10*, a key mesenchymal signal in inducing epithelial branching during lung development translates the positional information in the mesenchyme into the early branching pattern. High levels of *Fgf10* expression are detected in the mesenchymal cells some distance from the epithelium at the focal sites adjacent to where secondary branches later emerge (3). Regulated expression of heparan sulfate (HS) at sites of budding, with distinct patterns of HS sulfation in the epithelial-associated ►basement



Lung. Figure 2 Different stages of lung development. Histologically, lung development is divided into four stages; pseudoglandular (E9.5–16.5), canalicular (E16.5–17.5), saccular (E17.5–P5) and alveolar stages (P5–P30). The development of the mouse bronchial and respiratory tree during the pseudoglandular stage is illustrated by *SP-C* expression.

membrane and mesenchyme, influence lung pattern formation by mechanisms that include regulation of FGF10 protein distribution and binding to its epithelial receptor. FGF10 also controls the differentiation of the epithelium by inducing ► *Surfactant protein C* (*SP-C*) expression and by up-regulating the expression of *Bmp4*, a known regulator of lung epithelial differentiation (4). FGF10 induces the expression of its own antagonist, *Sprouty2* (*Spry2*), in the distal epithelium. Reduction of *Spry2* expression in cultured lungs using ► antisense oligonucleotides resulted in an increase in lung branching and up-regulated expression of specific lung epithelial maturation/differentiation markers. Conversely, targeted over-expression of *Spry2* in peripheral lung epithelium results in a lower level of branching, smaller lungs with a particular “moth-eaten” dysplastic appearance along the edges of the lobes and a decrease in epithelial cell proliferation.

The lung is enveloped by a thin serous membrane, the mesothelium, composed of a monolayer of flat cells of mesodermal origin. The mesothelium expresses *Fgf9*, which acts as a proliferative growth factor on the underlying mesenchyme, mostly through the receptor FGFR2-IIIc.

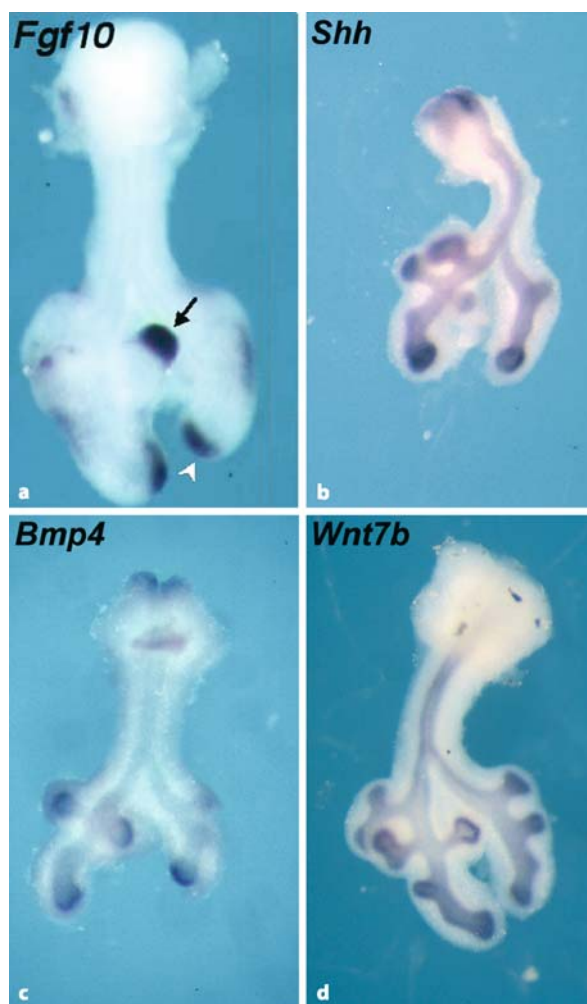
Sonic Hedgehog (SHH) Induces Mesenchymal Cell Proliferation

Sonic hedgehog (*Shh*) (Fig. 3b) is expressed throughout the endoderm with the highest levels at the most distal tips. By contrast, the gene encoding the SHH receptor (called *Patched* or *Ptc*), is expressed mostly in the adjacent mesoderm. In *Shh*^{-/-} embryos, the trachea and primary buds fail to separate from the esophagus. The primary buds also remain as small sacs and do not grow out or branch. *Fgf10* expression, which is highly spatially restricted in wild type, is up-regulated and widespread in the mesenchyme in contact with the epithelium of the *Shh* null mutant mouse lung. In addition, over-expression of *Shh* in the mouse lung using a transgenic approach leads to reduced *Fgf10* expression. Thus, *Shh* signaling to the mesenchyme negatively regulates *Fgf10* expression, and the resulting misregulation of *Fgf10* in *Shh* mutant lungs most probably accounts for the observed failure of secondary branching. Local suppression of *Shh* signaling by the induction of Hh-binding proteins such as *Patched* (*Ptc*) and *Hedgehog interacting protein* (*Hip*) in the mesenchyme at branch tips, may serve to facilitate FGF signaling locally (1).

Bone Morphogenetic Protein 4 (BMP4) Controls Proliferation and Proximal-distal Differentiation of the Lung Epithelium

Bmp4 (Fig. 3c) expression is first detected in the ventral mesenchyme of the developing lung when the primordial lung buds are merging from the foregut. This mesenchymal expression is maintained until E13.5. Expression of *Bmp4* is also detected in the distal endoderm of the developing lung bud. Over-expression of *Bmp4*, driven by the *Surfactant Protein-C* (*SP-C*) promoter in the distal endoderm of transgenic mice, causes abnormal lung morphogenesis, with cystic terminal sacs and inhibition of epithelial

proliferation. In contrast, *SP-C* promoter-driven over-expression of either the BMP antagonist *Xnoggin* or a dominant negative *Alk6* BMP receptor to block BMP signaling, results in severely reduced distal epithelial cell phenotypes and increased proximal cell phenotypes in the lungs of transgenic mice (4). However, the exact roles of BMP4 in early mouse lung development remain controversial. *Bmp4* is induced at the tip of the growing lung buds in response to mesenchymal *Fgf10*. In isolated E11.5 mouse lung endoderm cultured in ▶Matrigel™, addition of exogenous BMP4 prevents further budding in response to FGF10, therefore ensuring a single extending bud, rather than a cluster of buds (4). However, addition of BMP4 to intact embryonic lung explant cultures stimulates lung branching morphogenesis.



Lung. Figure 3 Gene expression in early mouse lung between E11.5 and E12. (a) *Fgf10* is expressed in the distal mesenchyme at sites of budding (arrows). (b) *Shh* is expressed throughout the epithelium. Note that its expression is up-regulated at sites of budding. (c) *Bmp4* is expressed at high levels in the distal epithelium. (d) Lung at E12, *Wnt7b* is expressed throughout the epithelium.

Wnt Signaling Controls Epithelial and Mesenchymal Differentiation

Wnt signaling plays also an important role in lung development. Between E10.5 and 17.5, β -catenin is localized in the cytoplasm and often also in the nucleus of the undifferentiated primordial epithelium (PE), differentiating alveolar epithelium (AE) and adjacent mesenchyme. *Tcf1* (*T-cell factor 1*), *Lef1* (*Lymphoid enhancer factor 1*), *sFrp1* (*secreted Frizzled-related protein 1*) and *sFRP2* are highly expressed in the mesenchyme adjacent to the PE and AE. *Tcf3* is highly expressed at the apical side of the PE and AE while *Tcf4* and *sFrp4* are detected in both the PE and adjacent mesenchyme. *Fz8* is highly expressed throughout the epithelium while *Fz2*, 3, 6 and 7 are expressed both in the epithelium and mesenchyme (5). These expression patterns suggest that Wnt signaling can originate from the epithelium and mesenchyme and can target both tissues in an autocrine and/or paracrine fashion. TOPGAL or BATGAL mice, which harbor a β -galactosidase gene under the control of a LEF/TCF and β -catenin inducible promoter, reveal that from E10.5 until E12.5, canonical Wnt signaling occurs throughout the epithelium and in the mesenchyme adjacent to the proximal airways where the bronchial smooth muscle cells (SMC) arise (5). From E13.5, TOPGAL activity is no longer present in the mesenchyme and the activity in the epithelium is reduced distally concomitant with the onset of expression of *Dkk1* in the distal epithelium (5). Conditional inactivation of the β -catenin gene in the epithelium of the developing mouse lung leads to neonatal death resulting from severe lung defects (6); branching of secondary bronchi is altered and the number of small peripheral alveolar ducts and terminal saccules is markedly reduced. In addition, the epithelium fails to undergo proper distal differentiation, lacking the expression of pro-*SP-C* protein and *Vascular endothelial growth factor a*, the latter correlating with a

reduction in alveolar capillaries. So far, inactivation of only two Wnt ligands has resulted in lung defects. *Wnt7b* (Fig. 3d) is normally expressed in epithelial cells of the lung periphery. *Wnt7b*^{-/-} mice exhibit perinatal death due to respiratory failure. Defects were observed in proliferation of the lung mesenchyme resulting in lung hypoplasia. In addition, *Wnt7b*^{-/-} embryos and newborn mice exhibit severe defects in the smooth muscle component of the major pulmonary vessels, with increased apoptosis of the vascular smooth muscle cells (VSMCs), resulting in rupture of the major blood vessels and hemorrhages in the lungs after birth. *Wnt5a* is expressed at high levels in the distal lung mesenchyme. *Wnt5a*^{-/-} mice die perinatally from lung defects including truncation of the trachea, overexpansion of the peripheral airways and delayed lung maturation. Absence of WNT5a activity in the mutant lungs leads to increased cell proliferation and up-regulation of the expression of *Fgf10*, *Bmp4*, *Shh* and *Ptc*.

On the other hand, hyperactive Wnt signaling changes the developmental potential of embryonic lung endoderm to express markers of gut endoderm differentiation. Inhibition of Wnt signaling throughout the lung by treatment of lung explants with exogenous DKK1, a potent and specific diffusible inhibitor of Wnt action, results in impaired branching and defects in smooth muscle differentiation and in the formation of the pulmonary vascular network as a result of decreased Fibronectin (FN) deposition (5). This points out the importance of correct deposition of the extracellular matrix in response to growth factor signaling and suggest that for branching/epithelial tip splitting, a solid bar of Fibronectin in the mesenchyme can act like a “rock in the stream” to divide the flowing stream of epithelium into two as it chemotaxes towards FGF10.

Regulatory Mechanisms

Proximal Distal Differentiation of the Lung Endoderm

Epithelial cell lineages are arranged in a distinct proximal-distal pattern in the airways. The larynx is lined with squamous epithelium and the upper airways are lined with ciliated columnar cells that express the forkhead gene *Hfh4* and mucus secreting cells. The lower airways are lined with Clara cells, which secrete a protein called Clara cell 10 kD protein or CC10. The alveoli are lined with alveolar type 1 and 2 epithelial cells (AEC 1 and 2). In the adult lung, the type 1 cells lie in close apposition to the capillaries of the alveoli, whereas the type 2 cells tend to be clustered at the junction between the alveolar ducts and the alveoli. Type 2 cells have the potential to transdifferentiate into type 1 cells. AEC 2 cells are characterized by their expression of the gene for surfactant protein C (SP-C). Pulmonary endocrine (PNE) cells are situated in small

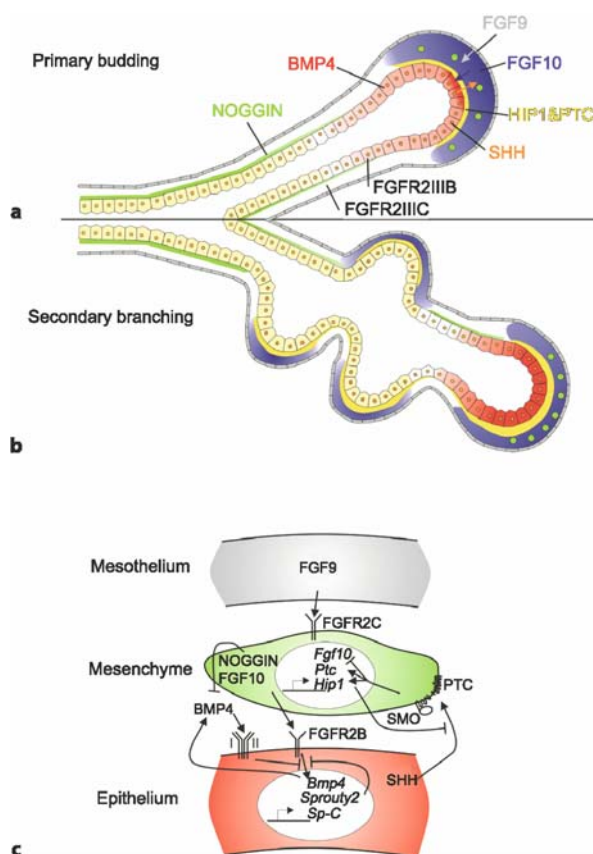
foci and are surrounded by other epithelial cells in the upper airways. PNE cells secrete peptides such as bombesin and calcitonin gene-related peptide, but their precise function is unknown. They also express high levels of *Delta1* (Dll1), encoding a surface ligand for Notch that may function in determining cell fate. The epithelial cells also express uniformly *Fgfr2* and *Hnf3* (*Foxa1*) and (*Foxa2*). The epithelial and mesenchymal cells at the very tips of the extending branches and buds have a specific pattern of gene transcription, related to intercellular signaling pathways (see molecular interactions). Genes expressed at high levels in the epithelium at the tips include *Shh*, *Bmp4*, *Wnt7b*, *Fgfr4* and *c-Fos*. By contrast, the mesenchyme around the tips expresses high levels of *Fgf10*, *Ptc* and *Pod1*. The distal tips are thus thought to be organizing centers that play key roles in regulating both the epithelial-mesenchymal interactions that drive branching morphogenesis and the proximal distal patterning of the lung epithelium.

Formation of a Functional Respiratory Unit (Alveolus)

The alveolar stage involves the formation and maturation of the alveoli. It involves the formation of primary and secondary septa that subdivide the saccules into multiple alveoli. The interstitial mesenchyme layer of each septum is initially quite substantial and contains a capillary network, which is not in very tight contact with the endodermal cells. As postnatal development in the mouse continues, the number of interstitial cells declines and the AEC 1 cells become very closely apposed to the capillary endothelial cells. In the mature lung, the tips of the septa contain smooth muscle cells that produce bundles of elastin fibers, which are important for alveogenesis. Knock out of several genes has resulted in failure of alveogenesis. In *Elastin*^{-/-} mice, perinatal development of airway branches is arrested and is accompanied by fewer distal air sacs that are dilated with attenuated tissue septa resulting in an emphysematous phenotype. In *Pdgfra*^{-/-} (*Platelet derived growth factor a*) mice, the smooth muscle cells and elastin fibers are absent and septae fail to form, leading to postnatal death (7). During the pseudoglandular stage, the growth factor PDGF-A is expressed by the distal epithelial cells and the receptor, PDGFR-A, by the surrounding mesenchyme cells from which the smooth muscle cells derive. Finally, inactivation of both *Fgfr3* and *Fgfr4* results in a postnatal lethal pulmonary phenotype in which excess elastin is laid down and alveoli fail to form.

Toward an Integrated Model of Branching Morphogenesis

As outlined earlier, different growth factors expressed either in the epithelium or mesenchyme are regulating



Lung. Figure 4 Integrated model of Budding. (a) The two primary buds initially elongate. FGF10 (blue), expressed in the distal mesenchyme, acts on the endoderm, through the FGFR2IIIB receptor isoform, to promote its proliferation (blue arrow). SHH (orange dots) is expressed in the epithelium and is up-regulated at the distal tips of the primary buds (darker orange dots). In response to FGF10 signaling, *Bmp4* (red) and *Sprouty2* expression are building up in the endoderm. Because *Fgf10* expression at the sites of future bud formation precedes *Bmp4* up-regulation and *Sprouty2* induction, lung bud outgrowth will occur before BMP4 and SPROUTY2 reach a critical concentration sufficient to inhibit FGF10. NOGGIN (a BMP4 inhibitor; green dots) is expressed in the distal mesenchyme in response to BMP4 signaling and at higher levels in the proximal mesenchyme near the proximal epithelium. The mesothelial layer expresses FGF9 and acts as a proliferative factor on the mesenchyme through receptor FGFR2IIIC (gray arrow). (b) Secondary bud formation is also induced by localized FGF10 (blue) expression in the mesenchyme corresponding to the future sites of secondary bud formation. Epithelial expression of SHH might be instrumental in restricting *Fgf10* expression to the sites of secondary bud formation by inhibiting *Fgf10* expression. SHH (orange dots) is up-regulated in the epithelium of the secondary buds after the activation of *Fgf10*. However, SHH signaling to the mesenchyme transcriptionally activates *Hedgehog-interacting protein 1* (*Hip1*) and *Patched 1* (*Ptc1*) (yellow), repressors of

the branching process. These factors are inducing cell proliferation, apoptosis and cell differentiation. The goal of this section is to integrate seemingly disparate data concerning the regulation of the branching process into a coherent overview. The complete cycle of budding consists of three phases, bud initiation, bud outgrowth and growth arrest. The molecular mechanisms controlling the initiation phase are still unclear and will certainly be the focus of future studies. In particular it will be extremely important to identify the first genes expressed in the region where the bud will form. These early genes are likely to regulate the expression of the genes that we have already identified. During the growth phase, FGF10 is so far the key molecule, and is expressed specifically in the mesenchyme adjacent to the forming epithelial buds. In the future it will be of vital importance to identify the positive regulators which allow such precise temporal-spatial expression. FGF10 expressed in the distal mesenchyme acts on the epithelium mostly through FGFR2-IIIb to promote cell proliferation. FGF10 also triggers chemotaxis on the epithelium (*via* a receptor which is not yet characterized) by modulating cell-cell and cell-extracellular matrix interactions to facilitate or promote cell migration. On the other hand, SHH secreted by the distal epithelium induces cell proliferation in the mesenchyme through PTC and GLI. This process would allow directional growth of the buds toward the localized domain of *Fgf10* expression until a negative feedback loop blocks bud outgrowth. Interestingly, the positive regulators of *Shh* and *Fgf10* expression are still unknown. For example, factors like the WNTs *Wnt7b* is expressed in the epithelium, could regulate *Fgf10* expression. During the growth arrest phase, sonic hedgehog in the epithelium negatively regulates *Fgf10* expression in the mesenchyme. The precise nature of this inhibition is unclear, and more experiments will have to be done to determine if the inhibition is directly promoted by SHH and at what threshold level the effect is elicited.

SHH signaling, to down-regulate the SHH pathway at sites of bud formation. As a result, *Fgf10* expression gets effectively inhibited only at the inter-bud regions. As outgrowth proceeds, the expression of *Bmp4* (red) increases in the distal endoderm, which prevents FGF10 inducing further budding from the growing lung bud. As a result, a single extending bud, rather than a cluster of buds, is generated. In response to lateral *Fgf10* expression, a new lateral, secondary bud is initiated, but only where the level of *Bmp4* expression falls below a threshold in the endoderm. The mechanism by which a new *Fgf10* expression domain is initiated is unknown. (c) Molecular interactions during lung development. The interactions described above are shown in more detail.

In parallel, FGF10 up-regulates *Bmp4* and *Sprouty2* expression in the epithelium, which in turn inhibit epithelial cell proliferation. This model, built essentially on signaling molecules, is obviously incomplete. Other molecules such as the extracellular matrix proteins, integrins and metalloproteases certainly play important roles in the bud outgrowth and growth arrest phases. In the future, we will take into consideration discoveries concerning signaling pathways, cell-cell and cell-extracellular matrix interactions to propose an integrated model of branching morphogenesis.

References

1. Chuang PT, McMahon AP (2003) Branching morphogenesis of the lung: new molecular insights into an old problem. *Trends in Cell Biology* 13:86–91
2. Warburton D, Bellusci S, Del Moral PM et al (2003) Growth factor signaling in lung morphogenetic centers: automaticity, stereotypy and symmetry. *Respir Res* 4
3. Bellusci S, Grindley J, Emoto H et al (1997) Fibroblast Growth Factor 10 (FGF10) and branching morphogenesis in the embryonic mouse lung. *Development* 124:4867–4878
4. Weaver M, Dunn NR, Hogan BLM (2000) *Bmp4* and *Fgf10* play opposing roles during lung bud morphogenesis. *Development* 127:2695–2704
5. De Langhe SP, Sala FG, Del Moral PM et al (2005). Dickkopf-1 (DKK1) reveals that fibronectin is a major target of Wnt signaling in branching morphogenesis of the mouse embryonic lung. *Dev Biol* 277:316–331
6. Mucenski ML, Wert SE, Nation JM et al (2003) Beta-catenin is required for specification of proximal/distal cell fate during lung morphogenesis. *J Biol Chem* 278:40231–40238
7. Bostrom H, Willetts K, Pekny M et al (1996) PDGF-A signaling is a critical event in lung alveolar myofibroblast development and alveogenesis. *Cell* 85:863–873

Lupus

► [SLE Pathogenesis, Genetic Dissection](#)

Luteinizing Hormone

Definition

Luteinizing hormone is an anterior pituitary hormone that stimulates the conversion of the ovarian follicle to a corpus luteum and maintains the corpus luteum. The primary target in females is the ovary and in males the

► [Leydig cells](#). In males, LH is the predominant regulator of testicular steroidogenesis.

► [Hypothalamic and Pituitary Diseases Genetics](#)

Lymphangiogenesis

Definition

Lymphangiogenesis designates the formation of the lymphatic system during development and in pathological conditions.

► [Angiogenesis](#)

Lymphangiomyomatosis

Definition

Lymphangiomyomatosis is a rare disorder seen almost exclusively in females, characterized by progressive proliferation of spindle cells, resembling immature smooth muscle, in the lung parenchyma and along lymphatic vessels in the chest and abdomen. The majority of patients present with dyspnoea. The disease is often associated tuberous sclerosis.

► [Tuberous Sclerosis](#)

Lymphocyte

Definition

Lymphocytes are cells of the immune system that arise from pluripotent stem cells in the lymphoid tissues (thymus, bone marrow), and are associated with all aspects of specific immunity. Lymphocytes can be T cells, B cells or NK (natural killer) cells.

► [Camel as a Model for Functional Genomics](#)

► [Interleukins and Their Receptors, a Role in Regulating Hematopoietic Stem Cells](#)

Lynch Syndrome

► [Hereditary Nonpolyposis Colorectal Cancer](#)

Lyonization

Definition

Lyonization (first described by Mary Lyon) refers to an X chromosome inactivation. One of the two X chromosomes in female cells is randomly switched off. This fact is interpreted to compensate for the existence of only a single X in male cells.

- ▶ SRY – Sex Reversal
- ▶ X–Chromosome Inactivation

Lysolecithin

Definition

Lysolecithin (Synonym: lysophosphatidylcholine) is generated by removal of a fatty acid residue from lecithin at stereochemical number–2 position. This process is catalysed by lecithin-cholesterin-acyl-transferase (LCAT). Lysolecithin is an excellent detergent, capable of solubilizing membranes, and causing cells (such as erythrocytes) to lyse.

- ▶ High-HDL Syndrome

Lysosomal Acid Hydrolases

Definition

Lysosomal acid hydrolases are hydrolytic enzymes within lysosomes that function at low pH optima.

- ▶ Glycosylation of Proteins

Lysosomes

Definition

Lysosomes are roughly spherical, subcellular organelles (0.25 – 0.5 μm in size) bounded by a single membrane. Lysosomes are involved in the degradation of macro-molecules, and for that purpose contain many different kinds of hydrolytic enzymes including proteases, lipases, nucleases, and polysaccharidases. Lysosomal storage diseases are caused by mutations in genes for one of those hydrolytic enzymes, resulting in accumulation of macromolecules (proteins, polysaccharides, and lipids) in the lysosome. Neurons of the central nervous system are particularly susceptible to damage.

- ▶ Peroxisomal Disorders

M Phase

Definition

M Phase defines a cell cycle stage during which the chromosomes are segregated (mitosis) and two daughter cells are formed (cytokinesis).

- ▶ [Cell Cycle – Overview](#)
- ▶ [G-Phase](#)

MA

Definition

MA (matrix) describes the retrovirus protein that forms the N-terminal cleavage product of the ▶ [Gag](#) poly-protein. It usually myristylates and interacts with the cell membrane.

- ▶ [Retroviruses](#)

Machado-Joseph Disease

Definition

Machado-Joseph disease (MJD) is an autosomal dominantly transmitted neurological disorder, often found in families of Portuguese Azorean extraction (descendants of William Machado). Although independently described as a seemingly separate disorder in Caucasians, ▶ [spinocerebellar ataxia–3 \(SCA3\)](#) is now known to be the same entity as MJD.

- ▶ [Repeat Expansion Diseases](#)

Machine Learning

Definition

Machine learning is a statistical theory of classification and clustering of high-dimensional data.

- ▶ [Computational Diagnostics](#)

Macromolecular Crowding

Definition

The concept of Macromolecular Crowding stems from the observation that the protein concentration in cells is very high and suggests that protein molecules are densely packed and literally touch each other.

- ▶ [Electron Tomography](#)

Macrophage

Definition

Macrophages develop from monocytes. They are phagocytic cells that help to initiate and are involved in all stages of immune responses. They recognize, take up (phagocytize) and digest all foreign antigens such as bacteria, viruses and cell debris. During wound healing, macrophages are present for several days. In addition to their immune function, they are a potent source of various growth factors and cytokines.

- ▶ [Inflammatory Responses](#)
- ▶ [Wound Healing](#)

Macula

Definition

Macula is the region of the yellow pigment at the center of vision within the fovea of primate retina. It is marked by a high density of cone photoreceptors, important for colour vision and high spatial resolution. Anatomically it is a small pit bordered by a wall, because the inner retinal layers connected to the photoreceptors are slightly excentrically positioned under the macular wall in order to ensure optimal spatial resolution.

- ▶ [Retinitis Pigmentosa](#)

MAD Phasing

JÜRGEN J. MÜLLER, UDO HEINEMANN

Max Delbrück Center for Molecular Medicine, Berlin, Germany

jjm@mdc-berlin.de

heinemann@mdc-berlin.de

Definition

► **Multi-wavelength anomalous diffraction (MAD)** is a method for phase determination in crystal structure analysis with X-rays. Tunable synchrotron beam lines permit measurement of X-ray diffraction data from crystals at two or more energies near the K- or L-► **absorption edges** of ► **anomalous scatterers** present in biological macromolecules with optimal signal-to-noise ratios. Anomalous scatterers may be naturally present in the crystalline molecule or introduced in various ways. Cryo-cooling extends the lifetime of the crystals and newly developed CCD or other X-ray detectors shorten the measuring time to ensure highly precise measurements with low radiation damage to the molecules. Highly redundant measurements at low noise level permit phase determination at one wavelength at or distant from the absorption edge (► **single-wavelength anomalous diffraction, SAD**).

Description

Diffraction experiments with crystals yield intensities of reflections at coordinates h, k, l in reciprocal space. To determine the crystal structure, i.e. the arrangement of atoms or molecules in the crystal lattice, the amplitudes and phases of the reflections are needed. Amplitudes are proportional to the square root of the measured intensities, but the phases are not revealed by the experiment. Crystal structures of small molecules or inorganic compounds are routinely determined by employing statistical, “direct” methods for solving the phase problem. These methods are suitable for the structure analysis of biological macromolecules only in exceptional cases, when the diffracting power of the protein or nucleic acid crystal is unusually high. To permit the crystal structure analysis of other biological macromolecules, several methods have been introduced that require either the presence of heavy atoms in the crystal (atomic number ≥ 15) or the availability of a model with similarity to the structure being determined. Heavy atoms are used for phasing the diffraction pattern through isomorphous replacement, by exploiting the anomalous effect or by a combination of both. Similar structures that can be used as search models

are a prerequisite for phasing through molecular replacement (for reviews see 1, 2). Here we will focus on modern phasing methods that rely exclusively on anomalous scattering.

MAD (multi-wavelength anomalous diffraction) phasing became the method of choice in macromolecular crystallography with the introduction of new techniques for (i) incorporating anomalous scatterers into biomolecules, (ii) cryo-crystallography, (iii) new fast X-ray detectors and (iv) the availability of tunable, highly brilliant synchrotron beam lines. Since 1990, more than 1800 new crystal structures have been determined using this method. The anomalous scattering is highly dependent on the energy of the X-rays interacting with the anomalous scatterer present in the biomolecule. Although anomalous scattering is not an absorption effect, the signal size is largest near the absorption edge of the scatterer and varies strongly with the X-ray energy at which the diffraction experiment is done. Some of the photons are absorbed and re-emitted by the sample electrons at the same energy (strong coupling to absorption edge energy), but retarded compared to a normally scattered photon. This retardation (phase shift) of the scattered photon causes the atomic form factor f to acquire an imaginary component.

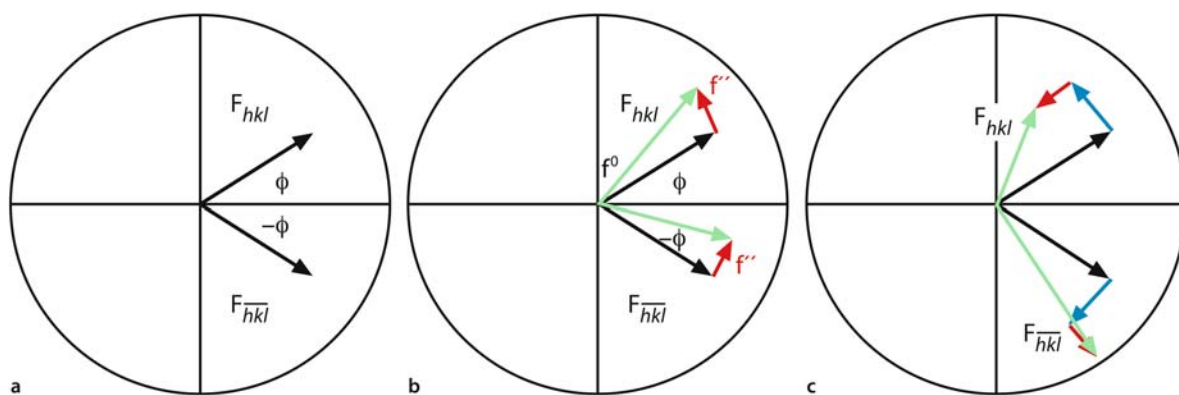
The total scattering factor of an atom is the sum of the ‘normal’ f^0 (scattering angle-dependent) and the ‘anomalous’ (wavelength/energy-dependent) component

$$f_{\text{total}}(\theta, \lambda) = f^0(\theta) + \{f'(\lambda) + if''(\lambda)\} \quad [1]$$

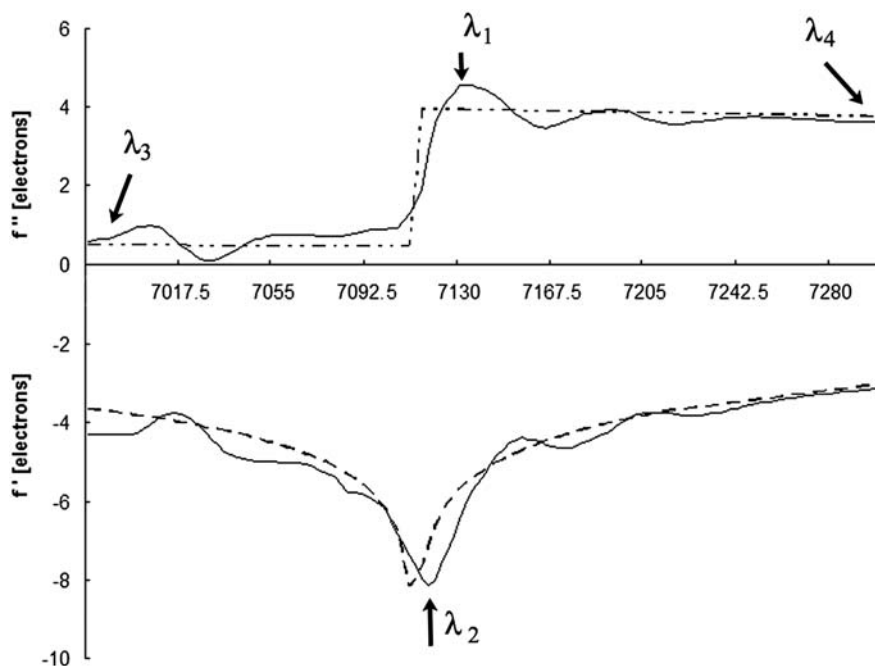
where θ is the scattering (Bragg) angle and λ is the X-ray wavelength.

In anomalous scattering, Friedel’s law is no longer valid, and this can be exploited to determine the phases of X-ray reflections and, furthermore, the absolute configuration of molecules. Reflections which follow Friedel’s law have equal amplitudes F_{hkl} and opposite phases φ (Fig. 1a). If all atoms scatter equally but anomalously, the amplitudes remain the same but the phase relationship no longer holds (Fig. 1b). This is because the f'' term is always positive. If normal and anomalously scattering atoms co-exist, the amplitude as well as the phase relationships are broken.

The imaginary component of the anomalous scattering, f'' , is approximately proportional to the atomic absorption coefficient and to the fluorescence at a chosen X-ray energy, and has its maximum near the absorption edge of an atom (Fig. 2, λ_1). The real component of the anomalous scattering, f' , has a minimum at the inflection point (λ_2) of the absorption curve. Reduction/oxidation and the local chemical environment of an anomalous scatterer modulate the



MAD Phasing. Figure 1 Vector diagrams illustrating anomalous scattering for the reflections (h, k, l) and $(-h, -k, -l)$. A reflection is presented as a vector with length equal to the scattering amplitude F and angle from the positive horizontal equal to the phase ϕ . (a) Friedel pairs of structure factors of normal atoms with equal amplitudes $|F(h, k, l)| = |F(-h, -k, -l)|$ and opposite phases $\phi(h, k, l) = -\phi(-h, -k, -l)$. (b) of anomalous scatterers where the amplitude law holds but the phase relationship is broken. (c) Normal and anomalously scattering atoms are mixed; the amplitude as well as the phase relationships are broken.



MAD Phasing. Figure 2 Anomalous scattering factors f' and f'' for Fe. The spectra are derived from fluorescence measurements near the Fe K-absorption edge at 7.111 keV (1.74346 Å). The wavelengths λ_1 (1.737 Å), λ_2 (1.744 Å), λ_3 (1.797 Å), λ_4 (0.92 Å) correspond to the wavelengths λ_{peak} , λ_{edge} , λ_{remote} , and λ_{native} in a MAD experiment because of maximal anomalous signals for those energies. The dotted lines are theoretical curves calculated using the server at http://www.bmsc.washington.edu/scatter/AS_periodic.html.

absorption compared with the free atom. The X-ray energy ($E_{\text{X-ray}} = 12.398 [\text{keV}]/\lambda [\text{\AA}]$) for an optimal anomalous signal is determined for each crystal by a fluorescence scan before a diffraction experiment is carried out at the synchrotron.

For a single type of anomalous scatterer, i.e. a MAD experiment at a single absorption edge, a minimum of two wavelengths are required to generate phases. MAD phasing is over-determined if one measures diffraction data at three or four wavelengths. Optimal anomalous

signals f'' and f' require great precision in setting up and controlling the wavelength during the diffraction experiment. Additional wavelengths (λ_3 and λ_4) may be chosen at points typically between 100 eV and 1000 eV distant from the absorption edge.

Estimation of the Size of the Anomalous Signal in Diffraction Data

Numerous elements have been used as anomalous scatterers in crystal structure analysis of biological macromolecules over the past fifteen years (3, 4; Table 1). Mn, Fe, Co, Ni, Cu and Zn ions are often natural components of metalloproteins. Selenomethionine as a substitute for methionine has been widely used as an anomalous-scattering label for proteins and brominated bases (5-bromouracil, 5-bromocytosine) may serve as labels for nucleic acids. Furthermore, elements used conventionally for [▶isomorphic replacement](#) like Pt, Au, Hg, Tl, Pb, Bi, As, Ga, Sr, Os, Ir, Th, Pa, U, Y, Sm possess K- or L-absorption edges within the energy region suitable for diffraction experiments (Table 1). Quick soaking with Br^- or K^+ , as well as high-pressure labeling with noble gases (basics, see (6)) have also been successfully used to incorporate anomalous scatterers into protein crystals. Although their absorption edges are outside the wavelength range of convenient X-ray diffraction experiments, the anomalous signal of sulfur and phosphorus can also be used for structure determination. Recently, an increasing number of macromolecular crystal structures have been determined by collecting very precise diffraction data at only one wavelength, establishing the single-wavelength anomalous diffraction method, SAD (4). Just as MAD, SAD, coupled with powerful phase improvement methods, may be based upon almost all atoms contained in Table 1 for phase determination, even if the diffraction experiments are performed in the home laboratory using X-ray generators with rotating copper anodes (f'' at $\text{CuK}\alpha$ wavelength, see column 7 in Table 1). The largest anomalous signals are available from lanthanides (good substitutes in Ca^{2+} binding sites) and actinides.

Because the anomalous signal is weak, special experimental efforts have to be made to obtain a high anomalous signal-to-noise ratio. The magnitude of the signal is estimated as the ratio of the dispersive, or the Bijvoet difference to the expected root-mean-square structure amplitude at zero scattering angle of N_m identical atoms within the macromolecule

$$\text{rms}|^0F|_{\text{macromolecule}} = \sqrt{N_m} \cdot f \quad [2]$$

The scattering factor f of an average non-hydrogen atom in a protein is usually taken as $6.7 e^-$, $7.2 e^-$ in

DNA and $7.26 e^-$ in RNA. The diffraction ratios for estimating the significance of the signal originating from N_a anomalous scatterers of a single type included in the macromolecule are

$$\frac{\sqrt{N_a/2} \cdot |f'_{\lambda 1} - f'_{\lambda 2}|}{\text{rms}|^0F|_{\text{macromolecule}}} \quad [3]$$

for the dispersive signal, and

$$\frac{\sqrt{N_a/2} \cdot 2f''_{\lambda}}{\text{rms}|^0F|_{\text{macromolecule}}} \quad [4]$$

for the Bijvoet signal. $|f'_{\lambda 1} - f'_{\lambda 2}|$ and f''_{λ} are listed in Table 1 for the most frequently used anomalous scatterers at their absorption edges or at $\text{CuK}\alpha$ wavelength.

Both expected relative changes in the diffraction intensities caused by the anomalous scattering can be compared directly with the R_{meas} (R_{sym} , R_{merge}) estimates of errors (2) in the raw data. If the expected anomalous signal is about 3%, the error estimates should be below this value.

As a rule of thumb, one ordered Se atom per 100 amino acid residues (with 7.7 atoms per residue on average) is usually assumed to permit the phasing of the corresponding X-ray diffraction data.

Algebraic Background of the Phase Determination

The structure factor equation for any reflection hkl derived from a crystalline macromolecule with N non-hydrogen atoms and N_{ano} anomalous scatterers can be written as

$$F_{\text{obs}}(\lambda, hkl) = \sum_{j=1, N} [f^0 + f'(\lambda) + if''(\lambda)]_j \exp[-B_j s^2(hkl)] \exp[2\pi i(hx_j + ky_j + lz_j)] \quad [5]$$

or as

$$\begin{aligned} F_{\text{obs}}(\lambda, hkl) = & \sum_{j=1, N} f_j^0 \exp[-B_j s^2(hkl)] \\ & \exp[2\pi i(hx_j + ky_j + lz_j)] \\ & + \sum_{m=1, N_{\text{ano}}} f'_m(\lambda) \exp[-B_m s^2(hkl)] \\ & \exp[2\pi i(hx_m + ky_m + lz_m)] \\ & + i \sum_{m=1, N_{\text{ano}}} f''_m(\lambda) \exp[-B_m s^2(hkl)] \\ & \exp[2\pi i(hx_m + ky_m + lz_m)] \end{aligned} \quad [6]$$

Eqs. [1] and [5] hold for all atoms, however, the magnitudes of f' and f'' are negligible for the light atoms H, C, N, O of biological macromolecules. This allows separation of the wavelength-independent and

MAD Phasing. Table 1 Selection of anomalous scatterers with absorption edges accessible by most synchrotrons. Atoms which occur naturally in some proteins, or belong to the lanthanides or actinides are colored in column 2. Roughly, the anomalous scattering factors f'' are 4 e⁻ near the K-edge and 10 e⁻ near L_{III}-edges. In reality, they may be higher because of resonance effects (white line = h{igh}, else l{ow}). For wavelengths far from the absorption edge, anomalous scattering may also be significant and anomalous effects are sometimes measurable at home X-ray sources using CuK α radiation (column 7). The [Protein Data Bank](http://www.bmrc.washington.edu/scatter/AS_periodic.html) entry (5) of a sample structure determined using the corresponding anomalous scatterer is provided in column 8. The calculated values are taken from http://www.bmrc.washington.edu/scatter/AS_periodic.html

Z	Atom	Edge	Energy (keV)/ λ (Å)	$ f'_{\lambda 1} - f'_{\lambda 2} _{\max}$ ($f'_{\text{edge calc}}$)	f''_{\max} ($f''_{\max \text{ calc}}$)	f''_{calc} at CuK α	White line / PDB entry
15	P	K	2.146/5.784	(-7.44)	(4.11)	0.43	[rev (4)]
16	S	K	2.47/5.0185	(-7.76)	(4.1)	0.56	h, 1CRN, 1EL4
17	Cl	K	2.822/4.3929	(-7.59)	(4.1)	0.70	[rev (4)]
20	Ca	K	4.0381/3.07	(-7.88)	(4.05)	1.3	[rev (4)]
23	V	K	5.4651/2.269	(-7.99)	(3.99)	2.13	1H2F
25	Mn	K	6.539/1.8964	8.7 (-8.12)	9.7 (3.96)	2.8	h, [rev (4)]
26	Fe	K	7.112/1.743	7.0 (-6.22)	4.0-5.0 (3.95)	3.2	l, 1AYF
27	Co	K	7.709/1.608	19.9 (-8.15)	3.9(3.94)	3.6	l, 1N2Z
28	Ni	K	8.333/1.488	(-7.7)	4.3(3.92)	0.51	l, 1Q0D
29	Cu	K	8.979/1.381	6.5 (-8.2)	4.2 (3.9)	0.59	l, 2CBP
30	Zn	K	9.659/1.284	5.7 (-8.11)	4.3 (3.9)	0.68	l, 1PTQ
31	Ga	K	10.367/1.196	(-7.95)	(3.89)	0.78	-
32	Ge	K	11.103/1.117	(-8.32)	(3.88)	0.89	-
33	As	K	11.867/1.045	(-8.25)	(3.87)	1.0	-
34	Se	K	12.658/0.979	7.5 (-8.32)	6.0 (3.85)	1.1	h, 1HCN
35	Br	K	13.474/0.920	6.1 (-8.29)	3.8 (3.83)	1.3	l, 1R71
36	Kr	K	14.326/0.865	(-8.26)	1.6-2.7 (3.8)	1.4	l, [rev (6)]
37	Rb	K	15.200/0.816	(-10.4)	(3.79)	1.6	1I0M
38	Sr	K	16.105/0.770	(-8.27)	(3.77)	1.8	1NVY
39	Y	K	17.038/0.728	(-8.26)	(3.75)	2.0	1LVA
42	Mo	K	20.0/0.6198	(-9.88)	(3.7)	2.7	-
48	Cd	K	26.711/0.464	(n.d.)	(n.d.)	4.7	l, 1DYP
53	I	L _I	5.188/2.388	(-9.4)	(13.4)	6.8	1BN2
54	Xe	K L _I	34.561/0.358 5.453/2.2738	(-9.5)	3.5 (13.4)	7.4	1C10
55	Cs	L _{II}	5.359/2.313	(-14.36)	(13.12)	7.9	1I0M
56	Ba	L _{II}	5.624/2.205	(-14.18)	(13.0)	8.46	1I0M
62	Sm	L _I	7.737/1.6025	(-19.8)	(12.9)	12.3	1A3C
64	Gd	L _{III}	7.243/1.712	<-19 (-20.2)	>10 (10.6)	12.0	h, 1H87
65	Tb	L _{III}	7.514/1.650	(-20.4)	19.9 (10.6)	9.24	h, 1M9S
66	Dy	L _{III}	7.790/1.592	(-20.2)	28(10.6)	9.75	h, -

MAD Phasing. Table 1 Selection of anomalous scatterers with absorption edges accessible by most synchrotrons. Atoms which occur naturally in some proteins, or belong to the lanthanides or actinides are colored in column 2. Roughly, the anomalous scattering factors f'' are 4 e^- near the K-edge and 10 e^- near L_{III}-edges. In reality, they may be higher because of resonance effects (white line = h{igh}, else l{ow}). For wavelengths far from the absorption edge, anomalous scattering may also be significant and anomalous effects are sometimes measurable at home X-ray sources using CuK α radiation (column 7). The [Protein Data Bank](http://www.bmcs.washington.edu/scatter/AS_periodic.html) entry (5) of a sample structure determined using the corresponding anomalous scatterer is provided in column 8. The calculated values are taken from http://www.bmcs.washington.edu/scatter/AS_periodic.html (Continued)

Z	Atom	Edge	Energy (keV)/ λ (Å)	$ f'_{\lambda 1} - f'_{\lambda 2} _{\text{max}}$ ($f'_{\text{edge calc}}$)	f''_{max} ($f''_{\text{max calc}}$)	f''_{calc} at CuK α	White line / PDB entry
67	Ho	L _I L _{III}	9.40/1.319 8.071/1.536	(-9.1) (-29.3)	(12.77) 21.7 (10.6)	3.7	h, 1MSB
68	Er	L _{III}	1.4835				h, 1F0J
70	Yb	L _{III}	8.95/1.386	16.0 (-31.2)	30.23 (10)	4.4	h, 2BOB
71	Lu	L _{III}	9.244/1.341	(-20.0)	10.7 (10.5)	4.75	1KHV
73	Ta	L _{III}	9.881/1.25	(-19.9)	10.5(10.51)	5.33	1E9R
74	W	L _{III}	10.21/1.214	17.5 (-24.4)	18.5 (10.5)	5.6	h, 1N7D
76	Os	L _{III}	10.871/1.141	(-19.5)	(10.2)	6.22	h, 1HC8
77	Ir	L _{III}	11.215/1.106	(-19.4)	(10.2)	6.57	h, 1VC7
78	Pt	L _{III}	11.564/1.072	(-19.3)	(10.2)	6.93	h, 1A2E
79	Au	L _{III}	11.919/1.040	(-19.2)	(10.2)	7.3	l, 1A8D
80	Hg	L _{III}	12.284/1.009	13.9 (-12.0)	10.1 (10.2)	7.69	l, 1GPM
81	Tl	L _{III}	12.658/0.980	(-18.9)	(10.2)	8.09	1LQK
82	Pb	L _{III}	13.055/0.950	(-19.1)	(10.1)	8.51	1NBS
83	Bi	L _{III}	13.419/0.924	(-18.9)	(10.1)	8.93	—
90	Th	L _{III} M _I	16.300/0.761 5.182/2.392	(-18.4) (-10.6)	(9.97) (22.7)	12.33	—
91	Pa	L _{III} M _I	16.733/0.741 5.366/2.310	(-18.9) (-10.5)	(9.96) (23.4)	12.88	—
92	U	L _{III} M _I	17.166/0.722 5.548/2.235	-21.5 (-34.7) (-10.3)	10.3 (9.96) (23.2)	13.41	h, 1NCG

wavelength-dependent terms in eq. [6]. The anomalous contributions for each type of scatterer can be expressed in terms of their normal structure factor components

$$F_{\text{obs}}(\lambda, hkl) = {}^0F_{\text{total}}(hkl) + \left\{ f'(\lambda)/f_0 + i f''(\lambda)/f_0 \right\} {}^0F_{\text{anops}}(hkl) \quad [7]$$

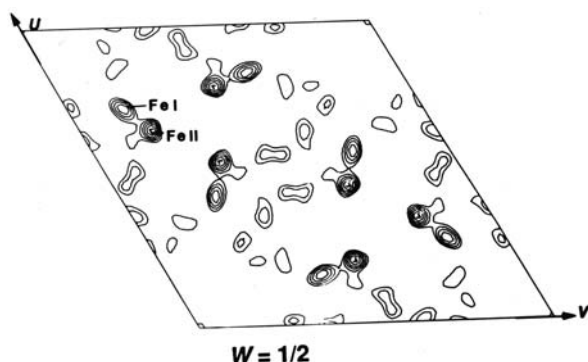
${}^0F_{\text{anops}}(hkl)$ is the normal structure factor of the partial structure representing the anomalous scatterers. Hendrickson (7) reformulated the magnitude of the measured intensity

$$|F_{\text{obs}}(\lambda, \pm hkl)|^2 = |{}^0F_{\text{total}}(hkl)|^2 + a(\lambda) |{}^0F_{\text{anops}}(hkl)|^2 + b(\lambda) |{}^0F_{\text{total}}(hkl)| |{}^0F_{\text{anops}}(hkl)| \cos(\varphi_{\text{total}} - \varphi_{\text{anops}}) \pm c(\lambda) |{}^0F_{\text{total}}(hkl)| |{}^0F_{\text{anops}}(hkl)| \sin(\varphi_{\text{total}} - \varphi_{\text{anops}}) \quad [8]$$

where the wavelength dependence is confined to the experimentally measurable coefficients

$$a(\lambda) = \frac{f'(\lambda)^2 + f''(\lambda)^2}{f_0^2} \quad b(\lambda) = \frac{2f'(\lambda)}{f_0} \quad c(\lambda) = \frac{2f''(\lambda)}{f_0}$$

For each reflection hkl and wavelength λ , one obtains two equations for the three unknowns F_{total} , F_{anops} , and $\delta\varphi = \varphi_{\text{total}} - \varphi_{\text{anops}}$. Measurement at two or more wavelengths results in over-determination of the



MAD Phasing. Figure 3 Harker section for the native anomalous difference Patterson map from the ferredoxin from *Anabaena* 7120 at 1.7 Å resolution (with permission from (8)). The section at $w = 1/2$ was calculated with coefficients $\|F(hkl) - F(-hkl)\|^2$. The map was contoured at 2σ and increasing in increments of 1σ . An entire unit cell in the u and v direction has been plotted. The positions of the iron atoms in the cluster were determined in the space group $P6_5$. Shown here are the positions of the Harker peaks arising at $u = 2x$, $v = 2y$, and $w = 0.5$. The corresponding positions of Fe I are $x = 0.365$, $y = 0.573$, $z = 0.018$ and of Fe II $x = 0.327$, $y = 0.593$, $z = 0.0$. The distance between the Fe in the [2Fe-2S] cluster is 2.68 Å as in the refined structure.

unknowns and renders possible least squares fitting procedures.

The maximal pure anomalous signal components can be experimentally determined from the so-called Bijvoet pairs of reflections $|F_{\text{obs}}(\lambda, 'hkl)|$, the space group symmetry equivalents to the two members of a Friedel pair (see above).

'Dispersive' phase information can be derived from differences between $|F_{\text{obs}}|$ at wavelengths having different f'

$$\Delta F_{\text{disp}} = |F_{\text{obs}}(\lambda_i, hkl)| - |F_{\text{obs}}(\lambda_j, hkl)| \quad [9]$$

the difference of amplitudes from Friedel pairs of reflections at wavelengths with significant f'' contributes to the 'Bijvoet' phase information

$$\Delta F_{\text{Bijvoet}} = |F_{\text{obs}}(\lambda_i, hkl)| - |F_{\text{obs}}(\lambda_j, -hkl)| \quad [10]$$

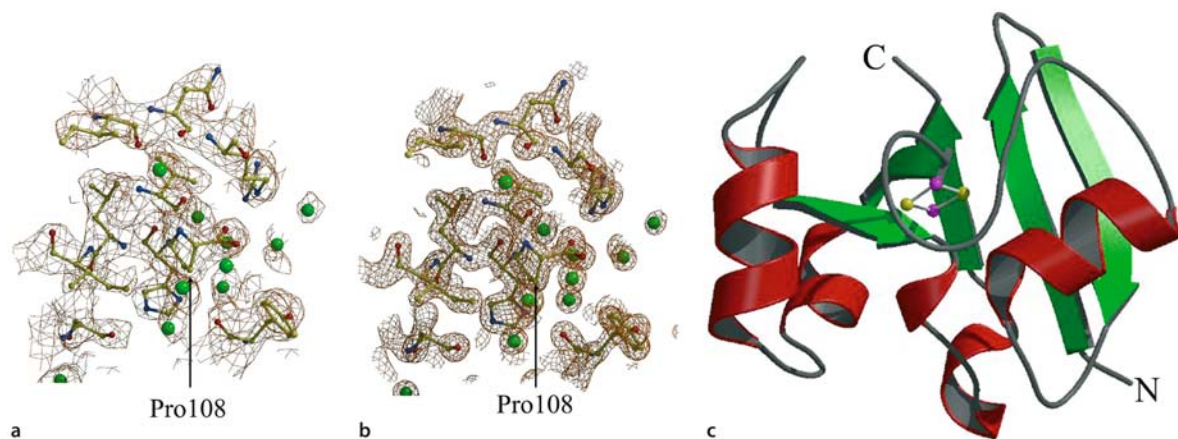
Taking equations [8, 9] and following (7)

$$\Delta F_{\text{Bijvoet}} = \frac{2f''(\lambda)}{f_0} |^0 F_{\text{anops}}(hkl)| \sin(\varphi_{\text{total}} - \varphi_{\text{anops}}) \quad [11]$$

the Bijvoet difference contributes to $\sin(\varphi_{\text{total}} - \varphi_{\text{anops}})$. The **▶ dispersive difference** contributes to the $\cos(\varphi_{\text{total}} - \varphi_{\text{anops}})$.

After calculation of initial estimates for F_{total} , F_{anops} , and $\delta\varphi$ using a phasing program, a phase estimate for the protein φ_{total} at each h, k, l

$$\varphi_{\text{total}} = \Delta\varphi + \varphi_{\text{anops}} \quad [12]$$



MAD Phasing. Figure 4 Structure determination of truncated adrenodoxin Adx(4-108) (Protein Databank entry 1AYF) by MAD phasing using the intrinsic Fe atoms as anomalous scatterers. (a) Vicinity of the C-terminal Pro108 of Adx(4-108) with the experimental electron density determined from the anomalous signal of the [2Fe-2S] cluster and contoured at 1σ (a), and (b), the final $2F_o - F_c$ electron density map at 1.85-Å resolution contoured at 1.5σ . (c) Secondary structure diagram of Adx(4-108) (9), the [2Fe-2S] cluster is shown in ball-and-stick representation with the anomalous scatterer Fe colored magenta (S in yellow). Water molecules are shown as green spheres.

can be calculated if the phase φ_{anops} of the anomalous scatterers' partial structure is known. One possible way to determine the partial structure is the use of the [▶Bijvoet-difference](#) [▶Patterson map](#) with coefficients $|\delta F_{\text{Bijvoet}}|^2$ when the number of sites is small (Fig. 3). In principle, the sub-structure can be determined using only the Bijvoet differences for single-wavelength data (4). For a larger number of sites direct methods are recommended (OASIS, SHELXS, SHELXD, SnB, (2, 10)). Several methods to overcome the phase ambiguity caused by the identity

$$\sin(\varphi_{\text{total}} - \varphi_{\text{anops}}) = \sin(180^\circ - \varphi_{\text{total}} + \varphi_{\text{anops}}) \quad [13]$$

when solving eq. [11] have been summarized in (2).

From the Anomalous-Scatterer Partial Structure to the Macromolecular Structure

After the determination of the heavy atom position, a phase refinement (e.g. MLPHARE, (10)) will follow, and after density modification (e.g. DM, (10)), the experimental electron density map is obtained (Fig. 4a). This process is widely automated in the programs SHARP and SOLVE (2). The main chain and side chains of the macromolecule can be traced interactively by hand (program O) or automatically (e.g. ARP/WARP, SOLVE/RESOLVE, for overview see (2)) depending on the quality of the phases and, thus, the electron density map. After several rounds of computer-assisted manual model correction, addition of water molecules, ions and other low-molecular-weight components and automatic refinement (e.g. REFMAC, (10)), the crystal structure (e.g. Figs. 4b, 4c) will be validated according to a set of quality criteria (2) including a reasonable stereochemistry, and the final atomic coordinates along with the experimental data will be deposited with the [▶Protein Data Bank](#) (5).

References

1. Drenth J (1999) Principles of Protein X-ray Crystallography, 2nd edn. Springer-Verlag, Berlin Heidelberg New York
2. Rossmann MG, Arnold E (2001) International Tables for Crystallography. Volume F: Crystallography of Biological Macromolecules. IUCr / Kluwer Academic Publishers, Dordrecht
3. Pappa HS, Tajbaksh S, Saunders AJ et al (1996) Incorporating anomalous scattering centres into macromolecules. *Curr Opin Struct Biol* 6:611–616
4. Dauter Z, Dauter M, Dodson E (2002) Jolly SAD. *Acta Crystallogr D* 58:494–506
5. Berman HM, Westbrook J, Feng Z et al (2000) The Protein Data Bank. *Nucleic Acids Res* 28:235–242
6. Quillin ML, Breyer WA, Griswold JJ et al (2000) Size versus polarizability in protein-ligand interactions: Binding of noble gases within engineered cavities in phage T4 lysozyme. *J Mol Biol* 302:955–977
7. Hendrickson WA (1979) Phase information from anomalous-scattering measurements. *Acta Crystallogr A* 35:245–247
8. Jacobson BL, Chae YK, Markley JL et al (1993) Molecular structure of the oxidized, recombinant, heterocyst [2Fe-2S] ferredoxin from *Anabaena* 7120 determined to 1.7-Å resolution. *Biochem* 32:6788–6793
9. Muller A, Muller JJ, Muller YA et al (1998) New aspects of electron transfer revealed by the crystal structure of a truncated bovine adrenodoxin, Adx(4-108). *Structure* 6:269–280
10. Collaborative Computational Project, Number 4 (1994) The CCP4 suite: Programs for protein crystallography. *Acta Crystallogr D* 50:760–763

Madin-Darby Canine Kidney Cells

Definition

Madin-Darby Canine Kidney (MDCK) cells are a model system of simple epithelia that, when cultured on a permeable support, form an electrically tight monolayer of polarized cells. MDCK cells have been extensively used since the 1970s to study the establishment of intercellular junctions, polarized sorting and trafficking of membrane proteins.

[▶Epithelial Cells](#)

Maguk Proteins

Definition

Maguk proteins are scaffold proteins that contain PSD-95–Discs Large A-Zonula occludens-1 ([▶PDZ](#)), [▶Src](#)-homology-3, and guanylate kinase domains.

[▶Tight Junctions](#)

Major and Minor Groove

Definition

Major and minor groove denote intrusion in the surface of double-stranded DNA fibers. In B-form DNA, the major and minor grooves are about 12 Å and 6 Å wide, respectively. They play an important role for the interaction of DNA with protein molecules.

[▶DNA Structure](#)

[▶Protein/DNA Interaction](#)

Major Histocompatibility Complex

Definition

Major histocompatibility complex (MHC) is found in all mammals whose products are primarily responsible for the rapid rejection of grafts between individuals, and function in signalling between lymphocytes and antigen presenting cells. The MHC (►HLA) complex is located at chromosome 6p21.

- DNA-based Vaccination
- Genetic Predisposition to Multiple Sclerosis

Major Locus

Definition

A single genetic locus that accounts for the majority of the variability in a trait.

- Cleft Lip Palate

Maladie de Gélineau

- Narcolepsy

Malar Hypoplasia

Definition

Malar hypoplasia is the underdevelopment of the cheekbones.

- Marfan Syndrome

MALDI

- Mass Spectrometry: MALDI
- Matrix-Assisted Laser Desorption/Ionisation

MALDI-MS

- Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry (MALDI-MS)

MALDI-TOF-MS

Definition

Matrix-assisted laser desorption/ionisation time-of-flight mass spectrometry.

- Mass Spectrometry: MALDI
- SNP Detection and Mass Spectrometry

Malignancy

Definition

The general term describes the tendency of (transformed) cells to divide in an uncontrolled fashion, to invade surrounding normal tissue and/or undergo metastasis, and to finally develop cancer. Malignant transformation describes the series of genetic changes that result in the development of cancer.

- Cap-Independent Translational Control
- Cancer Invasion and Metastasis: Cellular, Molecular and Clinical Aspects
- Splicing

Malignant Transformation

- Malignancy/Malignant Transformation

Mammalian Fertilization

EVELINE S. LITSCHER, PAUL M. WASSARMAN
Brookdale Department of Molecular, Cell &
Developmental Biology, Mount Sinai School of
Medicine, New York, NY, USA
eveline.litscher@mssm.edu
paul.wassarman@mssm.edu

Definition

Fertilization in Mammals

The essence of life is the ability to reproduce and perpetuate the species. ► **Fertilization** or sexual reproduction is defined as the union of two germ cells, ► **egg** (female) and ► **sperm** (male), whereby the somatic chromosome number is restored and the resulting offspring exhibit characteristics of their parents (1). Male and female gametes fuse and recombine inherited traits of the two parents to produce individuals with novel assortments of genes. Sexual reproduction, as opposed to asexual reproduction, a process that gives rise to offspring genetically identical to the parent organism, has great advantages. Consequently, a vast majority of plants and animals have adopted it.

Fertilization refers specifically to fusion of male and female gametes to form a zygote. However, the overall pathway to fertilization encompasses a host of events that precede gamete fusion. Among these, extensive transcription and translation, architectural rearrangements and modification of existing organelles, production of new organelles, meiotic maturation and capacitation and other events must occur in sperm and eggs prior to fertilization. In mammals, fertilization takes place internally in the protected environment of the ampulla of the oviduct, as opposed to external fertilization where sperm and egg meet outside the parent's body (e.g., as in fish, reptiles, amphibians and invertebrates).

Characteristics

Oogenesis in Mammals

► **Oogenesis** is initiated during fetal development and continues throughout the reproductive lifetime of females (3, 6). Primordial germ cells become the oogonia that populate fetal ovaries and proliferate to become non-growing oocytes. The transition from oogonia to oocytes involves a change from ► **mitosis** to ► **meiosis**. Shortly after birth, non-growing oocytes are arrested in prophase I of meiosis until stimulated to grow and resume meiotic progression at about the time of ovulation. The pool of small, non-growing oocytes is the sole source of unfertilized eggs in sexually mature females. At least fifty percent of oocytes present in the ovary at birth are lost during the first week following birth. In this context, it has recently been proposed that there is a pool of proliferative germ cells that sustains oocyte and follicle production in the postnatal mammalian ovary (Johnson J et al (2004) *Nature* 428:145–150).

Human oogonia undergo numerous rounds of mitotic divisions over a period of several months until shortly before birth (mice ~4 mitotic cycles). For example, 7 million oogonia are present at ~20 weeks gestation, but only 0.5–1 million oocytes, more or less evenly divided between each of two ovaries, remain at birth (mice

~10,000–15,000 oocytes/ovary). Following birth, the number of oocytes continues to decline even before the onset of puberty. After puberty, many oocytes are lost each month, with typically one oocyte undergoing ovulation out of the population recruited. During the reproductive lifespan of a woman, only 400–500 oocytes will be ovulated, while remaining oocytes slowly undergo atresia and are lost.

The non-growing human oocyte, ~35 µm in diameter (mice ~15 µm), is contained within a cellular follicle that grows concomitantly with the oocyte from a single layer of a few epithelial-like cells to three layers of cuboidal granulosa cells by the time the oocyte has completed growth (≥6 months; mice ~2 weeks). Non-growing oocytes progress through the pachytene stage of prophase I of meiosis during which time homologous chromosomes pair and crossing-over and recombination occur between paired chromosomes (mice ~60 h). During their growth phase characterized by high rates of transcription and translation, modification of existing organelles (e.g., Golgi and mitochondria) and production of new organelles (e.g., zona pellucida and cortical granules), oocytes are arrested at the diplotene stage of the first meiotic prophase. Oocytes then enter a prolonged resting phase, which is terminated before ovulation, which occurs for the first time at puberty. Oocytes progressively acquire meiotic competence (i.e., the ability to undergo ► **meiotic maturation**) and fertility potential during the final days (► **Graafian follicle** stage) before ovulation. Fully-grown human oocytes, ~120 µm in diameter (mice ~80 µm), in Graafian follicles resume meiosis and complete the first meiotic reductive division (meiotic maturation, separation of homologous chromosomes) just prior to ovulation. Meiotic maturation (~36 h, mice 12–14 h) is characterized by germinal vesicle (nuclear) breakdown, formation of the first meiotic spindle (metaphase I), expulsion of the first polar body and arrest at metaphase of the second meiotic division (metaphase II). It occurs in preovulatory Graafian follicles in response to a surge of gonadotropins and leads to a mature, ovulated oocyte or unfertilized egg. Unfertilized human eggs possess 22 somatic chromosomes or autosomes (mice, 19 autosomes) and 1 sex chromosome (total haploid content, 23 chromosomes), each composed of 2 chromatids aligned on the metaphase II spindle and a small polar body containing 22 autosomes and 1 sex chromosome. Ovulated eggs are released from Graafian follicles, enter the opening of the oviduct (Fallopian tube) and move to the lower ampulla region of the oviduct where fertilization can take place. Ovulated eggs complete meiosis (i.e. become haploid cells) with separation of chromatids and emission of a second polar body only upon fertilization by sperm (i.e. restoring the diploid state). Human eggs remain capable of being fertilized

and giving rise to offspring for ~24 h (mouse 8–12 h) following ovulation.

Spermatogenesis in Mammals

► **Spermatogenesis** is the process by which spermatogonial cells divide and differentiate into mature spermatozoa (2, 4). Spermatogonial cells, which are diploid, multiply mitotically and can differentiate to produce primary spermatocytes (meiotic cells). These do not divide until after puberty and are then produced continuously within the testis throughout the lifetime of sexually mature males, providing a steady-state turnover of male gametes. Pituitary gonadotropins, such as LH and FSH, are the major regulators of spermatogenesis and are required to initiate it at the time of puberty. Spermatogenesis is divided into three functionally distinct phases, (1) the proliferative phase, during which some spermatogonia divide to replenish the stem cell pool and others undergo mitotic divisions and become committed to further differentiation and produce primary spermatocytes, (2) the meiotic phase, in which each spermatocyte undergoes meiotic divisions to give rise to 4 haploid spermatids and (3) spermiogenesis, during which round spermatids differentiate into mature spermatozoa that are motile and become capable of fertilizing eggs.

Spermatocytes do not separate completely during meiosis and, as a result, a group of germ cells remain connected to one another *via* cytoplasmic bridges and undergo further development within this syncytium in a synchronous manner throughout the remainder of the spermatogenic process. Spermatocytes enter a prolonged meiotic phase of spermatogenesis, which in humans takes 22–24 days to complete. Primary spermatocytes enter prophase I of the first meiotic division (meiosis I) and homologous chromosomes pair and exchange genetic material by crossing-over and recombination during the pachytene stage of prophase I (~14 days). They divide to form two secondary spermatocytes, each containing a haploid complement of chromosomes. After a brief interphase, secondary spermatocytes enter the second meiotic division (meiosis II) and each divides to form two haploid spermatids. There is no period of arrested development during meiosis for spermatocytes as is the case for oocytes. RNA and protein synthesis is maximal during meiosis and decreases abruptly during early spermiogenesis and differentiation. Spermatids differentiate into mature spermatozoa that are motile and capable of fertilizing eggs, a process that involves morphological and functional changes. The major changes that occur are formation of the flagellum for motility, development of the ► **acrosome**, which contains proteolytic enzymes needed for sperm penetration through egg investments, condensation of chromosomes and nucleoproteins of the nucleus and removal of excessive cytoplasm.

Spermatozoa undergo minor morphological but significant functional modifications during transport in the male genital tract, especially in the epididymis where the ability of sperm to fertilize eggs is acquired. The final stage of maturation, called ► **capacitation**, occurs in the female genital tract following ejaculation. Ejaculated sperm have an oval head, which is composed of a nucleus and acrosomal cap (vesicle), which covers the anterior portion of the head. The neck is the region of connection of the sperm tail (flagellum) to the head. The flagellum, composed of microtubules, is the mobile part of the sperm and consists of a middle-piece and tail. The middle-piece is surrounded by a mitochondrial sheath and the tail is divided into two regions, the principal- and end-pieces. The sperm length of different mammals varies considerably, ranging from 50–250 μm (humans, 60–65 μm).

► **Sertoli cells** play a crucial role in spermatogenesis, providing a favorable environment for male germ cell differentiation. They are in direct contact with developing spermatogonia, spermatocytes and spermatids during an active period of macromolecular biosynthesis. When spermatogenesis is complete, cytoplasmic extensions with the Sertoli cells are broken and spermatozoa are released. The major regulators of Sertoli cell function are FSH produced by the pituitary and testosterone produced by ► **Leydig cells** of the testis in response to LH from the pituitary.

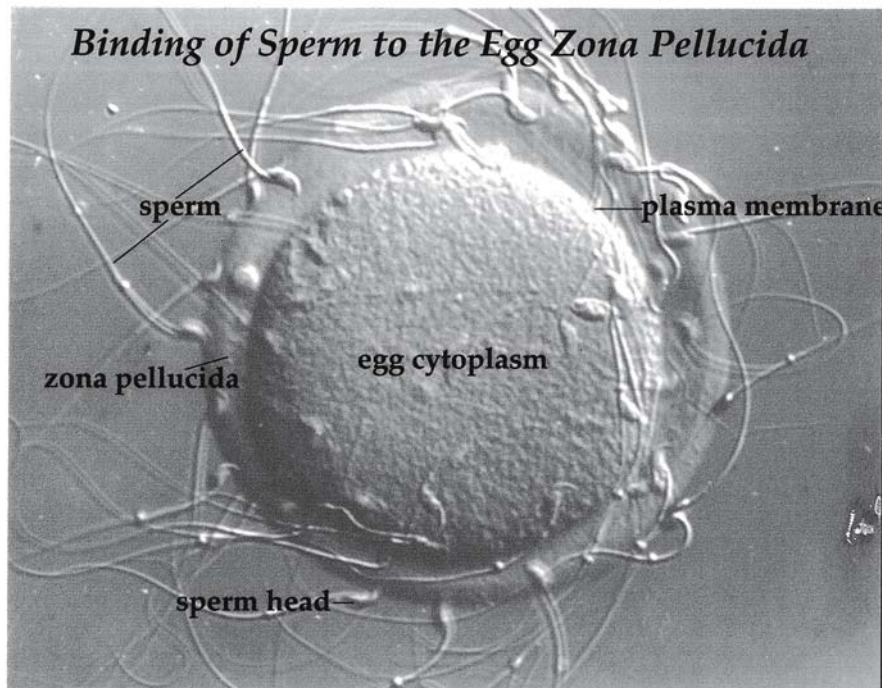
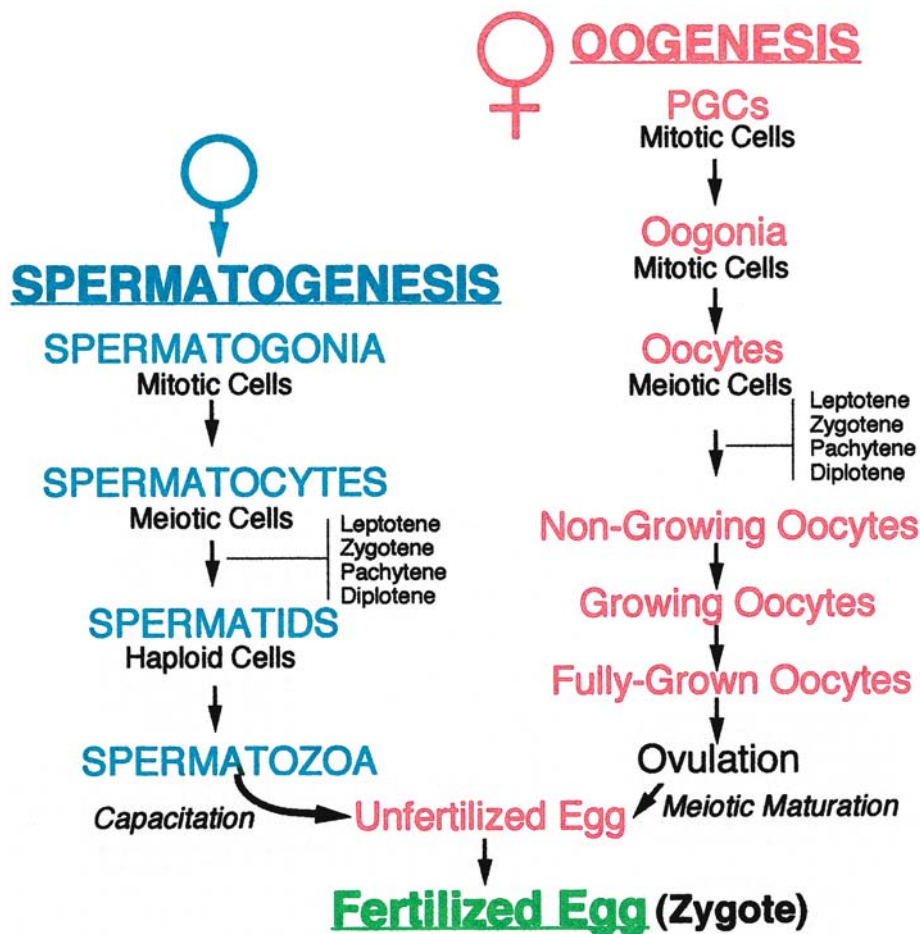
A common feature of spermatogenesis is that the majority of spermatocytes and spermatids fail to complete development. Degeneration of developing germ cells is a physiological phenomenon, with critical phases of extensive cell loss during spermatogonial development, meiotic divisions and spermiogenesis.

Molecular Interactions

Sperm-Egg Interactions

Eggs released from Graafian follicles enter the opening of the oviduct and move to the lower ampulla region where fertilization takes place (5, 7). Very few (~10) ovulated eggs are found in oviducts of mice (humans ~1) and, similarly, very few sperm are found at the site of fertilization (~100–150) as compared to the number of sperm deposited into the female reproductive tract (~10,000,000). It takes ~15 min for ejaculated mouse sperm (human ~30 min) to traverse the reproductive tract and reach the oviduct, possibly guided by chemotaxis regulated by follicular factors. Like eggs, spermatozoa have quite a short lifespan (~24 h) in the female reproductive tract.

Ovulated eggs are surrounded by a ► **zona pellucida** (ZP), an extracellular coat made up of three glycoproteins, called ZP1, ZP2 and ZP3. Follicle cells, interspersed in a matrix of hyaluronic acid, are coupled *via* gap junctions to fully-grown oocytes, with processes that traverse the ZP and contact and connect



Mammalian Fertilization. Figure 1 Top: Outline of mammalian oogenesis and spermatogenesis leading to production of mature eggs and sperm and to fertilization (see text for details). PGCs, primordial germ cells. Bottom: Photomicrograph of mouse sperm bound to the unfertilized mouse egg zona pellucida. Magnification approximately 775x.

with oocyte plasma membrane. The ZP is an oocyte/egg-specific organelle whose glycoprotein components are synthesized and secreted by growing oocytes. The ZP is a relatively thick and homogeneous meshwork of interconnected filaments that increases in thickness to $\sim 6.5 \mu\text{m}$ in mouse (human $\sim 13 \mu\text{m}$) as the oocyte grows and increases in diameter. The ZP of eggs from all mammalian species (including marsupials) consists of ZP1–3 and even egg coats from non-mammalian species (e.g., fish, amphibia, and birds) contain glycoproteins structurally related to ZP1–3.

Much of our knowledge about sperm-egg interactions comes from studies with mice. The ZP is responsible for species-specific sperm binding and induces bound sperm to undergo the acrosome reaction (i.e. cellular exocytosis). Removal of the ZP exposes egg plasma membrane directly to sperm and, as a result, virtually eliminates any barriers to *in vitro* fertilization between mammalian species. In response to fertilization, the ZP hardens and becomes impenetrable to sperm, a reaction referred to as the “slow block” to polyspermy. In addition, the ZP protects unfertilized eggs during ovulation as they are extruded from Graafian follicles and released into the oviduct. The ZP also protects growing embryos as they are transported along the oviduct and into the uterus. Just prior to implantation in the uterus, expanded blastocysts hatch from the ZP.

ZP2 and ZP3 are assembled into long filaments using non-covalent bonds and the filaments are interconnected by ZP1. ZP2 and ZP3 are monomers and are present in about equimolar amounts, whereas ZP1, a dimer composed of identical polypeptides held together by intermolecular disulfides, is present in minor amounts. The three glycoproteins, ZP1–3, are encoded by single copy genes located on different chromosomes (mouse, chromosomes 19, 7 and 5; human, chromosomes 11, 16 and 7). The primary structures of ZP polypeptides from different mammals are closely related to one another; for example, the polypeptide sequences of mouse and human ZP1, ZP2 and ZP3 are $\sim 40\%$, 56% and 67% identical, respectively. Since all three glycoproteins play structural roles in the egg ZP, they probably interact with one another in a similar manner regardless of species.

At fertilization, sperm must recognize and bind to the ZP, undergo the **acrosome reaction** and traverse the ZP in order to reach and fuse with egg plasma membrane. These events are mediated by ZP3 and ZP2 respectively. ZP3 serves as the primary receptor for sperm and initiates the acrosome reaction upon binding of acrosome-intact sperm to the ZP. Acrosome-intact sperm bind to specific O-linked oligosaccharides present on ZP3 in a region of ZP3 polypeptide near the carboxy-terminus (“sperm combining-site”) and, since sperm are highly motile cells, each must bind to tens of thousands of ZP3 molecules located at the surface of

the ZP. Acrosome-intact sperm bind to the ZP using one or more proteins associated with plasma membrane overlying the sperm head. These proteins recognize and bind to ZP3 and are often called “egg binding-proteins”. Although several candidates have been proposed, for example β -galactosyltransferase, sperm protein-56, zonadhesin and mannose- and galactose-binding proteins, the nature of the egg binding-protein remains problematic. In this context, it has recently been proposed that the mammalian sperm protein “Izumo” (named after a Japanese shrine dedicated to marriage) is necessary for fusion of sperm and egg plasma membrane (Inoue N et al (2005) *Nature* 434:234-238).

As a consequence of binding to ZP3, sperm acrosome-react and release their acrosomal contents into the ZP. In this context, the ZP is a highly porous structure that is penetrated by large macromolecules and even by small viruses. The acrosome is a large secretory vesicle that appears in spermatids as a product of the Golgi and is biochemically similar to a lysosome. During the acrosome reaction, multiple fusions occur between plasma membrane and outer acrosomal membrane at the anterior region of the sperm head and hybrid membrane vesicles are formed. As a result, the egg ZP is exposed to the acrosomal contents of bound sperm. ZP3 is the natural agonist that initiates the acrosome reaction following binding of sperm to the ZP. Apparently, multivalent interactions between ZP3 and egg binding-protein(s) on the surface of the sperm head trigger this Ca^{2+} -dependent reaction. ZP2 serves as the secondary sperm receptor to which ZP-associated, acrosome-reacted sperm bind, possibly by using a sperm proteinase called acrosin and/or its zymogen form, proacrosin.

Penetration of the ZP by sperm is achieved by a combination of sperm motility and enzymatic hydrolysis, the latter being catalyzed by acrosin and possibly other acrosomal proteases. Only acrosome-reacted sperm can penetrate the ZP and fuse with egg plasma membrane. The course taken by sperm is indicated by a narrow slit left behind in the ZP of the fertilized egg. In mice, it takes ~ 15 – 20 min for acrosome-reacted sperm to penetrate the ZP and reach the egg plasma membrane. The post-acrosomal plasma membrane of acrosome-reacted sperm fuses with egg plasma membrane. Fusion of the sperm head with the egg is followed by entry of the sperm tail into egg cytoplasm. Several sperm proteins have been implicated in binding and fusion of sperm with egg plasma membrane. One of these, called fertilin, is a member of the ADAM (a disintegrin and metalloprotease domain) family of transmembrane proteins. Fertilin is a heterodimer of α and β subunits, with disintegrin and metalloprotease domains. In mice, but not in humans, binding of acrosome-reacted sperm to egg plasma membrane may be supported by the disintegrin domain of fertilin and

an integrin receptor (perhaps integrin $\alpha 6\beta 1$) on egg plasma membrane. In addition to integrins, CD9, a member of the tetraspanin family, is thought to facilitate sperm-egg membrane interactions in mice through lateral associations with integrins and other egg membrane proteins. Although sperm ADAMa and egg integrins and CD9 are the best characterized proteins involved in mammalian gamete interactions, recent evidence from knock-out mice models suggests that fertilin α/β and integrin $\alpha 6\beta 1$ may not be required for sperm-egg fusion and/or that other, as yet unidentified proteins, participate as well.

Once fertilized, the egg becomes a zygote. Soon after passage of a sperm into the egg cytoplasm, sperm and egg form distinct nuclei, the male and female pronuclei. The two haploid pronuclei approach each other, but remain distinct until after the membrane of each pronucleus has broken down and their chromosomes assemble on a metaphase spindle and progress toward the first cleavage division of the 1-cell embryo (zygote).

Regulatory Mechanisms

Prevention of Polyspermy

Eggs react very rapidly to sperm entry, undergoing both structural and functional changes (5, 7). Among these are fusion of cortical granules with egg plasma membrane accompanied by release of their contents, resumption of the second meiotic division and emission of a second polar body and formation of the egg nucleus. Some responses of the egg to fertilization by a single sperm occur in order to prevent polyspermic fertilization, a lethal event.

Prevention of polyspermy is achieved by immediate changes (within seconds) in electrical properties of egg plasma membrane (fast block) and by slower changes in the properties of the ZP; the latter represent the so-called “**▶zona reaction**” (slow block). The zona reaction occurs within a few minutes of fertilization and is induced by release of **▶cortical granule** contents into the ZP (**▶cortical reaction**). Cortical granules are small secretory vesicles in the egg cortex that fuse with egg plasma membrane upon activation, probably through release of Ca^{2+} from egg cytoplasmic stores. As a consequence of the cortical reaction and release of enzymes, the ZP hardens (i.e., becomes less soluble), bound acrosome-reacted sperm fall off the ZP and free-swimming acrosome-intact sperm are prevented from binding to the ZP.

Cortical granule enzymes, which include a variety of proteinases and glycosidases, modify ZP2 and ZP3. As a result, ZP3 no longer acts as a sperm receptor for acrosome-intact sperm and ZP2 cannot maintain binding of acrosome-reacted sperm. For example, a proteinase is released from activated mouse eggs that cleaves ZP2 by limited proteolysis, converting ZP2 to ZP2_f, a reaction that modulates ZP structure and

function. For ZP3, there is no evidence for proteolysis following fertilization, but it is likely that one or more glycosidases released from cortical granules are involved in inactivation of ZP3 sperm receptor activity. After fertilization in the ampulla region of the oviduct, the egg is transported down the oviduct and into the uterus (human ~80 h, mice ~72 h) by ciliary action and muscular peristalsis. The ZP is essential during transport, as it surrounds and protects the early embryo preventing it from disintegrating or sticking to the walls of the oviduct. Once the embryo approaches the uterus, it hatches from the ZP and implants in the uterine wall, a process probably aided by a “hatching” enzyme originating from the embryo itself. Localized proteolysis of the ZP is likely to account for the hole through which an expanded blastocyst finally emerges to implant in the uterus.

Reference

1. Alberts B et al (1994) *Molecular Biology of the Cell*, Garland Publishing, New York, pp 1011–1035
2. Eddy EM, O'Brien DA (1994) The Spermatozoon. In: Knobil E et al (eds) *The Physiology of Reproduction*, vol 1. Raven Press, New York
3. Grudzinskas JG, Yovich JL (1995) *Gametes – The Oocyte*. Cambridge University Press, Cambridge p 441
4. Grudzinskas JG, Yovich JL (1995) *Gametes – The Spermatozoon*. Cambridge University Press, Cambridge p 316
5. Saling PM (1996) Fertilization: Mammalian Gamete Interactions. In: Adashi EY et al (eds) *Reproductive Endocrinology, Surgery, and Technology*, vol 1. Lippincott-Raven Publishers, Philadelphia
6. Wassarman PM, Albertini DF (1994) The Mammalian Ovum. In: Knobil E et al (eds) *The Physiology of Reproduction*, vol 1. Raven Press, New York
7. Yanagimachi R (1994) Mammalian Fertilization. In: Knobil E et al (eds) *The Physiology of Reproduction*, vol 1. Raven Press, New York
8. Primakoff P, Myles DG (2002) Penetration, adhesion, and fusion in mammalian sperm-egg interaction. *Science* 296:2183–2185
9. Wassarman PM (2003) Fertilization. *ChemTracts-Biochemistry and Molecular Biology* 16:117–204

Manic Depression

▶Bipolar Affective Disorder

ManNAc

▶N-Acetyl-D-Mannosamine

Mannose 6-Phosphate Receptor

Definition

Precursors of lysosomal enzymes are tagged with a luminal mannose 6-phosphate residue in the *cis*-Golgi. These residues are bound with high affinity by mannose 6-phosphate receptor (MPRs) which in turn mediate sorting from the *trans*-Golgi network (TGN) to the late endosome. Due to the low pH in the late endosomal lumen, the MPRs release their ligands and are recycled back to the TGN.

- [Molecular Chaperones and Cochaperones](#)
- [Protein and Membrane Transport in Eukaryotic Cells](#)

Map Distances

Definition

Maps of the human genome are generally of two types: genetic and physical. Genetic maps measure the amount of recombination between two loci and genetic map units, and are quantified as either % recombination (theta) or cM (centimorgan; Morgan unit(s)) between two loci. Physical maps aim to quantify the actual amount of DNA, usually in base pairs between two loci. One percent recombination is roughly equivalent to 1,000,000 base pairs of DNA.

- [Mendelian Forms of Human Hypertension and Mechanisms of Disease](#)

MAPK

- [Mitogen-Activated Protein Kinases](#)

MAPK Pathway

Definition

MAPK Pathway denotes one of the intracellular signaling pathways which may be initiated through activation of plasma membrane receptors of the ► [EGFR](#) family (► [HER1](#), ► [HER2](#), [HER3](#), [HER4](#)); this pathway involves sequential activation of several effector molecules including Ras, Raf-1, and ► [mitogen-activated protein kinase \(MAPK\)](#), resulting in

transcription of target genes associated with cellular proliferation.

- [Breast Cancer](#)

Mapping Function

Definition

Mapping function is the mathematical relationship between measured map distance and actual recombination frequency. The function transforms the recombination fraction between two loci into an additive distance measure.

- [Genetic Epidemiology](#)

MAPs

- [Microtubule Associated Proteins](#)

Marfan Syndrome

PHILIP F. GIAMPIETRO

Department of Medical Genetic Services, Marshfield Clinic, Marshfield, WI, USA
giampietro.philip@marshfieldclinic.org

Synonyms

MFS

Definition

Marfan syndrome is an autosomal dominant ► [pleiotropic](#) disorder of ► [connective tissue](#) with an incidence in the general population of approximately two to three in 10,000. There is no particular ethnic predisposition. The major clinical findings associated with Marfan syndrome include aortic root dilatation that may lead to dissection, ► [ectopia lentis](#) and ► [dural ectasia](#) (4).

Characteristics

► [Mitral valve prolapse](#), aortic root dilatation, mitral regurgitation, aortic valve regurgitation and dissection and rupture of the ► [aorta](#) are the most common cardiovascular findings in patients with Marfan syndrome. Additional ocular signs include myopia, increased axial globe length that may lead to retinal detachment and corneal flatness. Skeletal features,

which may collectively represent a single major criteria and lead to earlier recognition of Marfan syndrome, include increased height and arm span, reduced ▶[upper to lower segment ratio](#), ▶[pectus excavatum](#) or ▶[pectus carinatum](#), joint laxity, scoliosis, ▶[acetabular protrusion](#), ▶[pes planus](#), ▶[arachnodactyly](#), high arched palate, dental crowding and hammer toes. Craniofacial features associated with Marfan syndrome include down-slanting palpebral fissures, ▶[malar hypoplasia](#), ▶[enophthalmos](#) and retrognathia. A diagnosis of Marfan syndrome should be considered in individuals who present with spontaneous ▶[pneumothorax](#). Minor features in individuals with Marfan syndrome include inguinal hernia and ▶[striae atrophicae](#).

Evaluation of individuals suspected of having Marfan syndrome is a multidisciplinary process. The evaluation should consist of clinical genetics, cardiology, ophthalmology and possibly orthopedics consultations. For a clinical diagnosis of Marfan syndrome, in the absence of family history, an affected person should display major criteria in at least two organ systems and involvement of a third organ system (1). In the presence of a positive family history, an affected person should display one major criterion in an organ system and involvement of a second organ system. Approximately 30% of cases of Marfan syndrome occur in the absence of a positive family history and represent new mutations in the fibrillin-1 (*FBN1*) gene.

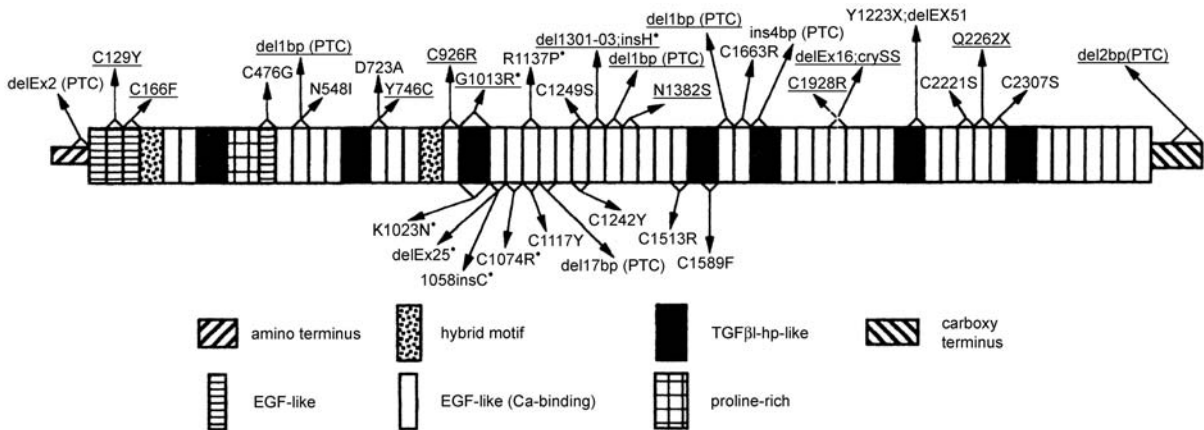
Marfan syndrome has clinical overlap with other conditions including congenital contractural arachnodactyly, familial aortic aneurysm, familial aortic dissection, familial ectopia lentis, MASS phenotype (myopia, mitral valve prolapse, mild aortic dilatation and skeletal involvement) and homocystinuria (1). Ehlers Danlos syndromes Type I and II are associated with joint hypermobility, skin fragility, mitral valve prolapse and aortic root dilatation. Individuals with ▶[Klinefelter](#) and ▶[fragile X syndromes](#) may have a marfanoid body habitus. Fragile X syndrome is associated with pectus excavatum, pes planus, mitral valve prolapse and joint hypermobility. Chromosomal studies, fragile X DNA testing and serum amino acid quantitation demonstrating elevated homocysteine levels may be instrumental in ruling out these conditions that have phenotypic overlap with Marfan syndrome.

Dural ectasia may be associated with back pain, lower radicular pain, leg weakness and headaches, or may be asymptomatic. Evaluation by computerized tomography or magnetic resonance imaging evaluation to determine if dural ectasia is present may be helpful in the diagnostic process. Posterior laminectomy may be indicated if the back pain is severe. Acetabular protrusion may also be asymptomatic. Radiological evaluation by the crossing of the acetabular and iliopectineal lines may also be of help when there is

diagnostic uncertainty. Clinical evaluation, including echocardiography, of both parents of a child in whom the diagnosis of Marfan syndrome is being considered may be necessary as part of the evaluation process. Developmental dysplasia of the hip occurs with an incidence of 2%. Osteoporosis and osteopenia have been reported in Marfan syndrome. No reported association exists between decreased bone mineral density and fracture in Marfan syndrome.

Life expectancy has increased among females (50% probability of survival to age 72) and males (50% probability of survival to age 65), largely due to prophylactic cardiothoracic surgical intervention and improved survival after cardiothoracic surgery (6). Since these data were reported in the mid 1990s, the life expectancy for patients with Marfan syndrome is probably greater. Early recognition and diagnosis of Marfan syndrome is important in order to enable implementation of treatment for cardiovascular symptoms. ▶[Beta-adrenergic blockers](#), such as atenolol, and calcium channel blockers are used to decrease the rate of aortic root enlargement and cardiovascular events. Prophylactic repair of thoracic aneurysms is recommended when the diameter reaches between 5.0 and 5.5 cm. The “composite graft” technique involves suturing a prosthetic aortic valve to the end of a Dacron conduit and attaching it to the aortic annulus. Patients who undergo this procedure will need to be maintained on coumadin in order to prevent ▶[thromboembolism](#). ▶[Endocarditis](#) prophylaxis should be implemented for all patients with valvular heart disease, including a composite graft repair or placement of an artificial valve. A “valve sparing” procedure may be recommended in certain instances. Patients should avoid contact sports and isometric exercise in order to minimize shearing forces on the aorta. Cardiac arrhythmias may occur in patients with Marfan syndrome that can be related to underlying left ventricular dysfunction, mitral valve prolapse and mitral regurgitation.

Periodic ophthalmologic evaluation is necessary to assess amblyopia, lens subluxation, anterior chamber abnormalities, cataracts, glaucoma and retinal detachment. Orthopedic evaluation should be sought for concerns related to scoliosis, joint hypermobility, pes planus and acetabular protrusion. Surgical correction of scoliosis may be necessary if bracing has failed or if the curve is greater than 40°. Estradiol has been demonstrated to induce epiphyseal closure and decrease final adult height when administered to premenarchal females. Periodic assessment of pulmonary function should be performed in adults and children due to the association of restrictive lung disease, emphysema, asthma and obstructive sleep apnea. Cardiac assessment and echocardiography should be performed every 6–8 weeks during pregnancy because there is an



Marfan Syndrome. Figure 1 *FBN1* gene containing 65 exons and various epidermal growth factor-like and transforming growth factor β -like domains. A multitude of *FBN1* mutations are illustrated. Reprinted with permission: Pyeritz RE (2002) Marfan syndrome and other disorders of fibrillin. In: Rimo DL, Connor MJ, Pyeritz RE, Korf BR (eds) Emery and Rimo's Principles and Practice of Medical Genetics. Fourth Edition. Churchill Livingstone, New York, p 4000 (Fig. 149.25).

increased risk of aortic dissection. As patients with Marfan syndrome survive longer, it is anticipated that degenerative arthritis of the hip, knees and other joints will represent new health concerns. The National Marfan Foundation provides a wealth of information and resources for patients and families (www.marfan.org).

Cellular and Molecular Recognition

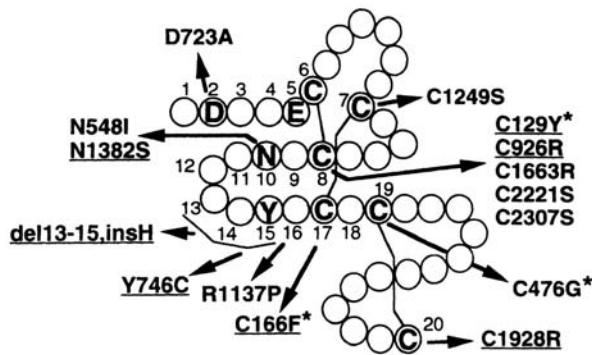
Marfan syndrome is caused by mutations in the *FBN1* gene on chromosome 15q21.1 (2). Fibrillin-1, a glycoprotein with a molecular weight of 350 kD, represents the major component of the 10–12 nm diameter microfibrils, the scaffolding on which tropoelastin is deposited (5). With rotary shadowing electron microscopy, microfibrils appear as “beads on a string,” having a periodicity of 50–55 nm. Microfibrils form linkages to elastin fibers, other components of the extracellular matrix and neighboring cells. They also are the principal component of the ocular zonules and are responsible for anchoring the lens. Microfibrils are composed of a variety of proteins, including emilin, fibrillin-1, fibrillin-2, lysyl oxidase and microfibril-associated proteins (MAP)-1, MAP-2, MAP-3 and MAP-4. *FBN2* deficiency is associated with congenital contractural arachnodactyly. Using electron microscopy, distinctive differences in microfibrils studied in patients with Marfan syndrome, as compared to unaffected patients, have been identified.

The *FBN1* gene is composed of 65 exons encoding 2871 amino acids and 4 additional alternatively spliced exons at the 5' end of the gene (Fig. 1). Fibrillin-1 is composed of a series of 47 epidermal growth factor (EGF)-like domains, each containing six cysteine-repeats.

Latent transforming growth factor- β (TGF- β) binding protein-like domains are interspersed between the EGF-like domains. Calcium binding is mediated by the presence of the EGF domain structure, three intradomain disulfide linkages, and negatively charged amino acid residues, including aspartic acid and glutamic acid (Fig. 2). The restriction of interdomain flexibility, which may facilitate protein-protein interactions and provide protection against proteolytic cleavage, is due to calcium binding. A series of 12 calcium binding EGF domains in the central portion of the fibrillin-1 molecule are thought to form a rigid rod-like structure that may be necessary for microfibril assembly. At the present time there is no evidence for a second Marfan syndrome locus.

Recent studies have provided evidence for an interaction between fibrillin-1 and a large latent complex in the extracellular matrix, composed of TGF- β binding protein and a covalently bound small latent complex composed of TGF- β and latency associated peptide dimers. Haploinsufficiency of fibrillin results in destabilization of this interaction and a liberation of the TGF- β from the matrix, thus enabling cytokine activation. The end results of this process hypothetically include impediment of the septation of distal airspace septation in lung tissue, myxomatous valve changes and bone overgrowth (8).

More than 500 mutations in the *FBN1* gene have been identified. The mutations are distributed among the 65 exons and usually represent a unique occurrence within each family studied. The majority of mutations are missense mutations that alter one of the conserved cysteine residues of the EGF or TGF- β domains. Other types of mutations include frameshift, insertion,



Marfan Syndrome. Figure 2 Mutations in fibrillin-1 EGF motifs in patients with Marfan syndrome. Many mutations affect calcium binding. Reprinted with permission: Pyeritz RE (2002) Marfan syndrome and other disorders of fibrillin. In: Rimoin DL, Connor MJ, Pyeritz RE, Korf BR (eds) Emery and Rimoin's Principles and Practice of Medical Genetics. Fourth Edition. Churchill Livingstone, New York, p 3998 (Fig. 149.24).

deletion and splice site alterations. Approximately 70% of *FBN1* mutations are identifiable in individuals who satisfy the diagnostic criterion for Marfan syndrome. Missense mutations in *FBN1* that affect cysteine residues important for proper EGF structure or amino acids involved in calcium binding act in a dominant negative fashion, usually resulting in a moderately severe phenotype. Some *FBN1* mutations that impair folding of EGF domains may cause protein misfolding and result in retention of mutant fibrillin-1 within the endoplasmic reticulum, thereby altering intracellular trafficking (7). The severe neonatal form of the disease appears to be caused by missense mutations in exons 24 to 26 or skipping of exons 31 or 32.

The large number of exons in the *FBN1* gene makes direct DNA sequencing of the entire gene impractical. Various methods have been utilized for *FBN1* analysis, including single strand conformation polymorphism analysis, heteroduplex analysis and long reverse transcription-polymerase chain reaction (►RT-PCR). Denaturing high performance liquid chromatography for each *FBN1* exon followed by DNA sequencing of exons with abnormal chromatogram profiles has shown promise as a mutation detection method due to its high sensitivity, specificity and relatively low cost. Identification of a *FBN1* mutation in a patient with Marfan syndrome is necessary for prenatal diagnosis. Linkage analysis may be incorporated for prenatal diagnosis and identification of unaffected family members in situations where a causative *FBN1* mutation has not been identified.

Type-1 fibrillinopathies is a term that is used to refer to Marfan syndrome and a series of related connective

tissue disorders that are caused by mutations in *FBN1*. *FBN1* mutations have been identified in patients with the MASS phenotype and in several patients with Shprintzen-Goldberg syndrome. The identification of an in-frame *FBN1* deletion in affected family members with autosomal dominant Weill-Marchesani syndrome (3) and a mutation causing a single base substitution at codon 1796 in members of a three-generation kindred with kyphoscoliosis have widened the phenotypic spectrum associated with *FBN1* gene mutations.

Clinical Relevance

Marfan syndrome is a relatively common disorder of connective tissue. Early recognition of the clinical features of Marfan syndrome by health professionals and patients may help prolong life and provide optimal treatment for patients. Understanding of the molecular basis of Marfan syndrome and other type-1 fibrillinopathies have led to an improved understanding of the properties of microfibrils, the major constituents of connective tissue. Genetic heterogeneity of *FBN1* mutations has been associated with varying phenotypic differences among patients with Marfan syndrome and other distinct syndromes. There is evidence that inherited disorders involving other connective tissue proteins are associated with the development of aortic aneurysms. For example, a locus for familial aortic aneurysm that is distinct from the *FBN1* locus has been identified at 11q23.2-q24. Further studies are being performed to identify genes that are associated with familial aortic aneurysm. At the present time no causative gene has been identified. Because of the complexity of connective tissue and the heterogeneity of familial aortic aneurysm, it is anticipated that other loci will be identified.

References

1. De Paepe A, Devereux RB, Dietz HC et al (1996) Revised diagnostic criteria for the Marfan syndrome. *Am J Med Genet* 62:417-426
2. Dietz HC, Cutting GR, Pyeritz RE et al (1991) Marfan syndrome caused by a recurrent *de novo* missense mutation in the fibrillin gene. *Nature* 352:337-339
3. Faivre L, Gorlin RJ, Wirtz MK et al (2003) In frame fibrillin-1 gene deletion in autosomal dominant Weill-Marchesani syndrome. *J Med Genet* 40:34-36
4. Pyeritz RE, McKusick VA (1979) The Marfan syndrome: diagnosis and management. *N Engl J Med* 300:772-777
5. Sakai LY, Keene DR, Engvall E (1986) Fibrillin, a new 350-kD glycoprotein, is a component of extracellular microfibrils. *J Cell Biol* 103:2499-2509
6. Silverman DI, Burton KJ, Gray J et al (1995) Life expectancy in the Marfan syndrome. *Am J Cardiol* 75:157-160

7. Whiteman P, Handford PA (2003) Defective secretion of recombinant fragments of fibrillin-1: implications of protein misfolding for the pathogenesis of Marfan syndrome and related disorders. *Hum Mol Genet* 12:727–737
8. Neptune ER, Frischmeyer PA, Arking DE et al (2003) Dysregulation of TGF-beta activation contributes to pathogenesis in Marfan syndrome. *Nat Genet* 33:407–411

Marker

Definition

Marker denotes a specific biomolecule related to the cause of a disease.

► [Proteomics in Microfluidic Systems](#)

Marker X Syndrome

► [Fragile X Mental Retardation Syndrome](#)

Martin Bell Syndrome

► [Fragile X Mental Retardation Syndrome](#)

MAS

► [McCune-Albright Syndrome](#)

Mask

Definition

A Mask denotes a transparency film (or other material) that selectively exposes and blocks light to a substrate that is placed below.

► [Proteomics in Microfluidic Systems](#)

Mass Action Law

The mass action law is the thermodynamic description relating the populations of unbound species and complex in interacting molecules; for example, if two molecules A and B form a reversible complex with the molar affinity constant K_A it takes the form $[AB] = K_A[A][B]$, with the molar concentrations $[A]$, $[B]$, and $[AB]$ of the unbound species A, B, and their complex, respectively.

► [Analytical Ultracentrifugation](#)

Mass Spectrometer

Definition

A mass spectrometer is a device designed to produce a mass spectrum of a sample to find out its composition. This is normally achieved by ionizing the sample and separating ions of differing masses and recording their relative abundance by measuring intensities of ion flux. A typical mass spectrometer comprises three parts: an ion source, a mass analyzer, and a detector. Mass spectroscopy allows detection of compounds by separating ions by their unique mass.

Techniques for ionization have been key to determining what types of samples can be analyzed by mass spectrometry. Two techniques which are often used with biological samples include electrospray ionization (John Fenn, Nobel Prize in Chemistry in 2002) and Matrix-assisted Laser Desorption Ionization (► [MALDI](#)) (for which Koichi Tanaka, Nobel prize the same year). The produced ions can then enter the mass analyzer.

Mass analyzers of many varieties have been produced.

► [The Time-of-flight](#) (TOF) analyzer is typically integrated with MALDI ion sources. Ions are boosted to the same kinetic energy by passage through an electric field, and the times they take to reach the detector are measured. The smallest ions (or those most highly charged) move most rapidly. Quadrupoles and quadrupole ion traps use electrical fields to selectively stabilize or destabilize ions falling within a narrow window of m/z values. Sector instruments change the direction ions are flying through the mass analyzer. Fourier Transform Mass Spectrometry uses perturbations in magnetic fields to measure m/z values extremely accurately. The best mass analyzer for an experiment depends upon the type of information to be gleaned from the experiment.

The final element of the mass spectrometer is the detector. Typically, some type of electron multiplier is used. Because the number of ions leaving the mass analyzer at a particular instant is typically quite small, significant amplification is necessary to get a signal.

► [Affinity Chromatography and *In Vitro* Binding \(Beads\)](#)

► [Mass Spectrometry: MALDI](#)

► [Protein Interaction Analysis: Chemical Cross-Linking](#)

► [Protein-Protein Interaction](#)

► [Proteomics in Human-Pathogen Interactions](#)

► [Proteomics in Cancer](#)

► [Two Hybrid System](#)

Mass Spectrometric Fragmentation

► [Mass Spectrometry: MS/MS](#)

Mass Spectrometry

Definition

Mass spectrometry is a technique for separating ions by their mass-to-charge (m/z) ratios.

► [Mass Spectrometry: ESI](#)

► [Mass Spectrometry: MALDI](#)

► [Mass Spectrometry: MS/MS](#)

► [Mass Spectrometry: Quantitation](#)

► [Mass Spectrometry: SELDI](#)

► [PNA Chips](#)

► [Proteomics in Cardiovascular Disease](#)

Mass Spectrometry: ESI

EVA-CHRISTINA MÜLLER

Max Delbrück Center for Molecular Medicine, Berlin, Germany

emu@mdc-berlin.de

Definition

► [Electrospray ionization \(ESI\)](#) is a method of generating ions directly from a solution. Different types of mass analyzers and detectors can be combined

with an ESI source to build a mass spectrometer for the determination of the molecular weights of peptides and proteins and the mass spectrometric sequencing of peptides. Since the ionization process takes place at atmospheric pressure with little thermal input, ESI is a soft ionization technique and is therefore especially suitable for biological samples such as proteins, non-covalent interaction complexes, peptides, drugs and DNA segments.

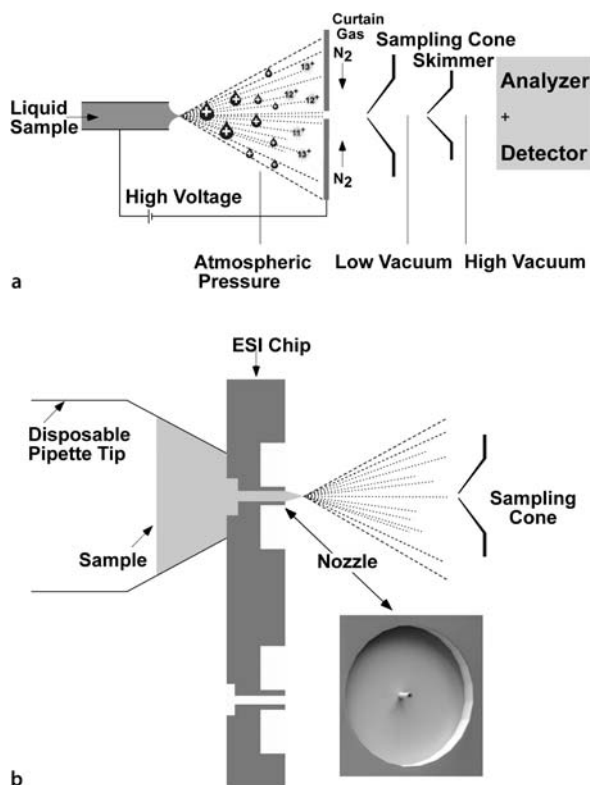
Characteristics

The breakthrough for ESI of biological samples came in 1988 at a symposium in San Francisco with the presentation of molecular weight determinations of polypeptides and proteins by John Fenn. Fenn showed that a molecular weight accuracy of 0.01% could be obtained by applying a signal-averaging method to the multiple ions formed in the ESI process (1). These multiply charged ions with low mass/charge ratio could be easily analyzed by any mass analyzer.

In the electrospray ionization (ESI) process, a flow of sample solution is pumped through a narrow-bore metal capillary held at an electric potential of a few kilovolts relative to a counter electrode (Fig. 1A). Drops of the solution, emerging from the tip of the needle, spray as a mist of very fine, charged droplets. As these charged droplets evaporate in a flow of a warm drying gas, the increasing coulombic repulsive forces of the ions within the droplets finally exceed the surface tension, causing the droplets to undergo fission. Successive fissions ultimately give rise to droplets containing a single solute molecule that retains the droplet charge, left as the remaining solvent evaporates. Multiply charged ions are commonly observed. To produce negative ions, which are often formed by proton abstraction, similar voltages of opposite sign are applied.

Suitable solvents for the analysis of peptides and proteins include water, methanol and acetonitrile (e.g. methanol/water/formic acid, 50/49.9/0.1, v/v/v). Non-volatile agents like salts (e.g. NaCl, KH_2PO_4 , Tris), detergents (e.g. Tween, Triton, SDS), chaotropic agents (e.g. urea, guanidinium salts) and solvents (e.g. DMSO, glycerol) interfere with the ESI process and suppress the ionization of the analyte. In most cases a desalting step is necessary for biological samples to obtain expressive ESI spectra.

The sample solution flow rate in an electrospray ion source is most commonly in the range of 3–20 $\mu\text{l}/\text{min}$. Experimental evidence has shown that the optimal sensitivities of detection using electrospray can be achieved by delivering lower flow rates. The observed ion current correlates with the concentration of the analyte rather than with the flow rate. Micro-electrospray ion sources with flow rates in the range of 0.2–1 $\mu\text{l}/\text{min}$ were developed and often used with



Mass Spectrometry: ESI. Figure 1 (a) Scheme for a nano-electrospray mass spectrometer. Droplets of the analyte emerge from the capillary and pass through the sampling cone and skimmer toward the mass analyzer. The droplets shrink by losing solvent molecules, thereby desolvating the ions they contain. (b) Scheme for an ion source with an ESI chip. A disposable pipette tip containing sample presses and seals against the inlet-side of the chip with the nano-electrospray plume spraying towards the mass spectrometer's orifice. A scanning electron micrograph pointing down towards a single nozzle is shown in the inset. The ESI chip has a 10×10 array of nozzles. (By courtesy of Sheng Zhang).

continuous flow in connection with a capillary liquid chromatography (CapLC) system, comprising a low flow capillary pump and an autosampler. With such a configuration automated data acquisition can be implemented.

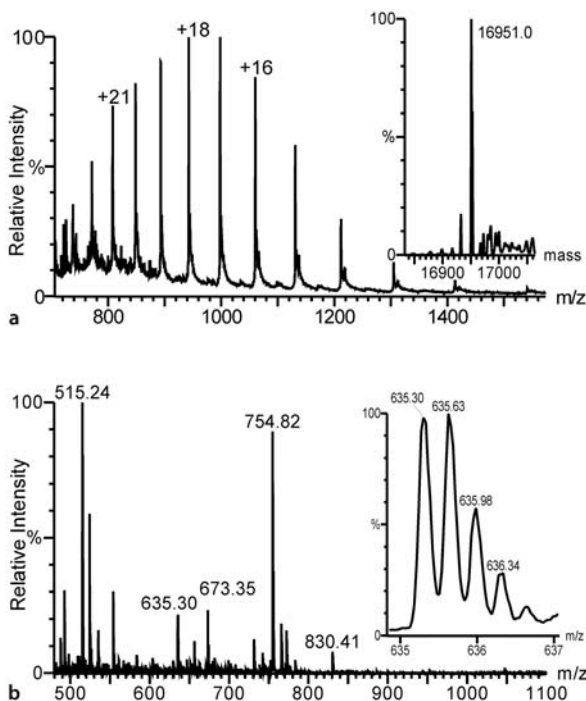
The nano-electrospray interface designed by Wilm and Mann (2) operates at low nl/min flow rates. The sample glass capillaries are metal coated and a backing gas pressure to aid flow stability but without pumping to impose a specific solvent flow is installed. ▶ **Tandem mass spectrometry MS/MS** characterization at the sub-femtomole level of samples present in 1–2 μl volumes can be accomplished with the nano-electrospray source. However, the low sample throughput is a disadvantage for ▶ **proteomics** and clinical approaches.

An automated nano-ESI-MS/MS method was described by Zhang et al. (3). It consists of a liquid handling robot (NanoMate) with a 96-well plate, a rack of 96 disposable, conductive pipette tips and a microchip containing a 10×10 array of nano-electrospray nozzles (ESI Chip). The nozzles are etched into the planar surface of a silicon microchip. Opposite to the surface containing the nozzles, a 10×10 array of through-chip channels is located. This is the inlet-side of the chip, where samples are introduced *via* a pipette tip (Fig. 1B). Typical flow rates are in the range of 100–200 nl/min. With the NanoMate, 100 relatively high sample throughput analyses can be performed without carryover.

ESI is compatible with many mass analyzers such as ion trap, quadrupole, hybrid quadrupole time of flight (TOF) and Fourier transform ion cyclotron resonance analyzers and is used for proteomics (4) and clinical analytics. Ion traps are robust, sensitive and relatively inexpensive and so have produced much of the proteomics data reported in the literature. More extensive and informative fragment ion (MS/MS) spectra are often generated in hybrid quadrupole TOF instruments. The combination of ESI and FTICR-MS is particularly powerful, because the trapping efficiency and resolving power increases with increasing charge (for a given mass) in the FTICR-MS and ESI generates multiply charged ions. The unsurpassed resolution and accuracy of the FTICR-MS technique facilitates the screening of complex biological samples without any extensive pre-purification (5).

As examples of protein and peptide molecular weight determination, mass spectra of horse heart apomyoglobin (Fig. 2a) and a peptide mixture of bovine adrenodoxin I peptides resulting from a digestion with the endoprotease GluC are shown. Myoglobin with the theoretical molecular weight of 16,951 D is analyzed with ESI. The process results in a typical charge pattern, an approximate Gaussian distribution of ion intensity. The resulting mass-to-charge ratio, from 700–2000 Th, can be easily measured in any mass analyzer. The charge pattern can be deconvoluted and the mass of the uncharged ion determined to a much higher accuracy than if the interpretation of data was based on a single ion. Averaging the results of multiply charged peaks leads to mass measurement with an accuracy of 0.01%.

The ability to retain fragile complexes during ESI enables the study of protein-protein and protein-drug complexes. In these cases special care must be taken in choosing the appropriate solvent for the analyses. Many proteins can be denatured and caused to unfold when subjected to extremes of pH, detergents or solutions containing high concentrations of organic solvents. An example of a noncovalent protein



Mass Spectrometry: ESI. Figure 2 (a) Positive ion electrospray mass spectrum of horse heart apo-myoglobin (myoglobin without the heme group). The insert shows the deconvoluted peak resulting from the transformation to the mass scale. The theoretical molecular weight is 16951.5Da. (b) Mass spectrum of a bovine adrenodoxin 1 peptide mixture resulting from a digestion with the endoprotease GluC. A triply charged peptide is shown in the inset with high resolution.

complex is myoglobin with a heme prosthetic group. The heme group is located in a hydrophobic crevice in the myoglobin molecules. Native myoglobin can be measured in water at a pH near physiological conditions. If acetonitrile is added to 50% by volume and at the same pH, the majority of the ion intensity converts to the apo-myoglobin form. If the pH of the water solution is dropped to pH 2.6, complete denaturation of the complex occurs. The automated nano-ESI system was used to investigate quantitative noncovalent interactions between ribonuclease A (RNase A) and cytidylic acid ligands, a well-characterized model protein-ligand complex and between an inactive endocellulase mutant and four oligosaccharide ligands (3).

The sensitivity and selectivity of analyte detection using electrospray MS and MS/MS are currently such that the limiting factors in trace detection frequently relate to sample preparation and presentation to the mass spectrometer.

Clinical Relevance

ESI with MS/MS is becoming an increasingly important analytical technology in the clinical laboratory environment (6). Immunoassays are very sensitive and can be very easily automated for clinical tests of one component. ESI-MS/MS is best suited for applications in which more than one compound must be measured simultaneously or tedious specimen preparation protocols or prohibitive interference problems occur.

This technology offers a new vision to newborn-screening programs that have the ability to screen for 30 or more metabolic disorders in a single analysis from one small disk of dried blood. It is evaluated as one of the most important advancements in neonatal screening since the introduction of a bacterial inhibition assay for phenylketonuria (7).

For patients at risk of inherited disorders of purine and pyridine metabolism, a rapid and specific method of diagnosis was developed based on ESI-MS/MS after chromatographic separation. This method proved to be effective in the detection of purine nucleoside phosphorylase deficiency, dihydropyrimidine dehydrogenase deficiency, ornithine transcarbamylase deficiency, molybdenum cofactor deficiency and adenylosuccinase deficiency.

In addition, ESI-MS/MS procedures to detect three immunosuppressant drugs (tacrolimus, sirolimus and cyclosporin) are used for the development of multiple-drug therapeutic regimens in specialized clinical laboratories.

Drug screening directly from urine, using HPLC ESI-MS/MS, without the need for prior sample extraction has been also demonstrated. Furthermore, the method is an essential technology for the laboratories that are charged with detecting drug doping among athletes and drug abuse.

References

1. Fenn JB, Mann M, Meng CK et al (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science* 246:64–71
2. Wilm M, Mann M (1996) Analytical properties of the nanoelectrospray ion source. *Anal Chem* 68:1–8
3. Zhang S, Van Pelt CK and Wilson DB (2003) Quantitative determination of noncovalent binding interactions using automated nanoelectrospray mass spectrometry. *Anal Chem* 75:3010–3018
4. Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. *Nature* 422:198–207
5. Bergquist J (2003) FTICR mass spectrometry in proteomics. *Curr Opin Mol Ther* 5:310–314
6. Dooley KC (2003) Tandem mass spectrometry in the clinical chemistry laboratory. *Clin Biochem* 36:471–481
7. Chace DH, Kalas TA, Naylor EW (2003) Use of tandem mass spectrometry for multianalyte screening of dried blood specimens from newborns. *Clin Chem* 49:1797–1817

Mass Spectrometry: MALDI

JOHAN GOBOM

Max Planck Institute for Molecular Genetics,
Berlin, Germany
gobom@molgen.mpg.de

Synonyms

Matrix-Assisted Laser Desorption/Ionization

Definition

Matrix-assisted laser [▶desorption/ionisation](#) (MALDI), an ionisation technique for mass spectrometric analysis of large biomolecules, was reported in 1988 by Franz Hillenkamp and Michael Karas of the University of Münster (1). Analytical samples are prepared for MALDI by embedding the analyte molecules in a condensed (most often solid) matrix of small molecules present at high molar excess ($>10^4$ – 10^7). The matrix has a very low vapour pressure and is highly absorbing at the wavelength of the desorption/ionisation laser. In a MALDI experiment a brief laser pulse irradiates a spot on the sample whereby the matrix molecules absorb the laser energy, resulting in ablation of a small volume of matrix and desorption of the embedded analyte molecules. In the expanding plume of desorbed material following the laser pulse, analyte molecules are ionised through processes that have not yet been fully elucidated (2), leading to the formation of predominantly singly charged positive and negative analyte ions.

Characteristics

MALDI – a Soft Ionisation Technique

Mass spectrometric analysis requires analytes in the form of gas phase molecular ions. These can be manipulated by magnetic and/or electric fields in a mass analyser and thus separated according to their mass-to-charge ratios (m/z), allowing one to determine their molecular mass. A long-time problem with mass spectrometry of large biomolecules was that the ionisation techniques known until the mid 80's were inefficient or led to decomposition of the analyte molecules, largely restricting mass spectrometry to molecules below 10 kilo-[▶Dalton](#) (kDa) and hampering the sensitivity of the technique. These limitations were overcome by the introduction of two new ionisation techniques, MALDI and electrospray ionisation (ESI), extending the accessible mass range to mega-Daltons. Due to the comparatively low degree of analyte decomposition observed with MALDI and ESI, these are known as soft ionisation techniques.

Desorption/Ionisation Lasers

The most commonly used lasers on commercial instruments are nitrogen lasers emitting in the ultra-violet (UV) at 337 nm. The beam is typically focused to a spot with a diameter of 10–500 μm , irradiating the surface with an effect of 10^6 – 10^7 W/cm^2 . The laser energy must be transferred to the sample in a very short time in order to avoid thermal decomposition of the analyte. Typically, pulse-widths of a few nanoseconds are used. The laser energy is an important parameter and is controlled by a laser beam attenuator. Best results are obtained with laser energies just above the threshold for desorption/ionisation. Excessive energy results in thermal decomposition of analyte molecules, increased chemical noise and saturation of the mass spectrometric detector. A current development is directed towards increasing the laser repetition rate in order to shorten the analysis time. For example, by using solid-state triple-wavelength Nd:[▶YAG](#) UV lasers emitting at 355 nm, kHz pulse rates can be achieved, thus making it possible to record over a thousand single-shot MALDI-TOF mass spectra per second. MALDI at different laser wavelengths is also being explored, for example using infrared (IR) lasers such as Nd:YAG emitting at 1065 nm and Er:YAG at 2940 nm.

MALDI Matrices

For successful MALDI MS analysis, the choice of the active matrix compound and selection of sample preparation technique are important parameters, both of which are dependent on the properties of the analyte. The MALDI matrix should fulfil the following requirements; be non-volatile in the high-vacuum of the [▶mass spectrometer](#), embed – or at least mix homogeneously with – the analyte molecules, have strong absorption at the desorption/ionisation laser wavelength and assist/enable ionisation of analyte molecules by proton transfer. Since the discovery of MALDI, several hundred compounds have been tested for their suitability as MALDI matrices but only about a dozen have proven efficient in fulfilling the above criteria. The most commonly used matrix compounds (for UV MALDI) and the analytes to which they are applicable are listed in Table 1. α -Cyano-4-hydroxycinnamic acid (CHCA) is arguably the most widely used matrix for peptide analysis. CHCA sample preparations can be used for samples containing a variety of contaminants such as buffer salts, chaotropic agents and detergents. Disadvantages of CHCA include strong background signals in the low-mass region (up to ca. 800 Da) and that it induces fragmentation of many large molecules. For proteins and fragile molecules, such as glycoconjugates, 2,5-dihydroxybenzoic acid (DHB) is the preferred matrix compound. Compared to CHCA, DHB yields fewer background

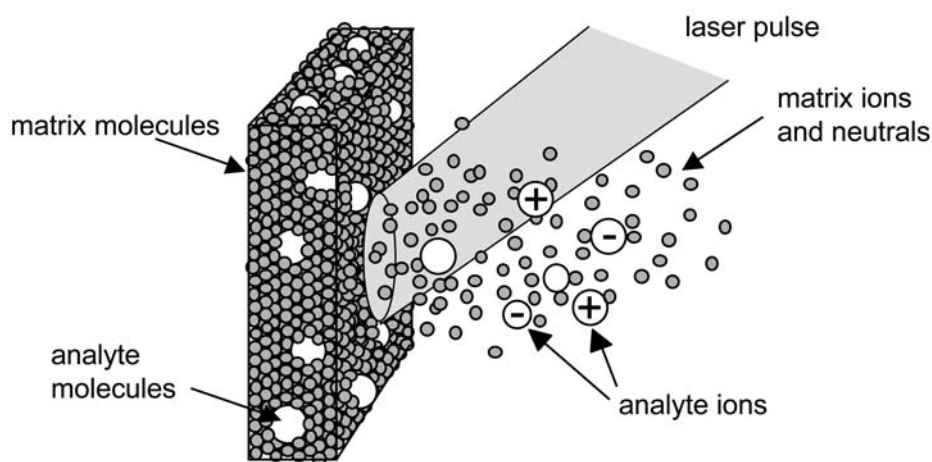
Mass Spectrometry: MALDI. Table 1 Some commonly used MALDI matrices, and examples of analytes to which they are applicable

matrix compound	analytes
α -cyano-4-hydroxycinnamic acid	peptides, small proteins
2,5-dihydroxybenzoic acid	peptides, proteins, glycoconjugates
sinapinic acid	peptides, proteins
2,4,6-trihydroxyacetophenone	glycoproteins, small oligonucleotides
3-hydroxypicolinic acid	oligonucleotides, small PCR products

signals and induces less analyte decomposition. A DHB sample preparation is, however, significantly more sensitive towards the above-mentioned contaminants, demanding efficient sample purification prior to the MALDI sample preparation. Compared to DHB, 2,4,6-trihydroxyacetophenone (THAP) and 3-hydroxypicolinic acid (HPA) result in even less analyte decomposition, at however, the cost of detection sensitivity. These matrices are the preferred choice for the analysis of nucleic acids, which are considerably more prone to undergo ion fragmentation reactions than peptides and proteins.

MALDI Sample Preparation

A variety of sample preparation techniques have been reported to date (3). The first and still much used, is the dried-droplet method (1). The aqueous sample is mixed with a matrix solution at a 10^4 – 10^7 molar excess of the matrix compound and a droplet of the analyte/sample mixture is placed on the MALDI sample plate, which is commonly made of stainless steel. After evaporation of the solvent, a crystalline mixture of analyte and matrix molecules is obtained. Another widely used sample preparation technique is the thin-layer method. First, a thin microcrystalline layer of matrix crystals is prepared on the sample plate. On top of this layer, an acidic aqueous droplet of the analyte solution ($\text{pH} < 2$) is placed and allowed to dry. As a result, the analyte molecules are exclusively incorporated into the outermost molecular layers of the matrix crystals. In this way, the analyte molecules are concentrated to a very small volume, thereby enhancing the detection sensitivity. In addition, the thin-layer preparation allows efficient washing of the sample, whereby salts, buffer components and chaotropic reagents such as urea can be removed. The resulting samples exhibit homogeneous morphology, resulting in good reproducibility for mass spectra acquired over the entire sample, thus facilitating automatic data acquisition. A disadvantage of the thin-layer method is that a sample spot is used up after only a few laser pulses, after which the sample plate has to be moved for the laser to irradiate a fresh spot. The thin-layer method relies on the analyte solution not dissolving the microcrystalline matrix layer, thus prohibiting the use of highly



Mass Spectrometry: MALDI. Figure 1 The MALDI process: Analyte molecules (white) are embedded in a condensed matrix of molecules (gray) exhibiting strong absorption at the desorption laser wavelength. A MALDI experiment is initiated by irradiating the sample with a brief laser pulse. The energy from the laser is efficiently absorbed by the matrix molecules resulting in ablation of a small sample volume, releasing analyte molecules into the gas-phase. In the rapidly expanding plume of desorbed analyte and matrix, analyte molecules are ionized.

water-soluble matrices such as DHB and precluding reversed-phase sample purification directly prior to sample preparation.

MALDI-TOF Mass Spectrometry

MALDI is used in conjunction with several different mass analysers (4), however the most commonly used in biological and medical laboratories is the time-of-flight (TOF) analyser. Fig. 2 shows the basic principle of a MALDI-TOF instrument. The ions generated following the laser pulse are accelerated by an applied electric field. All ions carrying the same charge acquire the same kinetic energy (E_k) after passage through the accelerating field:

$$E_k = \frac{mv^2}{2} = zU$$

where m = the mass of the ion, v = its velocity, z = the number of charges, and U = the accelerating voltage. Thus, their resulting velocity will be inversely proportional to the square root of their mass:

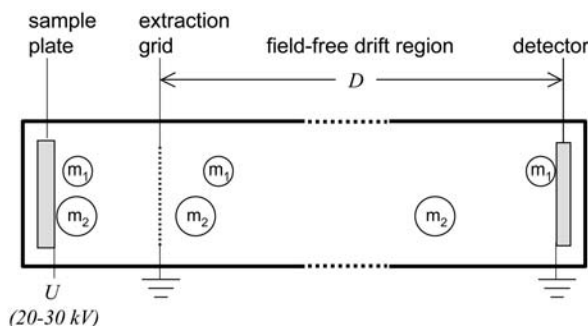
$$v = \sqrt{\frac{2zU}{m}}$$

The ions then pass into the field-free drift region, typically 1–2 m in length, where their velocity differences result in a spatial separation. Because the ions spend most of their time in this region, their time-of-flight measured at the detector is approximately:

$$t = D\sqrt{\frac{m}{zU}}$$

where D is the length of the field-free drift region. To obtain a high mass accuracy, the mass scale is usually calibrated by using well-defined molecular mass standards.

There are several reasons why MALDI MS has become a widespread technique in biological and medical research laboratories in the last 15 years. Pico- to femto-moles of analyte molecules can be routinely detected; amounts that are sufficiently low for the technique to be adequate for the detection of many compounds of interest when isolated from biological or clinical samples. The high mass accuracy of modern mass spectrometers (e.g., typically better than 50 ppm for MALDI-TOF MS peptide analysis) often makes it possible for one to derive detailed chemical structural information on the analyte molecules. The production of analyte ions by laser irradiation of a solid sample on a metal plate lends the technique well to automation. Several hundred samples can be automatically analysed on a single sample plate on currently commercially available MALDI-TOF instruments. In the following, a



Mass Spectrometry: MALDI. Figure 2 The principle of MALDI-TOF MS: In the figure, two positively charged ions of mass m_1 and m_2 , where $m_1 < m_2$, were generated following a laser pulse. The ions are accelerated by an electric field applied between the sample plate and an extraction grid. After passing through the acceleration region, the ions' velocities are proportional to the square root of the inverse of their respective masses. In the field-free drift region, the ions are separated by their different velocities and thus arrive at the detector at different time points. The impact of the ions on the detector results in a current that is recorded over time and constitutes the time-of-flight spectrum. By calibration, the time-of-flight spectrum is converted into a mass spectrum.

few examples are given of analytical strategies involving MALDI MS.

Protein Identification by MALDI MS Peptide Mass Fingerprinting

MALDI ▶peptide mass fingerprinting (PMF) is a widely used technique for identification of proteins isolated by, e.g. gel electrophoresis or chromatography. The unknown isolated protein is cleaved by a protease of high specificity. Commonly used is trypsin, which cleaves polypeptides at the C-terminal side of lysine and arginine residues, unless the next amino acid in the sequence is proline. MALDI MS analysis of the resulting mixture of peptides yields a peptide mass fingerprint; a set of measured molecular masses of tryptic peptides derived from the analysed protein. In a database search, these mass values are matched against sets of masses calculated for all protein sequences in a sequence database. The search returns a list of the sequence entries with the highest number of matching mass values, and various algorithms are used to rank the sequences and determine the probability that the highest-ranking sequence entry is the true identity of the analysed protein. Recording a peptide mass fingerprint instead of a measuring the molecular mass of the intact isolated protein has several advantages. For example, if a protein carries post-translational modifications, contains mutations or is processed, its

measured molecular mass will not match the mass calculated based on its sequence in the database and consequently its identification will fail. The same will happen if the protein's sequence in the database contains an error. With a peptide mass fingerprint in contrast, only signals corresponding to the peptides that carry the modified amino acid or containing the mutation site, will be affected. The recorded m/z values of the other signals will match the masses calculated based on the protein's amino acid sequence, thereby rendering the identification procedure rather robust. In addition, detection sensitivity as well as mass accuracy is higher for peptides than for proteins.

Structural Characterization by MALDI MS/MS

MALDI MS is not limited to measuring the masses of intact analyte molecules. Many instruments provide the possibility of isolating a selected analyte and, by different activation methods, transferring energy to the analyte, thereby inducing fragmentation of the molecule. Mass analysis of molecular fragments derived from a single selected analyte can provide detailed structural information and is termed tandem-MS or MS/MS analysis. For proteins and peptides, this information can for example be used to verify a postulated amino acid sequence and to pinpoint a protein modification to a certain amino acid. On MALDI mass spectrometers, analyte fragmentation can be induced by increasing the laser energy above the desorption threshold. Fragment ions produced by this method are termed post-source-decay (PSD) fragments.

Clinical Relevance

In the last 15 years, MALDI mass spectrometry has matured into an analytical technique with many applications in the clinical field. In the following, a few examples of established and emerging analytical strategies using MALDI MS are given.

2-DE and MALDI MS

Of particularly widespread use is the combination of two-dimensional gel electrophoresis (2-DE) and MALDI MS. 2-DE has the resolving power to separate and visualize over 10,000 proteins in crude cell or tissue extracts in a single experiment. 2-DE protein spot patterns for different samples, e.g. from patients and healthy controls, can be compared to reveal aberrances in normal protein abundance and modification. Proteins of interest can be hydrolysed by a protease *in situ*, i.e., while immobilized in the gel sample, and readily identified by MALDI MS PMF and MS/MS analysis of the produced peptides. The combination of 2-DE and MALDI MS has been used extensively for detection and identification of biomarkers for many different

diseases, including cancer, Alzheimer's disease, allergy and inflammatory diseases.

Protein Interaction Analysis

The combination of the above-described PMF and MS/MS techniques with various affinity enrichment methods or non-denaturing separation techniques has made it possible to study protein-protein, protein-DNA and protein-carbohydrate interactions. These experiments start with the isolation of the complex, followed by MS analysis of each constituent component. The relatively high throughput of MS makes it possible to perform such studies on a large scale (5).

Genotyping by MALDI MS

Although less established than gel-based methods, MALDI MS is capable of sequencing oligonucleotides of lengths up to 30–50 base pairs. Advantages of MALDI MS include short analysis times and that no chemical labels are required for detection. Furthermore, mass spectrometric detection is unaffected by secondary structure, which can pose problems with gel-based methods, e.g., for sequencing GC-rich regions. MALDI MS is therefore often used when large numbers of short sequences are to be determined. Several schemes for detection of single-nucleotide polymorphisms (SNPs) using MALDI MS have been developed (6). High-throughput genotyping of SNPs has the potential to become a routine technique for clinical applications.

MALDI MS Screening of Diagnostic Fluids

Due to the ability of MALDI MS to detect biomolecules in complex mixtures and in the presence of relatively high salt concentrations, several methods have been developed in which MALDI MS is used as a tool for rapid, direct screening and detection of modifications of proteins in small volumes of diagnostic fluids. For example, several proteins with links to Alzheimer's disease have been detected in cerebrospinal fluid in which they are present at nano-molar (nM) concentrations.

Analysis of Protease Activity in Diagnostic Fluids

While many enzymes in biological fluids are present at concentrations too low for their detection by MALDI MS, their activity can be measured indirectly. A substrate (protein or peptide) is added to a sample of the diagnostic fluid at a concentration at which it can be detected by MALDI MS. Samples of the mixture are then analysed at time points following the addition and the activity and specificity of the native proteases in the fluid can be monitored by their action on the added molecule. This technique has for example been employed in the study of the processing of neuropeptide Y in cerebrospinal fluid.

Fingerprinting of Whole Microorganisms

In microbiology, MALDI-TOF MS is employed to analyse samples prepared from whole viruses, bacteria and spores. The microorganisms are lysed when mixed with the MALDI matrix solution, allowing the direct desorption of peptides and proteins in the sample. The mass spectra acquired from such samples are complex and usually cannot be interpreted directly. However, with a reproducible sample preparation, the mass spectrum constitutes a characteristic fingerprint by which a microorganism can be distinguished from others. One application of clinical relevance is the detection of biomarkers for taxonomic identification of bacteria by MALDI-TOF MS.

MALDI-TOF Imaging Mass Spectrometry

MALDI-TOF imaging mass spectrometry is an emerging technique, made possible by the development of high-repetition rate desorption/ionization lasers and fast data acquisition systems. Thin slices of tissue samples are attached to the MALDI sample plate and coated by MALDI matrix. A MALDI-TOF instrument is then used to acquire mass spectra in a dense pattern on the sample surface. Parallel to the way in which a picture produced by an optical scanner consists of pixels each assigned an intensity value, the mass spectrometric image also consist of pixels, each represented by one mass spectrum. By assigning different colours to selected m/z values, the mass spectrometric image can be used to visualize the spatial distribution of selected compounds in the studied tissue. The spatial resolution, limited by the diameter of the desorption/ionisation laser, is approaching 1 pixel/ μm . MALDI-TOF imaging mass spectrometry can be used to map the distribution of selected compounds in tissue.

References

1. Karas M, Hillenkamp F (1988) Laser desorption ionization of proteins with molecular masses exceeding 10000 Daltons. *Anal Chem* 60:2301–2303
2. Zenobi R, Knochenmuss R (1998) Ion formation in MALDI mass spectrometry. *Mass Spectrom Rev* 17:337–366
3. Kussmann M, Nordhoff E, Rahbek-Nielsen H et al (1997) Matrix-assisted Laser Desorption/Ionization Mass Spectrometry Sample Preparation Techniques Designed for Various Peptide and Protein Analytes. *J Mass Spectrom* 32:593–601
4. Hoffmann ED, Stroobant V (2002) *Mass Spectrometry: principles and applications*. John Wiley, Chichester
5. Gavin AC, Bosche M, Krause R et al (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415:141–147
6. Tost J, Gut IG (2002) Genotyping single nucleotide polymorphisms by mass spectrometry. *Mass Spectrom Rev* 21:388–418

Mass Spectrometry: MS/MS

EBERHARD KRAUSE

Institute for Molecular Pharmacology, Berlin, Germany
ekrause@fmp-berlin.de

Synonyms

Tandem mass spectrometry; mass spectrometric fragmentation

Definition

Tandem mass spectrometry (MS/MS) is a method for obtaining sequence and structural information by measurement of the mass-to-charge ratios of ionized molecules before and after dissociation reactions within a mass spectrometer which consists essentially of two mass spectrometers in tandem. In the first step parent or precursor ions are selected for further fragmentation by energy impact and interaction with a collision gas. The generated daughter or product ions can be analyzed by a second scan step. MS/MS measurements of peptides can be performed using electrospray or matrix-assisted laser desorption ionization in combination with triple quadrupole, ion trap, quadrupole-TOF (time-of-flight), TOF-TOF or ion cyclotron resonance (FTICR) MS. Tandem mass spectrometry of peptides generated by proteolytic digestion of proteins provides amino acid sequence-specific ions and can be used to identify and characterize proteins (1).

Characteristics

Dissociation of Molecules in Mass Spectrometry

The development of so-called “soft” ionization techniques such as electrospray ionization (►ESI) and matrix-assisted laser desorption/ionization (►MALDI) has enabled the generation of gas-phase ions of non-volatile, sensitive biopolymers without fragmentation which allows mass analysis of e.g. proteins, peptides, carbohydrates and oligonucleotides. Using these techniques, intact molecular ions are measured, but no structural information is obtained. In order to obtain specific fragments during MS analysis which are interpretable in terms of molecular structure, MS/MS (or tandem MS) experiments can be performed. Valuable fragmentation requires that the energy transmitted to an ion exceed the vibrational energy of a specific bond. This can be achieved by accelerating the ion of interest into a collision cell where it can undergo multiple collisions with an inert gas (collision-induced dissociation, CID). Two types of CID are used. Triple quadrupole and ion trap instruments typically utilize low-energy collisions (<35 eV) whereas tandem time-of-flight instruments employ high-energy

collisions (keV range). Low-energy conditions are particularly useful in the analysis of protonated peptides and proteins. Such conditions produce fragmentation at the amide bonds and provide spectra which are characteristic of the primary structure (2). The recently described electron capture dissociation (ECD) method uses an entirely different fragmentation mechanism with the advantage that backbone cleavages are preferred and side-chain modifications, such as phosphorylation or glycosylation, remain unaffected.

MS/MS Instruments

Tandem MS experiments can be performed by different types of mass spectrometers, which consist of the four basic types of mass analyzer as well as various combinations thereof. Triple quadrupole (QqQ) instruments use three quadrupoles in series such that the first and third quadrupoles are operated as mass analyzers. The second quadrupole is used as a collision cell in order to generate various fragments. QqQ instruments can be utilized to perform any kind of MS/MS analysis, but are characterized by the relatively low resolution of the mass analyzer. In a quadrupole time-of-flight mass spectrometer the second mass analyzer is replaced by a TOF analyzer which allows extremely fast acquisition of spectra, high mass resolution (up to 25,000 FWHM), and low femtomole sensitivity. Because of the high mass resolution, the quadrupole TOF instrument enables exact charge state assignments. The recently introduced TOF-TOF analyzer equipped with a MALDI ion source combines very accurate and sensitive mass measurements with sequencing capability by high energy CID. Such an analyzer allows reliable peptide and protein analyses of complex samples and the process can be easily automated.

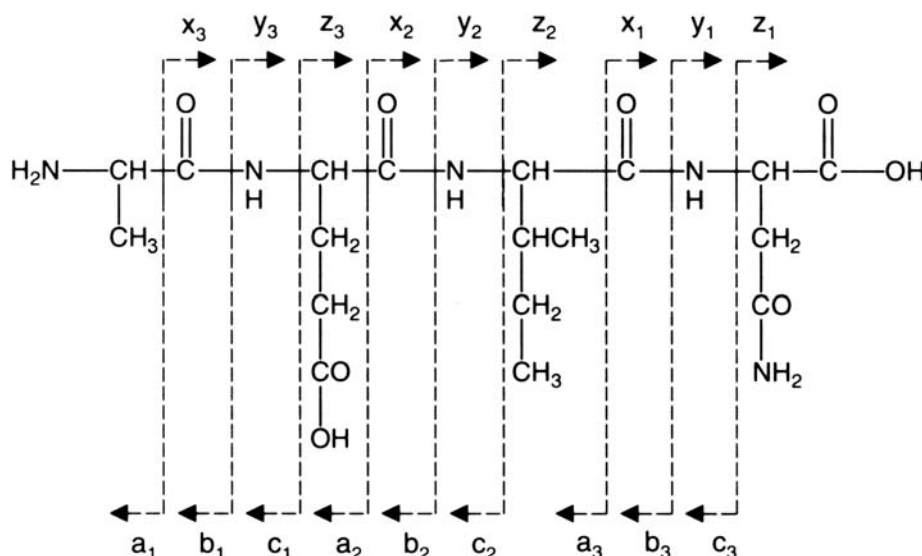
Quadrupole ion trap (IT) and Fourier-transform ion cyclotron resonance (FTICR) mass spectrometers use one mass analyzer and the tandem-in-time concept to perform MS/MS measurements. In the IT mass spectrometer, the ions of interest are first captured and then dissociated in the ion trap by low energy CID prior to acquisition of the fragment ion spectrum. Ion traps are robust, sensitive and enable multi-stage tandem MS (MS^n). However, they are considered to be low mass accuracy instruments. FTICR mass spectrometers capture ions at high magnetic fields and allow MS and MS/MS measurements to be made with high sensitivity and outstanding mass accuracies (better than 5 ppm). However, because of operational complexity and limited fragmentation efficiency, this kind of instrument is rarely used in proteome research. In general, most MS/MS mass analyzers can use both ESI- and MALDI-ion sources. ESI- and MALDI-MS/MS experiments are advantageously combined on-line and off-line respectively, with capillary liquid chromatography for analysis of complex mixtures.

Tandem MS Modes

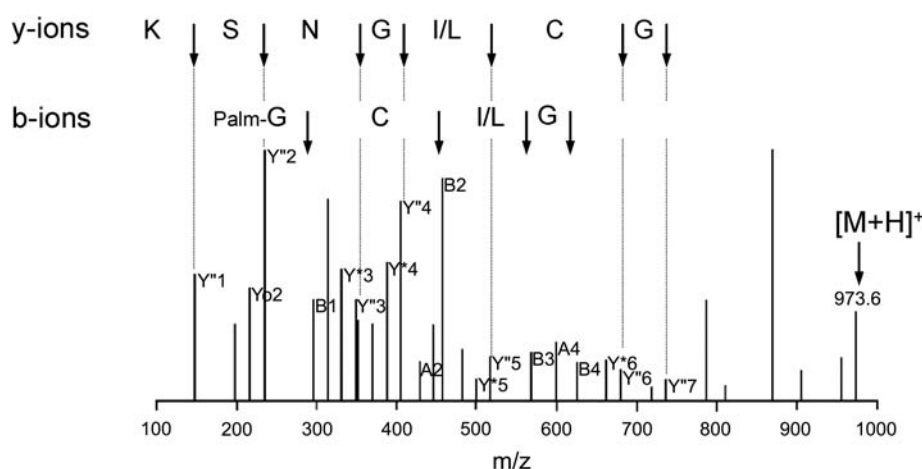
Product ion scanning is the most frequently used type of MS/MS experiment. A precursor ion, i.e. the ionized molecule of interest, is isolated by a first mass analyzer. After dissociation in the collision cell, a mass spectrum of fragment ions is acquired by a second mass analyzer. The most commonly used technique for the dissociation of peptides is collision-induced dissociation (CID) which involves the collision of accelerated ions with an inert gas (argon). Under these conditions, peptide ions are basically cleaved at the amide bonds along the amino acid sequence (Fig. 1). The product ion spectrum consists of ladders of sequence ions either from the N-terminus (b-type ions) or from the C-terminus (y-type ions). Amino acid sequences can be derived by subtracting the masses of adjacent ions. The fragment ion pattern depends strongly on the sequence and the presence of specific amino acids. In the case of tryptic peptides (from proteolytic digestion of proteins by trypsin and having the basic amino acids arginine or lysine at the C-terminus), the spectra are dominated by y-ions, thus facilitating the interpretation. In contrast, internal fragmentation, preferential cleavages and the loss of neutral groups increase the complexity and hamper the interpretation of product ion spectra. Product ion scanning of peptides generated by proteolytic digestion of proteins using both low- and high-energy collision conditions gives at least partial information about the amino acid sequence. In particular, this has led to the development of sequencing methods which can be used to identify proteins by searching protein sequence databases or nucleotide databases (EST).

For precursor ion scanning, the first mass analyzer is used to scan the entire mass range of interest and to transmit the ions into the collision cell. The second mass analyzer is used as a mass filter which allows only specific fragment ions to be transmitted to the detector. The detector only registers signals when fragment ions of interest (reporter ions) are split off from precursor ions. Using this MS/MS mode, ions which share a structural feature (a particular class of compounds) can be analyzed and resolved from other molecular ions. For [post-translational protein modifications](#) which can produce characteristic fragment ions (phosphorylation, glycosylation, Cys-palmitoylation, etc.), precursor ion scanning of the proteolytic digest of the protein can be used to identify the sites of attachment (3).

Similar to precursor ion scanning, neutral loss scanning is generally used to analyze peptide and protein modifications. This scan mode can be employed if characteristic neutral fragments are formed upon CID. The mass of the product ion differs from the mass of the precursor ion by the neutral fragment. Neutral loss can be detected by scanning simultaneously the first and the



Mass Spectrometry: MS/MS. Figure 1 MS/MS peptide fragmentation by collision induced dissociation.



Mass Spectrometry: MS/MS. Figure 2 MS/MS analysis of native G protein α_s . The protein was in-gel digested by trypsin and the resulting peptides analysed by electrospray tandem mass spectrometry. The mass peak with m/z 973.6 corresponded to the N-terminal modified sequence GCLGNSK of G_{α_s} . The b- and y-ions of the MS/MS spectrum show that native G_{α_s} carries an additional group at the N-terminal amino acid with a mass of 238 that corresponds to the palmitoyl moiety (5).

second mass analyzer with an offset which corresponds to the mass of the neutral fragment. For example, in order to determine the phosphorylation of proteins, the loss of phosphoric acid (neutral loss of 98 mass units) from phosphoserine or phosphothreonine is measured.

Tandem Mass Spectrometry in Proteomics

Proteome research, which includes (i) analysis of molecular networks in a cell following functional

stimulation, (ii) differential display effects for the comparison of protein expression and (iii) determination of post-translational modifications, requires efficient methods for the identification and characterization of proteins. In addition to methods for the separation, visualization and quantification of proteins, MS/MS techniques provide sensitive and efficient tools for the identification and characterization of proteins and thus play a central role in proteomics.

MS/MS in conjunction with capillary liquid chromatography (LC) has been used for identification of proteins excised from 1-D or 2-D gels. The automated nature of protein database searching using uninterpreted MS/MS data (4), the low femtomole sensitivity and the ability to identify a protein based on a single MS/MS product ion spectrum allow the identification of more than 100 proteins per day. Thus, extensive protein identifications have been performed for example to identify cancer-related proteins or to study human pathogens. During a study of MHC-associated peptides the use of LC-MS/MS methods for the analysis of complex peptide mixtures was developed.

In combination with novel stable isotope-tagging methods (metabolic labeling using amino acids, enzymatic incorporation of ^{18}O , chemical reactions using ▶ICAT) and multi-dimensional capillary LC (strong cation exchange/reversed phase), tandem mass spectrometry enables quantification of differences in protein expression. In order to study multi-protein complexes, techniques which include LC-MS/MS and tandem affinity purification (TAP) tags for identification of proteins have been established (1).

Since MS/MS-generated fragment ion spectra provide a “structural fingerprint” of particular peptides resulting from enzymatic degradation of the protein by specific endoproteases, MS/MS can be used to determine the site of post-translational modification. Numerous reports show that functionally important modifications such as phosphorylation, glycosylation, methylation, myristoylation and palmitoylation are accessible to analysis *via* tandem mass spectrometry in such a way that sequences and modified amino acid side chains can be deduced from C- and N-terminal fragment ions of product ion spectra (Fig. 2).

Clinical Relevance

Proteomics technologies which include highly sensitive and high-throughput MS/MS methods for efficient identification and characterization of proteins have the potential to increase our understanding of the molecular basis of diseases. This could lead to new disease-specific targets. Tandem mass spectrometry directly contributes to protein profiling, the identification of diagnostic markers and toxicity studies and thus has a profound impact on clinical diagnosis and drug discovery. In addition to its implications for proteomics, MS/MS is becoming an increasingly important method in therapeutic drug monitoring (immunosuppressants, antidepressants) and analysis of metabolic pathways. It is expected that substantial increase in MS/MS sensitivity as well as nano scale LC techniques will further expand the range of applications.

References

1. Aebersold R, Mann M (2003) Mass spectrometry-based proteomics. *Nature* 422:198–207
2. Tang XJ, Thibault P, Boyd RK (1993) Fragmentation reactions of multiple-protonated peptides and implications for sequencing by tandem mass spectrometry with low-energy collision induced dissociation. *Anal Chem* 65:2824–2834
3. Wilm M, Neubauer G, Mann M (1996) Parent ion scans of unseparated peptide mixtures. *Anal Chem* 68:527–533
4. Eng JK, McCormack AL, Yates JR (1994) An approach to correlate tandem mass-spectral data of peptides with amino acid sequences in a protein database. *J Am Soc Mass Spectrom* 5:976–989
5. Kleuss C, Krause E (2003) Gas is palmitoylated at the N-terminal glycine. *EMBO J* 22:826–832

Mass Spectrometry: Quantitation

OLIVER FIEHN¹, WOLFRAM WECKWERTH²

¹UC Davis Genome Center, Davis, CA, USA

²Max-Planck-Institute of Molecular Plant Physiology, Golm/Potsdam, Germany
ofiehn@ucdavis.edu

Definition

Mass spectrometry is widely used in proteomics for protein identification, mostly after tryptic digestion into peptides (1). For decades, mass spectrometry has also served for determining the absolute or relative amounts of components of complex matrices. There are countless reports on absolute quantification of target molecules by external or internal standards, often by use of stable isotope markers. For some application areas, semi-quantitative methods have been added that aim at comparative assessments of a wide range of compounds rather than ultimate precision for a very few targets. However, in biological contexts, such classical quantitation strategies have only recently been adopted. When sample preparation is aiming at metabolites, such approaches have been named ‘metabolomics’ (2) (and, for historical reasons, ‘metabonomics’ in biomedical applications). In the last couple of years there are also reports on quantitative approaches to proteomics using mass spectrometry. All quantitative methods must follow certain guidelines, termed method validation (3). Starting from the definition of the validity area (i.e. the exact description of the scope and objective of the analysis), each method must give certain minimum characteristics such as how compounds of interest can be distinguished from chemical or biological background noise (‘selectivity’),

the lower limit ('sensitivity') and the dynamic range where quantitation can be achieved, a robustness assessment and, eventually, the method reproducibility. For each of these prerequisites, mass spectrometric methods may face problems due to instrumental constraints such as ▶**duty cycles**, or lack of tandem mass spectrometric capabilities. More severe, however, are fundamental constraints of ionization suppression and adduct formation that may ultimately undermine any precise comparative quantitation.

Characteristics

There are more than 50 types of mass spectrometer used for biological determinations, each having specific characteristics limiting their use outside pre-defined applications. Some features of mass spectrometers, however, are common to all of them, enabling performance comparisons of the instruments and simultaneously, setting fundamental constraints on the analysis of complex mixtures. Firstly, and looking trivial at first sight, mass spectrometers can only detect ions, not neutral species. Therefore, the details of the ionization method are an integral part of any quantitative mass spectrometric method. The same argument is fundamental for another aspect of mass spectrometric detection; pure or 'naked' ions only survive a very short time if they undergo collision, for example at high gas pressures. Therefore, compounds must reach vacuum pressures immediately after, or even before, ion formation. Secondly, all mass spectrometers are able to distinguish different masses; this is called mass resolution $R = m/\delta m$. Since chemical and most physical properties of the molecules are almost identical, if only a few atoms are exchanged for stable isotopes (^{13}C , ^{18}O , ^2H , ^{15}N), such stable isotope marker compounds are the ideal reference for exact quantitation in mass spectrometry. This is especially true when they are added at the earliest possible time point in sample preparation in order to account for potential losses due to adsorption, precipitation or oxidation. Today, all mass spectrometers are able to distinguish nominal masses (e.g. m/z 212 from m/z 213). A few types of mass spectrometers, for example some time-of-flight instruments, have a higher ▶**resolution** of $R = 5,000$ to 50,000 or, ultimately, there are FT-ICR mass spectrometers with a resolution of up to $R > 1,000,000$. Such resolution enables the differentiation and quantitation of co-eluting isobaric compounds that have identical nominal masses but different exact masses due to small differences in the exact atomic weights ($^{16}\text{O} = 15,994915 \text{ u}$, $^1\text{H} = 1,008725 \text{ u}$, $^{14}\text{N} = 14,003074$).

GC/MS

One of the oldest and most classical methods is electron impact ionization of molecules in the gas phase. This technique gives robust ionization with reasonable

efficiency, but obviously requires volatility of the compounds. Therefore, it is restricted to small molecules such as metabolites and is usually coupled to ▶**gas chromatography**. In a single run, up to 20,000 components can be detected by GC x GC coupling prior to MS detection if time-of-flight mass spectrometers are utilized. Due to the large energy excess ('hard ionization'), the ionization efficacy is not hampered by co-eluting compounds over a wide range of concentrations. Therefore, quantitation can be reliably achieved even by external calibration. Higher precision can be achieved by adding stable isotope labeled compounds as outlined above and this technique has been used for over 40 years. A rougher estimation of differences in metabolite abundances uses direct comparisons of relative peak areas between two or many experiments; this is then called 'relative quantitation' for metabolomic use. This strategy is comparable to approaches undertaken in proteomics or transcriptomics where dye labeling or staining of 2D gel spots is compared without particular calibration curves. By relative quantitation, up to 1,000 metabolites can be compared between various biological tissues by GC/TOF analysis. Unlike 2D gel approaches in proteomics however, the sample throughput is much higher (and costs are lower). This enables a good estimation of reproducibility and its distinction from inherent biological variability – a field of the utmost importance, which has yet to be extensively covered in proteomics (see below). Quantitation in metabolomics by GC/MS regularly achieves 2–10%CV for identical extracts (technical error due to the instrument) or 20–30%CV for the total process (including errors in the extraction and sample preparation process), with typically some 3–4 orders of magnitude dynamic range, good robustness even in high-throughput operations, excellent selectivity due to the good chromatographic resolution (which is some 10-fold higher than in typical liquid chromatography applications) and good sensitivity due to low chemical noise. However, GC/MS is restricted to thermostable, small compounds (up to ~500 u) that can be made volatile, usually by chemical derivatization. Therefore, other approaches must complement GC/MS. As a complementary option, metabolomics by capillary electrophoresis coupled to MS detection detected 1,600 compounds from *B. subtilis* cultures with an overall method precision of 30▶%CV for 65 identified metabolites. Other approaches utilizing novel column types for LC/MS are lacking method validation so far.

LC/MS

The most difficult aspect of quantitation in LC/MS is the ionization process. Unfortunately, this has not always been sufficiently acknowledged, although there are numerous fundamental publications showing the

quenching effect of co-eluting compounds on the ionization efficiency of the target molecules. This phenomenon is called ion suppression and is fundamental to all 'soft' ionization techniques, especially for the most often used electrospray interface. Basically, the ionization efficiency of a particular molecule in a droplet eluting from the LC system is directly related to its hydrophobicity and its susceptibility to receiving a charge; more hydrophilic compounds will preferentially locate in the interior of the droplet, decreasing the likelihood of undergoing desolvation before reaching the skimmer of the mass spectrometer. If a slightly more hydrophobic molecule (or non-volatile component) is present at the same time as the more hydrophilic one, it will suppress ionization of the hydrophilic compound. Quantitation between experiments is inevitably hampered (especially when many peak abundances are different between the two experiments) and a prediction of the severity of ion suppression is impossible. The same effect results from differences in the ability of molecules to become charged (e.g. basicity/acidity). This and other factors impairing ionization efficiency and quantitation, like geometry and size of the nozzle tip, have been nicely demonstrated by Karas and co-workers using model metabolites (4). This phenomenon is specifically a problem whenever the total composition of a specimen (often called the 'matrix') is believed to be altered, i.e. in any typical biomedical experiment such as comparing \blacktriangleright WT to mutant tissues or healthy to diseased samples.

On top of the difficulty of assessing the effects of ion suppression come alterations in adduct formation. Most biomolecules are only ionisable after adduct formation. The simplest form is protonation or deprotonation, but dozens of further adducts are known in LC/MS, ranging from adducts to the LC eluent (e.g. $[M + \text{acetonitrile} + H]^+$, or $[M + NH_4]^+$ in ammonium acetate buffered LC eluents) to more complex and less predictable adducts like $[M + M + H]^+$, or, in case of co-elution or direct infusion, $[M_1 + M_2 + H]^+$. For quantitation, the preferred adduct species must be known for each target molecule and its ratio to less abundant adducts must be monitored, since subtle alterations in matrix composition (e.g. minor quantities of Na- or K-salts in the LC buffer eluents) may have caused significant alterations in adduct ratios. The presence of basic amino groups in peptide mixtures diminishes the danger of variable adduct formations in proteomics compared to metabolomic studies.

Due to the unpredictability of the effects of changes in overall matrix compositions on ion suppression and adduct formation, claims that are made for direct LC/MS quantitation of peptides from complex mixtures without internal standards have to be treated with great caution (5). Instead of direct quantitation, use of stable

isotope labeled samples has been extensively studied during the last four years. The idea is to combine a control sample with a test sample that consists of stable isotope labeled peptides. Then, both samples are combined and each peptide pair is quantitatively compared based on the abundance ratio of the labeled sample to the unlabeled control. Two basic approaches can be distinguished: either biological (*in vivo*) or chemical (*in vitro*) (6) incorporation of stable isotopes into the peptide test mixture. The advantage of the *in vivo* strategy is the uniformity of protein labeling which is independent of the protein primary structure but it must be seen as a disadvantage that *in vivo* labeling is restricted to use in growing cell cultures. *In vitro* labeling has gained more interest in the past few years. The idea here is to carry out chemical derivatisations with and without stable isotope linkers for specific peptide moieties (such as carboxylic acids, cysteine residues or amino groups). This labeling can also serve to reduce the complexity of mixtures before the LC/MS runs are carried out, without loss in proteomic information. For example, at least one cysteine is present in more than 90% of all predicted yeast \blacktriangleright ORFs and the thiol group of cysteines can readily be used for chemical derivatisations. A biotinylated derivative of iodoacetamide that included a stable isotope marked linker was used for quantitative proteomics. After derivatization, crude protein (or peptide) mixtures can then be purified over streptavidin columns, resulting in a much less complex mixture prior to LC/MS quantitation of each labeled/unlabeled peptide pair. However, disadvantages have been recognized in that thiol residues tend to be so reactive (specifically against oxidation), that the recovery from 2D gels may be irreproducible and that slight differences in the oxidative states of the control/test comparisons may result in high method errors. A variety of other reagents have been tested for proteomic surveys, for example by applying D₂O during tryptic digestion. Other approaches have been specifically focused on post-translationally modified targets such as phosphorylated proteins. However, important parts of the validation of quantitative proteomic methods are still lacking, including in-depth robustness testing. So far, no study has been published focusing on the overall method % CV from different matrices. Reproducibility testing has so far been restricted to the LC/MS errors alone, which have been found to be 10–20%CV. However, it is known that sample preparation usually accounts for the largest error in analytical methods and this can also be assumed to be true for proteomics.

If specific proteins are targeted instead of trying to gain proteomic overviews, the classical analytical method guaranteeing selectivity and sensitivity involved using antibodies for Western blots. It has been shown, that even in this area of protein quantitation, classical

strategies can be complemented by LC/MS approaches. The concomitant characteristics of specific retention times for peptides in LC as well as the unique mass and mass fragmentation have been exploited by a 'mass Western' method. This method utilizes the selectivity power of tandem mass spectrometers by carrying out the quantitation on the ►MS/MS level, i.e. on fragment masses that are highly specific for each compound. If the mass spectrometric detection is then set to the characteristic masses of a peptide and to one of its most abundant daughter ions at the typical LC retention time of this peptide, the increase in signal-to-noise ratios enables the reliable quantitation even of very low abundance proteins after tryptic digestion, because any chemical noise molecule is very unlikely to undergo the same fragmentation at the same fragmentation energies. However, due to potential ion suppression effects, the synthesis and spiking of a stable isotope form of this 'mass Western peptide' is still advantageous for exact quantifications.

Matrix Assisted Laser Desorption/Ionization-TOF

One of the most often applied techniques for identifying proteins after 1D or 2D gel separations is the in-gel digestion of proteins followed by peptide mass fingerprinting using matrix assisted laser desorption/ionization ►TOF mass spectrometry. For the MS detection, the sample is co-crystallized with a UV absorbing matrix on a plate. When a high-energy UV laser is then pulsed onto such crystals, the matrix immediately sublimates concomitantly with the desorption and ion formation of the target molecules. It is commonly believed that quantitation cannot be done using ►MALDI-TOF, due to differences in analyte/matrix distribution from crystal to crystal. However, modern MALDI-TOF instruments are capable of collecting and averaging spectra from many crystals per sample, enabling at least semi-quantitative fingerprint analysis in mutant screens. If internal standards were added, quantitation by MALDI-TOF was shown to result in identical data to that garnered through LC/MS applications, for example for toxic secondary plant metabolites like potato glycoalkaloids. If stable isotope labeling is included, MALDI-TOF was demonstrated to result in 5–20%CV on protein quantification after tryptic digestion from 2D gels.

Significance in Quantitation

Apart from the actual choice of direct or stable isotope labeling quantifications in LC/MS, the next and very practical question is how many samples do I need to run to find significant differences between control and test samples? Just a few studies have been published on this topic, although this ultimately should be done in any method validation. Finding statistically significant differences relates to comparing mean values and their

corresponding deviations from the means. These deviations are now composed of errors of the analytical method (instrument errors + sample preparation errors) and the inherent biological variation. Unfortunately, the latter is very often underestimated by biologists, and furthermore, it may vary from genotype to genotype and from organ to organ. Molley et al. have found a large contribution of biological variation on top of instrumental errors, which may reach a total error of up to 70%CV in proteomic applications (7). In some cases, appropriate sample pooling strategies may compensate for unwanted biological variation. On the other hand, however, there is inherent information in biological variation that is lost if only mean values are compared. For example, if co-regulation of gene expression transcripts, protein and metabolite abundances are to be analyzed in system biology approaches, the possibility of finding and the significance testing of such co-regulations require a certain amount of variability in the network data sets. For such cases, the analysis of biological snapshots from minute amounts of samples would be more suitable than pooling strategies.

Instrumentation

For many purposes, the choice of the actual type of mass spectrometer being used for proteomic (or metabolomic) studies is less important than is commonly believed. Each instrument has its own advantages and disadvantages, depending on the actual application. For example, if quantitation is supposed to be verified by peptide identifications, a fast switching between full scan MS and (data dependent) MS/MS fragmentations must be achievable, the so-called duty cycle. Some of the instruments do not allow a convenient way to do MS/MS experiments but are very fast at acquiring full scan spectra (like TOF MS), whereas others have good mass resolution and mass accuracy for *de novo* peptide identification (like QTOF hybrid instruments) but lack high duty cycles. Others like FT-ICR-MS instruments are known for ultimate sensitivity, mass resolution and mass accuracy, but again come with low duty cycles and additionally, prohibitive prices and technical problems causing long down times over the year. Further, mass accuracy may depend on frequency perturbations and charge repulsions and quantitation may also be very difficult by FT-MS if too many ions are present at one time point. MS/MS spectra may vary dramatically between different instrument types due to the differences in the methods used for compound fragmentations. The best comparable spectra across instruments from different manufacturers are gained by ion trap mass spectrometers, and compilation of MS/MS libraries have therefore started with ion traps. Ultimately, the chromatographic (or electrophoretic) resolution, the type and maintenance of the electrospray interface and the sample

preparation method including stable isotope markers are more important than the actual instrument that is eventually employed.

Clinical Relevance

Modern clinical research relies in all areas on the precision of quantitative and qualitative biochemical analysis for diagnostic purposes, biomarker validation and mechanistic studies. With current methodologies, quantitative metabolite and protein profiling may go far beyond classical diagnostic tests like automated recognition of metabolic diseases or newborn screening. It is the task of analytical chemists and biochemists to serve clinicians by accurately assessing standard operating procedures and critical steps in the process in order to achieve validated quantitative methods with low overall method %CVs. From a technical perspective, current mass spectrometers are able to fulfill such requirements.

References

1. Ferguson PL, Smith RD (2003) Proteome analysis by mass spectrometry. *Annu Rev Biophys Struct* 32:399–424
2. Weckwerth W (2003) Metabolomics in systems biology. *Annu Rev Plant Biol* 54:669–689
3. Krull IS (1999) Analytical method development and validation for the academic researcher. *Anal Lett* 32:1067–1080
4. Schmidt A, Karas M, Dülcks T (2003) Effect of different solution flow rates on analyte ion signals in nano-ESI MS, or: When does ESI turn into nano-ESI? *J Am Soc Mass Spectrom* 14:492–500
5. Bondarenko PV, Chelius D, Shaler TA (2002) Identification and relative quantification of protein mixtures by enzymatic digestion followed by capillary reversed-phase liquid chromatography-tandem mass spectrometry. *Anal Chem* 74:4741–4749
6. Gygi SP, Rist B, Gerber SA et al (1999) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 17:994–999
7. Molloy MP, Brzezinski EE, Hang J et al (2003) Overcoming technical variation in quantitative proteomics. *Proteomics* 3:1912–1919

Mass Spectrometry: SELDI

GURU REDDY, RICHARD RUBIN,
SCOT WEINBERGER
CIPHERGEN BIOSYSTEMS, INC., FREMONT, CA 94555, USA
sweinberger@ciphergen.com

Synonyms

SELDI; SELDI-TOF-MS; SELDI-ProteinChip technology; ProteinChip technology

Definition

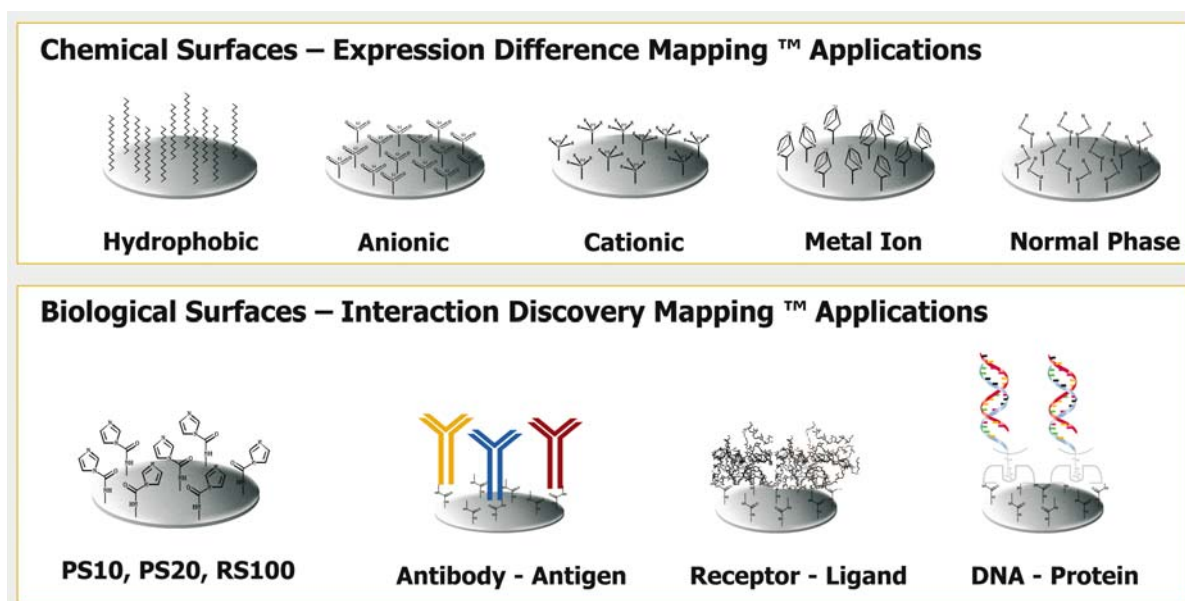
SELDI (surface-enhanced laser desorption and ionization) is a method of protein capture and enrichment on a chemically or bioaffinity active solid phase surface, often followed by selective washing steps and then followed by laser “elution” of the proteins into a detector, usually a time-of-flight mass spectrometer. The whole method is often referred to as SELDI-TOF MS or SELDI ProteinChip technology. With similarities to ►MALDI technology (in which the sample probe is not an active binding partner to protein analytes), an energy absorbing molecule or protein co-crystallization matrix is typically added on top of the captured proteins to assist ionization *via* laser excitation. SELDI is commonly used to discover disease associated protein ►biomarkers and to develop diagnostic assays.

Characteristics

The key components of SELDI ProteinChip technology are the ►ProteinChip Arrays, the ►ProteinChip Reader and the associated software. ProteinChip Arrays contain various chromatography or biological surfaces to capture proteins from complex biological mixtures. Chromatographic surfaces are composed of hydrophobic, hydrophilic, ion exchange, immobilized metal or other chemistries. These surfaces are often used for profiling of proteins from biological mixtures, for biomarker discovery and for assay development. The activated surfaces contain covalently immobilized specific bait molecules such as antibodies, receptors or oligonucleotides and are often used for bio-molecular interaction studies such as protein-protein and protein-DNA interactions. Users of the activated surface arrays are enabled to customize them with their own bait molecules *via* a simple incubation procedure.

The primary ►Expression Difference Mapping™ application of SELDI involves applying a biological sample such as serum, urine, cell lysates or tissue extracts to a spot on a ProteinChip Array and allowing proteins to bind to the surface. The unbound proteins and mass spectrometric interfering compounds are washed away and the proteins that are retained on the array surface are analyzed and detected by laser desorption/ionization TOF-MS using a ProteinChip Reader (Fig. 1). A variety of statistical techniques and bioinformatic software systems are then used to analyze the data and to detect the differences between the protein profiles of two sample sets (1).

The SELDI ProteinChip technology is a versatile platform that can be used for the discovery of disease associated protein biomarkers and for their purification and identification. In addition, the same platform can be used for ►Interaction Discovery Mapping™ applications to detect proteins that interact with the newly discovered biomarker and also to develop predictive



Mass Spectrometry: SELDI. Figure 1 ProteinChip® Array surfaces. (top) Chromatographic surfaces are composed of hydrophobic, ion exchange, immobilized metal affinity capture (IMAC) and normal phase surfaces that are used to capture proteins from biological samples. These surfaces are most useful for Expression Difference Mapping™ studies. (bottom) Biological surfaces are made from reactive surfaces that are used to covalently couple biological molecules such as antibodies, receptors or nucleic acids. These are used for Interaction Discovery Mapping™ studies.

diagnostic assays. In either case, a major benefit of the technology is that the discovery applications lead immediately to the creation of an assay that can be implemented in repetitive studies on large sample populations. This is in contrast to other discovery technologies that often require the search for or creation of a new antibody post-discovery to create an assay.

Clinical Relevance

One of the key applications of SELDI ProteinChip technology is the discovery of disease-associated proteins from crude biological samples. This is achieved by comparing the expression profiles of proteins from normal cells and diseased cells (or normal/diseased fluid samples) and selecting the proteins that are differentially expressed. Proteins can be either over- or under-expressed; SELDI allows the detection of both cases simultaneously. The proteins that are differentially expressed in the diseased cells, once validated on a clinically relevant sample set, can be used as protein biomarkers for disease diagnosis. Since many of the diseases are complex often more than one protein biomarker is needed for accurate diagnosis of the disease. In other words, each disease has specific pattern of proteins or a “protein fingerprint”. These disease specific protein fingerprints can

be analyzed and identified by pattern recognition software that is part of the SELDI ProteinChip technology or by a variety of other software algorithms. A major advantage of the SELDI platform is throughput, enabling a statistically significant population of samples to be analyzed rapidly to uncover specific disease-associated protein peaks and/or disease specific patterns. Alternative protein expression monitoring technologies (e.g. 2-D gel analysis) typically offer significantly lower sample throughput although the elucidation of protein identification often comes more quickly. The concept of the SELDI Expression Difference Mapping approach is that attention is drawn quickly to only the proteins that are relevant to the disease under investigation. Once the fingerprints establish the proteins of highest relevance, the SELDI analysis protocol provides a short list of relevant proteins to be identified as well as establishing a route to rapid purification for the purposes of protein identification.

Petricoin, Liotta and colleagues from FDA-NCI Clinical Proteomics Program published a landmark paper in February 2002 in which they used the SELDI ProteinChip technology to discover a signature pattern of proteins that are specific to ovarian cancer. These protein patterns were used for early detection of ovarian

cancer with more than 90% [specificity](#) and [sensitivity](#) (2). Similar studies in their laboratory and other laboratories discovered specific protein patterns that are used for early detection of breast and prostate cancer with more than 90% specificity and sensitivity. Detection of differentially expressed protein can also lead the discovery of protein based drug targets. For example, David Ho and colleagues at the Aaron Diamond AIDS Research Center published a paper in Science in November 2002 in which they used ProteinChip technology to discover the protein-based inhibitors of HIV replication that appear to be responsible for the fact that 2–5% of HIV-positive patients do not progress to AIDS even over many years (3).

Following Expression Difference Mapping procedures to detect disease specific protein biomarkers, it is important to identify the relevant proteins uncovered. SELDI assisted purification is the key to biomarker identification. Utilizing the wide range of chromatographic chemistries that are present on the ProteinChip Arrays, binding and elution conditions of proteins can be worked out right on the ProteinChip Array thus enabling high speed purification of proteins from crude samples. Once the protein of interest is enriched either “on spot” on a ProteinChip Array or in solution, it is digested with proteases and the masses of the resulting peptide fragments are measured and submitted to a database for protein identification. For confirmation of protein identification, ProteinChip Arrays can be coupled to tandem mass spectrometers through a ProteinChip Interface and the sequence of the peptides can be determined.

Most diseases are caused by functional dysregulation of protein-protein interactions. Understanding the role of protein-protein interactions or protein networks is essential for accurate diagnosis of the disease and for development of targeted therapeutics. SELDI ProteinChip technology can be used to discover the proteins that interact with the newly discovered biomarkers and to decipher the protein networks. Once the protein biomarker is discovered and identified, either the protein biomarker itself or its antibody can be covalently coupled to a ProteinChip Array and used as a bait to discover interacting proteins and protein complexes. Since many of the protein biomarkers are processed products of intact proteins or are post-translationally modified (phosphorylated, glycosylated etc), identification of proteins that interact with biomarkers can lead to the discovery of enzymes that either process or modify these biomarkers. Discovery of the interacting proteins can lead to a better understanding of the biology of the disease. Moreover, these interacting proteins can be the new targets for drug discovery.

References

1. Weinberger S, Dalmasso E, Fung E (2001) Current achievements using ProteinChip Array technology. *Curr Opin Chem Biol* 6:86–91
2. Petricoin EF, Ardekani AM, Hitt BA et al (2002) Use of proteomic patterns in serum to identify ovarian cancer. *Lancet* 359:572–577
3. Zhang L, Yu W, He T et al (2002) Contribution of Human α -Defensin 1, 2, and 3 to the Anti-HIV-1 Activity of CD8 Antiviral Factor. *Science* 298:995–1000

Maternal Transcript

Definition

Maternal transcript refers to mRNA that has been provided via the maternal oocyte into the embryo, and is transcribed prior to the onset of zygotic transcription.

[►Morpholino Oligonucleotides, Functional Genomics by Gene ‘Knock-down’](#)

Matrigel

Definition

Matrigel™ is a solubilized basement membrane preparation extracted from EHS mouse sarcoma, a tumor rich in ECM ([►extracellular matrix](#)) proteins. Its major component is [►laminin](#), followed by collagen IV, heparan sulfate proteoglycans ([►HSPGs](#)), and entactin. At room temperature, Matrigel(tm) Matrix polymerizes to produce biologically active matrix material resembling the mammalian cellular basement membrane. Cells behave very similarly, as they do *In Vivo* when they are cultured on Matrigel(tm) Matrix. It provides a physiologically relevant environment for studies of cell morphology, biochemical function, migration or invasion, and gene expression.

[►Lung](#)

Matrix

Definition

Matrix describes the aqueous compartment of mitochondria that is enclosed by the inner membrane. In the

context of proteomics, matrix denotes a stable material for affinity chromatography, which is coupled with the substance that binds a specific Tag.

- ▶ Mitochondria – Biogenesis and Structural Organization
- ▶ Mitochondrial Myopathies
- ▶ Protein Tags

Matrix Metalloproteinases

Definition

Matrix metalloproteinases (MMPs) are zinc-dependent endoproteinases with the ability to degrade nearly all components of the ▶ [extracellular matrix](#). All matrix metalloproteinases have a pro-domain acting as an internal inhibitor of activity, and activation occurs with the cleavage of this domain. They comprise of a large family with over 20 members. Individual members are products of different genes and show selectivity in the subset of matrix molecules. They are able to degrade and can thus profoundly and selectively remodel the extracellular matrix.

- ▶ [Extracellular Matrix](#)
- ▶ [Proteases and Inhibitors](#)

Matrix-Assisted Laser Desorption/Ionization Mass Spectrometry

Definition

MALDI-MS is the second most commonly used method for ionization of biomolecules in mass spectrometry, in addition to ▶ [ESI](#). Samples are mixed with a UV-absorbing matrix substance and are air-dried on a metal target. Ionization and desorption of intact molecular ions are performed using a UV laser pulse. The mass of the biomolecules of interest is determined by the time of flight to a detector in the electric field (TOF).

- ▶ [Mass Spectrometry: MALDI](#)
- ▶ [Mass Spectrometry: MS/MS](#)
- ▶ [Mass Spectrometry: Quantitation](#)
- ▶ [Proteomics in Cancer](#)

Matrix-CGH (Comparative Genomic Hybridization)

Definition

Matrix-CGH (comparative genomic hybridization) describes a method to detect chromosomal defects in tumor cells, by simultaneously hybridizing DNA from tumor and normal samples that are labeled with different colours onto a microarray with chromosomal DNA fragments.

- ▶ [Microarrays in Pancreatic Cancer](#)

Maturation Process

Definition

Maturation process describes the transformation of a primary RNA transcript realised by enzyme(s) and/or other RNA(s). When it is required, the biological function of an RNA is acquired only after maturation. This is also called processing.

- ▶ [Genetic Code](#)

Maximal Shortening Velocity

Definition

Maximal shortening velocity is a measure to describe the contractile response of individual muscles or muscle fibres. It describes shortening velocity at zero-load.

- ▶ [Muscle Contraction](#)

Maxizyme

Definition

Maxizyme stands for minimized, active, x-shaped (heterodimeric), and intelligent (allosterically controllable) ribozyme.

- ▶ [Catalytic RNA](#)

MBP

Definition

The abbreviation MBP either stands for maltose binding protein, which is a 42.5 kDa protein used to tag recombinant proteins that often increases the solubility of the protein. MBP is also the abbreviation for Myelin basic protein, a major component of myelin.

- ▶ [Glial Cells and Myelination](#)
- ▶ [Protein Tags](#)
- ▶ [Recombinant Protein Expression in Bacteria](#)

MCA/MR Syndrome

Definition

MCA/MR syndrome is the common medical term for a genetic disorder comprising of specific multiple congenital anomalies (MCA) and mental retardation (MR), for example, Down syndrome.

- ▶ [Microdeletion Syndromes](#)

McCune-Albright Syndrome

Definition

McCune-Albright Syndrome (MAS) is a complex endocrine disorder classically defined by the clinical triad of hyperpigmented skin lesions, premature puberty and polyostotic fibrous dysplasia of the bone. Causative molecular defects are activating mutations in *Gαs*.

- ▶ [G-Proteins](#)

MCD

- ▶ [Meesmann Corneal Dystrophy](#)

McKusick Number

Definition

All human genes and genetic diseases are assigned a number in the constantly updated database ▶ [OMIM](#)

(Online Mendelian Inheritance of Man) at ▶ <http://www3.ncbi.nlm.nih.gov/Omim/>.

- ▶ [Glial Cells and Myelination](#)

MCM

Definition

MCM is the abbreviation of minichromosome maintenance genes, which are involved in the initiation of DNA replication.

- ▶ [DNA Helicases](#)
- ▶ [Replication Origins](#)

MDCK Cells

- ▶ [Madin-Darby Canine Kidney \(MDCK\) Cells](#)

MDR

Definition

Multidrug resistance (MDR) is the simultaneous resistance to structurally and functionally unrelated natural product anticancer drugs. MDR in tumor cells can occur intrinsically without treatment, and/or can be acquired by therapy-related modalities, particularly by chemotherapy, followed by a cross-resistance to further chemotherapeutic drugs to which they had never been exposed before. MDR still represents one of the major causes for failure of chemotherapy.

- ▶ [Multi-Drug Resistance](#)

MDR Modulators

Definition

The majority of hydrophobic drugs and chemicals have been demonstrated to interact with membrane transporter proteins MDR1, MRP1 or MRP2, which produce resistance against chemotherapy by actively

extruding a wide range of anticancer drugs from the resistant tumor cells. MDR1-mediated multidrug resistance (MDR) can be reversed by the action of a group of compounds known as MDR modulators or chemosensitizers. First-generation MDR (multidrug resistance) modulators are molecules which affect the drug transport out of a multidrug resistant cell by competitive binding, and thereby inhibiting one or more [▶ABC transporters](#). These modulators are used as reversal strategy for cells and tumors possessing the multidrug resistance phenotype. Clinical administration was limited by adverse drug effects due to the requirement of high modulator concentrations.

[▶Multidrug Resistance](#)

MDR1

[▶Multidrug Resistance Gene 1 \(MDR1\) alias ABCB1](#)

Meconium Ileus

Definition

Meconium ileus designates intestinal obstruction by overly dense meconium; it occurs in the first days of life in 10–20% of cystic fibrosis patients.

[▶Epistasis in Cystic Fibrosis](#)

Medaka as a Model Organism for Functional Genomics

FELIX LOOSLI, JOACHIM WITTBRODT
Developmental Biology Programme, European
Molecular Biology Laboratory, EMBL, Heidelberg,
Germany
felix.loosli@embl.de
jochen.wittbrodt@embl.de

Definition

The recent completion of whole genome sequences provides the data that allow the study of gene function at a genomic scale. On the one hand, comparison of genomes is very important to confirm gene annotations and to identify novel genes on the basis of conserved open reading frames (1). On the other hand, sequence

comparison of whole genomes allows identification of conserved non-coding regions that may be functionally relevant. Thus the functional genomics of a given species profits from the availability of data that allow interspecies sequence comparison. Furthermore annotated genomes are the basis for tools that are essential for functional genomics such as [▶unigene libraries](#) and [▶cDNA](#) [▶microarrays](#) or [▶oligo microarrays](#).

In the near future, functional genomics of vertebrates will profit especially from the whole genome sequences (WGS) of three teleosts, namely puffer fish (*Fugu rubripes*), medaka (*Oryzias latipes*) and zebrafish (*Danio rerio*). Both medaka and zebrafish are well-established genetic vertebrate model systems and the sequencing of their genomes will be finished in the near future. The genome of the puffer fish is already available. This represents a unique situation for comparative genome analysis and for the functional analysis of vertebrate genomes.

Characteristics

Medaka as a Model System

Medaka is a small, egg-laying freshwater fish found primarily in Japan, which combines the advantages of a short generation time, small genome size, well-established genetics and easy husbandry (2). The eggs and the embryos are transparent and allow easy observation by standard microscopy. Under laboratory conditions, the generation time is 8 weeks. The genome size of medaka corresponds to approximately 800 Mb, 1/3 of the human genome and less than half of the zebrafish genome. Medaka is a well-established genetic system. Inbred lines of two highly polymorphic populations are available and protocols for large-scale mutagenesis and efficient [▶transgenesis](#) have been published. Furthermore, reverse genetic approaches for interference with gene function are established (see below).

Medaka is also of special interest with respect to comparative genomics. The genome sequence of the relatively closely related puffer fish (40 Myr) has been determined. The other teleost genetic model system, zebrafish is more distantly related (180 Myr) and its genomic sequence will be available in the near future. Thus, medaka provides an important complementary teleost genetic model system for both comparative sequence analysis (3 teleost genomes) and functional genomics (2 teleost genetic systems). This offers a unique possibility for addressing questions relating to the evolution of genomes, such as gene contents and organization and the identification of regulatory elements and combining these with a functional analysis (2).

Functional Analysis of Genomes

Functional genomics requires tools to test for possible functions of coding and non-coding elements of the

genome. Most widely used are systematic mutagenesis screens, such as have been carried out in invertebrates and vertebrates (zebrafish, see (3, 4); medaka, see (5) and mouse, see (6)). In most of these cases, these forward genetic approaches were chemical mutagenesis screens designed to induce point mutations that affect coding regions of genes or splicing of transcripts. Regulatory regions are not usually identified by such an approach. The advantage of this approach is the efficiency with which random mutations can be induced, which allows screening of large numbers of genomes in a short time. The disadvantage is the difficulty of identifying mutation afterwards, which often requires tedious and time-consuming positional cloning strategies.

Recently, a medaka large-scale mutagenesis screen using ENU as a mutagen has been carried out as a collaborative effort by a number of Japanese and European laboratories (H. Kondoh, M. Furutani-Seiki, pers. communication) (7). The conditions were previously established in a pilot screen (5). The aim of this screen was to identify recessive lethal mutations that affect various aspects of embryonic development. Morphological criteria as well as molecular markers were used to identify mutants.

Other approaches to induce mutations that are easier to identify involve viral or transposon mediated insertional mutagenesis. Insertional mutagenesis using a ►pseudotyped virus has been carried out in zebrafish (8). Currently, these vectors are also being tested in medaka. An alternative system is provided by the ►transposon mediated ►gene trap and ►enhancer trap experiments (9) that have successfully trapped genes and led to the generation of insertional mutants in medaka (Arenz and Wittbrodt, unpublished).

Reverse Genetics

Reverse genetic approaches have also been established for medaka (10). The most reliable method is the morpholino based gene knock-down. Morpholinos are small anti-sense oligonucleotides (about 25 bp) with a morpholine backbone that strongly reduces their enzymatic degradation. These oligos hybridize sequence specifically to their target mRNA and block translation over several days. As in zebrafish, RNA interference (RNAi) appears to result in mainly unspecific developmental defects (Loosli and Wittbrodt, unpublished), precluding the application of this method.

Genetic Tools

To identify the causative gene underlying a given mutant phenotype involves the establishment of linkage between the mutated locus and a polymorphic marker. This requires the use of polymorphic strains. Molecular analyses of different medaka wild catches

consistently indicated the presence of two highly polymorphic populations, the Northern and Southern Japanese populations (2). Recent sequence comparison of the respective loci from the Southern and Northern populations revealed single base pair polymorphisms as high as 3% for intronic regions and about 1% for coding regions. This means in practical terms that for any gene of interest, polymorphisms can be detected between the two.

Inbred Lines

Linkage analysis requires that polymorphisms within a given strain are reduced to a minimum, as achieved in inbred strains. By definition, an inbred strain is homozygous at 99% of all loci; it is isogenic. Such a strain can be obtained after at least 20 generations of successive brother-sister mating. One inbred strain, the *Cab* strain (Southern population) (2), has been established from medaka fish initially obtained through the Carolina Biological Supply Company (North Carolina). A second strain was established using a stock derived from the Northern population that was kept as a closed stock over 10 years (5) (Wittbrodt, unpublished). Further inbred lines from these two populations have been established by other laboratories and successfully used in mutagenesis and mapping experiments (2). Thus, uniquely for lower vertebrate genetic systems, at least two inbred strains with high fecundity and easy breeding are available for the two polymorphic populations of the medaka.

Transgenesis

For conventional production of transgenic fish the foreign DNA is, as in other lower vertebrates (fish and amphibia), injected into the cytoplasm of the one to two cell stage embryo. Although the injected DNA is distributed in a mosaic fashion over the whole embryo, the analysis of a sufficient number of embryos gives a clear picture of a temporal and spatial expression pattern by using reporter genes such as ►GFP or β -gal. The ratios reported for obtaining stable transgenic lines by injecting circular plasmid DNA into the cytoplasm of the two-cell stage embryo vary between 1–5%.

For medaka, a novel approach has been developed to integrate DNA into the genome efficiently (11). It is based on *in ovo* linearization of circular plasmid DNA, containing the transgene of interest flanked by two ►I-Sce I meganuclease recognition sites. Co-injection of this construct together with the I-Sce I meganuclease at the early one-cell stage leads to integration of the transgene at the 1–4 cell stage. The promoter dependent expression pattern can be studied in more than 80% of the injected embryos. Furthermore a striking increase in germline transmission was observed (more than 30%). Therefore, both expression analysis by transient transgenesis and generation of stable transgenic lines

are greatly facilitated by the meganuclease method. Combined with the available genomic resources and comparative approaches this method is mandatory for the efficient analysis of non-coding DNA (promoter and enhancer/silencer elements) of medaka in particular and in comparative approaches with other species in general.

Enhancer/Gene-Trapping

At a very low rate, the meganuclease protocol leads to enhancer trapping effects, where novel patterns are detected that are most probably due to enhancer elements in the vicinity of the integration site (11).

A much higher frequency of enhancer trapping in medaka has recently been described using a transposable element (9). This transposon, an artificially reconstructed member of the *Tc-1* superfamily, *Sleeping Beauty* (*SB*), mediates reporter construct integration in vertebrate tissue culture and in medaka facilitates highly efficient reporter gene integration into the targeted genome (35% transgenesis) at early cleavage stages (1–16 cell stages). The transgene is stably transmitted to and expressed in the subsequent generations. In contrast to the meganuclease approach outlined above, the SB system leads to a high rate of specific spatial and temporal expression patterns (30% of stable lines) indicative of integration in the vicinity of regulatory genomic elements.

The encouraging experimental approaches described above have been used to generate a large number of stable transgenic lines expressing GFP in specific patterns. The advantages of these lines are manifold. They allow the following of the fates of cells and tissues *in vivo* in wild type or mutant embryos and they allow the isolation of the labeled cells from dissociated embryos by ►FACS to either propagate them in culture or to use them as a specific mRNA source for subtractive cloning or expression profiling experiments. Furthermore, ubiquitously expressing fish lines have been successfully used as donor lines for cell transplantation experiments (2).

Libraries, Microarrays and Databases

Systematic sequencing of the medaka genome in a whole genome shotgun approach is under way and will be anchored on a BAC contig map. In this contig the key markers from the genetic map (12) are physically localized. In addition several approaches have been followed to provide tools for functional genomics.

On the transcriptome level, the approach is the systematic identification of all transcripts in two parallel but complementary approaches. Firstly, large numbers of expressed sequence tags (ESTs) are generated by a number of laboratories and deposited in a public database (NCBI). Secondly, a “unigene” set

is generated using oligo fingerprinting on cDNA libraries (Himmelbauer, Berlin).

The combination of these approaches so far has led to the identification of more than 20000 individual transcripts/genes. This information was used to establish cDNA macro-arrays and oligo (70 mer) micro-arrays. It is planned to complete the entire “►transcriptome” of medaka within the next year.

The expression patterns of all genes represented in the “unigene” set are systematically determined by whole mount *in situ* hybridization in a collaborative effort and the information is deposited in a database (13) (►<http://medaka.dsp.jst.go.jp/MEPD/>). The aim is to annotate the genome not just by name and potential function, but also with the expression patterns of the respective genes.

The unigene cluster established allows further applications for functional studies on a genomic scale as the cDNAs represented are enriched for full length clones and the cloning vector allows the *in vitro* generation of mRNAs as well as expression in cell culture, thus extending the possible range of application of this library.

Comparative Approaches

The rate of divergence in neutrally evolving genomic regions is relatively low. Therefore it is necessary to compare genomes of diverged vertebrates to detect functionally conserved regions. The existing teleost model systems provide the required material for such an approach.

Another important source for comparative approaches will be the growing number of mutations that are isolated and identified in medaka and zebrafish. Of special interest will be mutations in these two species that affect ►orthologous genes. This will allow directly comparison between the similarities and divergence of gene function. Recently, such a pair of mutations has been identified in medaka and zebrafish. The medaka *eyeless* mutation has been shown to affect the homeo-domain transcription factor *rx3* (14) and the orthologous zebrafish *rx3* gene is mutated in the recently identified *chokh* mutation (15). Sequence and expression of both *rx3* genes are highly conserved and morphologically both mutations are indistinguishable in their consequences. However at the molecular level, these mutations result in significant differences in that other orthologous genes are differentially affected by loss of *rx3* function. This indicates that *rx3* function has diverged in these species. Thus these mutations offer a unique paradigm for study of the divergence of gene function by changes in gene regulation by cross-species comparison at the molecular level. The ongoing whole genome sequencing projects of medaka and zebrafish provide the tools to compare the regulatory

elements of these genes and to address the evolution of regulatory interactions in vertebrates molecularly .

Clinical Relevance

Taken together the combination of the genomic and genetic tools available in medaka in particular and fish in general facilitate approaches that have not so far been possible in vertebrates. It puts fish into a position that is ideal (and unique) for functional and comparative genomics and as a model system to study human diseases.

References

1. Ureta-Vidal A, Ettwiller L, Birney E (2003) Comparative genomics: genome-wide analysis in metazoan eukaryotes. *Nat Rev Genet* 4:251–262
2. Wittbrodt J, Shima A, Scharl M (2002) Medaka – A model organism from the Far East. *Nat Rev Genet* 3:53–64
3. Driever W, Solnica-Krezel L, Shier AF et al (1996) A genetic screen for mutations affecting embryogenesis in zebrafish. *Development* 123:37–46
4. Haffter P, Granato M, Brand M et al (1996) The identification of genes with unique and essential functions in the development of the zebrafish *Danio rerio*. *Development* 123:1–36
5. Loosli F, Koster RW, Carl M et al (2000) A genetic screen for mutations affecting embryonic development in medaka fish (*Oryzias latipes*). *Mech Dev* 97:133–139
6. Hrabe de Angelis MH, Flaswinkel H, Fuchs H et al (2000) Genome-wide, large-scale production of mutant mice by ENU mutagenesis. *Nat Genet* 25:444–447
7. Shima A, Himmelbauer H, Mitani H et al (2003) Fish genomes flying. *EMBO Rep* 4:121–125
8. Golling G, Amsterdam A, Sun Z et al (2002) Insertional mutagenesis in zebrafish rapidly identifies genes essential for early vertebrate development. *Nature Genetics* 31:135–140
9. Grabher C, Henrich T, Sasado T et al (2003) Transposon mediated enhancer trapping in Medaka. *Gene* 322:57–66
10. Carl M, Loosli F, Wittbrodt J (2002) Six3 inactivation reveals its essential role for the formation and patterning of the vertebrate eye. *Development* 129:4057–4063
11. Thermes V, Grabher C, Ristoratore F et al (2002) I-SceI meganuclease mediates highly efficient transgenesis in fish. *Mech Dev* 118:91–98
12. Naruse K, Fukamachi S, Mitani H et al (2000) A detailed linkage map of medaka, *Oryzias latipes*: comparative genomics and genome evolution. *Genetics* 154:1773–1784
13. Henrich T, Ramialison M, Quiring R et al (2003) MEPD: A medaka gene expression pattern database. *Nucleic Acids Res* 31:72–74
14. Loosli F, Winkler S, Burgtorf C et al (2001) Medaka *eyeless* is the key factor linking retinal determination and eye growth. *Development* 128:4035–4044
15. Loosli F, Staub W, Finger-Baier K et al (2003) Loss of eyes in zebrafish caused by mutation of *chokh/rx3*. *EMBO Rep* 4:894–899

Medial/Lateral Ganglionic Eminence

Definition

The medial and lateral ganglionic eminence are specific regions within the ventricular zone of the brain where the majority of inhibitory interneurons are generated. From there they migrate rostrally to their target region in higher brain areas.

► [Neurons](#)

Medicinal Product

Definition

Fairly equivalent terms of medicinal products are medicines or drugs. In the broader context, the term is also used for biological or biotechnology-derived products applied in or on humans.

► [Clinical Gene Transfer](#)

Medulloblastoma

Definition

Medulloblastoma are severe childhood brain tumours which arise in the cerebellum but can spread throughout the central nervous system. Medulloblastomas are classified as primitive neuroectodermal tumours (► [PNET](#)). Although medulloblastomas are generally sporadic, they have also been associated with hereditary syndromes including Gorlin's syndrome, Turcot syndrome and Rubinstein-Taybi syndrome.

► [Hedgehog Signaling](#)

Meesmann Corneal Dystrophy

Definition

Meesmann corneal dystrophy is an autosomal dominantly inherited eye disorder. The disorder affecting the cornea has been mapped to human chromosome 17, and is caused by mutations in the genes encoding the cornea-specific keratins K3 and K12. Meesmann

corneal dystrophy causes fragility of the anterior corneal epithelium resulting in superficial corneal irritations (►keratitis).

►Heritable Skin Disorders

MEF

Mouse Embryonic Fibroblasts.

►Senescence

Megalin

Definition

Megalin is a member of the low density lipoprotein receptor family most similar to LRP1. Megalin is a 600-kDa endocytotic receptor for a multitude of heterogeneous ligands. Ligands include elements of lipoprotein metabolism, of the blood clotting and fibrinolytic systems, calcium, polybasic drugs, and others. Megalin is expressed in a number of resorptive, often polarity-differentiated epithelia, which are heavily engaged in receptor-mediated endocytosis.

►Wnt/Beta-Catenin Signaling Pathway

Meiosis

Definition

Meiosis is the type of cell division by which germ cells (eggs and sperm) are produced from oogonia or spermatogonia. Meiosis involves a reduction in the amount of genetic material to one half. In a two-step process, comprising of two successive nuclear divisions with only one turn of ►DNA replication, one parent cell produces four daughter cells that have half the number of chromosomes found in the original parent cell. Due to cross-over events and recombination, the genetic content of all descendants is not identical (as opposed to Mitosis).

►Cell Division

►Fragile X Syndrome

►Meiosis and Meiotic Recombination

Meiosis and Meiotic Recombination

CHARLES TEASE

Department of Biological Sciences, University of Warwick, Coventry, UK

ctease@btinternet.com

Definition

Meiosis is a cell division process that occurs only in germ cells (oocytes in females, spermatocytes in males) and is central to the formation of gametes. The principal purpose of this specialized cell division is to produce gametes that are haploid, i.e. with a single set of chromosomes. Thus, in humans, meiosis generates eggs and sperm with 23 chromosomes (a haploid set) from a diploid oocyte or spermatocyte that contained 46 chromosomes (i.e. 2 sets of 23 chromosomes). Germ cells achieve this reduction in chromosome number by undergoing two rounds of nuclear division after DNA replication, rather than one as is normal in somatic cells undergoing mitosis. In the first meiotic division, the two sets of chromosomes separate and one complete set passes to each of the resulting daughter cells. This division therefore achieves the halving of chromosome number. In the second meiotic division, the 2 chromatids of each replicated chromosome separate and segregate to different daughter cells. The end product of meiosis is therefore cells with a haploid number of unreplicated chromosomes. Restoration of the normal diploid number occurs after fertilization of an egg by a sperm.

In addition to reduction in chromosome number, meiosis also increases the genetic variation in the population. This increased variation is achieved in two ways, random assortment of maternally- and paternally-derived chromosomes at the first meiotic division and recombination, a process of reciprocal exchange of material between a pair of chromosomes.

Characteristics

The Stages of Meiosis

By convention, cellular division is described as a series of stages that are defined by the appearance and behavior of the chromosomes. Thus prophase describes condensing chromosomes, metaphase the period when the chromosomes are attached to the microtubules of the spindle and are oriented on the spindle equator, and anaphase the period when the chromatids separate. Meiosis follows this convention, the two division cycles being termed meiosis I and meiosis II; the principal features of the different stages of meiosis are described in Table 1 and Fig. 1. The longest stage of

Meiosis and Meiotic Recombination. Table 1 Chromosome events of meiosis

Stage	
Premeiotic DNA synthesis	Replication of DNA. Cohesin laid down between the newly formed sister chromatids.
Prophase I	Prophase I takes up by far the largest part of the time that a cell spends in meiosis. The crucial events of homologous chromosome pairing and crossing-over occur at this time. Prophase I is conventionally divided into 5 sub-stages:
Leptotene	Chromosome condensation commences. Axial elements are formed at the cores of the chromosomes. Telomeres bind to the nuclear membrane and also tend to congregate in a restricted part of the nucleus to form a chromosomal bouquet. Homologue searching is initiated.
Zygotene	Homology testing is undertaken using a recombination-like process. Synapsis is initiated between the axes of homologous chromosome segments to form the synaptonemal complex (SC).
Pachytene	Synapsis is at its maximum extent. Recombination intermediates, presumed to form during the homology testing process, mature into crossovers.
Diplotene	Synapsis begins to break down.
Diakinesis	Synapsis has completely disappeared and homologues are held together only at positions of meiotic recombination (chiasmata).
Metaphase I	A microtubular spindle is formed; chromosome pairs (bivalents) attach to and align on the spindle.
Anaphase I	Sister chromatid cohesion breaks down except at centromeres and homologous chromosomes segregate to opposite spindle poles to produce haploid nuclei.
Metaphase II	Chromosomes align on the spindle.
Anaphase II	Sister chromatid cohesion lapses at the centromeres, chromatids segregate to opposite spindle poles.

meiosis is prophase I when the chromosomes undergo two characteristic and important features of chromosome behavior, namely homologous chromosome pairing and recombination (4, 7).

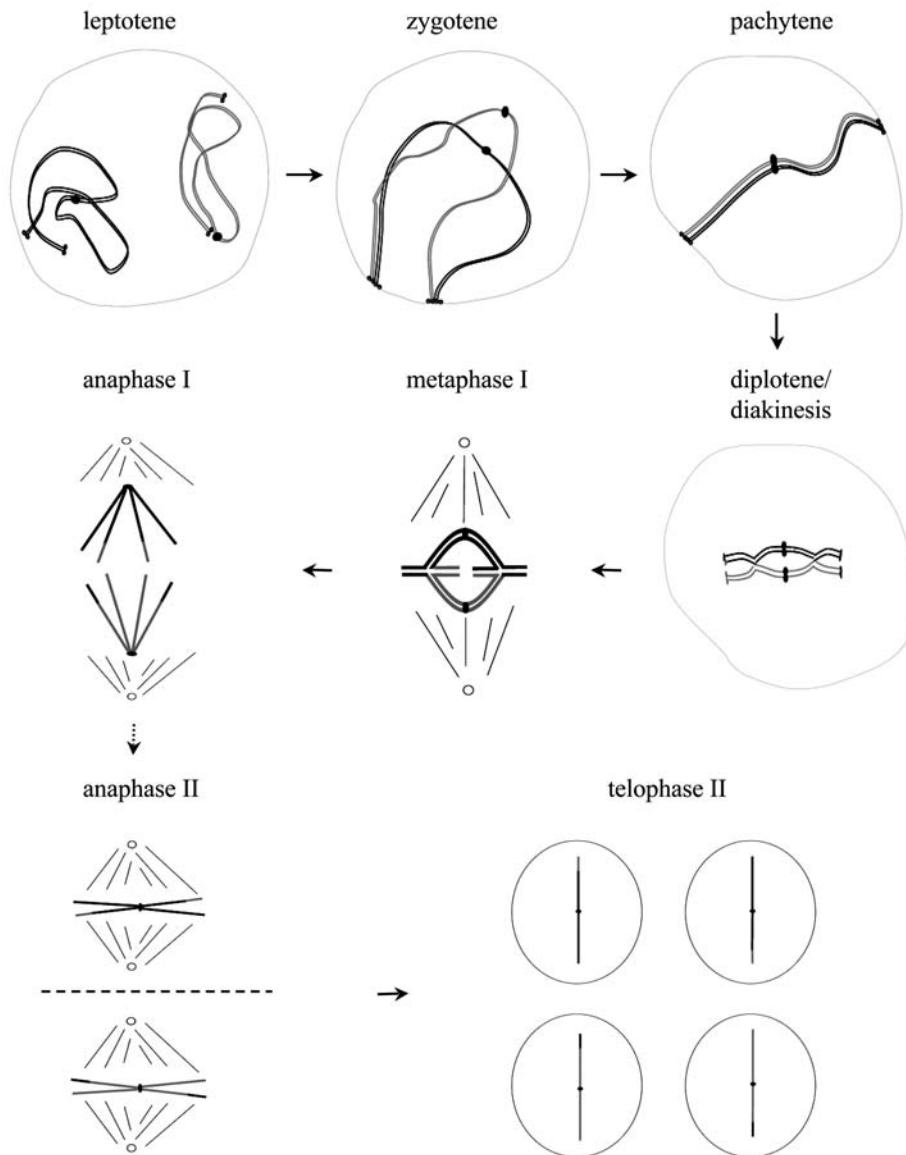
Pre-Meiotic DNA Replication and Sister Chromatid Cohesion

The basic mechanisms of [▶DNA replication](#), for example the origin of replication sites used and the enzymes involved, appear to be similar in mitotic and premeiotic cells. Nevertheless, the time taken to complete replication is considerably longer in the premeiosis S phase. The reason for this has yet to be established but may be related to the need to modify chromatin organization to accommodate the unique aspects of chromosome behavior that occur in meiosis. In addition to replicating the genome, a vital structural re-organization of the chromatin occurs at premeiotic S phase. Following replication, a protein complex called [▶cohesin](#) is laid down between sister chromatids. In somatic cells, the cohesin complex is formed by 4 protein subunits called SMC (structural maintenance of chromosome) 1 α and 3, and SCC (sister chromatid cohesion) 1 and 3. Meiosis-specific variants have been found for 3 of these subunits, SMC1 β (SMC1 α), REC8

(SCC1) and STAG3 (SCC3). These variants replace some or all of the proteins present in somatic cells. This meiotic cohesin complex has important roles in meiotic chromosome pairing, recombination and in the correct segregation of chromosomes at the first meiotic division (2, 6). These roles are described more fully in the sections on chromosome pairing, recombination and segregation.

Homologous Chromosome Pairing

During leptotene, the chromosomes condense and form a continuous proteinaceous axis (Fig. 2). This is in part formed from the cohesin axis and also by the addition of particular meiotic proteins of the [▶synaptonemal complex](#) (SC), the meiosis-specific pairing structure (Fig. 3). To date, three SC proteins have been identified, namely SCP (synaptonemal complex protein) 1, 2 and 3. Two of these proteins, SCP2 and 3, associate with the cohesin axis. When the homologous partner is identified, then synapsis occurs at the axial interface (Fig. 2). A third SC protein, SCP1, stretches between the axes and is the binding protein of synapsis. SCP1 is a long filamentous protein with two coiled domains at the amino and carboxyl termini. The carboxyl end of the protein binds to the chromosomal



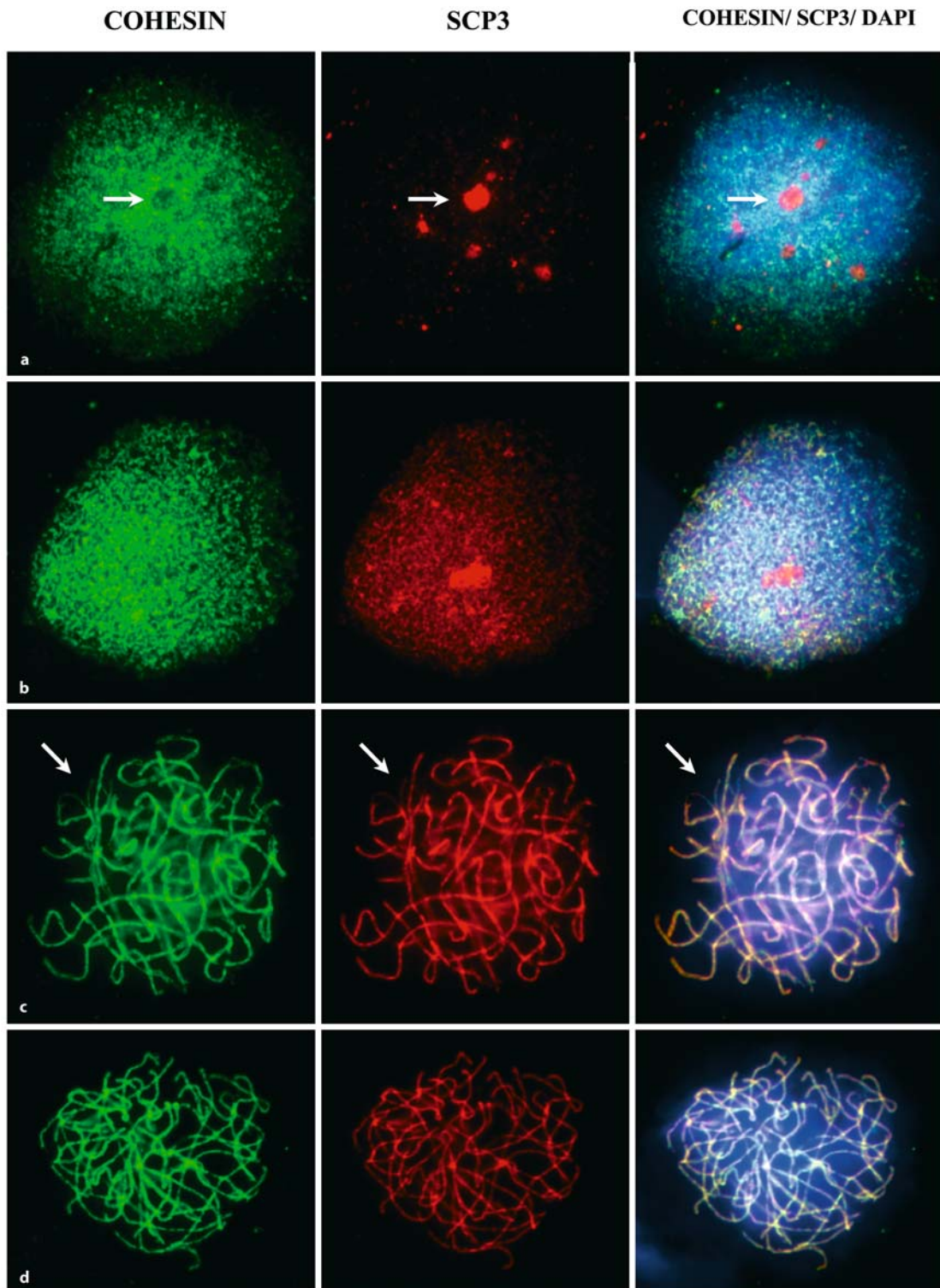
Meiosis and Meiotic Recombination. Figure 1 A schematic drawing of an idealised pair of chromosomes to illustrate chromosome pairing, recombination (chiasma formation) and chromosome and chromatid segregation. The characteristic features of each meiotic stage are described in Table 1.

axes, the amino terminal end to that of another SCP1 protein stretching out from the axis of the homologue (Fig. 3). Homologous chromosome synapsis is thus marked by the formation of a proteinaceous structure that, by electron microscopy, has a tripartite appearance comprising two ►lateral elements and a central region containing a ►central element. This synaptic structure, comprising the cohesin axes and the SC, is crucial for the control of the numbers and distributions of meiotic recombination events.

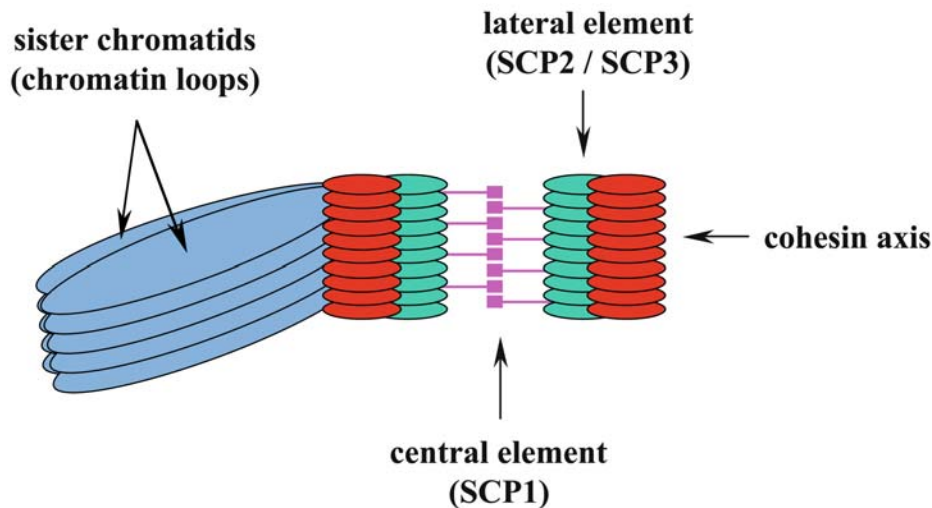
One of the unique features of meiosis is that pairs of homologous chromosomes actively seek out and

intimately associate with their partners during prophase I. This whole process falls within the general heading of pairing, but can be considered to include a number of distinct steps, homology searching and testing, alignment and synapsis.

Homologous chromosomes are usually located in separate regions (sometimes called domains or territories) of the nucleus. Consequently, homologous chromosomes need to undertake an active search for their partner in order to initiate pairing. The mechanism of this search has yet to be detailed, but there is clear evidence that chromosomes move around the nucleus



Meiosis and Meiotic Recombination. Figure 2 Human fetal oocytes at different stages of early prophase I. The cells have been stained simultaneously for the cohesin protein REC8 (green), the axial element (AE) protein SCP3 (red) and chromatin (blue). (a) Preleptotene: the cohesin protein is present throughout the nucleus, while the AE protein is largely confined to nucleoli (arrow). (b) Leptotene: the chromosomes have begun to condense and form linear fibrils composed of both cohesin and AE proteins. (c) Zygotene: homologous pairs of chromosomes have commenced synapsis; this is not complete, however (arrow). (d) Pachytene: the homologues are now fully associated along their axes to form linear synaptonemal complexes (SCs).

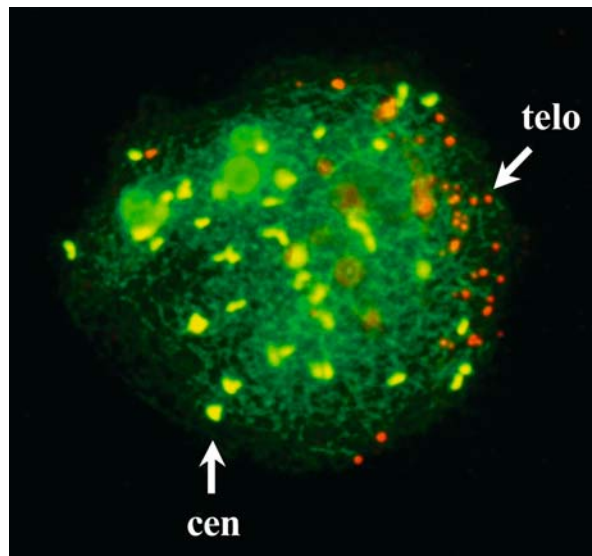


Meiosis and Meiotic Recombination. Figure 3 A cartoon illustrating the axial nature of synapsed chromosomes. Although not shown here, there is evidence that the bases of the chromatin loops extend into the central region of the synaptonemal complex. The bulk of the chromatin, however, lies outside the complex and apparently is not available for recombinational interaction with DNA from the non-sister-chromatids of the homologous partner.

during this process. One of the first identifiable events is the binding of telomeres to the nuclear membrane during leptotene (7). This is accompanied by condensation of the chromosomes into extended linear structures. In most organisms, the telomeres tend to congregate in a restricted region of the nuclear membrane in late leptotene to early zygotene (Fig. 4). This behavior is termed ‘bouquet’ formation and is believed to be part of the homology search process. Bouquet formation tends to be transitory.

Given that there are simultaneously many chromosomes in the nucleus attempting to pair, the question arises as to how the chromosomes distinguish their partners. The current consensus is that homology testing is achieved through a process involving recombination (see below). Where homology is present, then the pair of chromosomes associates (synapse) intimately and forms an SC.

The SC acts as the interface between homologous chromosomes during synapsis. It is important to realize that the bulk of the chromatin that comprises each chromosome is excluded from this synaptic interface. Most of the chromatin is present as long chromatin loops emanating from the chromosomal axis (Fig. 3). For the autosomes and the X chromosome pair in oocytes, synapsis occurs along the entire chromosomal axes. In the male, however, the X and Y chromosomes share homology only over short segments, a 2.6 Mb region at the ends of their short arms, and a 0.23 Mb region at the distal end of the long arm. Synapsis is therefore restricted between these two chromosomes.



Meiosis and Meiotic Recombination.

Figure 4 A leptotene stage, human fetal oocyte with fibrillar axial element formation (green). The centromeres are stained an intense green (arrow) and are located throughout the nucleus. In contrast, the telomeres (red, arrow) are concentrated peripherally in one part of the nucleus as a consequence of the chromosomes forming a bouquet.

Meiotic Recombination

Recombination and pairing/synapsis are intimately associated. The initial step in the pairing process is the formation of double strand breaks (DSBs). Recent

studies have shown that a DNA topoisomerase II-like enzyme, called SPO11, is central to the formation of DSBs. Organisms in which SPO11 is mutated, and therefore not functional, invariably show disrupted chromosome pairing. Once the DSB is formed, exonucleolytic enzymes digest one strand from the 5' end to leave a 3' overhang of single-strand DNA (Fig. 5). DNA repair proteins, such as RAD51 and DMC1, bind to this single-strand DNA and initiate strand invasion of double-stranded DNA, the initial step of homology testing and formation of meiotic crossovers (Fig. 5). In recent years, a number of proteins have been identified in mammalian cells that play a role in meiotic recombination (Table 2). One of the current challenges in meiotic research is trying to piece together the detail of the metabolic pathway of meiotic recombination.

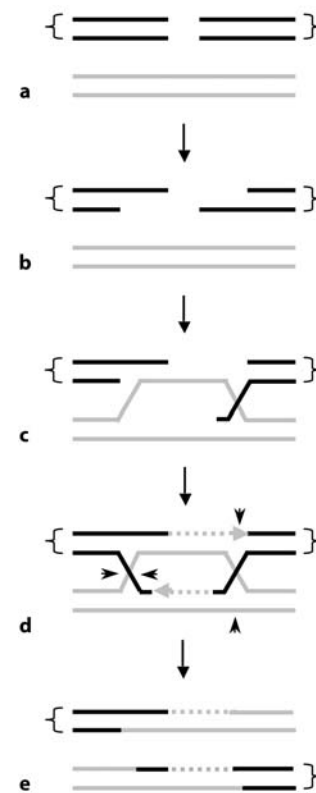
It has long been supposed that the SC is the place at which **crossing-over** occurs between homologous chromosomes and that the SC also plays a crucial role in the control of numbers and distributions of crossovers. This supposition is based on the observation that where SC formation is abnormal, then there is a correlated effect on crossing-over. Recently it was shown that a DNA mismatch repair protein called MLH1 performed an essential role late in the process of meiotic recombination. This protein co-localizes with the SC (Fig. 6) as anticipated if this structure is the site of crossing-over.

After intimate synapsis between homologous chromosomes breaks down in diplotene, sister chromatid adhesion around the position of recombination maintains crossovers between homologous chromosomes. As the chromosomes condense, crossovers become visible cytogenetically as **chiasmata** (Fig. 7). Chiasmata hold the pair of homologues together through to metaphase I. Chiasmate pairs of homologous chromosomes are termed bivalents; if chiasma formation fails, the homologues separate and are present as **univalents**.

Chiasma numbers and distributions have been used to describe patterns of meiotic recombination in human spermatocytes. Various studies have shown that human males have on average about 50 chiasmata per spermatocyte and that these crossovers are biased in distribution towards the distal (telomeric) ends of chromosome arms.

Interestingly, human oocytes show a different pattern of recombination to spermatocytes; a significantly higher number of crossovers, more than 70 per cell on average, are formed and these tend to be more interstitially located than in the male (Figs. 6, 8). The mechanisms that produce such disparate distributions in the two sexes remain unclear.

One aspect of the control of crossover numbers has recently become somewhat clearer. Total SC lengths



Meiosis and Meiotic Recombination. Figure 5 An illustration of the double strand break (DSB) model for meiotic recombination. (a) A DSB is formed by SPO11 in combination with other proteins. (b) Exonucleolytic enzymes digest back (resect) the 5' ends of the DSBs to leave 3' overhangs. (c) Strand exchange proteins, such as RAD51 and DMC1, are recruited to the single strand overhangs. The single-strand DNA invades a duplex on the homologous, non-sister chromatid, displacing a DNA strand to form a D loop structure. (d) After both single strand ends have invaded, DNA synthesis initiates from the 3' ends to replace the digested base pairs. The resulting structure is called a double Holliday junction. (e) The Holliday junction is resolved into a crossover by cutting the Holliday structure in opposite orientations.

can vary between cells within an individual. It has been shown that the numbers of crossovers show a correlated variation. In other words, the length of the physical platform (the SC) available for recombination influences the numbers of crossovers. This effect is also clear when one compares the two sexes. The total length of SC in human oocytes is over 500 μm compared with 250–300 μm in the male and likewise the rate of recombination is higher in female meiosis. It is still unclear what produces this effect on axial chromosome lengths in the two sexes but there is a suspicion that sex-related differences in chromosome condensation are involved. Of course, this still leaves

Meiosis and Meiotic Recombination. Table 2 The timing of appearance and possible function of some recombination-associated proteins

Protein	Stages present	Role
SPO11 (sporulation)	leptotene – zygotene	Produces DSBs, essential for synapsis
RAD51	leptotene – pachytene	RecA homologue that binds to single-stranded DNA; may mediate single end invasion during recombination
DMC1 (disrupted in meiosis)	leptotene – pachytene	RecA homologue that binds to single-stranded DNA; colocalizes with RAD51
RPA (Replication protein A)	zygotene – pachytene	Binds to single-stranded DNA; possibly involved in early steps of recombination
BLM (Bloom syndrome mutated helicase)	leptotene – pachytene	Helicase, colocalizes with RPA
MLH1 (MutL-homologue)	zygotene – diplotene	Required for completion of crossing-over
MLH3	pachytene	Involved in DNA mismatch repair
MSH4/5 (MutS homologue)	zygotene – pachytene	Required for chromosome pairing, may interact with MLH1 to control crossing-over

unanswered the questions as to why such inter-sex variation in meiotic chromosome organization is present in the early part of meiosis and what may be the purpose of the different patterns of recombination in the two sexes.

Chromosome Segregation at Meiosis I

Although the SC begins to break down at the end of pachytene and is lost completely during diplotene, the cohesin axes appear to be maintained. Thus sister chromatids remain closely associated. This characteristic is essential for the maintenance of chiasmata between homologous chromosomes.

During diplotene, diakinesis and metaphase I, it is believed that a cohesin protein complex acts as a “glue” to bind together sister chromatids. It is not clear whether only the meiosis-specific cohesin proteins are involved or if other cohesin proteins, such as those present in somatic cells, also play some role. This will undoubtedly be clarified over the next few years. It is also unclear whether the cohesin proteins remain along the whole chromatid axis (as at early prophase I) or if binding is restricted to only some regions as occurs in mitotic metaphase chromosomes. What is clear, is that absence of meiotic cohesin proteins is incompatible with normal progression through meiosis and that affected germ cells generally die early in prophase I.

At metaphase I, a microtubular spindle forms; the bivalents attach to the microtubules *via* their

►kinetochores and move to the spindle equator. Unlike mitosis, the two kinetochores of each chromosome behave as a single unit and attach to microtubules from the same spindle pole. The kinetochores of each homologue, within a bivalent, orient to opposite spindle poles. When the bivalents are correctly oriented (and there appears to be a cell checkpoint that monitors this process) a signal pathway is activated that eventually results in the removal of the cohesin proteins from between sister chromatids, except at the kinetochores, and the homologous chromosomes segregate from each other (anaphase I). In somatic cells, breakdown of the cohesin complex is achieved through lysis of SCC1 by an enzyme called separase; it is presumed that the same process occurs in meiosis, although in this instance the target of lysis is REC8, the meiosis-specific version of SCC1. One of the principal differences between mitosis and meiosis is that at mitotic anaphase sister chromatids separate along their complete lengths, including ►centromeres, allowing chromatids to pass to opposite spindle poles. In meiosis, in contrast, cohesion breaks down along chromatid arms but remains in place at the centromeres. There is some evidence that the cohesin complex between chromatid arms may be constitutionally different from that in the centromere and that this differentiation may be fundamental to the protection from lysis of REC8 in the centromere at anaphase I of meiosis.

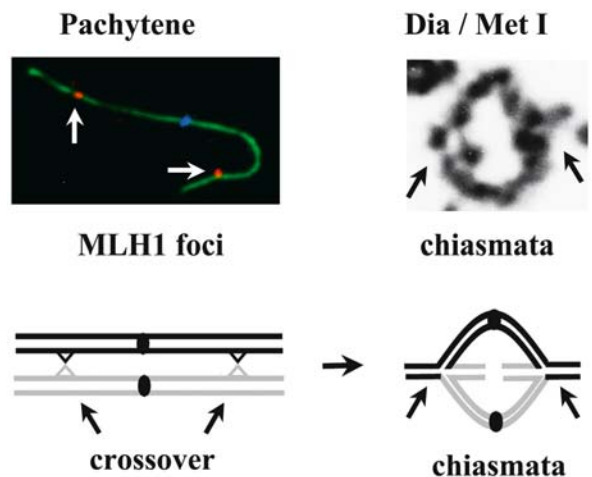


Meiosis and Meiotic Recombination. Figure 6
A pachytene stage spermatocyte and oocyte. The synaptonemal complexes (SCs) are stained red; in the male, centromeres are also stained (blue). The cells illustrate the localization of foci of the recombination protein, MLH1, to the SCs. The cells show the inter-sex variation in number of crossovers: in the male, there are 53 MLH1 foci, in the female, 78. Moreover, in the male, foci are often located close to chromosome ends (white arrows), while in the female, they are generally more centrally positioned (arrow). In the male, the position of the lightly staining X and Y chromosomes is indicated (green arrow).

Chromatid Segregation

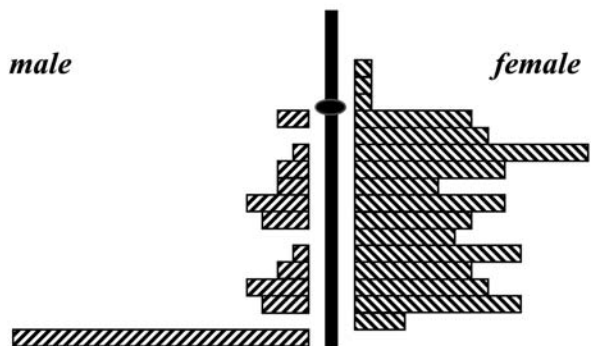
At metaphase II, the two chromatids of each chromosome are held together only at their centromeres. These chromatids are now genetically dissimilar because of recombination, i.e. they are a mixture of sister and non-sister chromatids. At anaphase II, therefore, genetic segregation occurs to produce products that are genetically different. This is in contrast to mitosis where sister chromatids segregate to produce genetically identical daughter cells.

When the chromosomes are attached to the spindle, the cohesin at the centromeres is removed allowing the chromatids to segregate to opposite poles. Currently it is not clear if the cohesin complex is modified after meiosis I to make it responsive to lysis by separase or if



Meiosis and Meiotic Recombination. Figure 7
A single bivalent at pachytene and metaphase I. The photographs (top) show a pachytene synaptonemal complex with 2 MLH1 foci (arrows), and a metaphase I bivalent with 2 chiasmata. Below, is a cartoon to illustrate the formation of recombinant chromatids at metaphase I following crossing-over at pachytene.

Chromosome 21



Meiosis and Meiotic Recombination. Figure 8
A histogram displaying the positions of meiotic crossovers along chromosome 21 in male and female human meiosis. The male shows preferential location of crossovers towards the distal end of the long arm compared with a tendency for more central location in females.

a different mechanism of cohesin breakdown is employed in meiosis II.

Sex Differences in Germ Cell Development

The developmental profiles of oogenesis and spermatogenesis are very different. In most mammals, including humans, female meiosis is initiated during fetal development whereas in the male it begins

postnatally. In oocytes, the early stages of prophase I occur *in utero* with germ cells progressing to diplotene when they enter a cell division arrest phase that lasts until shortly before ovulation in the adult ovary.

Another important sex difference is that there are no self-replenishing oogonial stem cells in ovaries in obvious contrast to the testis that has spermatogonial stem cells. The lack of stem cells in the ovary has the important consequence that female mammals have a finite population of germ cells that is depleted during reproductive life. Female germ cells are also subject to considerable loss during development; late in gestation approximately 70% of oocytes die (a process termed atresia). It is thought that errors in meiotic chromosome pairing may be one of the factors that contribute to the likelihood of cells becoming atretic.

Clinical Relevance

The crucial importance of meiosis to gamete formation is emphasized by the consequences of errors in the process. Meiotic errors can have a variety of adverse outcomes; impaired fertility in the affected individual, generation of *de novo* gene and chromosome mutations, and production of aneuploid gametes.

There is evidence that the complex sequence of chromosome behavior in meiosis is subject to monitoring systems termed cell checkpoints (1, 5). Failure of chromosome pairing or recombination can trigger such checkpoints and cause affected cells to arrest in development. These checkpoints appear to be considerably more efficient in male than female meiosis, as adjudged by the fact that errors in pairing and recombination appear to be more likely to result in arrested cell development in the male. A number of studies have shown that errors of meiotic chromosome behavior are present in germ cells of a higher proportion of men with impaired fertility than in the general population. Notwithstanding this observation, meiosis appears to be more error-prone in human female than male meiosis. For example, one study of synaptonemal complexes in fetal oocytes recorded anomalies in over 30% of the cells. Although most of these cells are believed to be lost through atresia, there is a suspicion that a proportion may survive to contribute to the pool of oocytes in the adult ovary. Pairing errors have an adverse impact on chiasma formation; given the crucial role of chiasmata in chromosome segregation, these cells would be expected to have an increased risk of chromosome [▶nondisjunction](#) at anaphase I.

In human oocytes, the monitoring system at metaphase I for chromosome orientation appears either to be absent or somewhat inefficient. As a consequence, oocytes are able to progress into anaphase I with mal-oriented chromosomes resulting in a higher likelihood of the incorrect segregation of chromosomes (termed

chromosome non-disjunction) compared with spermatocytes. More than 90% of autosomal trisomies (i.e. conceptions with an abnormal chromosome number) are the result of chromosome nondisjunction in female meiosis (3). It is also worth noting that human oocytes also appear to be more likely than the equivalent male cells to have failure of crossing-over, resulting in unpaired univalent chromosomes at metaphase I. Once again, the reason(s) for this considerable sex disparity in meiotic chromosome behavior is unknown.

▶Cell Division

▶Mammalian Fertilization

▶Mitosis

References

1. Abraham RT (2003) Checkpoint signaling: epigenetic events sound the DNA strand-breaks alarm to the ATM protein kinase. *Bioessays* 25:627–630
2. Hagstrom KA, Meyer BJ (2003) Condensin and cohesin: more than chromosome compactor and glue. *Nat Rev Genet* 4:520–534
3. Hassold T, Hunt P (2001) To err (meiotically) is human: the genesis of human aneuploidy. *Nat Rev Genet* 2:280–291
4. Page SL, Hawley RS (2003) Chromosome choreography: the meiotic ballet. *Science* 301:785–789
5. Roeder GS, Bailis JM (2000) The pachytene checkpoint. *Trends Genet* 16:395–403
6. Uhlmann F (2003) Chromosome cohesion and separation: from men and molecules. *Curr Biol* 13:R104–R114
7. Zickler D, Kleckner N (1999) Meiotic chromosomes: integrating structure and function. *Annu Rev Genet* 33:603–754

Meiotic Mapping

Definition

Meiotic mapping is a method of genetic mapping based on the segregation pattern in the progeny of genetic crosses.

▶Mutagenesis Approaches in Yeast

Meiotic Maturation

Definition

Meiotic maturation describes the final step of oogenesis, regulated by hormones. It includes nuclear

breakdown and separation of homologous chromosomes, and results in ovulation of an egg capable of being fertilized by sperm.

► [Mammalian Fertilization](#)

► [Meiosis and Meiotic Recombination](#)

Meiotic Mismatch

Definition

Meiotic mismatch refers to errors in DNA replication and comprises any non-complementary base pair in double stranded DNA. Mismatched DNA molecules are designated heteroduplexes. Their replication produces daughter molecules with different sequences at the site of unpaired bases.

► [Mendelian Forms of Human Hypertension and Mechanisms of Disease](#)

Meiotic Recombination

► [Meiosis and Meiotic Recombination](#)

Melanocytes

Definition

Melanocytes are specialized cells in the basal layer of the epidermis that synthesize and transfer melanin pigments to surrounding keratinocytes.

► [Proteomics in Ageing](#)

Melanosomes

Definition

Melanosomes comprise of a class of membrane-bound organelles containing the pigment melanin, a polymer

of the amino acid tyrosine. It is very abundant in pigment cells called melanocytes.

► [Molecular Motors](#)

MELAS

Definition

MELAS is the acronym for Mitochondrial Encephalomyopathy with Lactic Acidosis and Stroke-like episodes. MELAS is a rare, maternally-inherited mitochondrial disease, due to point-mutations in the mitochondrial tRNA^{Leu(UUR)} gene. The disease is characterized by stroke-like episodes and a mitochondrial myopathy. Organ systems included in the multi-systemic involvement are the central nervous system, skeletal muscle, eye, cardiac muscle, and more rarely the gastrointestinal system.

► [Mitochondrial Myopathies](#)

► [tRNA](#)

Melatonin

Definition

Melatonin is the trivial name for N-Acetyl-5-methoxytryptamin. That is a hormone which is produced by the pineal gland during the night and contributes to the regulation of seasonal rhythms. The suprachiasmatic nucleus (SCN) of the hypothalamus expresses melatonin receptors, and melatonin may have a direct action on SCN to also influence circadian rhythms.

► [Circadian Clocks](#)

Melting Temperature

Definition

Melting temperature (T_m) is operationally defined as the temperature at which half of a temperature-induced transition has occurred. It is not a thermodynamic quantity. The melting temperature (T_m), for instance, characterizes the stability of the DNA hybrid formed between an oligonucleotide and its complementary strand. T_m is critical for determining the optimal temperature at which to use an oligonucleotide as a primer in PCR applications.

► [Thermodynamic Properties of DNA](#)

Membrane

Definition

- Biochemical Engineering of Glycoproteins
- Biological Membranes

Membrane Protein

Definition

Membrane protein refers to a protein that is associated with a membrane, rather than found free in the cytoplasm of a cell. A membrane protein may be integral (embedded or buried) in the membrane, or peripheral (attached more or less loosely, by interactions with either lipid or intergral membrane proteins).

- Biochemical Engineering of Glycoproteins
- Biological Membranes

Memory of Addiction

Definition

Memory of addiction means the neuroplastic changes that took place during drug-taking. Following repeated drug use, various brain structures store memories in various circuits. In the absence of reinforcement, the behavioral response will gradually diminish to finally be extinct. With regular use, tolerance develops and the abuser must use more drugs to achieve the same intensity or effect. Over time, physical dependence and addiction develop.

- Addiction, Molecular Biology

MEN1

- Multiple Endocrine Neoplasia Type 1

Mendelian (Inheritance)

Definition

Mendelian inheritance describes the process by which individuals inherit and transmit to their offspring one of the two alleles present in homologous chromosomes, indicating that variants in single genes are responsible for the trait (elaborated by Gregor Mendel). Several modes of inheritance are recognized e.g. autosomal recessive, autosomal dominant and X-linked inheritance (see also ► Multifactorial Inheritance).

- Cleft Lip Palate
- Common (Multifactorial) Diseases
- COPD and Asthma Genetics
- Mendelian Forms of Human Hypertension and Mechanisms of Disease
- Mitochondrial Myopathies
- Repeat Expansion Diseases

Mendelian Disorder

Definition

Mendelian disorder refers to single gene disorders following simple inheritance patterns as elaborated by Gregor Mendel.

- Atopy Genetics
- Heritable Skin Disorders

Mendelian Forms of Human Hypertension and Mechanisms of Disease

FRIEDRICH C. LUFT

Franz Volhard Clinic, HELIOS Clinical Center Berlin, Max Delbrück Center for Molecular Medicine, Medical Faculty of the Charité, Berlin, Germany
luft@fvk-berlin.de

Synonyms

Mendelian hypertension; monogenic hypertension

Definition

The genetic defects causing Mendelian forms of hypertension are attributable to a defect in a single gene locus. The elucidation of Mendelian forms of hypertension is of great interest in identifying mechanisms that elevate blood pressure. All Mendelian forms of hypertension to date have been autosomal-dominant with a single exception, ►[apparent mineralocorticoid excess](#), which is autosomal-recessive. Complex genetic studies involving populations are underway to map genes that may influence blood pressure and to test associations with genetic variations of known genes important for blood pressure regulation.

Characteristics

The single strongest characteristic found to date in Mendelian hypertension is the role of mechanisms influencing volume homeostasis. All except one form have been so-called “salt-sensitive” forms of hypertension. In these forms, changes in dietary salt intake have a great influence on blood pressure. Thus, the influence of the ►[mutation](#) requires an environmental attribute, namely relatively generous salt intake. However, not all Mendelian forms of hypertension are salt-sensitive: In persons with ►[essential primary hypertension](#), about half are salt-sensitive while half are not.

Glucocorticoid-remediable Aldosteronism (GRA)

Patients with ►[glucocorticoid](#) remediable aldosteronism have an autosomal dominant hypertension and are usually suspected of having ►[primary aldosteronism](#) (►[aldosterone](#)). They have a volume expansion, a salt-sensitive form of hypertension, tend to ►[metabolic alkalosis](#) with ►[hypokalemia](#) and respond to both thiazide ►[diuretics](#) and spironolactone. The latter fact is a clinical clue that ►[mineralocorticoid](#) products may be involved. Their renin values are low while the aldosterone values are elevated. The patients also have 18-hydroxy- and 18-oxo-cortisol, steroids not normally found in urine in large amounts. Recognizing these abnormal products (an intermediate phenotype) led to solving the mystery. Replacement quantities of prednisone ameliorate the hypertension, cause the abnormal steroids to disappear and give the syndrome its name. The abnormal cortisol derivatives and the favorable effects of glucocorticoid treatment suggested that inner cortical zones, which express the gene for 17 β -hydroxylase (CYP17) and are ACTH-responsive, were the source of the excess mineralocorticoids. Two distinct gene products, 11 β -hydroxylase (CYP11B1) and aldosterone synthase (CYP11B2), perform the terminal steps in glucocorticoid and mineralocorticoid biosynthesis, respectively. ►[Linkage](#) analysis in a large ►[pedigree](#) localized the responsible gene to chromosome 8q, precisely at the site where the genes for 11

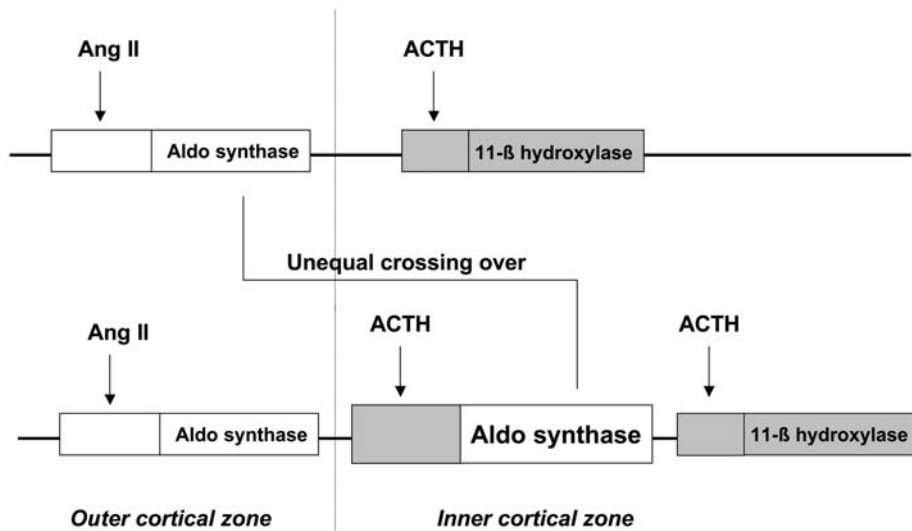
β -hydroxylase and aldosterone synthase also reside. In affected individuals, a chimeric gene consisting of the promoter-regulatory region of 11 β -hydroxylase and the structural portion of aldosterone synthase is located between CYP11B2 and CYP11B1. The chimeric gene results from a meiotic mismatch and unequal crossing over (Fig. 1). The protein product performs all reactions required for aldosterone production; however, the protein is ACTH rather than angiotensin (Ang) II-dependent. Ectopic expression of this protein in the inner cortical zones permits the formation of 18-hydroxy- and 18-oxo-cortisol, the biochemical hallmarks of GRA. Finally, suppressing steroidogenesis in the inner cortical zones with exogenous glucocorticoids alleviates the hypertension.

Liddle's Syndrome

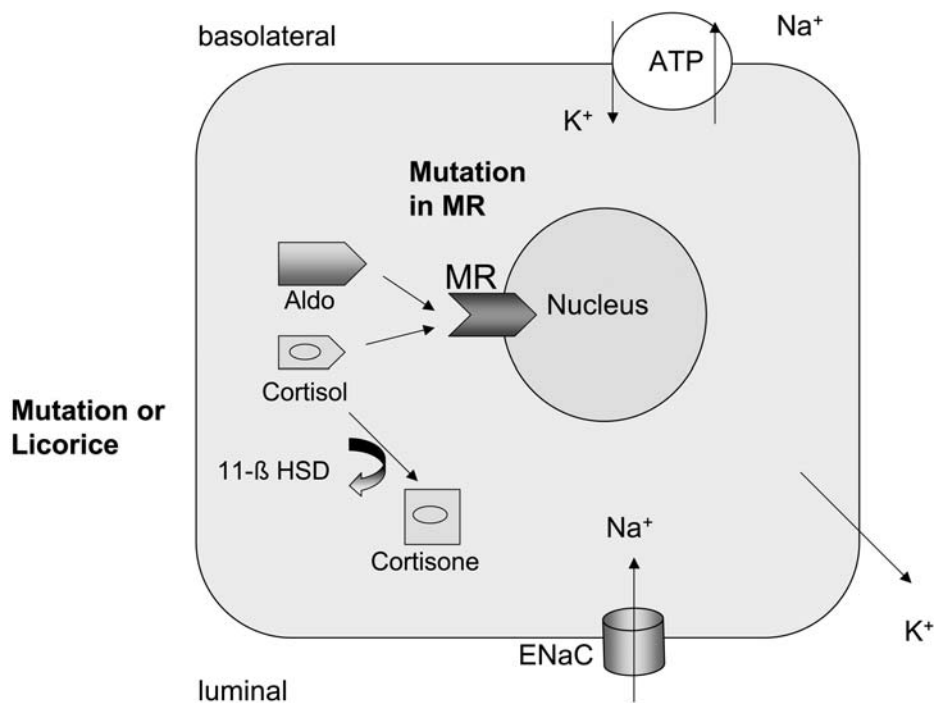
Liddle described patients with autosomal dominant Mendelian hypertension who also tended to metabolic alkalosis with hypokalemia. His patients had low renin and low aldosterone values; however, they did not respond to spironolactone, while thiazides and triamterene reduced the blood pressure. This observation convinced Liddle that they probably did not have a form of mineralocorticoid excess. Liddle speculated that they would show a distal tubular defect of enhanced sodium and chloride reabsorption. A renal transplant performed on a patient with ►[Liddle's syndrome](#) who developed renal failure cured the disease, providing strong evidence that the problem resided within the kidneys rather than in a humoral regulatory system. The responsible gene in a family with Liddle's syndrome was assigned to chromosome 16p. The gene encodes the β -subunit of the epithelial sodium channel (ENaC). The channel is amiloride and triamterene sensitive, explaining the efficacy of these drugs in the syndrome. The channel remains inappropriately permeable even in the face of high salt intake, thereby explaining the salt sensitive hypertension. The molecular mechanisms of Liddle's syndrome involve alteration or deletion in the cytoplasmic tails of the β - or γ -subunits. As a consequence, the channels are not internalized (clathrin-coated pits pathway) or degraded (Nedd4 pathway) and instead remain activated on the cell surface.

Apparent Mineralocorticoid Excess (AME)

Genetic AME resembles the syndrome observed in persons ingesting large amounts of licorice. Licorice gluttony and treatment with carbenoxolone both cause a volume expansion, low renin, low aldosterone, salt-sensitive form of hypertension, which may also feature metabolic alkalosis and hypokalemia. Interestingly, the hypertension responds to both thiazide and spironolactone, but no abnormal steroid products are present in the urine. Both licorice and carbenoxolone contain



Mendelian Forms of Human Hypertension and Mechanisms of Disease. Figure 1 A chimeric gene is formed by meiotic mismatch and unequal crossing over with the promoter region of the 11-β hydroxylase gene (dark box) and the coding region of the aldosterone synthase gene (white box). As a result, the aldosterone synthase gene is under control of ACTH in the inner cortical zone. Aldo, Aldosterone; Ang II, Angiotensin II.

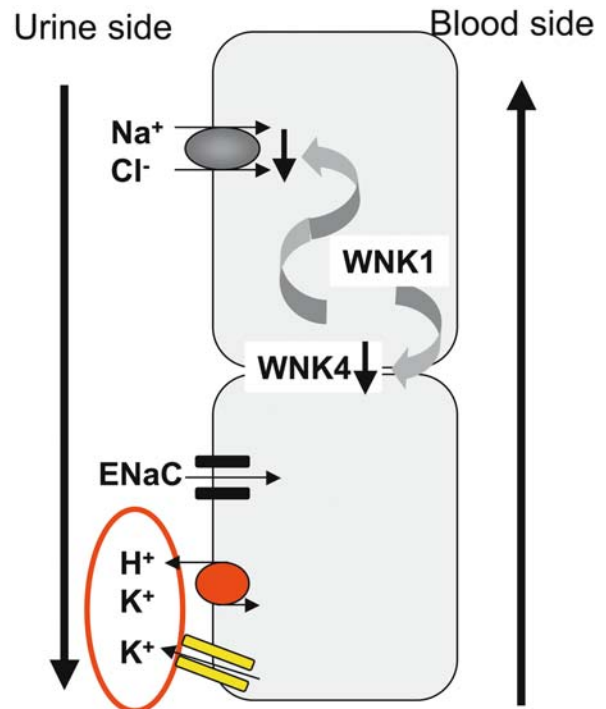


Mendelian Forms of Human Hypertension and Mechanisms of Disease. Figure 2 Cortical collecting duct cell: The mineralocorticoid receptor (MR) has the same affinity for cortisol as for aldosterone (Aldo). The enzyme 11-β hydroxysteroid dehydrogenase (11-β HSD) “protects” the MR by metabolizing cortisol to cortisone, which has no affinity. A mutated or an inhibited enzyme results in an increased intracellular concentration of cortisol and an increased activation of the MR. Mutated MR can result in an altered configuration so that the MR is activated by sterols not containing a 21-hydroxyl group. Increased MR activity causes enhanced Na⁺ reabsorption (ENaC, Na-K-ATPase) and K⁺ excretion.

glycyrrhetic acid, which inhibits 11 β -hydroxysteroid dehydrogenase. The enzyme is responsible for converting cortisol to cortisone. In the distal renal tubule, this step is crucial for protecting the mineralocorticoid receptor, which has the same affinity for cortisol as for aldosterone. This step protects us all from developing AME. Inhibition of 11 β -hydroxysteroid dehydrogenase results in AME. Interestingly, AME may also occur as a rare, autosomal recessive form of hypertension. The 11 β -hydroxysteroid dehydrogenase gene, which has a renal-specific isoform, was a hot [▶candidate gene](#) for this condition. The clinical clues included volume dependent salt sensitive hypertension, tendency to hypokalemia and metabolic alkalosis, low renin and low aldosterone values, responsiveness to both thiazides and spironolactone despite absence of aldosterone or any abnormal mineralocorticoid products and resemblance to licorice gluttony. In 8 of 9 families, mutations in the renal-specific isoform gene for 11 β -hydroxysteroid dehydrogenase were found which indeed rendered the product incapable of converting cortisol to cortisone (2) (Fig. 2). Thus, the mineralocorticoid receptor is unprotected from cortisol in these patients and cortisol functions to occupy the mineralocorticoid receptor.

Mineralocorticoid Receptor

A new Mendelian form of hypertension caused by an activating mutation in the mineralocorticoid receptor was revealed by screening for mutations in the mineralocorticoid receptor in seven unrelated patients referred for possible monogenic hypertension. A mutation in heterozygous state at codon 810 in the mineralocorticoid receptor gene results in a leucine for serine substitution. This residue lies in the hormone-binding domain. The index case had severe hypertension, as did four relatives. Four other relatives had no hypertension. Affected persons all exhibited the leucine for serine substitutions, had low plasma renin activities and low aldosterone concentrations. Since the phenotype resembles Liddle's syndrome, the presence of ENaC mutations had to be ruled out. It was speculated that the mineralocorticoid receptor gene mutation is an activating mutation in the receptor. Interestingly, affected women exhibit worsening hypertension during pregnancy, suggesting that progesterone occupancy of the receptor results in activation rather than inhibition of aldosterone-like effects. Similarly, spironolactone makes the blood pressure elevation worse. The consequences of the mutation MR-S810L allow mineralocorticoid receptor activation by steroids lacking 21-hydroxyl groups (3). The L810 residue in helix 5 of the ligand-binding domain makes a new van der Waals interaction with A773 in helix 3. This interaction eliminates the requirement for the 21-hydroxyl group of aldosterone to interact with N770 in helix 3. The



Mendelian Forms of Human Hypertension and Mechanisms of Disease. Figure 3

Distal nephron: The thiazide-sensitive Na, Cl cotransporter is over-active in pseudohypoaldosteronism type 2. As a result, less Na⁺ is available for the ENaC and fewer K⁺ and H⁺ ions are excreted in the cortical collecting duct. Therefore the syndrome features volume expansion, hypertension, hyperkalemia and mild hyperchloremic metabolic acidosis. "With-no-lysine" (WNK) kinase 4 down-regulates the cotransporter. When the kinase is mutated, the cotransporter is hyperactive. WNK1 regulates WNK4 downward. Gain-of-function mutations in WNK1, would down-regulate WNK4, causing cotransporter hyperactivity.

modification explains why compounds that are normally antagonists now are agonists for the receptor.

▶Pseudohypoaldosteronism Type II (PHA Type II)

PHA type II ([▶Gordon's syndrome](#)) is characterized by familial hypertension with hyperkalemia, slight hyperchloremic metabolic acidosis and otherwise normal renal function. Thiazide diuretics are highly effective in this syndrome, commensurate with salt sensitivity. Multilocus linkage analysis yielded a lod score of 8.1 for linkage to chromosomes 1q and 17p. Interestingly, the chromosome 17p locus overlaps with a syntenic interval in the rat that contains a blood pressure quantitative trait locus. Two genes cause pseudohypoaldosteronism type II (4). Both genes encode members of the WNK family of serine-threonine



Mendelian Forms of Human Hypertension and Mechanisms of Disease. Figure 4 Hand roentgenogram of a 6 year-old Turkish boy with autosomal dominant hypertension (blood pressure 150/90 mmHg) and brachydactyly type E is shown. The white arrow indicates the shortened metacarpal bones, which define this form of brachydactyly. The phalanges are also shortened. Additionally, cone shaped epiphyses are present (marked by arrowhead in the right panel).

kinases. Disease-causing mutations in WNK1 are large intronic deletions that increase WNK1 expression. The missense mutations in WNK4 cause loss of function. The mutations cluster in a short, highly conserved segment of the encoded protein. Both proteins are expressed in the distal nephron (kidney). WNK1 is cytoplasmic, while WNK4 localizes to tight junctions. The *Xenopus* oocyte system was used to show that WNK4 suppresses the thiazide-sensitive sodium chloride cotransporter (5). However, WNK4 does not interfere with cotransporter protein synthesis. Instead, the cotransporter is not incorporated into the cell membrane and cannot come to the surface. WNK1 expression, on the other hand, prevents the WNK4 suppression of the cotransporter. Thus, the WNK kinases serve as a sodium regulatory pathway in the distal nephron (Fig. 3). ▶ **Gain-of-function mutation** in WNK1 would result in highly active cotransporter function. Loss of function mutation in WNK4 would have the same effect.

Autosomal Dominant Hypertension with Brachydactyly

Affected family members with this syndrome featuring ▶ **brachydactyly** type E, have a dramatic increase in blood pressure with age and die before the age of 50 years from multiple strokes. The gene(s) responsible

were mapped to chromosome 12p (6). The hypertension can be easily distinguished from other Mendelian hypertensive syndromes described thus far. The patients are not salt-sensitive and have normal renin, angiotensin, aldosterone and catecholamine responses. By measuring plasma renin activity and plasma aldosterone supine and upright, other conditions can be conveniently excluded. The phenotyping efforts showed that the patients do not respond to any particular form of medication. Beta-blocker, calcium antagonists, alpha-blocker and ACE inhibitor all improve blood pressure without significant differences. A multi-drug therapy is required for the treatment of patients. The mechanism of the hypertension is unknown. However, an additional phenotype was discovered, which may provide a clue, namely anomalous vessels in the posterior fossa that may impinge upon the brain stem. Detailed autonomic testing revealed that the ability of the ▶ **baroreflex** to buffer changes in vascular tone was severely impaired (7). The hypertension could therefore be related to abnormal ▶ **baroreceptor reflex** function in these subjects.

Molecular Diagnostics

Glucocorticoid-remediable aldosteronism can be most easily identified from the clinical picture, response to minimal quantities of glucocorticoids (prednisone 5 mg

daily) or by Southern blot hybridization. Documenting the presence of mutations in the two subunits responsible can show Liddle's syndrome. Apparent mineralocorticoid excess also requires mutation screening of the responsible gene. Only a single mutation in the mineralocorticoid receptor gene has been found to date. Other loci for Gordon's syndrome have been found, but the genes have not yet been identified. Autosomal-dominant hypertension with brachydactyly has a typical type E brachydactyly phenotype. Since the gene(s) are not yet known, we have relied on linkage analysis, cosegregation of microsatellite markers, to obtain confirmatory evidence in additional families.

Common Disease

References

1. Chobanian AV, Bakris GL, Black HR et al (2003) The Seventh Report of the Joint National Committee on Prevention, Detection, Evaluation, and Treatment of High Blood Pressure: the JNC 7 report. *JAMA* 289:2560–2572
2. Mune T, Roberson FM, Nikkilä H et al (1995) Human hypertension caused by mutations in the kidney isozyme of 11 β -hydroxysteroid dehydrogenase. *Nat Genet* 10:394–399
3. Geller DS, Farhi A, Pinkerton N et al (2000) Activating mineralocorticoid receptor mutation in hypertension exacerbated by pregnancy. *Science* 289:119–123
4. Wilson FH, Disse-Nicodeme S, Choate KA et al (2001) Human hypertension caused by mutations in WNK kinases. *Science* 293:1107–1112
5. Yang CL, Angell J, Mitchell R et al (2003) WNK kinases regulate thiazide-sensitive Na-Cl cotransport. *J Clin Invest* 111:1039–1045
6. Schuster H, Wienker TF, Bähring S et al (1996) Severe autosomal dominant hypertension and brachydactyly in a unique Turkish kindred maps to human chromosome 12. *Nat Genet* 4:98–100
7. Jordan J, Toka H, Heusser K et al (2000) Severely impaired baroreflex buffering in patients with monogenic hypertension and neurovascular contact. *Circulation* 102:2611–2618

Mendelian Hypertension

►Mendelian Forms of Human Hypertension and Mechanisms of Disease

Mendelian Inheritance in Man

►MIM

Mental Retardation

Definition

Mental retardation is defined by the US “Individuals with Disability Education Act” as “...significantly subaverage general intellectual functioning existing concurrently with deficits in adaptive behaviour and manifested during the developmental period that adversely affects a child's educational performance”.

►Fragile X Syndrome

Mesenchymal Stem Cells

Definition

Mesenchymal stem cells are adult stem cells that occur in various organs (bone marrow, skin, adipose tissue), and physiologically regenerate mesenchymal cell types including fibroblasts, muscle, bone, tendon, ligament adipose tissue.

►Stem Cells - Overview

Mesenchyme

Definition

Mesenchyme is the part of the embryonic ►mesoderm that consists of loosely packed, unspecialized cells of mesodermal and ectodermal origin, from which connective tissue, bone, cartilage, and the circulatory and lymphatic systems develop.

►Limb Development

►Lung

Mesoderm

Definition

Mesoderm is the middle embryonic germ layer, lying between the ectoderm and the endoderm, from which connective tissue, muscle, bone, and the urogenital and circulatory systems develop.

►Lung

Mesonephros

Definition

Mesonephros [derived from Greek terms *mesos* – middle; *nephros* – the kidneys] is the second excretory organ formed in the mammalian embryo, caudal to the pronephros. The pronephric ducts become mesonephric or Wolffian ducts in later development. The mesonephric duct portion persists in the adult male as sperm ducts.

►SRY – Sex Reversal

Messenger RNA

Definition

Messenger RNA (mRNA) refers to a ribonucleic acid that is transcribed from one strand of the DNA duplex, into a strand of mRNA by the action of DNA-dependent RNA Polymerase II, and serves as a template for protein synthesis. After transcription, the mRNA is processed, e.g. to excise introns, and then transported to the cytoplasm. There, the ORF (open reading frame) is translated into protein. Only 1–5% of the total cellular RNA is mRNA, the rest is ribosomal RNA, transfer RNA and functional RNA.

►Cap-Independent Translational Control

►DNA Chips

►Fragile X Syndrome

►Full Length cDNA Sequencing

►Morpholino Oligonucleotides, Functional Genomics by Gene ‘Knock-down’

►Recombinant Protein Production in Mammalian Cell Culture

►tRNA

Messenger RNA Stability

SHIRLEY R. BRUCE, MILES F. WILKINSON
Department of Immunology, The University of Texas,
MD Anderson Cancer Center, Houston, USA
mwilkins@mdanderson.org

Definition

Genetic information flows from DNA to messenger RNA (mRNA) to protein (DNA→mRNA→protein).

The total amount of protein produced by a cell depends upon the efficiency of this information flow, that is (i) the rate of mRNA synthesis from the DNA template (transcription) and (ii) the rate of protein synthesis from the RNA template (translation). This chapter focuses on another critical variable that dictates the amount of protein produced – the stability of the mRNA template. While sometimes underappreciated, this level of regulation is extremely important since the half-lives of eukaryotic mRNAs can vary from a few minutes to more than a day. The decay rates of most mRNAs encoding housekeeping proteins are constant, whereas the half-lives of mRNAs encoding many cell type specific proteins change in response to specific extracellular and intracellular signals. This allows the amount of protein translated from these regulated mRNAs to vary in response to specific circumstances. The regulation of mRNA stability offers several advantages to a cell. First, rapid changes in mRNA stability can permit transient expression of proteins. This is essential in the case of cytokines and proto-oncoproteins, which have undesirable side effects (including developmental defects and lethality) when expressed for long periods. Second, regulation of mRNA stability works together with the regulation of the rate of mRNA synthesis to maximize the flexibility of gene expression. For example, an mRNA that is intrinsically unstable will reach its steady-state level after transcriptional induction more rapidly than will a stable mRNA. Conversely, the templates for many constitutively expressed proteins are stable mRNAs, presumably because this is energetically more favorable for the cell.

Because mRNA stability is a critical variable that determines the amount of protein ultimately produced, a multitude of factors control this event. Below we describe some of the regulatory elements and RNA-binding proteins that dictate both intrinsic and regulated mRNA stability.

Characteristics

Life of an mRNA

The generation of a translationally competent eukaryotic mRNA requires a series of events coordinated by ►RNA polymerase II (Pol II). A 7-methylguanosine-cap structure (►5' cap) is added in a three-step reaction to the 5' end of the mRNA as it emerges from Pol II (1). This reaction is mediated by a guanylyl transferase and a methyltransferase, both of which are bound to the carboxy-terminal domain (CTD) of Pol II. The 5' cap serves to protect mRNAs from decay by impeding 5'→3' exonucleases from processively degrading the 5' end of mRNAs (discussed below). After 5' capping, approximately 200 adenosine residues (a ►poly(A) tail) are added to the 3' end of the mRNA by

polyadenylation factors bound to the Pol II CTD, including poly(A) polymerase (1). Like the 5' cap, the 3' poly(A) tail protects the mRNA from decay. In addition, these 5' and 3' modifications are crucial for efficient mRNA splicing, nuclear export and translation.

After the mRNA 5' and 3' ends are modified, the mRNA is transported to the cytoplasm. In the cytoplasm, the mRNA is thought to form a circle. This circularization is accomplished by a tight interaction between the eukaryotic translation initiation factor 4F (►eIF-4F) and ►poly(A)-binding protein (PABP), which are bound to the 5' cap and the poly(A) tail, respectively. This closed circle conformation stabilizes the mRNA by sequestering the 5' and 3' ends within a multiprotein complex. In addition, this circular conformation may maximize protein production, as it is thought to allow ribosomes to circle around the mRNA many times to mediate multiple rounds of translation efficiently.

Alterations that break the stable mRNA circle lead to mRNA decay. A common means of breaking the mRNA circle is to remove the poly(A) tail. This ►deadenylation event releases PABP from the 3' end of the mRNA, thereby abolishing the interaction of the 5' and 3' ends of the mRNA. Opening of the mRNA circle exposes the mRNA to rapid attack by RNA cleavage enzymes that recognize the 5' and 3' ends of the linearized mRNA (►exonucleases). As discussed below, the speed of deadenylation is a major rate-limiting step dictating the rate of mRNA decay.

Deadenylation

Two deadenylation complexes have been identified in yeast. The first complex, PAN, contains two exonucleases, Pan2p and Pan3p (2). PAN is probably responsible for the initial trimming of the poly(A) tail that occurs in the nucleus. In addition, PAN contributes to complete mRNA deadenylation in the cytoplasm, which ultimately leads to decay of the mRNA body. The other yeast deadenylation complex consists of Ccr4p and Ccr4-associated factor 1 (Caf1p). Evidence suggests that these two proteins form a complex that mediates most of the mRNA deadenylation in the cytoplasm. While repression of either Ccr4p/Caf1p or PAN activity inhibits mRNA deadenylation, inhibition of both activities is required to prevent deadenylation completely. Therefore, these two complexes may have partially redundant functions.

Several deadenylases have been identified in mammalian cells. The best characterized one is ►poly(A) ribonuclease (PARN). So far, a direct homologue of PARN has not been found in yeast. Interestingly, PARN interacts not only with the poly(A) tail but also with the 5' cap. This suggests that binding of PARN with the 5' cap disrupts the interaction between the 5' cap and the

poly(A) tail, thereby opening up the RNA circle, which stimulates both the rate and processivity of deadenylation.

5'→3' Decay

In yeast, 5'→3' decay is the major pathway that degrades mRNA. For 5'→3' decay, the 5' cap must be removed from the mRNA by decapping protein 1 (Dcp1p) (2). However, Dcp1p can remove the 5' cap only after the mRNA has been deadenylated down to a single adenosine residue. Thus, deadenylation and decapping are coupled. After removal of the 5' cap, the mRNA body is degraded by the cytoplasmic 5'→3' exonuclease Xrn1p (Fig. 1). While most yeast mRNAs are degraded in the cytoplasm, for reasons that are presently unclear, some mRNAs are degraded in the nucleus by the nuclear 5'→3' exonuclease Xrn2p (Rat1p).

Mammalian cells also use the 5'→3' decay pathway to degrade some mRNAs, although *in vitro* studies suggest that it is not the major pathway used to degrade most mammalian mRNAs. Like yeast cells, mammalian cells have decapping enzymes (called ►DCP1 and ►DCP2 in humans) that remove the 5' cap from deadenylated mRNAs to make them susceptible to 5'→3' decay. 5'→3' decay is mediated by murine and human homologues of yeast Xrn1p and Xrn2p that have been identified (called ►XRN1 and ►XRN2 in humans).

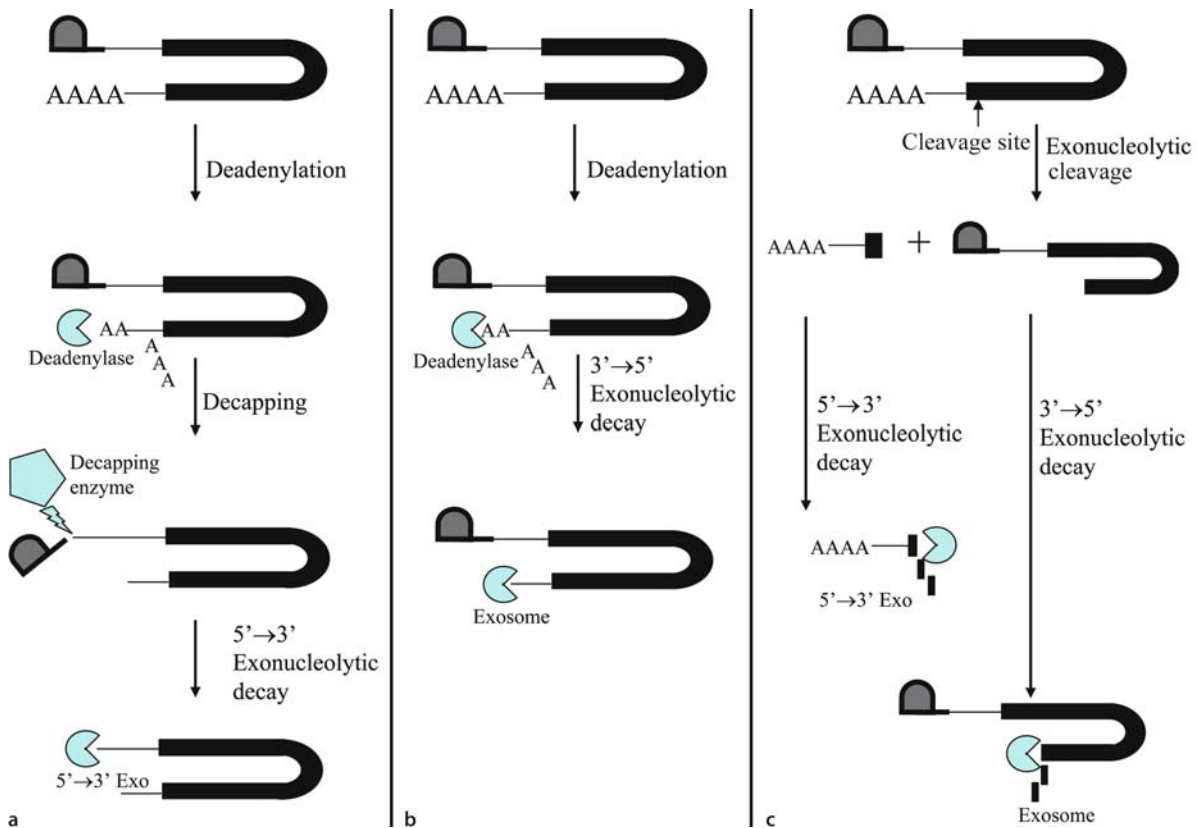
3'→5' Decay

A large ►exosome complex is responsible for 3'→5' decay (Fig. 1). The core exosome is a complex of 10 proteins, most of which are thought to be 3'→5' exoribonucleases. While 3'→5' exonucleolytic decay is the minor ►mRNA decay pathway in yeast, it appears to be the major mRNA decay pathway in mammalian cells (2).

Unlike 5'→3' mRNA decay, exosome-mediated mRNA decay can occur without decapping. However, in some cases decapping may be an obligate downstream consequence of exosome-mediated decay. This decapping is mediated by the scavenger decapping enzyme (►DcpS), which associates with exosome components to form a multi-enzyme degradation complex. The coupling of 3'→5' mRNA decay and decapping, the latter allowing for 5'→3' mRNA decay, provides an efficient and coordinated means of destroying mRNAs.

Endoribonucleolytic Decay

Some mRNAs are degraded by a pathway that is initiated by ►endoribonucleases, enzymes that cleave an mRNA at an internal site. These enzymes generate free 5' and 3' ends so that the mRNA is subject to rapid decay by the 5'→3' and 3'→5' exonucleases just



Messenger RNA Stability. Figure 1 mRNA decay pathways. (a) The 5'→3' exonucleolytic decay pathway. The mRNA is deadenylated, which is thought to open the mRNA circle, thereby exposing the 5' cap to hydrolysis by the decapping enzyme complex. Following decapping, the mRNA is degraded by 5'→3' exonucleases. (b) The 3'→5' exonucleolytic decay pathway. Following deadenylation, the mRNA is degraded by 3'→5' exonucleases present in the exosome. This pathway does not require decapping. (c) The endonucleolytic decay pathway. A site-specific endonuclease cleaves an mRNA at a specific site, which generates free ends that are recognized by 3'→5' and 5'→3' exonucleases, leading to rapid decay of the mRNA body.

described (2). Importantly, the initiation of this endoribonucleolytic decay pathway does not require either deadenylation or decapping (Fig. 1). Examples of mRNAs that are degraded by site-specific endoribonuclease activities are those encoding the growth-promoting proteins *c-myc* and transferrin receptor and the red blood cell protein β -globin. To date, the only mammalian transcript-specific endoribonuclease that has been purified is G3BP, which specifically cleaves between cytosine and adenine residues in the 3' untranslated region (UTR) of *c-myc* mRNA. Although little is known about how endoribonucleolytic decay is regulated, in theory it could be blocked when the cleavage sites are covered by sequence-specific RNA-binding proteins. Conversely, endonucleolytic decay could be reinitiated when the RNA-binding proteins are degraded or removed. Well-studied examples of regulated mRNA decay are described in the next section.

Regulated mRNA Decay mRNA-Stability Regulatory Proteins

The rates of decay of many mRNAs are altered to suit a new circumstance. For example, environmental cues such as stress, growth stimulation and metabolic changes alter the half-lives of many mRNAs. Common mRNA targets of this type of regulation are those encoding cytokines, proto-oncoproteins and growth factors. All of these need to be rapidly increased in levels when there is a need (e.g. to induce cell proliferation or to generate molecules to combat infection). This type of regulation also has the advantage that an mRNA can be quickly degraded when it is no longer needed. If this were not possible, the proteins encoded by these mRNAs could cause serious side effects, including malignancy and anti-host immune activity (i.e. autoimmune disease). One of the most intensely studied classes of elements that regulate mRNA stability is the **▶AU-rich elements**

(AREs). These important *cis* elements are present in the 3' UTRs of proto-oncogene, growth factor and cytokine mRNAs (3). AREs bind to a large group of regulatory proteins, some of which are shown in Table 1, that are collectively called ARE-binding proteins (AUBPs). When bound by their cognate elements, AUBPs alter the stability of the mRNA (Fig. 2). Some AUBPs destabilize their target mRNAs, others stabilize them and still others can either stabilize or destabilize mRNAs, depending on the circumstance. An example of a destabilizing AUBP is tristetraprolin (TTP), which decreases the stability of transcripts encoding the cytokines tumor necrosis factor- α and granulocyte-macrophage colony-stimulating factor. An example of an mRNA-stabilizing AUBP is HuR. Interestingly, HuR not only stabilizes ARE-containing mRNAs but can also promote their translation. AUBPs may also have other functions; for example, HuR may also participate in transporting some ARE-containing mRNA out of the nucleus, as HuR shuttles between the nucleus and the cytoplasm. Exactly how AUBPs regulate mRNA stability is currently under investigation. The AUBPs that promote mRNA decay may act by recruiting the exosome and/or enhancing deadenylation.

The stability of mRNAs can also be regulated by elements other than AREs. These elements can be in the 3' UTR (like AREs), the 5' UTR or the coding region. One of the better-characterized non-ARE elements is in the coding region of the proto-oncogene *c-fos* mRNA. This element, the major protein-coding determinant (mCRD), binds to a complex of at least five proteins, PABP, PABP-interacting protein, hnRNP D, UNR and NS1-associated protein 1 (Table 1) (4). Exactly how the mCRD complex destabilizes *c-fos* mRNA is not known but it must be at least 450-nt upstream of the poly(A) tail to function and it appears to act by promoting deadenylation. Because the mCRD complex contains PABP, one hypothesis is that the complex initially protects mRNAs from decay by sequestering the poly(A) tail from exonuclease attack. Then, after the first round of translation, the ribosome may either displace or reorganize the mCRD complex, thereby exposing the poly(A) tail to exonucleases, which ultimately causes the mRNA to be degraded by the enzymes described in earlier sections.

Nonsense-Mediated Decay (NMD)

Exposure to environmental chemicals and radiation causes DNA mutations, including nonsense mutations and frameshift mutations, both of which can generate ►premature termination codons (PTCs). PTCs also arise because of biosynthetic errors, including errors in transcription and mRNA splicing. Two types of genes that are particularly susceptible to acquiring PTCs are the T-cell receptor and immunoglobulin family genes.

These genes undergo programmed DNA rearrangements that frequently generate frameshifts and thus downstream PTCs (5). The truncated proteins produced from these and other PTC-bearing genes can have dominant-negative or deleterious gain-of-function effects on the cell. These undesirable consequences are prevented or reduced by ►nonsense-mediated decay (NMD), an evolutionarily conserved mechanism that recognizes mRNAs harboring PTCs and degrades them (5, 6).

Two signals are required to elicit NMD. The first signal is a stop codon in frame with the start ATG. In yeast, the second signal is a sequence element found in the 3' end of most *S. cerevisiae* mRNAs. In mammals, the second signal is typically a downstream spliceable intron (5, 6). This second signal serves to restrict the NMD response to mRNAs with premature, not normal, stop codons, as normal stop codons are usually in the 3' terminal exon and thus have no intron downstream. The downstream intron probably does not directly trigger NMD but rather creates a "mark" that is the direct signal eliciting NMD. This mark is a large complex of proteins called the exon junction complex (EJC) that is deposited approximately 20–25 nts upstream of every intron after RNA splicing (Fig. 2). The presence of an EJC after a stop codon triggers NMD, whereas EJCs before the stop codon do not, thereby restricting NMD to mRNAs with PTCs, not those with normal stop codons.

The EJC contains several proteins that have been shown to be essential for NMD, including UPF2 and the nuclear-cytoplasmic shuttling proteins UPF3 and UPF3X (Table 1). UPF2 serves as an adaptor protein that binds to UPF3 and UPF3X, as well as the DEAD-box RNA helicase UPF1 (SMG-2), which also shuttles between the nucleus and the cytoplasm. UPF3X also interacts with Y14, another shuttling EJC component essential for NMD. It remains to be determined how the EJC is assembled, including the order in which its individual components are recruited, as well as exactly how the EJC proteins trigger mRNA decay. Other proteins essential for NMD include SMG-1, -5, -6, and -7, which function in NMD by controlling the phosphorylation status of UPF1 (SMG-1 phosphorylates UPF1, while SMG-5, -6, and -7 promote UPF1's dephosphorylation).

Clinical Relevance

Alterations in mRNA stability can either prevent or promote disease. One example of the former is destabilization of aberrant transcripts harboring PTCs by the NMD pathway (8). NMD reduces the levels of mutant PTC-bearing transcripts encoding truncated proteins with dominant-negative or deleterious gain-of-function activities, thereby improving the phenotype of

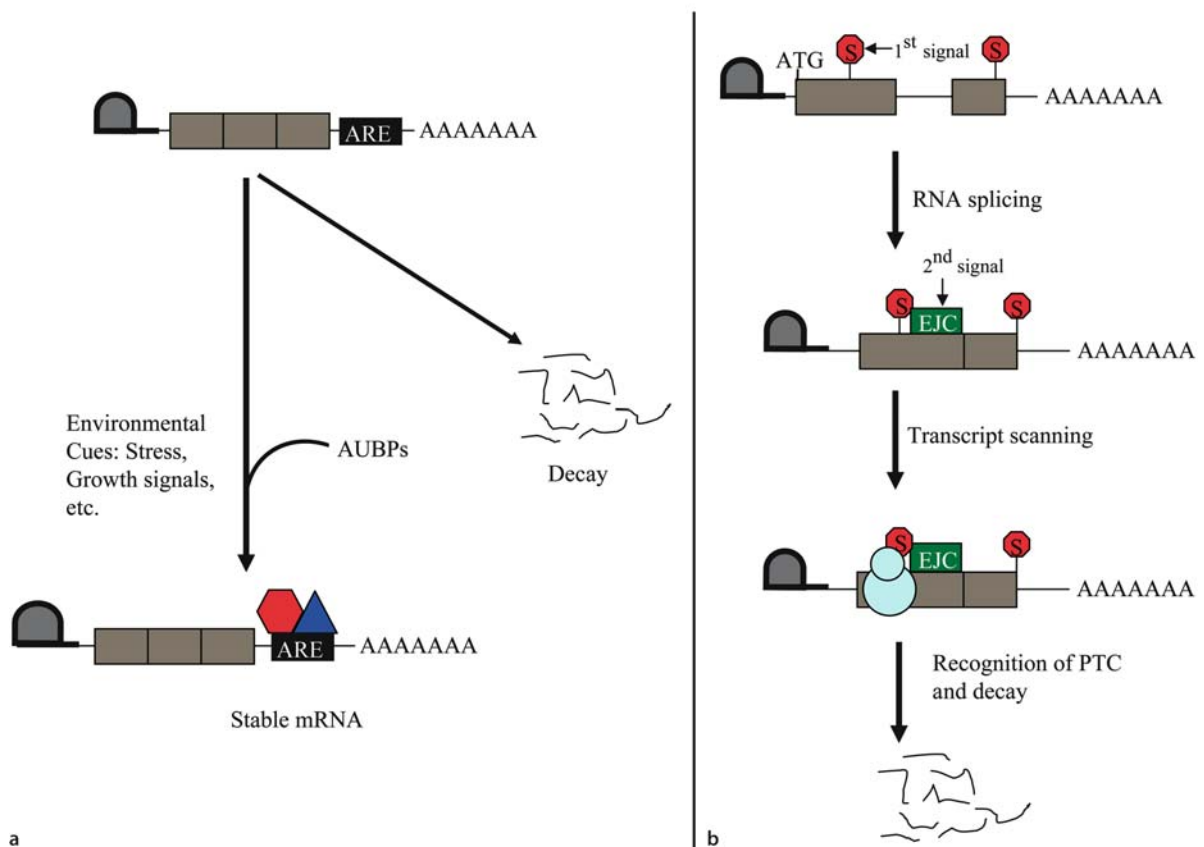
Messenger RNA Stability. Table 1 Factors that mediate or regulate mRNA Decay

Property	Factor	Family	Functions
General	PABP		binds poly(A) tail; promotes mRNA stabilization
	eIF-4F		trimer that binds the 5' cap
	Lsm	Sm-like	required for deadenylation-dependent decay
Nucleases	PARN	RNase D	deadenylase
	PAN (PAN2/3)	RNase D	deadenylase complex
	Caf1p	RNase D	forms complex with Ccr4p
	Ccr4p		deadenylase (catalytic subunit)
	XRN1	Exonuclease	5'→3' exonuclease, cytoplasmic
	XRN2 (RAT1)	Exonuclease	5'→3' exonuclease, nuclear
	Exosome		≥ 10-subunit 3'→5' exonuclease complex
	G3BP	Endonuclease	sequence-specific endonuclease
Decapping	DCP1/DCP2		dimeric complex that hydrolyzes the 5' cap
Regulated mRNA Stability	HuR	ELAV	ARE-binding protein, mRNA stabilizer
	AUF1	hnRNP (D)	ARE/mCRD-binding protein, stabilizer/destabilizer
	TTP	zinc finger	ARE-binding protein, mRNA destabilizer
	NSAP1	hnRNP (R)	mCRD complex-associated protein
	UNR	cold shock	purine-rich sequence-binding mCRD protein
NMD	UPF1 (SMG2)	DEAD box	RNA helicase, ATPase
	UPF2 (SMG3)	unknown	EJC protein that binds UPF1 and UPF2
	UPF3 (SMG4)	unknown	EJC protein, putative 2nd signal for NMD
	UPF3X (UPF3B)	unknown	EJC protein, putative 2nd signal for NMD
	SMG1	PI3-kinase	phosphorylates UPF1
	SMG5	TTP repeats	essential for SMG2 dephosphorylation, associates with PP2A phosphatase
	SMG6	TTP repeats	essential for SMG2 dephosphorylation
	SMG7	TTP repeats	essential for SMG2 dephosphorylation
	Y14	RNA binding	EJC protein that binds to Magoh
	Magoh		EJC protein
	Hrp1p	hnRNP-like	2nd signal that elicits NMD in yeast

many human genetic diseases (5, 6, 8). Disruption of the NMD pathway is embryonic lethal in mice, suggesting that mutations and biosynthetic errors that create dominant-negative or deleterious gain-of-function mutants may be relatively common. There are many cases of mRNA dysregulation that contribute to or cause disease (7). For example, the growth-promoting transcription factor c-myc is over-expressed as a result of mRNA stabilization in some

tumor cell types. This stabilization has been shown to be caused either by mutation of the *c-myc* 3' UTR or by amplification of the gene encoding CRD-BP, a KH-domain protein that binds to and stabilizes *c-myc* mRNA. Another nuclear growth-promoting factor that is over-expressed in tumors as a result of mRNA stabilization is cyclin G1. Rearrangements of the cyclin G1 gene in mantle cell lymphomas cause expression of truncated transcripts that lack AREs and are therefore

Regulated mRNA Decay



Messenger RNA Stability. Figure 2 Regulated mRNA decay. (a) mRNAs containing AU-rich elements (AREs) in their 3' untranslated region are unstable unless environmental cues such as stress, growth factors or antigenic insults cause these mRNAs to be stabilized. ARE-binding proteins (AUBPs) are responsible for regulating the stability of these mRNAs. Some AUBPs elicit mRNA stabilization (shown), while others trigger mRNA destabilization (not shown) (see Table 1). (b) Nonsense-mediated decay (NMD) is a putative RNA surveillance pathway that degrades mRNAs harboring a stop codon at a premature position (a nonsense codon). This RNA decay pathway is triggered when an mRNA contains a stop codon (first signal) followed by the EJC (second signal), a large complex of proteins deposited by the RNA splicing machinery just upstream of exon-exon junctions. Most normal mRNAs only contain a stop codon in the final exon, and thus there is no second signal, rendering these mRNAs impervious to NMD.

stabilized. Secreted factors that stimulate cell growth are also sometimes over-expressed in tumors as result of mRNA stabilization. Examples of this type of dysregulation include transcripts encoding the cytokines IL-1, IL-3, IL-6, GM-CSF and M-CSF. Cytokine mRNA stabilization not only contributes to tumor formation but can also cause inflammation. This has been demonstrated in a transgenic mouse model in which TNF α mRNA stabilization (as a result of removal of the AREs from the 3' UTR) causes chronic inflammatory arthritis and bowel disorders mimicking naturally occurring human conditions. A human disease sometimes associated with

mRNA destabilization, rather than stabilization, is α -thalassemia, a red-blood cell disorder. Naturally occurring mutations in the normal α -globin stop codon cause read-through of ribosomes into the 3' UTR, which dislodges specific RNA-stabilization factors, thereby causing α -globin mRNA destabilization. Dysregulation of mRNA stability is also associated with many other human diseases, including Alzheimer's disease. However, while many of these examples are intriguing, in most cases it remains to be determined whether such alterations in mRNA stability are merely markers of disease or whether they actually cause the disease.

References

1. Shatkin AJ, Manley JL (2000) The ends of the affair: capping and polyadenylation. *Nat Struct Biol* 7:838–842
2. Wilusz CJ, Wormington M, Peltz SW (2001) The cap-to-tail guide to mRNA turnover. *Nat Rev Mol Cell Biol* 2:237–246
3. Bevilacqua A, Ceriani MC, Capaccioli S et al (2003) Post-transcriptional regulation of gene expression by degradation of messenger RNAs. *J Cell Physiol* 195:356–372
4. Grosset C, Chen CY, Xu N et al (2000) A mechanism for translationally coupled mRNA turnover: interaction between the poly (A) tail and a c-fos RNA coding determinant via a protein complex. *Cell* 29:29–40
5. Li S, Wilkinson MF (1998) Nonsense surveillance in lymphocytes? *Immunity* 8:135–136
6. Wagner E, Lykke-Anderson J (2002) mRNA surveillance: the perfect persist. *J Cell Sci* 1:3033–3038
7. Hollams EM, Giles KM, Thomson AM et al (2002) mRNA Stability and the Control of Gene Expression: Implications for Human Disease. *Neurochem Res* 27:957–980
8. Holbrook JA, Neu-Yilik G, Hentze MW et al (2004) Nonsense-mediated decay approaches the clinic. *Nat Genet* 36:801–808

Meta-Analysis

Definition

Meta-analysis is a statistical procedure for combining a number of separate data sets (often obtained in different laboratories) as a single sample.

- [Diabetes Mellitus](#)
- [Genetic Predisposition to Multiple Sclerosis](#)
- [Manic Depression](#)
- [Microarray Data Analysis](#)

Metabolic Alkalosis

Definition

Metabolic alkalosis refers to disturbance(s) in the acid/basic equilibrium indicated by an increased pH-value above 7.44. It is characterized by an increase of bicarbonate ions, loss of H⁺-Ions (exchange of H⁺ for Na⁺ in the renal tubuli), hypokalemia, endocrine imbalances (hyperaldosteronism), and treatment with corticoids.

- [Mendelian Forms of Human Hypertension and Mechanisms of Disease](#)

Metabolite Fingerprinting

Definition

Metabolite fingerprinting describes a multi-parallel metabolic analysis aimed at sample characterisation, which does not require a priori metabolite identification.

- [Metabolomics](#)

Metabolite Phenotype

Definition

Metabolite phenotype characterizes the qualitative and quantitative metabolite composition of a biological sample.

- [Metabolomics](#)

Metabolite Profiling

Definition

Metabolite profiling refers to a multi-parallel metabolic analysis aimed at understanding metabolite as well as sample characteristics, which consequently requires a priori metabolite identification of detector readings.

- [Metabolomics](#)

Metabolome Profiling

- [Metabolomics](#)

Metabolomics

JOACHIM KOPKA
Max Planck Institute of Molecular Plant Physiology,
Golm, Germany
kopka@mpimp-golm.mpg.de

Synonyms

Metabolome Profiling; Metabolite Profiling

Definition

Metabolomics is defined as the science of comprehensive monitoring of the metabolic complement in biological systems. In this sense, metabolomics adds the fourth Rosetta stone to the previously developed systems wide approaches towards global analysis of biological processes, such as genomics, transcriptomics and proteomics. Metabolomics encompasses cellular and secreted metabolites of organisms and includes [▶xenobiotic substances](#), when internalised or subject to biochemical conversion. The chemical nature of metabolites comprises low molecular weight gases, volatile compounds, inorganic and organic ions, highly polar to lipophilic substances and biopolymers, except DNA, RNA and proteins. Multi-parallel metabolomic studies will be instrumental in understanding gene function at the metabolic level and will allow insight into the function of the multitude of primary and secondary metabolites. Moreover, metabolome profiles will contribute to unravelling the mode of action of drugs and increase our knowledge about diseases.

Characteristics

Metabolomics is a young field of analytical science in dynamic development. Currently no single technology platform that would allow the multi-parallel analysis of the complete metabolome is conceivable (1). This is in contrast to genome, transcriptome and proteome profiling technologies, which monitor molecules of highly similar chemical properties, such as DNA, RNA and proteins, respectively. Major obstacles in metabolome analyses are the high diversity of substance properties a metabolome analysis has to accommodate and the vast range of metabolite concentrations. These may range from highly abundant, dominating compounds to equally important trace compounds, which may carry biological signals.

Thus the field of metabolomics may be best discussed starting from technology platforms, which are recruited for comprehensive metabolite analysis, through experimental set-up and designs, to major areas of application and appropriate bioinformatics tools for data mining.

Technology Platforms

Technology platforms in metabolomic analysis utilise universal properties of chemical compounds, such as the mass of molecules and molecular fragments, nuclear magnetic resonance, chromatographic interaction and spectrophotometric properties. No single technology will provide sufficient insight into the metabolome. Instead combinations of technologies will be developed, which have a minimum overlap of

assessed metabolites and which are adjusted to the expected composition of the biological sample under investigation. Four aspects of analytical technologies need to be taken into consideration, the [▶sensitivity](#) and [▶dynamic range](#) of quantitative analysis, the [▶selectivity](#) of detection, the speed of analysis and the potential to unravel the chemical identity of metabolites. In general, so-called hyphenated analytical technologies exhibit the highest potential for use in effective metabolomic studies (1, 2). Hyphenated technologies link compound separation technology such as liquid chromatography ([▶HPLC](#)), gas chromatography (GC), or capillary electrophoresis (CE) to highly specific means of detection, such as optical, spectroscopic detectors using infrared (IR), near infrared (NIR) and ultraviolet (UV) wavelength ranges, surface enhanced Raman spectroscopy (SERS) or induced fluorescent detection (LIF), nuclear magnetic resonance ([▶NMR](#)) or [▶mass spectrometry](#) (MS).

Non-Hyphenated Separation Technologies

GC, LC, and CE have longstanding applications in quantitative chemical analysis and are routinely coupled to single wavelength channel detectors. Single channel detector devices are mostly non-selective, i.e. flame ionisation detection (FID) or single wavelength UV detection. For this reason substance identity has to be established by highly specific, laborious and time-consuming pre-purification or chemical derivatisation steps. Extensive pre-processing is in fundamental conflict with the comprehensiveness required in metabolomic analysis. However, such analyses have the potential to provide highly exact quantitative metabolite information about single compounds or a small number of compounds of the same compound class. Multi-dimensional chromatographic methods, like two-dimensional thin layer chromatography (2D-TLC) or two-dimensional gas chromatography (GC-GC), which both exploit differential separation principles in consecutive steps, appear promising.

Non-Hyphenated Detection Technologies

NMR, MS, or spectrophotometric analysis were first applied to highly purified chemical compounds and used for structural elucidation of novel chemical compounds and metabolites. The development of 2D-NMR, multi-dimensional MS (MS^n), and Fourier transform (FT) spectroscopic applications, i.e. Fourier transform infrared spectroscopy FTIR, allows the application of these methods with a minimum of pre-purification required. Thus 2D-NMR, MS^n , and FTIR are under intense scrutiny for direct applications in metabolomic studies. This set-up of metabolome analysis promises high sample throughput due to the high speed of analysis, but is limited by three

fundamental aspects. Firstly, MS does not provide information about the identity or presence of structural isomers. MSⁿ and MS coupled to highly reproducible molecular fragmentation, such as electron impact ionisation (EI), improves this deficiency but ultimately fails at the level of stereoisomers and epimers. Nevertheless, MS allows extremely high selectivity at the level of mass resolution. Technologies range from mass unit resolution provided by bench-top GC-MS systems to the unsurpassed mass resolution and mass accuracy of Fourier transform ion cyclotron mass spectrometry (FT-ICR-MS) (2). Secondly, NMR and 2D-NMR provide exact insight into the conformation of molecules and even allow non-invasive application to living organisms or cell cultures. However, NMR spectra of complex mixtures are difficult to deconvolute into sets of signals that belong to a single metabolite. Signals of trace compounds are often obscured by predominant metabolites. Thirdly, resolution of spectroscopic properties is usually relatively poor. Thus changes in respective spectra, UV, NIR, FTIR or SERS, from complex samples are usually difficult, if not impossible, to relate to changes in specific metabolites.

Hyphenated Technologies

Hyphenated technologies allow one to counterbalance deficiencies introduced by limitations of the chosen separation or detection technologies. Thus hyphenated methods, e.g. HPLC-NMR, HPLC-UV with photodiode array detection (PAD), HPLC-MS, GC-MS, CE-MS, HPLC-LIF or CE-LIF, currently dominate metabolomic technology. Approximate sensitivities of several methods were estimated (1) at 10⁻⁶ mol (NMR), 10⁻⁹ mol (HPLC-UV), 10⁻¹² mol (GC-MS), 10⁻¹⁵ mol (LC-MS) and 10⁻²³ mol (CE-LIF). In general, analytical technologies appear to gain sensitivity at the expense of selectivity. Technologies aimed at highly concentrated compounds allow generic sample preparation, whereas analysis of trace compounds requires more targeted sample preparation and enrichment from higher amounts of sample. Four technologies will be shortly described.

NMR

NMR is least sensitive of the current metabolomic methods, but the only non-destructive technology available for *in-situ* monitoring. In non-hyphenated applications, high automation and sample throughput can be obtained. Hyphenation to HPLC greatly increases NMR potential by successive analysis of HPLC peaks in stop-flow experiments. Gain in sensitivity is at the expense of greatly increased time required for each sample. The focus of HPLC-NMR is the structural elucidation of non-identified metabolites. Especially effective systems in structural elucidation interface HPLC-NMR with MSⁿ.

HPLC-MS

HPLC-MS is one of the most wide spread technologies. HPLC systems aimed at separation of lipophilic as well as hydrophilic compounds are available. However, the separation properties of HPLC result in broad peaks, a high number of co-eluting compounds and comparatively long times of analysis per sample. In addition, the means of ionisation, for example electrospray ionisation (ESI) are subject to severe ion suppression. This effect suppresses signals of co-eluting and easy to ionise compounds. Suppression in ESI leads to artefacts, such as highly variable measurements or even complete loss of signal. The strength of HPLC-MS is the broad scope of metabolites that can be monitored without need of chemical derivatisation and the availability of a range of different MS detectors. Ion trap ESI-MS allows generation of fragmentation pathways for substance elucidation, triple-quadrupole detectors or quadrupole-time of flight detectors are applied in multi-parallel quantitative analysis. Finally, Fourier transform ion cyclotron mass detectors (FT-ICR-MS) combine high mass resolution with accuracy and optional in-source fragmentation for quantitative and structural analysis, unfortunately at very high cost.

GC-MS

GC-MS is the first relatively low cost but high-throughput technology routinely applied in diagnostic metabolite profiling and plant metabolomic studies. GC provides high separation efficiency, which allows ideal separation of complex biological mixtures. However, GC requires samples to be volatile, a requirement which can readily be accommodated by chemical derivatisation. A toolbox of generic as well as highly specific derivatisation schemes is available from the long history of quantitative GC analysis. Hyphenation to electron impact ionisation MS allows highly reproducible mass spectral fragmentation without the occurrence of suppression effects. Quadrupole GC-MS systems are widespread in quantitative analysis; hyphenation to ion trap MS systems allows structural elucidation and analysis of trace compounds. Novel time of flight (TOF) mass analysers allow rapid detection, enhanced automation and increased sample throughput in fast GC-TOF-MS applications (3).

Experimental Design

Two major approaches have become evident in recent metabolomic studies (4).

► Metabolite Fingerprinting

Fingerprinting approaches in metabolomics – the term metabonomics is more frequently used in this context – are aimed at high throughput classification and description of biological samples. Analysis of the changes in detector readings is sufficient for the fingerprinting of

samples. Unravelling the underlying changes in metabolite composition is either optional or performed off-line after isolation of those detector readings that are discriminatory for the types of samples and conditions under investigation. Fingerprints are often performed with minimal sample preparation, direct infusion and analytical systems with fast duty cycles. Applications may be mainly in FTIR, NIR, SERS and NMR.

Metabolite Profiling

In contrast, metabolite profiling requires *a priori* knowledge about metabolite identity. Ideally, known metabolites are monitored by multi-parallel quantification. Quantification may be absolute or – as routinely performed in gene expression profiling – relative to a pre-defined control sample. This multi-targeted quantitative approach may be extended to those components of biological samples that are yet to be chemically identified, but can be reproducibly and unambiguously recognised. Means of reproducible recognition are combinations of molecular features, such as molecular mass, mass spectral fragmentation, chromatographic retention, specific shifts in NMR and optical spectra. Consequently, metabolic profiling is mostly performed with hyphenated technologies, for example HPLC-MS and GC-MS. In these cases separation of isomers by chromatographic interaction ideally complements high resolution mass detection of molecules and identification by mass spectral fragmentation.

Applications and Bioinformatics Tools

Metabolomic studies, even though currently restricted to relatively small portions of the metabolic complement of microbial organisms, plants, animals and humans (see clinical relevance), are applied in descriptive as well as in functional analysis. Descriptive approaches aim at microbial, plant and animal characterization by analysis of the so-called ►**metabolite phenotype** of samples. The focus of interest is the metabolic diversity in different organs under biotic and abiotic stress conditions, as well as changes in response to nutrient availability or upon genetic modification (2, 5). Applied bioinformatics and statistical tools, which support sample description and classification, are hierarchical. K-means cluster analysis, multivariate principal component analysis, self-organizing maps and probability testing (this overview allows the presentation of only a small selection of the multitude of reported bioinformatics applications) are some of the tools used. Metabolomics contributes to functional genomics approaches and is used for comparatively low cost screening of systematic gene silencing or over-expression projects, which can only be performed if supported by customized metabolic databases and access to public genomic and metabolomic databases (1, 2). Analysis of drug and pesticide modes of action

leads to attempts to assign function to metabolites and xenobiotics. These novel multi-parallel approaches are made possible only by metabolite profiling. Changes of metabolite concentrations may now be understood at systems level in interaction with other metabolites or measurements of protein and mRNA levels. Studies start to attempt metabolic and regulatory pathway reconstruction, metabolic modelling using flux studies, metabolite network analysis and large-scale co-response analysis, e.g. the simultaneous change of metabolite, protein or transcript concentrations (1, 2, 6).

Clinical Relevance

Metabolite profiling was first suggested and applied in the field of clinical chemistry, before metabolomics was discovered for the field of functional genomics. Body fluids, especially blood plasma and urine are subjected to chemical screening *via* metabolite profiling technologies. The aim of these studies is to discover novel biomarkers and indicative metabolite patterns for tumour diagnosis and monitoring of reproductive and other human diseases (2). Metabolomic approaches are currently revisited, because metabolite profiling may represent relatively low cost alternatives to proteomic and transcriptomic clinical analysis. In addition recently improved data mining tools, for example automated mass spectral deconvolution and identification software for GC-MS profiles of complex biological samples facilitate metabolome analyses (1, 2, 3). Increased numbers of novel metabolic markers and patterns will improve diagnosis of human diseases, help monitoring of drug treatment and aid therapy. Pharmacokinetic studies, which currently predominantly utilise radioactive tracing of applied drugs, may be complemented by metabolomic analysis for the detection of unexpected side effects in endogenous human metabolism. Drug development may benefit from discovery and characterisation of early pre-clinical and clinical safety markers and markers for clinical efficacy.

References

1. Sumner LW, Mendes P, Dixon RA (2003) Plant metabolomics: large-scale phytochemistry in the functional genomics area. *Phytochemistry* 62:817–836
2. Harrigan GG, Goodacre R (2003) *Metabolic Profiling: Its role in Biomarker discovery and gene function analysis*. Kluwer Academic Publishers, Dordrecht
3. Wagner C, Sefkow M, Kopka J (2003) Construction and application of a mass spectral and retention time index database generated from plant GC/EI-TOF-MS metabolite profiles. *Phytochemistry* 62:887–900
4. Fiehn O (2002) Metabolomics – the link between genotypes and phenotypes. *Plant Mol Biol* 48:155–171
5. Roessner U, Luedemann A, Brust D et al (2001) Metabolic profiling allows comprehensive phenotyping of genetically or environmentally modified plant systems. *Plant Cell* 13:11–29

6. Urbanczyk-Wochniak E, Luedemann A, Kopka J et al (2003) Parallel analysis of transcript and metabolic profiles: a new approach in systems biology. *EMBO Rep* 4:989–993

Metacentric Chromosome

Definition

Metacentric chromosome is a chromosome on which the centromere is located in the middle, so that the two arms flanking the centromere are of equal length.

► [Centromeres](#)

Metachromatic Leukodystrophy

Definition

Metachromatic Leukodystrophy is a lipid storage disease (sulfatide lipidosis) that is caused by the deficiency of arylsulfatase A. The neuronal accumulation of sulfatide causes progressive degeneration of neurons in the CNS, leading to mental retardation and death within the first decade of life.

► [Polyadenylation](#)

Metalloprotease

Definition

Metalloprotease refers to a type of peptidase that has a metal ion at its active site.

► [Adherens Junctions](#)

Metanephric Blastema

Definition

Metanephric blastema designates loosely organised mesenchymal cells that have the property to form nephrons in the kidney.

► [Kidney](#)

Metaphase

Definition

Metaphase describes the period during cell division when the chromosomes are fully condensed into discrete bodies, and are attached to the Mitotic Spindle, but prior to the segregation of sister chromatids into each of the two daughter cells.

► [Centromeres](#)

► [Meiosis and Mitotic Recombination](#)

► [Nuclear Compartments](#)

Metaphase Plate

Definition

Metaphase plate refers to the plane midway between two spindle poles to which chromosomes align in

► [metaphase](#) cells.

► [Mitotic Spindle](#)

Metastable Disulfide Species

Definition

Metastable disulfide species comprise of a subset of disulfide-protected species, in which both the disulfide bonds and thiol groups are protected, inhibiting all disulfide bond reactions (oxidation, reduction, and reshuffling). Thus, a local or global unfolding step must precede disulfide bond reactions in such metastable species. Two kinds of metastable species exist: disulfide-secure and disulfide-insecure.

► [Protein Disulfide Bonds](#)

Metastasis

Definition

Metastasis describes the ability of tumor cells to leave their site of origin and migrate to other locations in the body, where a new colony is established.

► [Cancer Invasion and Metastasis: Cellular, Molecular and Clinical Aspects](#)

► [Microarrays in Colorectal Cancer](#)

Methotrexate

Definition

Methotrexate is a folate analog which inhibits dihydrofolate reductase. This enzyme is involved in the synthesis of precursors for DNA replication. Methotrexate thus inhibits DNA replication and cell division, and is therefore used as an anti-cancer drug.

► [DNA Amplification](#)

5-Methylcytosine

Definition

5-methylcytosine is a cytosine that is modified by the addition of a methyl group. In vertebrate genomes, 5-methylcytosines are found mainly within CpG dinucleotides.

► [CpG Islands](#)

Methyl Ester

Definition

Methyl ester is a functional group in an organic molecule in which the -OH group in a carboxylic acid is replaced with an -OCH₃ group.

► [Protein Prenylation](#)

7-Methyl-Guanosine

Definition

7-Methyl-guanosine is a guanosine nucleoside with the N-7 position methylated, resulting in a positive electric charge.

► [RNA Capping](#)

Methyl-AcylCoA Racemase Deficiency

Definition

Methyl-acylCoA racemase (AMACR) deficiency is a peroxisomal disorder resulting from mutations in the *AMACR* gene, which codes for the peroxisomal enzyme 2-methyl-acylCoA racemase. It is a slowly progressive neurological disorder resembling ► [infantile Refsum disease](#).

► [Infantile Refsum Disease](#)

► [Peroxisomal Disorders](#)

Methylation of Proteins

MARK T. BEDFORD

The University of Texas M.D. Anderson Cancer Center, Science Park – Research Division, Smithville, MO, USA
mtbedford@mdanderson.org

Definition

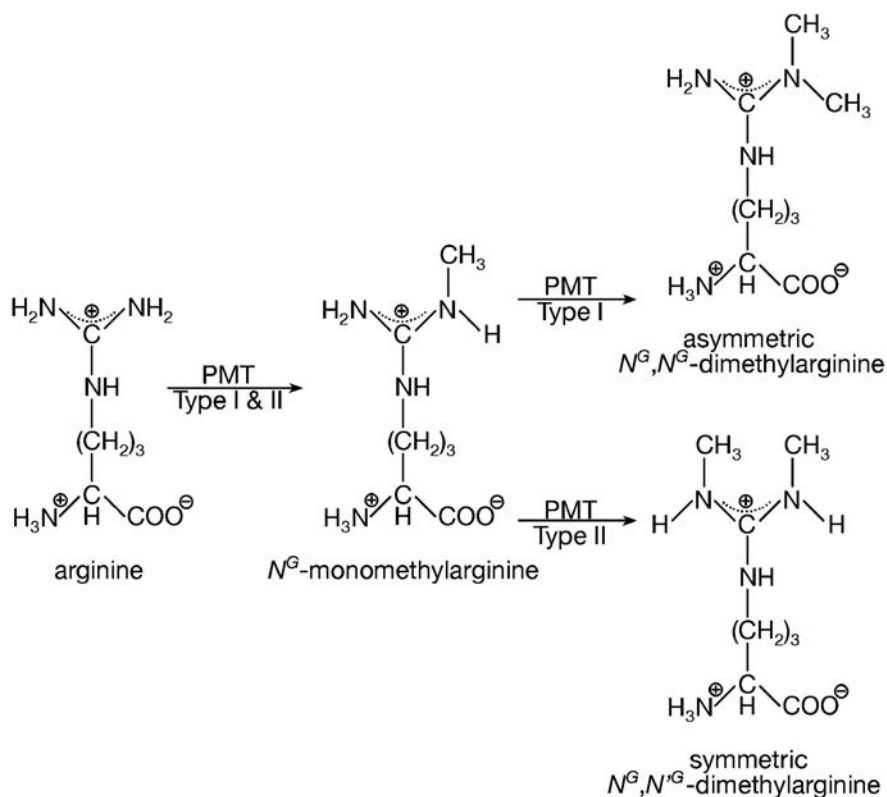
Proteins can be methylated on the side-chain nitrogens of arginine and lysine residues or on carboxy-termini. Protein methylation is a way of subtly changing the primary sequence of a peptide so that it can encode more information. This process of covalently altering a protein after its synthesis is called posttranslational modification. Apart from methylation, proteins can be posttranslationally modified by phosphorylation, acetylation, ubiquitination, sulfation and hydroxylation. Most of these modifications are on the whole reversible; arginine and lysine methylation is relatively stable.

Characteristics

Protein methyltransferases utilize the methyl donor – *S*-adenosyl-L-methionine (AdoMet) – as a cofactor. Lysine and arginine residues are the primary acceptors of methyl groups within the body of a protein. A smaller group of proteins are methylated at their C-terminal ends, either in the absence of additional modification or after processing.

Arginine Methylation

Arginine methylation is a very common posttranslational modification, that occurs predominantly in the nucleus and the main pool of proteins that are modified



Methylation of Proteins. Figure 1 Chemical structure of the methylated forms of arginine. Arginine residues within different sequence motifs and different substrates are methylated by different protein methyltransferases (PMT). The Type I enzymes catalyze the formation of asymmetric $\text{N}^{\text{G}},\text{N}^{\text{G}}$ -dimethylarginine residues and the Type II enzyme catalyzes the formation of symmetric $\text{N}^{\text{G}},\text{N}^{\text{G}}$ -dimethylarginine residues. N^{G} -monomethylarginine is an intermediate that is generated by both enzyme types.

in such a way are those proteins with RNA binding properties (1). In addition, enzymes that facilitate histone acetylation ([▶CBP/p300](#)) and histones themselves are also arginine methylated – thus implicating this modification in [▶transcriptional regulation](#). The methylation of arginine residues is catalyzed by at least two different classes of protein arginine methyltransferase (PRMT) enzymes (Fig. 1). The Type I enzymes catalyze the formation of asymmetric $\text{N}^{\text{G}},\text{N}^{\text{G}}$ -dimethylarginine (aDMA) residues and the Type II enzyme catalyzes the formation of symmetric $\text{N}^{\text{G}},\text{N}^{\text{G}}$ -dimethylarginine (sDMA) residues. Both enzyme types generate N^{G} -monomethylarginine (MMA) intermediates. In mammals, five Type I enzymes (PRMT1, 3, 4, 6 & 8) and a single Type II enzyme (PRMT5) have been described. Regions of proteins that are methylated are often, but not always, glycine and arginine-rich and are referred to as [▶GAR motifs](#). The direct biological consequences of arginine methylation are at the level of protein-protein interactions. More broadly, arginine methylation has been implicated in a variety of cellular processes, including protein trafficking, signal transduction and transcriptional regulation (1).

Lysine Methylation

Histones are proteins that create a hub around which DNA is wrapped to form [▶chromatin](#). The N-terminal tails of the core histones are exposed, and are targets for posttranslational modification and protein recognition that regulate gene expression. The pattern of modifications on the histone tails is referred to as the “[▶histone code](#)” and different combinations of posttranslational modifications (methylation, acetylation and phosphorylation) are recognized by distinct proteins. Lysine methylation plays a central part in the “histone code”. Most lysine methyltransferases contain an evolutionarily conserved [▶SET domain](#) (named after three *Drosophila* proteins that harbor this domain, Su(var), Enhancer-of-zeste and Trithorax). SET domains possess AdoMet-dependent methyltransferase activity, but are structurally distinct from other AdoMet binding proteins (2). Lysine residues can accept up to three methyl groups forming mono-, di-, and trimethylated derivatives (Fig. 2). Different SET domain-containing proteins possess distinct substrate specificities. For example, SET1, SET7 and MLL methylate lysine 4 (K4) of histone H3, whereas Suv39h1, ESET and G9a

specifically methylate lysine 9 (K9) of histone H3. Methylation at K4 and K9 are mutually exclusive and the consequences of site-specific methylation are diametrically opposed. Methylation at K4 correlates with an active state of transcription whereas methylation at K9 is strongly associated with transcriptional repression and [▶heterochromatin](#). Additional lysine residues on histone H3 (K27 and K36) and histone H4 (K20) are also significant sites of methylation by specific SET domain-containing enzymes. Although the histones are the prime target of lysine methyltransferases, other cellular proteins carry *N*-methyllysine residues including elongation factor 1A and the calcium sensing protein calmodulin.

Prenylcysteine Methylation

Eukaryotic proteins with C-termini that end in a CAAX motif are often subjected to a series of posttranslational modifications. This CAAX-tail processing takes place in three steps. First, a prenyl lipid anchor is attached to the cysteine through a thioether linkage. Then endoproteolysis occurs to remove the last three amino acids (-AAX) of the protein to expose the prenylcysteine α -COOH group. Finally, the exposed prenylcysteine group is methylated. The targeted disruption of the methyltransferase for mouse CAAX proteins, isoprenylcysteine carboxyl methyltransferase (ICMT), resulted in mid-gestation lethality (3), thus emphasizing the importance of this modification. The biological function of prenylcysteine methylation is to facilitate the targeting of CAAX proteins to membrane surfaces within cells. Prenylcysteine can be demethylated and this reverse reaction is catalyzed by isoprenylcysteine carboxyl methyltransferases (ICME). CAAX box containing proteins that are prenylcysteine methylated include Ras, GTP-binding proteins, nuclear lamins and certain protein kinases. Many of these proteins participate in cell signaling, and they utilize prenylcysteine methylation to concentrate them in two dimensions on the cytosolic surface of the plasma membrane where they are functional. It has been hypothesized that [▶small molecule inhibitors](#) of ICMT will retard the growth of tumors caused by activated Ras by restricting its

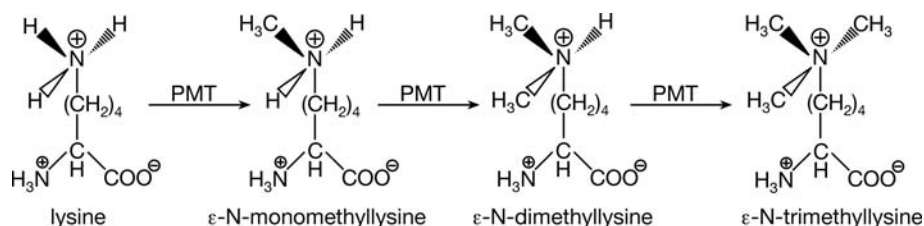
cellular localization to the cytosol and effectively incapacitating the oncogene, but such drugs have yet to be developed.

Protein Phosphatase 2A Methylation

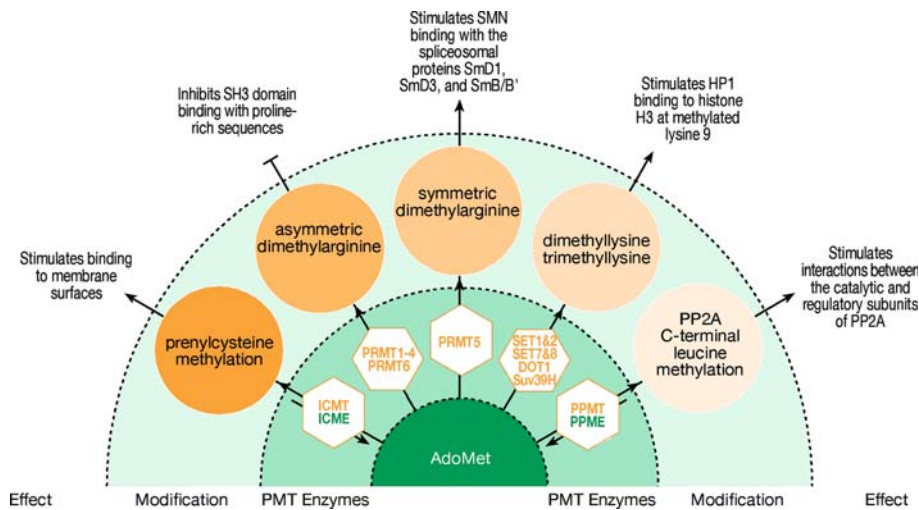
In eukaryotic cells, phosphatases catalyze the removal of phosphate groups from tyrosine, serine and threonine phosphoproteins. The catalytic subunit of the major serine/threonine phosphatases, PP2A, is covalently modified by the reversible methylation of its C-terminus to form a leucine carboxy methyl ester. Unlike CAAX motif methylation, no C-terminal processing is required to facilitate methylation. This C-terminal methylation event regulates the recruitment of regulatory proteins into PP2A complexes through the stimulation of protein-protein interactions, thus indirectly regulating the activity of the PP2A complex (4). Methylation is catalyzed by a unique protein phosphatase methyltransferase (PPMT). The methyl group is removed by a specific protein phosphatase methyltransferase (PPME). These two opposed enzymes make PP2A methylation a dynamic process that fluctuates during the [▶cell cycle](#) and in response to stimuli like cAMP.

L-Isoaspartyl Methylation

As proteins age within a cell they undergo spontaneous damage, including the formation of isoaspartyl peptide bonds. The accumulation of isoaspartyl within proteins is deleterious to the cell as it causes protein instability, loss of biological activity and stimulation of autoimmune responses. A methyltransferase dependent pathway exists for the conversion of L-isoaspartyl back to L-aspartyl. The spontaneous time/age-dependent degradation of L-aspartyl residue results in the formation of a succinimidyl intermediate. This intermediate is spontaneously hydrolyzed either back to L-aspartyl or, in a more favorable reaction, to abnormal L-isoaspartyl. To prevent the accumulation of L-isoaspartyl, this residue is methylated by the protein L-isoaspartyl *O*-methyltransferase (PIMT), which catalyzes the formation of a methyl ester, which in turn is converted back to a succinimidyl intermediate (5). Loss



Methylation of Proteins. Figure 2 Chemical structure of the methylated forms of lysine. Lysine residues within different sequence motifs and different substrates are methylated by different SET domain containing protein methyltransferases (PMT). Lysine residues can be mono-, di-, and trimethylated.



Methylation of Proteins. Figure 3 Graphic depiction of the broad spectrum of biological effects caused by protein methylation.

and gain of function mutations have unmasked the biological importance of PIMT in age-related processes. Mice lacking this enzyme die young as a result of fatal epilepsy, whereas flies engineered to over-express PIMT exhibit an increase in life span of over 30%.

Functional Importance of Protein Methylation

Regulation of Protein-Protein Interactions

A common theme with methylated proteins, as is also the case with phosphorylated proteins, is the role this modification plays in the regulation of protein-protein interactions (Fig. 3). The arginine methylation of proteins can either inhibit or promote protein-protein interactions depending on the type of methylation. The asymmetric dimethylation of arginine residues in close proximity to proline-rich motifs can inhibit the binding to ►SH3 domains (6). The opposite effect is seen with interactions between the survival of motor neurons protein (SMN) and the snRNP proteins SmD1, SmD3 and SmB/B', where binding is promoted by symmetric dimethylation of arginine residues in the snRNP proteins (7).

A well-characterized example of a methylation dependent protein-protein interaction is related to the selective methylation of lysine 9, by Suv39h1 on the N-terminal tail of the histone H3 (2). Di- and trimethylation of this lysine residue facilitates the binding of heterochromatin protein HP1. Because HP1 and Suv39h1 interact, it is thought the binding of HP1 to histone H3 is maintained and even allowed to spread along the chromatin. The HP1 protein harbors a ►Chromo domain that is responsible for the methyl-dependent interaction between HP1 and lysine 9 of histone H3. It is likely that additional Chromo domain

containing proteins will bind the same site as HP1, as well as to other lysine methylated positions on histones H3 and H4.

C-terminal methylation has also been implicated in the regulation of protein-protein interactions. The methylation of the protein phosphatases 2A (PP2A) catalytic subunit enhances the binding of the regulatory B subunit and facilitates PP2A holoenzyme assembly (4). Although the posttranslational addition of methyl groups to amino acids does not alter the charge of the modified residue, it does add one to three fairly bulky groups that are used for structural recognition by protein domains and protein folds.

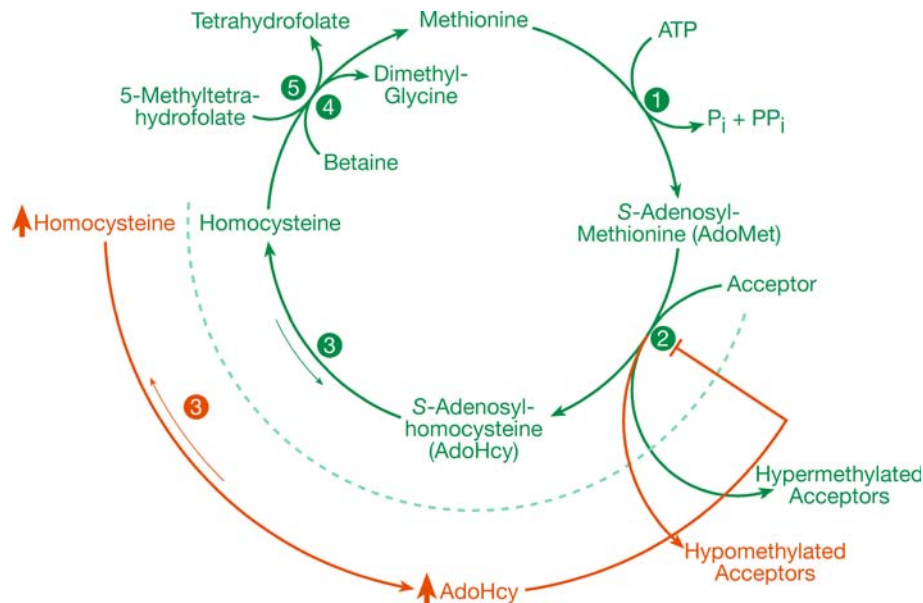
Clinical Relevance

Myelination

Human myelin basic protein (MBP) is symmetrically dimethylated at arginine 107. The methylation of MBP is involved in the maintenance of myelination—the process that insulates axons—carried out by the plasma membrane of a Schwann cell. Arginine methylation augments the interaction between MBP and myelin lipids. A shortage of vitamin B₁₂ causes a decrease in AdoMet levels and an ensuing reduction in methylated MBP, which in turn results in demyelination. Thus, the arginine methylation of MBP is central for the formation and maintenance of compact myelin.

Spinal Muscular Atrophy

Mutations in the survival of motor neurons (SMN) protein cause the neuromuscular disease—spinal muscular atrophy. SMN is involved with the assembly of RNA-protein complexes called small nuclear ribonucleoproteins (snRNPs). snRNPs play a key role in the ►splicing of nuclear pre-mRNA in eukaryotes.



Methylation of Proteins. Figure 4 The activated methyl cycle. The numbers represent enzymes: 1 – methionine adenosyltransferases, 2–AdoMet-dependent methyltransferase, 3–AdoHcy hydrolase, 4–betaine homocysteine methyltransferase, 5–methionine synthase. With elevated homocysteine levels the orange pathway becomes more active.

SMN binds the spliceosomal core proteins SmD1, SmD3 and SmB/B', and this binding is contingent on the symmetric methylation of arginine residues in the C-terminal regions of these splicing factors. Whether protein methylation levels play a role in spinal muscular atrophy has yet to be determined.

Regulation of Nitric Oxide Synthases

Nitric oxide (NO), which is synthesized by nitric oxide synthase (NOS), plays multiple roles in the cardiovascular system. L-arginine is the substrate for endothelial NO generation. The arginine analogues, N^G -monomethylarginine (MMA) and asymmetric N^G, N^G -dimethylarginine (aDMA), but not symmetric N^G, N^G -dimethylarginine (sDMA) (Fig. 1), inhibit the activity of NOS. Proteolysis of methylated proteins provides the cellular source of free methylarginine residues. This methylarginine pool size is controlled by at least two dimethylarginine dimethylaminohydrolase (DDAH) enzymes, which specifically hydrolyze MMA and aDMA but not sDMA. The NOS inhibitory properties of free methylarginine are thus under tight enzymatic control.

Cancer

Arginine and lysine methylation play a central role in the regulation of transcription. The arginine methyltransferases, PRMT1 and PRMT4 are transcriptional coactivators of nuclear steroid receptors (►nuclear hormone receptors) and as such are involved in hormone

responsive tumor progression. With regard to lysine methylation and SET domain containing proteins, the best link to cancer has been established by translocations at the *MLL1* locus. The *MLL1* protein contains a SET domain, and ►chromosomal translocations involving the *MLL1* gene occur in about 80% of infant leukemias. Another SET domain containing protein, *RIZ1*, has been implicated in a broad spectrum of human cancers. The *RIZ* gene maps to the distal short arm of chromosome 1, a region that is often rearranged in cancers. On a molecular level, frameshift and missense mutations of *RIZ1* have been found in human tumors, and reduced *RIZ1* expression is associated with many types of cancer. The many associations between protein methylation and cancer are consistent with these enzymes playing important regulatory roles in the cell.

Hyperhomocysteinemia

Hyperhomocysteinemia is caused by disturbances of the methyl cycle (Fig. 4), which is responsible for the conversion of homocysteine into methionine and then the generation of the methyl donor—S-adenosyl-L-methionine (AdoMet). Methylation reactions that use AdoMet as a co-factor produce a methylated product (proteins, DNA, RNA, lipids, polysaccharides or small molecules) and S-adenosyl homocysteine (AdoHcy). AdoHcy is then broken down by AdoHcy hydrolase into adenosine and homocysteine. If the remethylation of homocysteine is blocked, then accumulated homocysteine will drive AdoHcy hydrolase to synthesize

AdoHcy in an energetically favorable reverse reaction (Fig. 4, highlighted in orange). AdoHcy is a potent product inhibitor of cellular methyltransferases and its elevated levels result in hypomethylation of a broad spectrum of substrates. Epidemiological and clinical studies have identified hyperhomocysteinemia as a risk factor for heart disease, neurodegenerative diseases (like Alzheimer's disease) and neural tube defects. Elevated levels of homocysteine can be caused either by enzyme perturbations at the genetic level or by vitamin deficiency.

References

1. McBride AE, Silver PA (2001) State of the arg: protein methylation at arginine comes of age. *Cell* 106:5–8
2. Kouzarides T (2002) Histone methylation in transcriptional control. *Curr Opin Genet Dev* 12:198–209
3. Bergo MO, Leung GK, Ambroziak P et al (2001) Isoprenylcysteine carboxyl methyltransferase deficiency in mice. *J Biol Chem* 276:5841–5845
4. Tolstykh T, Lee J, Vafai S et al (2000) Carboxyl methylation regulates phosphoprotein phosphatase 2A by controlling the association of regulatory B subunits. *EMBO J* 19:5682–5691
5. Clarke S (2003) Aging as war between chemical and biochemical processes: Protein methylation and the recognition of age-damaged proteins for repair. *Ageing Res Rev* 2:263–285
6. Bedford MT, Frankel A, Yaffe MB et al (2000) Arginine methylation inhibits the binding of proline-rich ligands to Src homology 3, but not WW, domains. *J Biol Chem* 275:16030–16036
7. Friesen WJ, Massenet S, Paushkin S et al (2001) SMN, the product of the spinal muscular atrophy gene, binds preferentially to dimethylarginine-containing protein targets. *Mol Cell* 7:1111–1117

Mf

► [Mobile Fraction](#)

M-FISH

Definition

M-FISH stands for multiplex- or multicoloured-FISH (► [Fluorescent in-situ hybridisation](#) (FISH)). It refers to the use of chromosome-specific probes, each labelled with a distinct combination of fluorochromes, which give each individual chromosome a unique spectral

signature. This allows a precise determination of the karyotype within a single experiment.

► [Nuclear Compartments](#)

MFS

► [Marfan Syndrome](#)

MGMT

► [Alkyltransferases](#)

MHC

► [Major Histocompatibility Complex](#)

MIAME

Definition

MIAME stands for Minimum Information About a Microarray Experiment. It refers to a set of guidelines that have been developed by the Microarray Gene Expression Data Society, and describe the information that needs to be reported about a microarray experiment to enable its unambiguous interpretation and potential reproducibility.

► [Microarray Data Analysis](#)

Microarray Data Analysis

ALVIS BRAZMA

European Bioinformatics Institute, European Molecular Biology Laboratory Outstation-Hinxton, Cambridge, UK
brazma@ebi.ac.uk

Definition

Understanding and utilizing massive amounts of data generated by microarray experiments require the use of a variety of nontrivial data analysis methods. Microarray data can help in understanding gene function and regulation, in finding the differences between normal and diseased states of a living system and in understanding the disease mechanisms. The choice of a particular data analysis method depends on the type of microarray used and on the goals of experiment, such as studies of gene expression, protein-DNA binding locations, genome rearrangements or single nucleotide polymorphisms.

Characteristics

There are two relatively well-defined parts in microarray data analysis. The first part, the raw data analysis, transforms the digital images generated by microarray hybridisation scans into measurements such as gene expression levels. The second, the higher-level analysis, aims at finding new insights into the biological processes studied in the respective experiments by analysing these measurement data. The first part mostly depends on the particular microarray platform used in the experiment, while the second depends on the goal of the experiment. To describe microarray data analysis methods first we need to consider how data is generated.

Data Generation

► **Microarrays** are a common name for a variety of technologies; here we will concentrate exclusively on ► **DNA microarrays**. These are used to measure the absolute or relative abundance of DNA or RNA molecules of different nucleotide sequences in a mixture (sample) of such molecules. There can be tens of thousands of features (spots) on such an array, each containing a large number of identical or related DNA molecules of lengths from about twenty to hundreds of nucleotides. After hybridising the sample to the array, by the rules of preferential binding each molecule in the sample will tend to bind to the spot on the array containing DNA of complementary sequence. If the molecules in the sample are labelled for detection, e.g. by fluorescent dyes, then the abundance of the molecules bound to a particular spot can be assessed by measuring the fluorescence intensity at the particular location on the array. Consequently, if we know the DNA sequences on the array, we can estimate the abundance of different molecules in the sample, which in turn can be used for measuring gene expression levels.

In practice the relationship between the fluorescence intensity, which we can measure directly and the abundance of a given DNA molecule is not straightforward.

Many technology specific features have to be taken into account in data analysis to transform the first value into the second. To improve the measurement accuracy, instead of measuring the molecule abundance directly, microarrays are often used for comparison of two samples, or of a sample to a control (for instance a sample from a diseased tissue can be compared to a normal one). In this case the sample and the control molecules are usually labelled with different fluorescence dyes, for example a red dye (Cy5) for the DNA from the sample and a green dye (Cy3) for the control. By measuring both intensities for each spot and calculating the intensity ratio, we can tell whether a particular molecule is abundant in the sample or in the control. This approach is known as ► **two channel microarrays**.

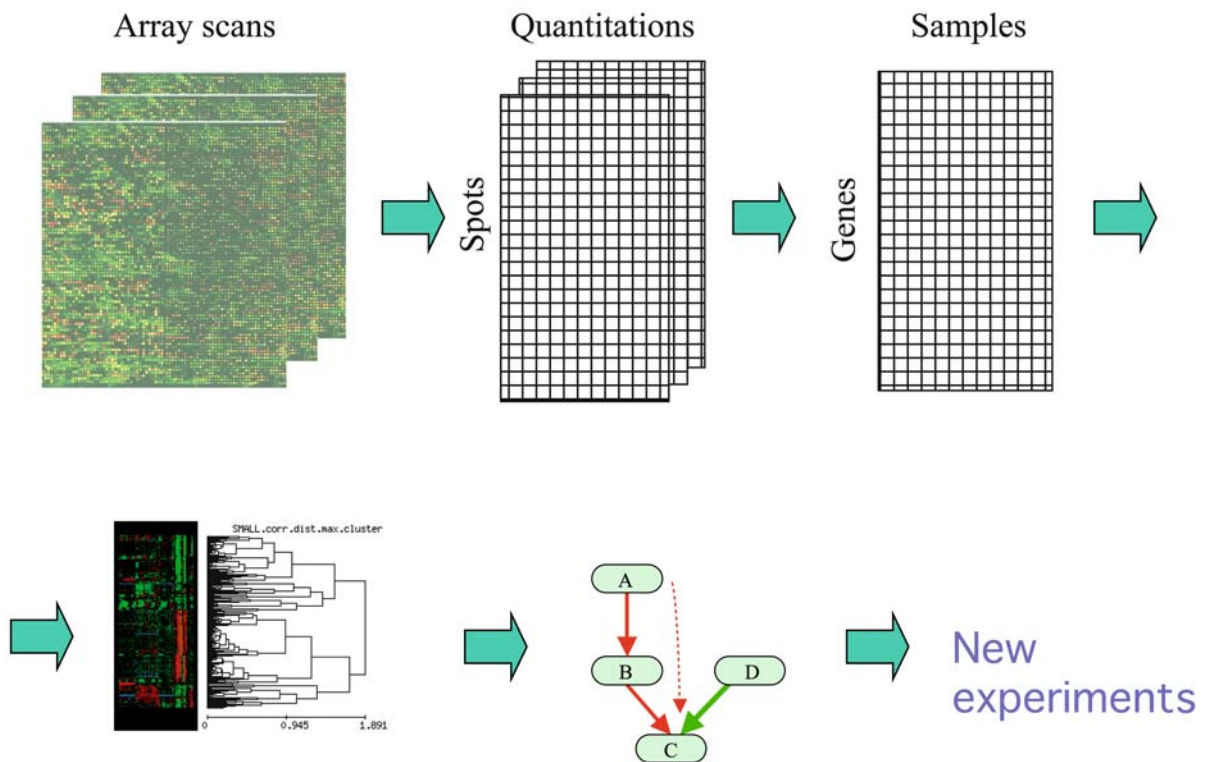
Alternatively each DNA sequence to be measured can be represented on the array by a set of different fragments, some of which may contain ► **mismatch nucleotides**. Comparing the intensities for perfect and mismatch sequence spots, the abundance of particular DNA molecules can be assessed. The popular Affymetrix GeneChip(technology exploits this principle. The exact methods of raw data analysis will obviously depend on the particular technology type.

Microarray data generation results in a digital image obtained by scanning the hybridized arrays. In fact, a microarray experiment will usually consist of a series of hybridizations, producing a series of images. For instance, if the goal of the study is to compare the gene expression in different cell types, the experiment will produce a set of images, one (or several) for each cell type. This set of images will be the starting point for data analysis (Fig. 1).

For simplicity, from now on we will assume that we are dealing with data from two channel microarrays used in gene expression studies, unless stated otherwise.

Raw Data Analysis – from Hybridisation Images to Gene Expression Data Matrix

First we have to transform each image into a numeric representation, which we call the spot quantitation matrix. In this matrix, each row represents a spot (feature) on the array, while each column represents a particular quantitation type, such as the mean of the pixel intensities in a spot, the median, the standard deviation or the mean (median) intensity of the pixels in the local background, as well as various spot quality indicators. This transformation step is performed by ► **image analysis** software, which identifies and delineates each spot on the array and extracts from each spot a set of quantitations (values). Various free and commercial microarray image analysis software packages exist; most of them produce from about 10–20 different values per spot.



Microarray Data Analysis. Figure 1 The starting point for microarray data analysis is a set of images from microarray scans. The first step is to transform each image into a spot quantitation matrix, where each row corresponds to one spot on the array, while each column contains different spot characteristics. The next step is to summarize these matrices in a gene expression matrix, where each row represents a gene, each column represents a particular experimental condition and each position in the matrix characterizes the expression level of a particular gene under particular experimental conditions. This completes raw data analysis. The analysis steps are rather variable, depending on the particular goals of the experiment, but typically they involve the reduction of the dimensionality of the data, e.g., by hierarchical clustering and trying to understand the biological mechanisms underlying the system under study. This may lead to new hypotheses, which can be verified in new experiments.

The spot quantitation matrix is the starting point for the next data transformation step, which has to summarize the set of spot quantitation matrices into one [gene expression data matrix](#). Note that a gene can be represented by several features on the array, containing the same or different sequences, while the same experimental condition can be studied on several arrays. In the gene expression matrix, each row represents a gene (as opposed to a feature on the array), each column represents a particular experimental condition and each position contains one or more quantities characterizing the abundance of the particular gene product. For two channel technologies the ratio or a logarithm of the ratio of such abundances are often used. Ideally we would also like to have one or more additional quantities, characterizing the estimated error level or the reliability of the particular data point. To transform fluorescence intensity measurements into related nucleotide abundances, we have to remove systematic noise originating from a variety of sources.

Different arrays may have different amounts of DNA on different spots; moreover the arrays may have imperfections introducing a noise component dependent on the location on the array. The behavior of different dyes may be different in altering the nucleotide hybridization kinetics, in the response to scanning and in other factors. The transformations that correct for the systematic noise and derive the abundance measurements are known as [data normalisation](#) (1).

The data normalisation methods can be based either on external controls (specific control features spotted on the array in combination with control molecules added to the sample) or on making various assumptions about the measurement data itself. One of the most popular assumptions used is the equality of the total amount of nucleic acids hybridised in all the samples. Such assumptions may or may not be reasonable for the particular experiment and one has to be careful in applying them. Among the most popular normalisation

methods is the locally weighted linear regression Lowess. Variance stabilising normalisation methods are now gaining popularity and have the advantage of being based on a general well-defined model. External control based normalisation is necessary in some cases, particularly if we do not know how the total amount of nucleic acid changes in the sample.

In addition to normalisation, there are different transformations we may wish to apply to the data either before, during or after the normalisation. For instance, we may wish to retain only the ratios that are obtained from spots with a fluorescence signal higher than a given threshold. Such filtering may result in missing data points, which we may need to impute from other related measurements. Sometimes we may want to discretize the measurements from continuous values into discrete ones, such as 'expressed', 'not expressed', 'up-regulated', 'down-regulated' or 'marginal'. Various nontrivial statistical methods have been developed to make such discretization meaningful.

There are a number of free and commercial software tools available for raw data analysis. We mention only a few here, since these tools are rapidly developing and new ones are becoming available. At this time, one of the most popular free image analysis software packages is ScanAlyze from Stanford University and MIDAS from TIGR. Bioconductor is an open source software library, which is popular for data normalization.

Higher Level Analysis – Extracting Biological Meaning from Data

The gene expression matrices contain information about gene expression levels under different conditions. We can use such a matrix to look for groups of genes that have similar expression profiles, i.e. that are expressed under similar experimental conditions in a similar way, and then predict the function of unknown genes in the group from the known ones. If a comparison between diseased and healthy states is the goal of the experiment, we can look for the genes that are differentially expressed in these two conditions, map this information onto the known metabolic or signalling pathways and try to derive new hypotheses about the disease mechanisms.

The size of the gene expression matrices is usually too large for a human researcher to study genes individually (for instance, for a human microarray there can be over 25,000 rows in the matrix – a row per gene). Therefore automated data analysis methods have to be used first. These methods can be roughly classified into two categories – supervised and [unsupervised analysis](#) methods (2).

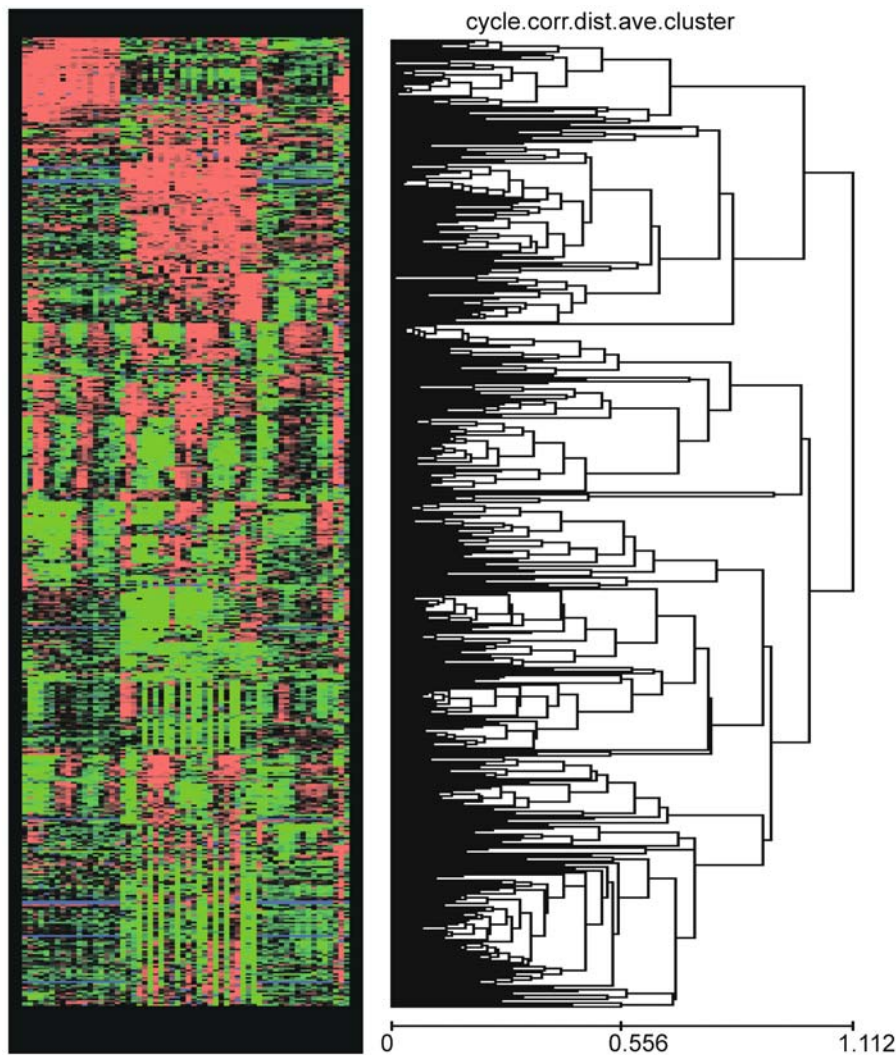
The goal of unsupervised data analysis is to effectively reduce the dimensionality of gene expression data matrix to a manageable size. For instance hierarchical or other [clustering](#) methods are used to group genes

of similar expression profiles. Coupled with visualizing the gene expression matrices in colour coded expression intensity maps (3), these methods provide a powerful tool for reducing thousands of profiles into several clusters characterizing the patterns of behavior (Fig. 2). To apply clustering methods, first we need to define a measure of similarity or distance between expression profiles. Each expression profile can be treated as a vector in a multidimensional space, therefore distance measures developed for vectors, such as Euclidian or correlation distance, can be used. Alternative unsupervised analysis methods are the principal component and correspondence analyses. Specialized methods, such as auto-correlation or Fourier analysis, can be applied to time-series data.

After reducing thousands of profiles into a few clusters of genes exhibiting 'interesting' expression behavior, a human investigator can study these genes in the context of prior knowledge to draw conclusions about the underlying biological processes and propose new, experimentally testable hypotheses. Automated software tools can help in finding clusters of genes sharing the same gene annotation or in analyzing their regulatory regions by looking for shared sequence patterns. Such approaches have been successful, for instance in identifying transcription factor binding sites in yeast.

Supervised data analysis methods exploit our prior knowledge about the genes or experimental conditions right from the beginning. For instance, if we are given a set of gene expression measurements for diseased and healthy samples, we can look for expression values in the matrix that allow the prediction of the 'disease' or 'normal' state of a new sample by its expression profile. This can be used in diagnostics, but moreover, if the predictors can be described by simple rules, these can also help us in understanding the biological mechanisms of the disease. For instance, if we observe that a small subset of genes are highly expressed in the disease state and poorly expressed in the normal state or *vice versa*, these can be hypothesized to be involved in the mechanisms of the particular disease. This approach has been successfully used, for instance, in cancer [classification](#).

There are a number of different techniques used in supervised gene expression data analysis, including linear regression or linear discriminant analysis, nearest neighbourhood analysis, support vector machines, neural networks and decision trees. In principle any [supervised analysis](#) or machine learning techniques can be applied. However it has to be noted, that in gene expression analysis the number of objects that we want to classify (samples) is typically much smaller than the number of parameters that can be used in classification (genes). This situation is driving the development of new data analysis methods.



Microarray Data Analysis. Figure 2 Hierarchical clustering. A gene expression data matrix has been clustered by a hierarchical clustering method, joining genes and groups of genes of similar expression profiles. The 'dendrogram' (hierarchical tree) on the right side of the image shows the order of joining. The left side of the image shows a so-called 'heat map' of the clustered gene expression data matrix, by visualizing the positive values in red and the negative values in green with the brightness representing the intensity of these values.

Automated methods can go beyond simple clustering or classification approaches. For instance the problem of reverse engineering of gene regulation networks has recently become popular. One can build Bayesian networks 'explaining' the gene expression matrices, or networks describing the changes in gene expression measured under different perturbations (such as gene knockouts), or one can use microarrays to study the transcription factor binding locations (microarray based chromatin immunoprecipitation experiments) and build networks of transcription factor interactions (4). Though it is unlikely that reverse engineering of gene regulation networks is possible purely from microarray data, nevertheless these attempts are

providing valuable insights into different properties of gene regulation networks (5).

There exist a large number of commercial and free software tools implementing various analysis methods. Some of the most popular free software tools are Cluster, TreeView and XCluster from Stanford, TIGR MultiExperiment Viewer from TIGR and Expression Profiler from the EBI.

Data Meta-Analysis and Microarray Databases

Microarray experiments are relatively expensive, but on the other hand they generate data that can continue producing new knowledge long after the experiments have been completed. Large, systematic, high quality

datasets are proving to be a valuable resource for the research community, similarly to genome sequence data. ► **Meta-analysis** of such datasets combining them with other relevant datasets and using new analysis methods may bring new insights, which go beyond the scope of the original studies. For instance, the comparison of the expression profiles of homologous genes across a range of organisms can help in predicting the orthologous genes, or comparing similar processes, such as the cell-cycle in different organisms. Gene expression data analysis, supervised or unsupervised, can help a biologist only if in addition to the gene expression data, information about the genes and experimental conditions reflected in the matrix are also provided. For instance, if the goal of the experiment is to study the cell-cycle, we will need to know at which time-points the measurements have been taken and how the cells have been synchronized and grown. We will also need to use the prior knowledge about the genes, such as what are the functions of the known genes, and what are the orthologs of these genes in other organisms. This means that the gene expression data matrix has to be well annotated if it is to be of use to the researchers.

The Microarray Gene Expression Data (MGED) Society has developed a standard for annotating microarray experiments known as the Minimum Information About a Microarray Experiment (► **MIAME**) (6). This has been adopted by most of the major scientific journals as a requirement for publication, and public databases have been recently established, accepting data annotated according to the MIAME standard. ArrayExpress (► <http://www.ebi.ac.uk/arrayexpress>) is such a public repository at the European Bioinformatics Institute. A researcher can ask the database for all the experiments of a given type for a given organism and retrieve the relevant data, which then can be combined with the researcher's own data, or used for designing new experiments. With amounts of microarray data growing, quality improving and new analysis methods being developed, such databases are becoming an essential resource for functional genomics research.

References

1. Quackenbush J (2002) Microarray data normalisation and transformation. *Nat Genet Suppl* 32:496–501
2. Brazma A, Vilo J (2000) Gene Expression Data Analysis. *FEBS Letters* 480:17–24
3. Eisen MB, Spellman PT, Brown PO et al (1998) Cluster analysis and display of genome-wide expression patterns. *PNAS* 95:14863–14868
4. Lee T et al (2002) Transcriptional Regulatory Networks in *Saccharomyces cerevisiae*. *Science* 25:799–804
5. Schlitt T, Brazma A (2002) Learning about gene regulatory networks from gene deletion experiments. *Comparative and Functional Genomics* 3:499–503

6. Brazma A et al (2001) Minimum information about a microarray experiment (MIAME)—toward standards for microarray data. *Nat Genet* 29:365–371
7. Causton H, Quackenbush J, Brazma A (2003) *Microarray Gene Expression Data Analysis – a beginners guide*. Blackwell Publishing, p 192

Microarray Technology

Synonyms

Biochip Technology; Gene Chip Technology; DNA Chip-Technology

Definition

Microarray technology describes an analytical procedure that is based on the orderly arrangement of multiple, micron-sized spots of biomolecule probes (i.e. nucleic acids with known sequences such as oligonucleotides, cDNA or genomic DNA) on a carrier (e.g. a glass slide), hybridized with a complex, labelled nucleic acid target in order to identify gene mutations or the expression level (abundance) of genes (RNA). The underlying principle of the procedure is nucleobase-pairing. A single array may represent parts of a genome or the entire set of genes of a given organism/cell. The most common gene expression microarrays fall into two categories: Affymetrix Genechips (a proprietary commercial array using 25 base oligonucleotides probe sets), and spotted DNA microarrays in which long cDNA probes or oligos are deposited and bound to a substrate, usually a coated glass microscope slide.

► **C. Elegans as a Model Organism for Functional Genomics**

► **Multifactorial or Common Diseases**

► **Rheumatism Related Genes, Identification**

► **RNA Interference in Mammalian Cells**

► **Xenopus as a Model Organism for Functional Genomics**

Microarrays

Definition

Microarrays contain a multitude of orderly arranged, micron-sized samples bound on a carrier material (usually a glass substrate, sometimes nylon). Depending on the type, the samples can contain DNA sequences (as in gene or ► **DNA microarrays**), tissue fragments (tissue microarrays), protein (protein microarrays), or others.

They provide a medium for matching known and unknown samples, which can be measured as a fluorescent, luminescent or radioactive signal. Microarrays are often also referred to as “chips” (DNA chips, biochips, etc.), because their fabrication resembles the production of microchips.

► *C. Elegans* as a Model Organism for Functional Genomics

► DNA Chips

► Mendelian Forms of Human Hypertension and Mechanisms of Disease

► Microarray Technology

► Microarrays in Colorectal Cancer

► Monoclonal Antibodies

► Multifactorial or Common Diseases

► Protein Chips

► Protein Microarrays as Tool for Protein Profiling and Identification of Novel Protein-Protein Interactions

► Proteomics in Microfluidic Systems

► Rheumatism Related Genes, Identification

► RNA Interference in Mammalian Cells

► *Xenopus* as a Model Organism for Functional Genomics

Microarrays in Colorectal Cancer

JUDITH M. BOER

Center for Human and Clinical Genetics, Leiden
University Medical Center, Leiden, The Netherlands
j.m.boer@lumc.nl

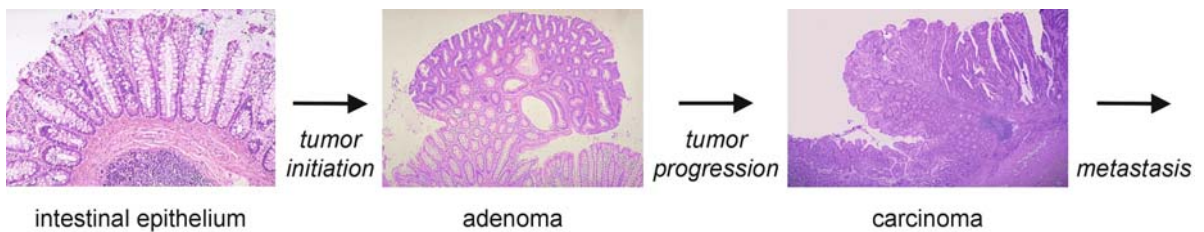
Synonyms

Colorectal cancer (CRC) is one of the most prevalent cancers in Western Europe and North America. At least 50% of the Western population develops a colorectal adenoma by the age of 70 and in about 1 in 10 of these individuals progression to malignancy ensues (Fig. 1) (1). The prognosis in advanced cases is poor, as more than one-third of the patients will die from progressive disease within five years. Currently, only a small fraction of the patients respond to treatment and its effectiveness is best in early stages of the disease.

The left panel shows normal colonic crypts in transverse section. Epithelial cells are aligned along the basement membrane. The center panel shows the dysplastic morphology typical of an adenoma, a benign lesion. Note the increased nuclear/cytoplasmic ratio, lack of uniform architecture, and many nuclei that are no longer lined up along the basement membrane. The

right panel shows a carcinoma, a malignant lesion that invades the basement membrane and infiltrates the subserosal layer. Cancer cells that migrate to distant locations, in many cases the liver, give rise to metastases. CRC was one of the first major epithelial cancers in which genetic alterations were described to occur in a systematic fashion during disease progression (2). The molecular mechanisms have been under intensive study for the last two decades. Normal cell growth depends on a balanced expression of growth-promoting and growth-suppressing genes. If growth-promoting genes are activated to a state of hyperfunction, either by mutation or quantity, they are termed “oncogenes”, which exert a positive effect on cell growth. One common example is *K-ras*, which is activated in CRC and many other tumors. Growth-suppressing genes are defined as “tumor suppressors” and are commonly lost following Knudson’s “two-hit” hypothesis (loss of one allele and inactivation of the other allele by mutation or promoter methylation). The best-characterized tumor suppressors inactivated in CRC are ► *APC* (*adenomatous polyposis coli*), which is involved in tumor initiation, and *Smad4/DPC4* (*deleted in pancreatic cancer locus 4*) and *p53*, which are involved in tumor progression to the late adenoma and carcinoma stages, respectively. Additionally, abrogation of mismatch repair systems contributes to some colorectal cancers (1). Nevertheless, exactly how those genetic alterations bring about the development and progression of colorectal adenomas to carcinomas remains to be resolved. To complicate the picture, accumulation of other genetic and epigenetic changes can occur, including loss of heterozygosity, gene inactivation, methylation, loss of imprinting and/or gene amplifications, all of which may alter gene expression. Therefore, genome-wide monitoring of gene expression is of great importance to disclose the numerous and diverse events associated with colorectal carcinogenesis.

This is where ► **gene expression analysis using DNA microarrays** comes into play, by making it possible to quantify the expression levels of thousands of genes in a tissue sample in one experiment. Messenger RNAs (mRNAs) are extracted from a biological sample, converted into DNA, labeled with fluorescent dyes, and then washed over a glass slide bearing a grid spotted with DNA sequences from tens of thousands of genes. The labeled sequences bind to spots representing the genes from which the mRNAs were transcribed. By analyzing the location and intensity of the fluorescent signals, one can determine the level of activity for each gene. The expression levels of all genes tested within a sample constitute its ► **gene expression profile**, which is then compared with other tumor or normal tissue expression profiles. In some cases, it is possible to analyze two samples on the same chip by using different colored dyes (3).



Microarrays in Colorectal Cancer. Figure 1 Adenoma-carcinoma sequence in colorectal tumorigenesis.

Definition

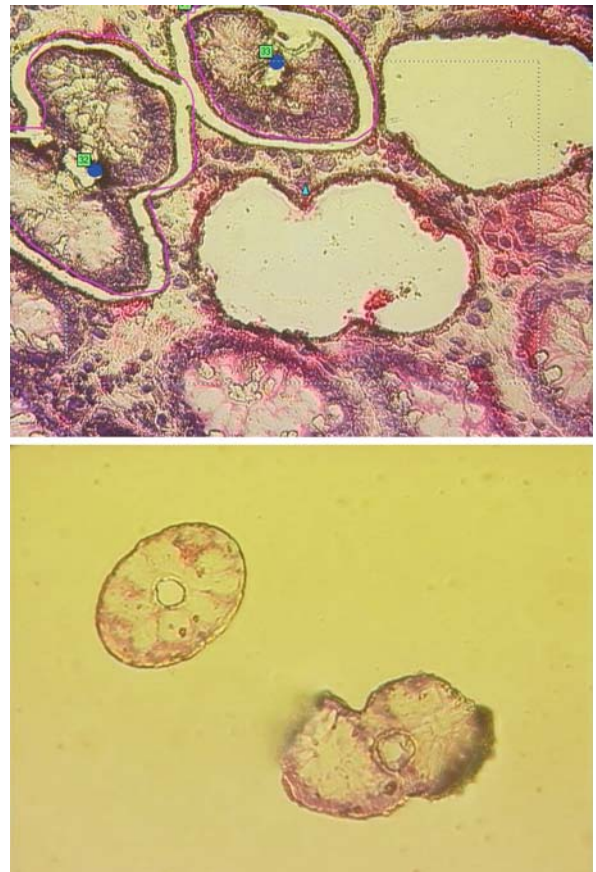
An important goal of expression profiling is the molecular [classification](#) of tumors into homogeneous groups that can predict their clinical behavior. The identification of genes whose expression is altered in colorectal cancer compared to normal colon epithelium is of tremendous importance because it may allow a better understanding of pathways that regulate colorectal carcinogenesis and yield new diagnostic biomarkers or targets for therapy.

Characteristics

Gene Expression Analysis in Colon Cancer

Several DNA microarray gene expression studies were undertaken with the aims of differentiating benign from malignant colorectal tumors and learning how the progression from normal tissue through adenoma and adenocarcinoma is mirrored in changes in gene expression (4, 5, 6). In these studies, typically around 20 tumors with paired normal colon were subjected to microarray analysis using high-density oligonucleotide microarrays (4, 5, 6) or complementary DNA microarrays (5). Generally, around 2% of the genes detected on the microarrays differ between tumor and normal colorectal tissue. Many of these changed genes were already known to be abnormally expressed in neoplastic tissue in general, or colon cancer in particular, but a large number had no previously characterized function. The comparison of gene expression data across tumor and normal samples from different studies revealed a large variability, probably representing a combination of assay heterogeneity, true biological difference and different percentages of tumor cells in microdissected samples (Fig. 2) (5) versus bulk tissue (4, 5, 6). The genes presented here showed tumor-specific expression changes in at least two microarray gene expression studies or were confirmed by independent methods (Table 1).

Unsupervised two-way hierarchical clustering performed by Levine et al. successfully distinguished adenocarcinoma from adenoma and normal tissue, generating a tree that appropriately represented the clinical relationship between the three tissue types (Fig. 3) (6). This supports the concept that genome-wide



Microarrays in Colorectal Cancer. Figure 2 Laser-capture microdissection of colon tumors. Tumors are complex tissues containing other cell types such as stroma cells and infiltrating immune cells (top panel). To enrich for tumor cells, they may be microdissected from briefly fixed and stained frozen sections using laser-capture technology. Usually, 5-10,000 cells are collected for RNA isolation (bottom panel). The RNA is subjected to two rounds of linear amplification to yield enough material for microarray hybridization.

expression profiling may permit a molecular classification of adenomas and carcinomas. Indeed, Nakamura and coworkers established a scoring system to separate

Microarrays in Colorectal Cancer. Table 1 Selection of genes identified in microarray studies as differentially expressed in colorectal cancer compared to normal colon

Accession Number	Gene Name	Gene Symbol	Change	Biological Process	Ref	Validation
Metabolism						
NM_002046	glyceraldehyde-3-phosphate dehydrogenase	GAPD	up	energy: glycolysis	(7)	RT-PCR
NM_005566	lactate dehydrogenase A	LDHA	up	energy: glycolysis	(5)	
NM_002300	lactate dehydrogenase B	LDHB	up	energy: glycolysis	(5, 7)	
NM_001428	enolase 1, (alpha)	ENO1	up	energy: glycolysis	(7)	
NM_000687	S-adenosyl homocysteine hydrolase	AHCY	up	energy	(4-6)	
NM_006907	pyrroline-5-carboxylate reductase 1	PYCR1	up	energy	(5, 6)	
NM_002591	phosphoenolpyruvate carboxykinase 1 (soluble)	PCK1	down	energy: gluconeogenesis	(4, 5)	
NM_001823	creatine kinase, brain	CKB	down	energy	(4, 5)	SAGE
NM_004046	ATP synthase, H+ transporting, mitochondrial F1 complex, alpha subunit, isoform 1, cardiac muscle	ATP5A1	down	energy	(4, 5)	
NM_000669	alcohol dehydrogenase 1C (class I), gamma polypeptide	ADH1C	down	energy	(4-6)	
NM_003011	SET translocation (myeloid leukemia-associated)	SET	up	DNA replication	(6, 7)	
NM_002128	high-mobility group box 1	HMGB1	up	transcription	(5, 6)	
NM_014390	EBNA-2 co-activator (100kD)	p100	up	transcription	(6)	RT-PCR
NM_002097	general transcription factor IIIA	GTF3A	up	transcription	(4-6)	
NM_003472	DEK oncogene (DNA binding)	DEK	up	RNA processing	(5, 7)	
NM_002157	heat shock 10kD protein 1 (chaperonin 10)	HSPE1	up	protein folding	(5, 6)	
NM_001738	carbonic anhydrase I	CA1	down	detoxification	(5, 6)	SAGE
NM_000067	carbonic anhydrase II	CA2	down	detoxification	(5, 6)	
NM_000717	carbonic anhydrase IV	CA4	down	detoxification	(4, 6)	
Cell growth						
NM_001827	CDC28 protein kinase regulatory subunit 2	CKS2	up	cell cycle checkpoint	(5, 6)	
NM_021874	cell division cycle 25B	CDC25B	up	cell cycle checkpoint	(4-6)	
NM_005239	v-ets erythroblastosis virus E26 oncogene homolog 2 (avian)	ETS2	up	cell growth	(6, 7)	
NM_001511	chemokine (C-X-C motif) ligand 1 (melanoma growth stimulating activity, alpha)	CXCL1	up	growth factor	(4, 6)	RT-PCR

Microarrays in Colorectal Cancer. Table 1 Selection of genes identified in microarray studies as differentially expressed in colorectal cancer compared to normal colon (Continued)

Accession Number	Gene Name	Gene Symbol	Change	Biological Process	Ref	Validation
NM_001316	CSE1 chromosome segregation 1-like (yeast)	CSE1L	up	proliferation/apoptosis	(4-6)	RT-PCR
NM_004417	dual specificity phosphatase 1	DUSP1	down	proliferation	(5)	SAGE
NM_001275	chromogranin A (parathyroid secretory protein 1)	CHGA	down	proliferation/differentiation	(4, 6)	RT-PCR
Tumor invasion						
NM_000358	transforming growth factor, beta-induced, 68kDa	TGFB1	up	cell-matrix adhesion	(4-6)	RT-PCR; SAGE
NM_001712	carcinoembryonic antigen-related cell adhesion molecule 1 (biliary glycoprotein)	CEACAM1	down	cell adhesion	(4-6)	RT-PCR; SAGE
NM_006890	carcinoembryonic antigen-related cell adhesion molecule 7	CEACAM7	down	cell adhesion	(4, 5)	
NM_003278	tetranectin (plasminogen binding protein)	TNA	down	cell adhesion	(4, 6)	RT-PCR
NM_006149	lectin, galactoside-binding, soluble, 4 (galectin 4)	LGALS4	down	cell-cell/cell-matrix adhesion	(5)	SAGE
NM_006082	tubulin, alpha, ubiquitous	K-ALPHA-1	up	cytoskeleton	(5)	
NM_001069	tubulin, beta polypeptide	TUBB	up	cytoskeleton	(5, 7)	
NM_001101	actin, beta	ACTB	up	cytoskeleton	(7)	
NM_002276	keratin 19	KRT19	down	cytoskeleton	(4, 5)	SAGE
NM_015515	keratin 23 (histone deacetylase inducible)	KRT23	down	cytoskeleton	(4)	RT-PCR
NM_000177	gelsolin (amyloidosis, Finnish type)	GSN	down	cytoskeleton	(5, 6)	SAGE
NM_005940	matrix metalloproteinase 11 (stromelysin 3)	MMP11	up	extracellular matrix breakdown	(4)	
NM_002421	matrix metalloproteinase 1 (interstitial collagenase)	MMP1	up	extracellular matrix breakdown	(5, 6)	
NM_002423	matrix metalloproteinase 7 (matrilysin, uterine)	MMP7	up	extracellular matrix breakdown	(6)	
NM_000584	interleukin 8	IL8	up	angiogenesis/cell adhesion	(4)	RT-PCR
Other and unknown						
NM_023009	MARCKS-like protein	MLP	up	signal transduction	(5, 6)	
BM988196	guanylate cyclase activator 2A (guanylin)	GUCA2A	down	signal transduction	(4, 6)	RT-PCR
NM_005080	X-box binding protein 1	XBP1	up	immune response	(5, 6)	
NM_002090	chemokine (C-X-C motif) ligand 3	CXCL3	up	immune response	(6)	

Microarrays in Colorectal Cancer. Table 1 Selection of genes identified in microarray studies as differentially expressed in colorectal cancer compared to normal colon (Continued)

Accession Number	Gene Name	Gene Symbol	Change	Biological Process	Ref	Validation
NM_001928	D component of complement (adipsin)	DF	down	immune response	(4, 6)	
NM_003890	Fc fragment of IgG binding protein	FCGBP	down	immune response	(4, 5)	
NM_001293	chloride channel, nucleotide-sensitive, 1A	CLNS1A	up	regulation of cell volume	(5, 6)	
NM_007102	guanylate cyclase activator 2B (uroguanylin)	GUCA2B	down	excretion	(4, 6)	RT-PCR; SAGE
NM_000111	solute carrier family 26, member 3	SLC26A3	down	excretion; colon mucosa	(4, 6)	
XM_045277	KIAA0746 protein	KIAA0746	up	unknown	(7)	Northern
AB033025	KIAA1199 protein	KIAA1199	up	unknown	(4)	RT-PCR

Summary of confirmed changes in gene expression identified in three microarray studies analyzing primary tumors (4-6) and one studying metastatic cell lines (7). Genes included in the list were found in at least two studies, were validated by an independent analysis method, or in the literature. Functional annotation for biological process was performed using the SOURCE web search (9).

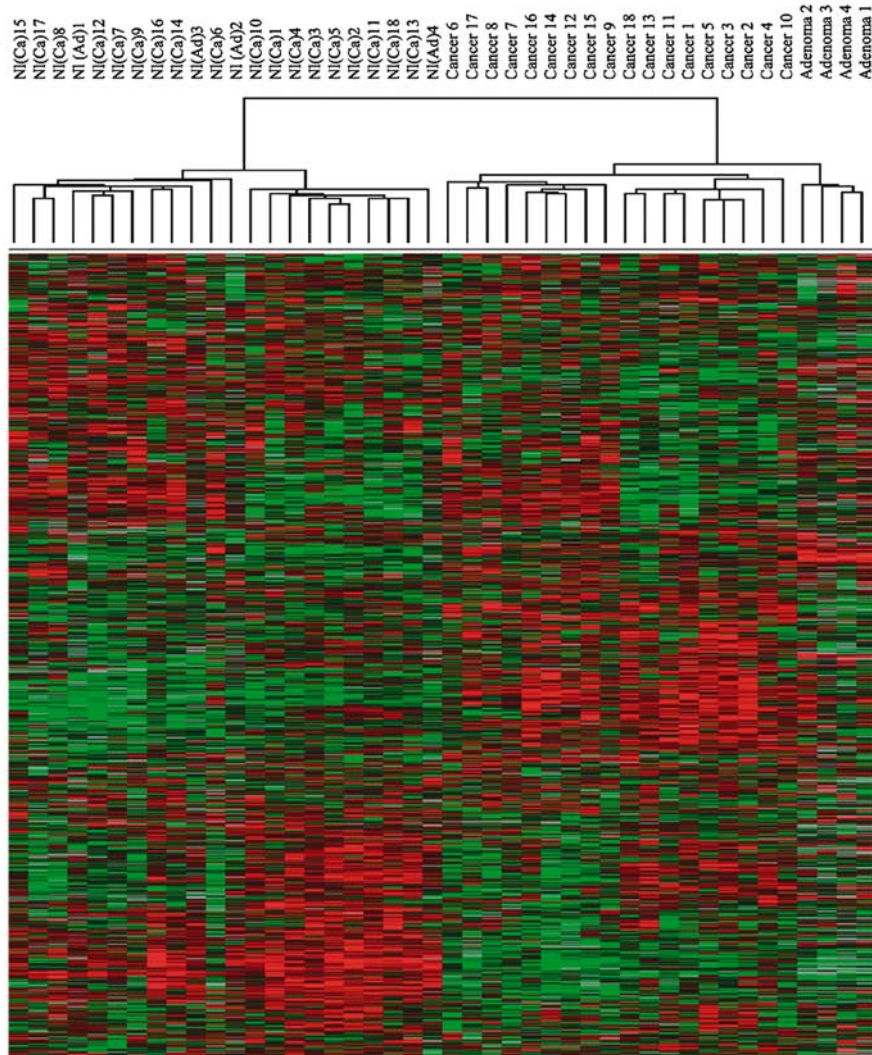
adenomas from carcinomas based on the expression levels of 50 discriminating genes. Application of this scoring system for evaluating five additional colorectal tumors correctly predicted their histological features (5).

However, many of the expression changes noted in the cancers were presaged by those in premalignant adenomas (4, 5, 6), which is not surprising since adenomas are considered the pathological and genetic precursors to adenocarcinoma of the colon (1). Examples of genes that are upregulated in both adenoma and adenocarcinoma are *EBNA-2 co-activator (100kD)*, *transforming growth factor beta-induced gene product (TGFB1)*, *CDC28 protein kinase regulatory subunit 2 (CKS2)*, *chemokine (C-X-C motif) ligand 1 (CXCL1/melanoma growth stimulating activity, alpha)*, *transcription factor IIIA (GTF3A)* and *matrix metalloproteinase 7 (MMP7/matrilysin)*. These genes may play a role at a relatively early stage of carcinogenesis. *MMP7* is detected in over 80% of adenomas and is not expressed in most normal epithelial cells. This metalloproteinase mRNA may be regulated by beta-catenin (also over-expressed in adenoma and CRC) and is an emerging target for cancer therapy (6). Adenomas also displayed attenuated expression of several genes identified as down-regulated in the carcinomas, including the colonic epithelial cell product *guanylin (GUCA2A)*, colon mucosa-associated *solute carrier family 26, member 3*

(*SLC26A3*) and *tetranectin (TNA)*. Several genes showed intermediate expression between adenoma and carcinoma, including the down-regulated genes *ATP synthase 5A1 (ATP5A1)*, *alcohol dehydrogenase 1 C (ADH1C)* and *phosphoenolpyruvate carboxykinase 1 (PCK1)* and the up-regulated genes *MMP7* and *CXCL1*.

Orntoft and coworkers studied different clinical stages and found that 70% of the alterations in gene expression occurred during the transition from normal to early-stage Dukes' A tumors (4). Far fewer genes are changing their level of expression during the progression through the different Dukes' stages. This indicates that the basic properties of tumor cells are acquired in the early tumor stages and that only minor changes, perhaps those involving the stromal components are needed later on. From a therapeutic point of view, this is important because the same targets seem to be present at most Dukes' stages (4). Some of the selected candidate genes can be regarded as progression markers that change in at least two consecutive stages, or Dukes' classifiers that show major change in one stage only (4). Examples of such progression markers are *phosphoenolpyruvate carboxykinase (PCK1)* and *interleukin 8 (IL8)*.

The majority of CRC related deaths are secondary to liver ►metastasis. One of the challenges to effective treatment lies in understanding the complex biology of metastasis in CRC, thus allowing for new therapeutic and genetic intervention. Quackenbush and coworkers



Microarrays in Colorectal Cancer. Figure 3 Hierarchical clustering differentiates colorectal adenomas, adenocarcinomas and normal samples. Expression levels of approximately 1800 genes in 18 carcinomas, 4 adenomas and 22 matched normal samples were combined into a single matrix. Both genes and samples were analyzed for similarity using two-way hierarchical clustering. Each color in the resulting visual map represents the expression level of the associated gene in that tissue sample, with a continuum of expression levels from dark green (lowest) to bright red (highest). Missing values are coded as silver. The carcinomas and their benign precursors, the adenomas, are placed on an entirely different trunk of the tree from the paired normal samples. Strikingly, the adenomas and carcinomas are also separated from each other, occupying different branches of the same trunk. Adapted from: Notterman *et al.*, 2001, *Cancer Research* 61, 3124–3130.

(7) compared gene expression profiles in two cell line models for human CRC metastasis. They identified genes that appeared to be differentially expressed in all highly metastatic cell lines relative to their reference. Included in the set are those previously known to possess altered patterns of expression in (colon) cancer as well as a large number with previously uncharacterized functions. Two

newly identified transcripts were shown to be differentially expressed in paired tumor and normal tissue samples, demonstrating the utility of cell line analysis as a model for understanding metastasis. The transcript *KIAA0746* was identified as a potential metastasis marker that showed increased expression correlated to progressing tumor stage.

Biological Processes Involved in Colorectal Carcinogenesis

Tumor Metabolism

Numerically, the most prominent group of genes that change expression during colon cancer progression encode proteins related to cell metabolism, in particular energy metabolism, nucleotide and nucleic acid metabolism (including gene products involved in mRNA transcription and processing) and protein synthesis and modification (4, 5, 6, 7). Examples of up-regulated genes involved in RNA/protein processing are ribosomes, general transcription factors, translation elongation/initiation factors, and chaperones. Several mRNAs encoding glycolytic enzymes are increased. Nuclear-encoded mitochondrial proteins showed a distinct behavior because the genes encoding these were all down-regulated in at least one of the tumor stages (4). The above data shows that an accelerated metabolism is a common feature of adenomas and carcinomas. Rather than a primary transforming event, the changes in metabolism may be a secondary effect of proliferation rate. Even so, they represent a neoplastic signature that can be used for classification.

Cell Growth and Tumorigenesis

The epithelium of the colon is renewed continuously by proliferating stem cells that are located at the bases of the colonic crypts. Their progeny migrate up the crypt, stop dividing and differentiate into mature cells. When the differentiating cells reach the top of the crypt, they undergo apoptosis and are shed into the lumen. Dysregulation of genes involved in these processes can tip the balance to hyperplasia and cancer. Cell-cycle progression is driven by positive and negative regulators that ultimately direct the fate of a cell either to form two daughter cells or to enter into the resting state. Deregulated cellular proliferation, arising from abnormal expression of genes that control cell cycle checkpoints such as *cell division cycle 25B* (*CDC25B*), and *CDC28 protein kinase regulatory subunit 2* (*CKS2*), plays a crucial role in tumorigenesis. Cell-cycle pathways are critical for the initial steps of tumorigenesis, and were found to be continuously involved in late stages of cancer progression and metastasis (4, 5, 6, 7). Among the deregulated genes are several associated with cell death, which may imply that broad repression of programmed cell death pathways is a crucial step in colorectal tumorigenesis (5).

Tumor Invasion

For a CRC tumor to become invasive, it must pass through the muscularis mucosa and infiltrate into the subserosal layer in which terminal lymphatics reside. Subsequently, genes that are involved in cell motility, breaking the barriers of cellular adhesion and remodeling

of the extracellular matrix play important roles in tumor invasiveness (4, 5, 6, 7). The cytoskeleton is the basic machinery that makes cells motile, a characteristic property of invasive cells. The structural cytoskeleton genes *actin beta* (*ACTB*), *tubulin alpha* (*K-ALPHA-1*), and *tubulin beta* (*TUBB*), and *gelsolin* (*GSN*), which is involved in actin filament polymerization, are up-regulated in CRC. Among the down-regulated genes encoding structural components were several keratins, which are intermediate filament proteins responsible for the structural integrity of epithelial cells. Cell adhesion molecules up-regulated in adenomas and carcinomas include *transforming growth factor beta-induced* (*TGFB1*), which is a negative regulator of cell-matrix adhesion, and *cadherin 3* (*CDH3*). Among the down-regulated genes involved in cell adhesion are the carcinoembryonic antigen-related cell adhesion molecules *CEACAM1* and *CEACAM7*, *galectin 4* (*LGALS4*) and *TNA*. TNA, a stromal component of tumors and a participant in proteolytic processes through its binding to plasminogen, is regarded as a tumor suppressor; a low plasma TNA level is related to a shortened survival (4). The microarray studies also indicate the involvement of several matrix metalloproteinases (MMP) in CRC; *Stromelysin 3* (*MMP11*), *interstitial collagenase* (*MMP1*) and *MMP7* were significantly more highly expressed in colonic neoplasia than in normal tissue. Proteins of the MMP family are involved in the breakdown of extracellular matrix in normal physiological processes, such as embryonic development and tissue remodeling, as well as in disease processes, such as arthritis and metastasis.

Hypoxia-induced Gene Expression

Neovascularization and increased glycolysis, two universal characteristics of solid tumors, represent adaptations to a hypoxic microenvironment that are correlated with tumor invasion, metastasis and lethality. In CRC expression studies, up-regulation of glycolytic enzymes, including *glyceraldehyde-3-phosphate dehydrogenase* (*GAPD*), *enolase 1* (*ENO1*) and *lactate dehydrogenases A* (*LDHA*) and *B* (*LDHB*), was found. Conversely, gene products involved in oxygen-dependent mitochondrial energy metabolism were down-regulated. A number of genes discriminating carcinoma from adenoma were relevant to hypoxia. Conceivably, carcinoma cells may be more exposed to starved and hypoxic conditions, where carbohydrate/oxygen homeostasis is impaired, than are adenoma cells (5). Although these metabolic changes are not the fundamental defects that cause cancer, they might confer a common advantage on CRC and many different types of cancers, which allows the cells to survive and invade. The only angiogenesis-inducing factor found up-regulated in CRC microarray studies

is *IL8*. This angiogenic cytokine was also shown to be produced by tumor cells in a CRC cell line-based study (6).

Validation of Microarray Expression Data

Expression changes found in microarray studies are usually verified by comparison with other gene expression studies using similar samples. The possibilities of doing this *in silico* are still expanding, with gene expression information available in several web-based gene annotation tools (8, 9). Also, microarray results can be compared to databases (10) containing several hundreds of tissues that have been analyzed by cDNA sequencing or ►serial analysis of gene expression (SAGE), a tag-sequencing method for measuring absolute gene expression levels. Experimental approaches to validate microarray results at the single gene level in additional tumor samples include quantitative ►RT-PCR and ►Northern blot analysis. Immunohistochemistry using antibodies directed against tumor-specific gene products gives additional information about protein expression in individual cell types.

Although the precise magnitude of expression change is not always recapitulated, the direction and order of magnitude of change predicted by the array is generally confirmed (4, 5, 6, 7). The good correspondence between the different microarray studies and traditional mRNA and protein expression data in colorectal cancer indicates the validity of the large number of new alterations that were found in the large-scale studies.

Clinical Relevance

As described above, DNA microarray studies have made significant contributions to the identification of genes that are differentially expressed during the development and progression of colorectal cancer. The characterization of these genes may give fundamental insights in the molecular processes involved in tumorigenesis in the colon. Although it is tempting to interpret changes in tumor-related pathways as important for tumorigenesis itself, the direct involvement of these changes in the initiation and progression of colon cancer has to be demonstrated. Correlation with disease mutations or interference in biological model systems may provide supporting evidence. However, both causative and secondary molecular alterations in CRC may serve as diagnostic biomarkers and new targets for therapy.

Because the clinical behavior of tumors cannot be accounted for completely by morphology, it is the hope of medicine that a molecular taxonomy based on specific expression profiles will provide a more accurate prognosis and prediction of response to therapy. Studies on other cancer types have shown

that the expression of a group of 20–70 selected genes defines a pattern with prognostic importance (reviewed in 3). As a first step to ensure that this is also possible in CRC, the global gene expression studies summarized here recapitulate the present classifications based on histology (4, 5, 6, 7). Future studies including tumors with clinical follow-up information should identify expression signatures that discriminate patients in various risk groups. This will allow a better-tailored treatment that may involve an aggressive therapy for the high-risk group and a limited operative procedure avoiding morbidity and disability in an early, low-risk subtype of a previously uniform tumor entity. Very promising in this respect are two studies in breast cancer and medulloblastomas that showed that gene expression features present in primary tumors at the time of presentation predict the course of the disease. This implies that properties such as the tendency to metastasize or treatment response might be diagnosed from analysis of the primary tumor and used as the basis of clinical decision-making (3).

Microarrays will continue to be used in research to find molecular portraits of clinical significance, but may not be practical for routine clinical diagnostics. The limited number of genes that define a signature facilitates the use of other techniques such as real-time quantitative PCR, which conserves samples and is faster, easier and less expensive than DNA microarrays.

References

1. Kinzler KW, Vogelstein B (1996) Lessons from hereditary colorectal cancer. *Cell* 87:159–170
2. Fearon ER, Vogelstein B (1990) A genetic model for colorectal tumorigenesis. *Cell* 61:759–767
3. Chung CH, Bernard PS, Perou CM (2002) Molecular portraits and the family tree of cancer. *Nat Genet* 32:533–540
4. Birkenkamp-Demtroder K, Christensen LL, Olesen SH et al (2002) Gene expression in colorectal cancer. *Cancer Res* 62:4352–4363
5. Lin YM, Furukawa Y, Tsunoda T et al (2002) Molecular diagnosis of colorectal tumors by expression profiles of 50 genes expressed differentially in adenomas and carcinomas. *Oncogene* 21:4120–4128
6. Notterman DA, Alon U, Sierk AJ et al (2001) Transcriptional gene expression profiles of colorectal adenoma, adenocarcinoma, and normal tissue examined by oligonucleotide arrays. *Cancer Res* 61:3124–3130
7. Hegde P, Qi R, Gaspard R et al (2001) Identification of tumor markers in models of human colorectal cancer using a 19,200-element complementary DNA microarray. *Cancer Res* 61:7792–7797
8. GeneCards [►<http://www.genecards.org>]
9. SOURCE [►<http://genome-www5.stanford.edu/cgi-bin/sourceSearch>]
10. SAGEmap [►<http://cgap.nci.nih.gov/Tissues>]

Microarrays in Pancreatic Cancer

MALTE BUCHHOLZ, THOMAS M. GRESS

Department of Internal Medicine I, University of Ulm,
Ulm, Germany

malte.buchholz@medizin.uni-ulm.de

thomas.gress@medizin.uni-ulm.de

Definition

Pancreatic cancer is the fifth leading cause of cancer related deaths in industrialized countries. With a 5-year survival rate of less than 5% and a median survival of less than 6 months, pancreatic cancer carries the most dismal prognosis of all solid tumors. Due to the lack of early specific symptoms, the disease is usually only detected at very advanced stages. In addition, pancreatic tumors tend to be aggressively invasive and are largely insensitive to chemotherapy and/or radiotherapy, leaving surgical resection as the only curative treatment option. Although considerable progress has been made in recent years in elucidating the molecular basis of ►**tumorigenesis** in the pancreas and a number of disease-associated genes have been identified, the biology of pancreatic tumors and especially the chain of early molecular events leading up to tumor formation remain poorly understood. The use of microarray technology, including comparative genomic analyses (►**Matrix-CGH**) for the identification of chromosomal imbalances, single nucleotide polymorphism (►**SNP**) analyses for the detection of mutations and genetic polymorphisms and gene expression profiling for the simultaneous assessment of ►**transcript** levels of up to several tens of thousands of genes in a single experiment, is rapidly adding to our understanding of pancreatic cancer genetics, laying the basis for the development of new diagnostic tests as well as new curative treatment modalities.

Characteristics

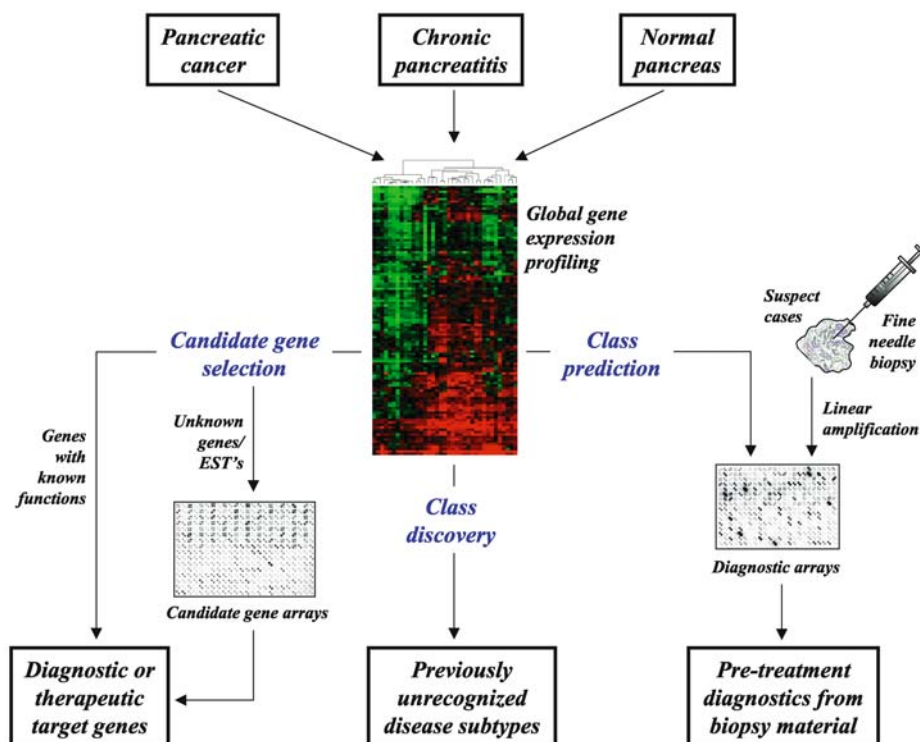
More than 90% of pancreatic tumors can be histologically defined as ductal adenocarcinomas, while all other types of pancreatic neoplasms (acinar-cell carcinomas, endocrine tumors, serous cystadenomas, carcinomas of the Papilla Vateri) are very rare. The term “pancreatic cancer” therefore exclusively refers to ductal adenocarcinomas throughout this article.

Pancreatic cancer is particularly difficult to diagnose accurately, both because of the lack of early symptoms and the fact that malignant tumors in the pancreas are often hard to differentiate from benign tumors arising in the context of a ►**chronic pancreatitis**. Inflammatory as well as malignant tumors are typically associated with

the production of massive amounts of ►**extracellular matrix** (ECM) components within the tumor, seriously compromising cytological analyses of biopsy samples. Up to 95% of the volume of a malignant pancreatic tumor can be made up of ECM deposits, a situation commonly referred to as the desmoplastic reaction of pancreatic cancer.

The ►**etiology** of pancreatic cancer remains poorly defined. The risk of developing this disease clearly increases with age, while the influence of environmental factors is still a matter of debate. An estimated 10% of pancreatic cancers are due to an inherited predisposition, but unlike other familial cancer syndromes, familial pancreatic cancer has a low ►**pene- trance** (<10%) and the age of onset is comparable to that observed in ►**sporadic cases**. In recent years, a number of genetic lesions have been identified that are linked to familial pancreatic cancer. Among them are ►**germline mutations** in the CDKN2A (which encodes two tumor suppressors – INK4A and ARF), BRCA2, LKB1 and MLH1 genes (for an overview, see 1). ►**Somatic mutations** in some of these genes are also implicated in the development of sporadic pancreatic cancer (see below). Another inherited syndrome predisposing for pancreatic cancer is hereditary pancreatitis (HP), a chronic inflammatory condition of the pancreas, which is most often caused by mutations in the cationic trypsinogen gene (PRSS1). Interestingly, longstanding sporadic chronic pancreatitis (spCP), which can arise from a number of different causes, is an established pancreatic cancer risk factor as well. It is not known if the different risks associated with HP and spCP are due to the length of the illness or some other factor.

Sporadic cases of pancreatic cancer have been demonstrated to be frequently associated with somatic mutations in the KRAS, CDKN2A, TP53, BRCA2 and SMAD4/DPC4 genes. However, since early stages of the disease are rarely detected, information about the sequence of mutational events and their contribution to the evolution of pancreatic cancer is scarce. Likewise, it is still unknown what cell type(s) pancreatic cancer arises from. Pancreatic adenocarcinoma cells resemble pancreatic duct cells in shape, ►**antigen expression** and formation of tubular structures. However, the focal expression of non-ductal lineage markers, including endocrine factors and pancreatic enzymes, and observations from animal model systems suggest that there are several routes to pancreatic adenocarcinoma, involving putative pancreatic stem cells and ►**trans- differentiation** of ►**endocrine** and ►**exocrine cells**. The use of microarray technology is expected to substantially improve our understanding of the biological principles of pancreatic cancer development. Especially the analysis of so-called pancreatic intraepithelial



Microarrays in Pancreatic Cancer. Figure 1 Application of microarray technology in pancreatic cancer diagnostics.

neoplasias (PanIN), which are thought to represent the precursor lesions of pancreatic cancer, will yield valuable information about the early steps in pancreatic cancer development and progression.

Clinical Relevance

Gene expression profiling using different types of DNA arrays is the most advanced and most widely used application of microarray technology in cancer research to date. The use of expression profiling in cancer diagnostics and therapy can broadly be divided into three general categories, candidate gene selection, class prediction and class discovery. While the first aims at identifying individual genes which may serve as new diagnostic or therapeutic targets, the latter two applications comprise the analysis of complex gene expression patterns to assign clinical samples to known disease entities or to define new disease subtypes which were previously not recognizable using classical diagnostic procedures (Fig. 1).

Candidate Gene Selection

From a data analysis point of view, this category is the most straightforward application of microarray technology. ▶**Expression profiles** of cancer tissues are compared with profiles of non-cancerous tissues from

the same organ using simple statistical tests to identify genes that are predominantly or exclusively expressed in the malignant state. However, exploitation of the results is complicated by the fact that several hundreds of genes may change their expression levels upon malignant ▶**transformation** of a normal cell. This also holds true for pancreatic cancer, where more than 360 differentially expressed genes were detected by our own group in the first ever expression profiling analysis of a solid tumor (2) and more genes have been added to the list since. While well-described genes can be screened for known characteristics enabling the development of new diagnostic tests or new treatment modalities, unknown genes or ▶**ESTs** have to be characterized in additional series of *in vitro* and/or *in vivo* experiments to select the most promising candidates. To this end, the genes are assembled on specialized candidate gene arrays and screened in serial experiments using model systems modulating important characteristics of tumor cells (invasive and metastatic potential, escape from growth control, resistance to chemotherapeutic agents, escape from programmed cell death etc.) for their involvement in crucial steps in tumorigenesis. Applying this strategy to pancreatic cancer, we were for example able to identify a novel RNA-binding protein (KOC) which in the

healthy organism is exclusively expressed in embryonic tissues, but is aberrantly re-expressed in pancreatic cancer as well as a number of other malignant diseases (3). PCR-based diagnostic assays for the presence of this gene product detect malignant cells in biopsy samples with 91% specificity and 97% sensitivity, thereby greatly enhancing the accuracy of 'traditional' cytological diagnostics.

An example of a novel therapeutic target selected on the basis of known characteristics is the tight junction protein claudin-4, which we found to be ectopically expressed by pancreatic cancer cells in many patient samples. This protein is normally located on the surface of intestinal epithelial cells and has previously been demonstrated to serve as the receptor of the *Clostridium perfringens* enterotoxin (CPE), the causative agent of many food poisonings. CPE exerts an acute cytotoxic effect on claudin-4 expressing cells by disrupting the cell membrane and destroying ion gradients, while cells not expressing claudin-4 remain completely unaffected. Since the protein is absent from all healthy pancreatic cells, CPE can be locally administered to selectively kill claudin-4 expressing cancer cells, as has been successfully demonstrated in animal experiments using subcutaneously implanted pancreatic tumors in immunodeficient nude mice (4).

Class Prediction and Class Discovery

A principally different approach to using expression profiling for diagnostic purposes is to try to identify complex patterns within gene expression profiles that are associated with the presence or absence of known diseases or disease subtypes (class prediction) or that will even allow definition of previously unrecognized disease subtypes associated with a characteristic prognosis, response to therapy, risk of relapse etc (class discovery). The analysis of the large amounts of data that are generated in the expression profiling experiments and their combination with clinical patient data to extract relevant expression patterns represents a challenging task that can only be met with the help of advanced computer sciences. Pattern recognition and data classification algorithms are being adapted and augmented for this purpose, but generally accepted standards are not yet emerging. Computational analysis of microarray data is an active area of research in bioinformatics and will undoubtedly see much advance in the near future.

A way to reduce the complexity (and the costs) of microarray analyses in diagnostics is to build specialized arrays with limited numbers of selected genes that have a high probability of yielding useful information. In order to improve the so far unsatisfactory accuracy of pre-operative pancreatic cancer diagnostics, we have developed such a specialized "diagnostic cDNA array" based on results of our own large scale expression

profiling analyses (see above) as well as public gene expression databases (►<http://www.ncbi.nlm.nih.gov/SAGE/>, ►http://www.ncbi.nlm.nih.gov/UniGene/info_ddd.shtml) and reports from the literature. The "diagnostic array" is composed of 487 features (corresponding to ca. 300 genes, some of which are represented by more than one cDNA clone) "spotted" in duplicate on nylon membranes. Since nylon membrane arrays are hybridized with radioactively labeled samples, this system offers a greater sensitivity and a broader ►dynamic range than glass slide microarrays, which are hybridized with fluorescently labeled samples. The feasibility of using this array to differentiate between pancreatic cancer and chronic pancreatitis was tested using surgically resected patient material. Since the biopsy samples which are available for pre-operative diagnostics in the clinical setting only yield very little mRNA material for analysis, linear amplification procedures as described for example by Eberwine et al. (5) have to be employed which allow the production of additional copies of every transcript from the sample prior to the labeling step without altering the relative abundance of each mRNA species in the original preparation. In the case of pancreatic fine needle biopsy samples, the situation is further complicated by the fact that the pancreas is the primary site of ►RNase production in the gastrointestinal tract, so that additional measures have to be taken to isolate and linearly amplify intact mRNA successfully from these samples. Using protocols developed in our lab, it is possible to obtain relevant expression profiles from as few as 500 cells from pancreatic tissues, a number which is far exceeded even in fine needle biopsy samples. In a preliminary series of experiments, we were able to demonstrate that our diagnostic array differentiates between benign and malignant processes in the pancreas with >90% accuracy using both surgically resected material and clinical fine needle biopsy samples.

In addition to this class prediction type of use of the diagnostic array, we expect to extend the analysis of the array data in a class discovery type of scenario to correlate characteristic expression patterns to clinical parameters such as tumor stage, prognosis, risk of relapse, response to therapy etc. The definition of distinct subtypes of pancreatic cancer based on characteristic expression patterns will provide much more detailed information about each individual tumor that can be used to select the most promising treatment option, setting the stage for therapy schemes custom tailored to the individual patient.

References

1. Bardeesy N, De Pinho RA (2002) Pancreatic cancer biology and genetics. *Nat Rev Cancer* 2:897–909

2. Gress TM, Müller-Pillasch F, Geng M et al (1996) A pancreatic cancer-specific expression profile. *Oncogene* 13:1819–1830
3. Müller-Pillasch F, Lacher U, Wallrapp C et al (1997) Cloning of a gene highly overexpressed in cancer coding for a novel KH-domain containing protein. *Oncogene* 14:2729–2733
4. Michl P, Buchholz M, Rolke M et al (2001) Claudin-4: a new target for pancreatic cancer treatment using *Clostridium perfringens* enterotoxin. *Gastroenterology* 121:678–684
5. Phillips J, Eberwine JH (1996) Antisense RNA Amplification: A Linear Amplification Method for Analyzing the mRNA Population from Single Living Cells. *Methods* 10:283–288

Microarrays in Plant Genomics

NIKOLAUS L. SCHLAICH, ALAN J. SLUSARENKO
Institute Bio III (Plant Physiology), RWTH Aachen,
Aachen, Germany
schlaich@bio3.rwth-aachen.de
alan.slusarenko@bio3.rwth-aachen.de

Definition

Microarrays consist of glass microscope slides (initially they were also made on membranes) with tiny, covalently attached spots of DNA (1.28 cm² can carry up to 500,000 DNA spots!). Each spot is called a “probe set” or “feature” and contains millions of identical single- or double-stranded DNA molecules from a specific plant gene. Hybridisation of the array with labelled cDNAs made from a population of mRNAs gives a “snapshot” picture of the plant transcriptome at the instant of tissue sampling. Thus, genome-wide alterations in gene expression between different states, e.g. leaves in the dark *versus* leaves in the light, can be catalogued in one experiment. If the identity of the genes making up the spots is known or suspected, the results of a microarray experiment can give an insight into the complexities and interrelationships of the gene expression associated with the response under investigation. This is an important first step into integrating the ►systems biology approach to close the gap in our understanding of how genotype “becomes” phenotype.

Characteristics

Whole genome studies became feasible in the late 1990s as completely sequenced genomes from several different organisms became available. Techniques were developed to study several thousand gene fragments,

called expressed sequence tags (or ►ESTs) and, in a few cases, even all of the genes of an organism in single experiments. Microarrays were among the first tools to be used in this way. Initially, the target gene sequences were spotted on to the support by hand but manual application was far too laborious and imprecise to yield reliably reproducible results. Therefore, robots were developed which could spot fractions of a microliter onto membranes or glass slides at very precise intervals. Another improvement has been the synthesis of the DNA strands directly on surface-modified glass slides.

Initially, PCR amplified double-stranded DNAs, made from plasmids containing inserts of variable length were used to prepare arrays. However, the different sizes of the DNA molecules resulted in different signal strengths due not only to transcript abundance in the probe DNA population but also to different hybridisation kinetics. The results from array experiments have been continually made more accurate and reliable by standardising the length of the target DNA sequence of each feature. Thus, today, homogenous single-stranded oligomers of 25, 60 or 80 nucleotides are used, bringing one of the two molecular species involved in hybridization to standard behaviour. Unfortunately, size differences in the complex cDNA probe cannot be controlled.

The application of DNA microarray technology to plant genomics in recent years has largely been with the model dicotyledonous plant *Arabidopsis thaliana*, which was the first plant whose entire genome was sequenced. Since *Arabidopsis* has “only” approximately 25,000 genes, DNA molecules designed to match specifically each of these genes can be arrayed on a single glass slide similar to the ones used in light microscopy. Since the position of each feature and its corresponding gene sequence on the microarray is known, it is possible to specifically detect the intensity of hybridisation signals from each spot and compare it to a reference array in a fully automated way. Thus, there are commercially available arrays in which all of the approx. 25,000 genes of the model plant *Arabidopsis* are represented. Whole genome arrays for plants of agricultural importance, e.g. rice and barley, are in production. For plants, in which the entire genome is not (yet) sequenced, there is no alternative but to resort to large EST-DNA arrays.

At the other end of the scale, arrays of a limited number of genes whose expression is associated with a particular stimulus can be prepared and serve to give a “fingerprint” or “signature” for a plant in a particular state. For example, about 150 key genes respond to the volatile signals released in plants by wounding. Reymond et al. (3) prepared an array with these genes and compared the wounding signature with that produced by aphid feeding and came to the conclusion

that aphids were able to minimise the expression of a subset of defence genes.

How Does an Array Experiment Work?

Microarray experiments aim at revealing the molecular events underlying dynamic processes. If you are interested in comparing plants in state A with plants in state B (let's say plants exposed to pathogens compared to control plants not exposed), you need to extract RNA from both plant populations. This RNA will be enzymatically converted to cDNA. In the conversion step labelled nucleotides are also incorporated into the cDNA molecules (this can be radioactive label or, in microarray-experiments, more commonly fluorescent label). The labelled cDNA from plant population A will then be hybridised to one array, while the labelled cDNA from plant population B will be hybridised to a duplicate array. The absolute (fluorescent or radioactive) signals from each spot on the array will then be analysed and compared between arrays A and B. This identifies genes that are up- or down-regulated by the different conditions used.

In this way, numerous studies have been carried out to date. The different conditions analysed reflect the current interest in different topics in plant development and responses to various stimuli.

At the time of writing, searching PubMed for "Arabidopsis" and "microarray" returns 87 hits of which at least 41 are relevant, i.e. they report microarray results. Obviously, all aspects of plant biology are covered by this flood of studies, including the development from seed to the mature, flowering plant and changes in transcription during a day/night cycle. Different abiotic stress conditions like hypoxia, ozone, heat, cold, salt, osmotic challenge, high light and de- or re-hydration have been studied with microarrays, as well as the response of plants to biotic stress using pathogens such as viruses, bacteria and fungi and pests such as nematodes and aphids. Similarly, plant nutrition/primary metabolism is a frequently studied topic, investigations being reported with plants grown under low nitrate, phosphate or iron conditions. Moreover, responses to hormone treatments were monitored with microarrays after exposing plants to auxin, abscissic acid or brassinosteroid (Table 1).

What Is the Purpose of Microarray Experiments?

Up- or down-regulation of genes points to a potential role of those genes in the process under analysis. For example, if a gene is differentially regulated in the population of plants that was exposed to the pathogen, this might reflect this gene having something to do with defence.

A danger in microarray studies is that they can lead to the production of long lists of up- or down-regulated

genes, without any real integration into a framework of biological relevance. Refinements to facilitate this step will be the next challenge for using microarray data in functional genomics. Attempts to utilise microarray data at this functional level have been made. Thus, Maleck et al. (2) showed in 2000 that a combination of cluster-analysed microarray data and data from genome sequencing projects identified potential regulatory elements in promoters of genes showing a similar expression pattern in the microarray. They used the *PR1* gene, known as a marker for systemic acquired resistance in plants, and grouped 45 ESTs from approximately 31 different genes showing a similar expression pattern into a *PR1* cluster. They identified within the 26 available 1.1 kb sequences upstream of the predicted translation start sites (ATG) a pentameric sequence, called the W box, which is present in all these 26 promoters and highly over-represented as compared to promoters of genes found in non-*PR1* clusters.

In another study, Scheideler et al. (4) were able to verify biochemical data gained over several years of research in plant pathophysiology in many labs with a microarray. They analysed in detail a set of 51 genes from a 13,000 EST microarray for up- or down-regulated genes encoding enzymes involved in primary or secondary metabolism at four time points after plants were exposed to an avirulent pathogen. They showed that changes in gene expression for genes encoding enzymes involved in primary metabolism were much less dramatic than changes in expression of genes encoding enzymes involved in pathways typically associated with pathogen defence, like the shikimate pathway or secondary phenolic metabolism starting from phenylalanine.

Moreover, arrays may help in the elucidation of a problem particularly pertinent in plants i.e. assigning a specific function to a given member of a multi-gene family or, on a larger scale, gene super families. For example, in *Arabidopsis* there are around 300 members of the supergene family encoding cytochrome P450 haem-thiolate proteins. What do they do? Why are there so many? One might get a first clue by looking at organ or tissue specific expression. This can be analysed by generating a "candidate" array containing only the genes under investigation (Table 1).

Another way to make use of array technology is the comparison of the transcriptome of a mutant with that of the wild type. This is particularly useful for mutants with a lesion in a regulatory gene. Understanding which downstream genes are misregulated in a given mutant under a specific condition can point to a potential function of that regulatory gene, which had otherwise been obscured by a non-penetrant phenotype. A further innovative use of micro array

Microarrays in Plant Genomics. Table 1 Examples of microarray studies in plants. For further reading we compiled a, by no means comprehensive, list of studies published in the various subdisciplines in plant biology that address a specific biological phenomenon with the use of microarrays. Our comments refer to the specific experimental conditions described in the publication

Biological phenomenon	Comment	Reference
plant nutrition	phosphate starvation: 29% of almost 6,200 genes tested were differentially regulated by 72 h after the onset of phosphate depletion	Wu et al., 2003. Plant Physiol. 132, 1260–71.
	phosphate starvation	Hammond et al., 2003. Plant Physiol. 132, 578–596.
	Nitrate: 5% of root genes and less than 1% of the 22,500 genes in leaves responded with differential expression after addition of nitrate	Wang et al., 2003. Plant Physiol. 132, 556–567.
plant hormones	gibberellin treatment	Ogawa et al., 2003. Plant Cell 15, 1591–1604.
	auxin treatment	Sawa et al., 2002. Plant J. 32, 1011–1022.
	abscissic acid treatment: 3.5% of 7,000 genes responded	Seki et al., 2002. Funct. Integr. Genomics 2, 282–291.
	brassinosteroid treatment	Goda et al., 2002. Plant Physiol. 130, 1319–1334.
	auxin and cytokinin treatment: 2% to 3% of 8,000 genes responded to	Che et al., 2002. Plant Cell 14, 2771–2785.
plant development	shoot growth in species distantly related to <i>Arabidopsis</i>	Horvarth et al., 2003. Plant J. 34, 125–134.
	senescence in autumn leaves of <i>Populus</i>	Bhalerao et al., 2003. Plant Physiol. 131, 430–442.
	germination	Dean Rider et al., 2003. Plant J. 35, 33–43.
	seedling development	Ma et al., 2003. Development 130, 969–981.
	far-red light stimulus	Wang et al., 2002. Plant J. 32, 723–733.
	unstable mRNAs	Gutierrez et al., 2002. PNAS 99, 11513–11518.
	light quality responses: 30% of >6,000 genes showed altered expression	Ma et al., 2001. Plant Cell 13, 2589–2607.

Biological phenomenon	Comment	Reference
	diurnal and circadian expression was shown by 11% and 2% of 7,800 genes, respectively	Schaffer et al., 2001. <i>Plant Cell</i> 13, 113–123.
	circadian rhythms: 6% of 8,000 genes were differentially regulated by circadian rhythms	Harmer et al., 2000. <i>Science</i> 290, 2110–2113.
organ specific transcription	pollen transcriptome	Hony & Twell 2003. <i>Plant Physiol.</i> 132, 640–652.
	seed transcriptome	Girke et al., 2001. <i>Plant Physiol.</i> 124, 1570–1581.
abiotic stress	rehydration: 2% of 7,000 genes were differentially expressed under conditions	Oono et al., 2003. <i>Plant J.</i> 34, 868–887.
	water-deficit stress	Bray 2002. <i>Ann. Bot. (London)</i> 89, 803–811.
	high-light irradiance: 1.5% of 7,000 genes showed a response by 3 hours	Kimura et al., 2003. <i>Photochem. Photobiol.</i> 77, 226–233.
	high-light stress	Rossel et al., 2002. <i>Plant Physiol.</i> 130, 1109–1120.
	low-oxygen tension: 6% of 3,500 genes responded	Klok et al., 2002. <i>Plant Cell</i> 14, 2481–2494.
	iron-deficiency	Thimm et al., 2001. <i>Plant Physiol.</i> 127, 1030–1043.
	H2O2 treatment: >2% of 8,000 genes showed changes	Desikan et al., 2001. <i>Plant Physiol.</i> 127, 159–172.
	drought, cold and salt stress: 4%, less than 1% and 3% of 7,000 genes responded, respectively	Seki et al., 2002. <i>Funct. Integr. Genomics</i> 2, 282–291. Seki et al., 2002. <i>Plant J.</i> 31, 279–292. Seki et al., 2001. <i>Plant Cell</i> 13, 61–72.
	salt, osmotic and cold stress	Kreps et al., 2002. <i>Plant Physiol.</i> 130, 2129–2141.

Microarrays in Plant Genomics. Table 1 Examples of microarray studies in plants. For further reading we compiled a, by no means comprehensive, list of studies published in the various subdisciplines in plant biology that address a specific biological phenomenon with the use of microarrays. Our comments refer to the specific experimental conditions described in the publication. (Continued)

Biological phenomenon	Comment	Reference
biotic stress	Alternaria infection: almost 10% of 7,000 genes were induced	Narusaka et al., 2003. <i>Plant Cell Physiol.</i> 44, 377–387.
	cyst nematode infection: 1.5 % of 8,200 genes were differentially regulated	Puthoff et al., 2003. <i>Plant J.</i> 33, 911–921.
	aphid infection	Moran et al., 2002. <i>Arch. Insect Biochem. Physiol.</i> 51, 182–203.
	virus infection	Whitham et al., 2003. <i>Plant J.</i> 33, 271–283.
	Pseudomonas infection: about 10% of 8,000 genes showed altered expression levels	Scheideler et al., 2002. <i>J. Biol. Chem.</i> 277, 10555–10561.
Mutant Analysis	systemic acquired resistance	Maleck et al., 2000. <i>Nature Genet.</i> 26, 403–410.
	dst mutant	Perez-Amador, et al., 2002. <i>Plant Cell</i> 13, 2703–2717.
arrays with selected genes	Chemically-induced cell death with approx. 100 genes	Swidzinsky et al., 2002. <i>Plant J.</i> 30, 431–446.
	transcription factors response to various stresses	Chen et al., 2002. <i>Plant Cell</i> 14, 559–574.
	cytochrome P450 array	Xu et al., 2001. <i>Gene</i> 272, 61–74.
	defence to various pathogens with an array containing >2,300 genes	Schenk et al., 2000. <i>PNAS</i> 97, 11655–11660.
	responses to wounding and aphid feeding on a 150 genes array	Reymond et al., 2000. <i>Plant Cell</i> 12, 707–720.
	ozone, drought or wounding	Matsuyama et al., 2002. <i>Environ. Pollut.</i> 117, 191–194.

technology to detect ►single nucleotide polymorphisms between genomes was suggested by Hazen & Kay (1).

As the above examples illustrate, microarray analysis has a great future potential as a tool in plant functional genomics.

References

1. Hazen SP, Kay SA (2003) Gene arrays are not just for measuring gene expression. *Trends Plant Sci* 8:413–416
2. Maleck K, Levine A, Eulgem T et al (2000) The transcriptome of *Arabidopsis thaliana* during systemic acquired resistance. *Nature Genet* 26:403–410
3. Reymond W, Weber H, Damond M et al (2000) Differential Gene Expression in Response to Mechanical Wounding and Insect Feeding in *Arabidopsis*. *Plant Cell* 12:707–720
4. Scheideler M, Schlaich NL, Fellenberg K et al (2002) Monitoring the switch from housekeeping to pathogen defence metabolism in *Arabidopsis thaliana* using cDNA arrays. *J Biol Chem* 277:10555–10561

Microarrays in Rheumatoid Diseases

THOMAS HÄUPL

Department of Rheumatology, Charité, Humboldt-University, Berlin, Germany
thomas.haeupl@charite.de

Definition

The Greek word “rheuma” or “rheumatism” has the meaning of “flow”. It relates to the ancient concept of the four “humors” flowing through the body and thought to be unbalanced in painful afflictions, resulting in swelling with fluids in tissue and joints.

Today, our understanding of rheumatic diseases is very different, leaving only the association of rheumatism with pain, which “flows” in the musculoskeletal system from joint to joint. Although rheumatic diseases comprise several hundred different entities of chronic disease, mostly recognized by their clinical phenotype, etiologic and pathophysiological knowledge is still limited. There are rheumatic diseases with systemic inflammation, degenerative joint and vertebral diseases, soft tissue diseases and metabolic diseases associated with musculoskeletal pain. The inflammatory rheumatic diseases with their main representatives listed in Table 1 present with a prevalence of about 5%. Disabilities early in life and life-threatening complications are of high socio-economic relevance.

Rheumatoid arthritis with about 1% prevalence is the most important representative.

In rheumatic diseases, gene expression screening is expected to improve pathophysiological understanding and therapeutic target identification. Furthermore, serologic autoreactivities provide important parameters for disease identification and classification. Based on familial predisposition, the search for markers of inheritance has also become increasingly relevant. Therefore, multiparameter analyses in rheumatic diseases comprise not only arrays for mRNA but also for protein and SNP analyses (►Microarrays).

Phenotypic clinical signs dominate in defining rheumatic diseases (Fig. 1). Only a few molecular parameters are currently established for routine application and qualify for diagnostic and classification purposes. These include autoantibodies like rheumatoid factor (RF), anti-citrullinated peptide antibodies and antinuclear antibodies (ANA) as well as unspecific markers of inflammation like C-reactive protein, erythrocyte sedimentation rate or differential blood count. For some of these diseases, such markers and targets are summarized in Table 2. Beyond these, many molecular parameters including cytokines, intracellular signaling molecules and transcription factors, markers of matrix destruction or genetic markers especially of the ►HLA locus have been investigated. However, insufficient disease specificity as well as variability between individual patients when collected as single parameters have discouraged usage for routine diagnostic purposes. These diagnostic limitations also apply to the most recent therapeutic targets, tumor necrosis factor alpha (TNF- α) and interleukin-1 (IL-1). Nevertheless, several of these parameters may reflect the heterogeneity of rheumatic diseases in stage, activity and possibly not yet characterized sub-entities, which seem to be recognized only by molecular means using multiparameter analysis with microarrays.

Characteristics

Samples for Microarray Analysis

The main expectations in microarray analyses are to gain an overview of the different pathological processes and to identify new or predominant pathways. Although genome wide screens have become available, interpretation of data is limited by the development of bioinformatics tools that sort the differentially expressed genes into meaningful data. For example, tissue analysis in synovitis may provide insight into the leading pathological processes of organ specific manifestations. However, variability of cellular composition due to infiltration in these complex samples will confuse the molecular interpretation. Analysis of purified cell types in contrast may provide insight into cell type associated pathologic profiles and gene regulation but

Microarrays in Rheumatoid Diseases. Table 1 Representatives of inflammatory rheumatic diseases

arthritis	connective tissue disease	vasculitis	myositis
rheumatoid arthritis (RA) spondylo-arthropathies (SpA) psoriatic arthritis reactive arthritis	systemic lupus erythematosus (SLE) scleroderma mixed connective tissue disease	Wegener's granulomatosis Churg-Strauss-Takayasu vasculitis	Polymyositis dermatomyositis inclusion body myositis

**Microarrays in Rheumatoid Diseases. Figure 1** Radiograph of joint destruction in rheumatoid arthritis: metacarpophalangeal, proximal interphalangeal and carpal joints show multiple bone erosions.

may introduce artefacts during separation. Both approaches, complex samples and purified cell types, have advantages and disadvantages and will provide different and complementary information.

Tissue Pathology

► **Rheumatoid Arthritis** (RA) is a good example to demonstrate the characteristics of tissue pathologies in

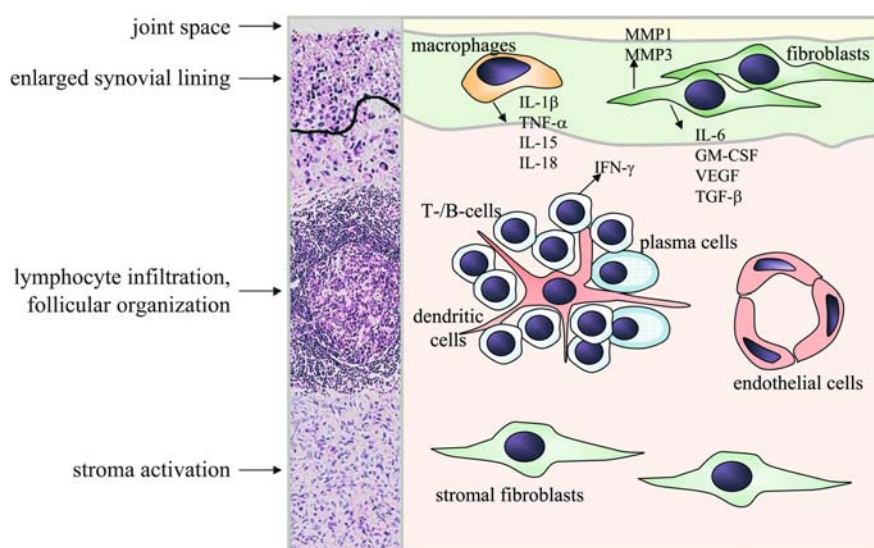
rheumatic diseases. Although classified as a systemic inflammatory disease, synovitis and joint destruction represent the central pathology in this disease (Fig. 2). The typical histological finding is infiltration of synovial pannus tissue into joint cartilage and subchondral bone, resulting in progressive destruction of the articulating surfaces. The synovial lining, normally only one or two cell layers that separate joint space from stromal tissue, is transformed in RA into a tremendously enlarged lining layer with macrophage accumulation and fibroblast proliferation. The sublining tissue displays extensive neovascularisation, formation of lymphocyte aggregates and stromal activation (1). Performing microarray analyses, these changes need to be attributed to their molecular equivalent to identify underlying pathomechanisms. Furthermore, it is difficult to identify in the large number of differentially expressed genes those regulated by disease specific molecular processes and separate others that only accumulate by infiltration of immune cells.

Gene Expression Profiling of Tissues and Tissue Cells

Based on a customized array with 96 pre-selected genes, Heller et al. demonstrated applicability of array technology for synovial tissue analysis to detect genes involved in joint destruction (MMPs stromelysin 1, collagenase 1, gelatinase A and human matrix metallo-elastase, TIMP 1 and 3) and inflammatory processes (IL-6, VCAM, MCP-1, MIF and RANTES) (2). More detailed analyses of synovial tissues from RA and osteoarthritis (OA) patients were performed on a 24,000 cDNA spotted array platform (3). These studies suggest a sub-classification into at least two different molecularly defined groups of RA, one with signatures of high inflammation, B- and T-cell activation (immunoglobulins, lck, STAT1, SDF1, HLA class II, IFI30 etc.) and the other with low inflammation but processes of tissue remodeling including expression of different types of collagen (II, XI) and activation of signaling pathways known from embryonic morphogenesis (Wnt5a). These results reflect the heterogeneity described in established histological scoring and grading of RA based on cellular composition and structural changes (1). Our own data on synovial tissues, established on a different technology platform,

Microarrays in Rheumatoid Diseases. Table 2 Molecular Markers and targets

Disease type/group	autoantibodies	genetic markers	therapeutic targets
RA	RF, anti-citrulline	HLA-DR4	TNF- α , IL-1, CD20
SpA	none characteristic	HLA-B27	TNF- α
SLE	ANA (anti-dsDNA, -Ro, -La, -Sm, etc.)	HLA-DR2	CD20
vasculitis	ANCA	variable	TNF
myositis	Jo-1, PM-Scl, Mi-2	HLA-DR3	miscellaneous

**Microarrays in Rheumatoid Diseases. Figure 2** Characteristics of synovial tissue inflammation in rheumatoid arthritis.

confirm many of these findings (4). A similar tissue-based analysis of periprosthetic membranes from prosthetic loosening in RA and OA patients also revealed up-regulation of various inflammatory genes (5). To enable association of differential gene expression with characteristic histomorphological structures or cell types, array analysis of laser micro-dissected synovial tissues with RNA from as few as 600 cells was shown to be feasible (6).

Alternatively, studies on purified and cultivated synovial fibroblasts are well established and several characteristics are sustained for many culture passages, reflecting an imprinted phenotype of molecular changes. Thus, Pierer et al. investigated profiles of synoviocytes on a functional basis by stimulation *via* toll-like receptor 2. With focus on chemokines, preferential activation of GCP-2, RANTES, MCP-2, IL8 and GRO2 was identified. Induction of MCP-2, RANTES and GCP-2 were dependent on NF- κ B.

Monocyte migration was influenced by RANTES and MCP-2, T-cell migration only by RANTES. Expression of GCP-2 and MCP-2 was identified not only in synovial tissue but also in synovial fluid, confirming the relevance of these results.

Gene Expression Profiling of Peripheral Blood Cells

Peripheral blood contains a mixture of different leukocytes, which may reflect systemic inflammatory processes. In a search for common and differentiating expression profiles in autoimmune diseases, Maas et al. investigated peripheral blood mononuclear cells (PBMC) from patients with rheumatoid arthritis, ▶systemic lupus erythematosus (SLE), type I diabetes and ▶multiple sclerosis. Comparing microarray expression profiles of these autoimmune diseases with profiles of healthy donors and their response to influenza vaccination, they characterized genes discriminating between normal immune and autoimmune responses.

However, no patterns were found to distinguish between the different autoimmune diseases. Candidate genes of autoimmunity were involved in apoptosis, cell-cycle progression, cell differentiation and cell migration but were not typical immune related genes. The algorithm that characterized patients with autoimmune diseases also identified healthy relatives of these patients, suggesting that the selection of genes may reflect a genetic trait rather than the disease process.

Analyses of PBMC in SLE revealed activation of interferon dependent pathways by comparing the findings with profiles induced by IFN- α in PBMC from healthy donors *in vitro* (7, 8). However, in one of these studies differential expression of granulopoietic genes was identified as well and turned out to originate from cells of the myeloid lineage. These cells co-separated, obviously upon disease associated changes, only in SLE. These findings again demonstrate the difficulties and risks of misinterpretation when microarray analyses are performed on complex cellular mixtures of unknown composition.

PBMC were also investigated in spondylo-arthropathies, rheumatoid arthritis and psoriatic arthritis on a 588-gene commercial array platform (9). Identified candidate genes included a myeloid nuclear differentiation antigen (MND1), two members of the S100 family of proteins (calgranulin A and B), JAK3 and MAP-kinase p38, TNF receptors, the chemokine receptors CCR1 and CXCR4 and IL-1 β and IL-8. As SDF-1, the ligand of CXCR4, was found to be increased in the synovial fluids of arthritides, the authors suggested an important role of this chemotactic axis in SpA and RA.

Genomic Markers of Disease

Investigations of genetic mutations are currently advancing into screening technologies for as many as possible identified [▶single nucleotide polymorphisms](#) (SNPs) in the human genome. In the past, some rheumatic diseases were found to be associated with certain HLA haplotypes. Focused investigations as well as genome wide approaches now suggest a growing number of polymorphisms as potential markers for rheumatic diseases. Such recently suggested polymorphisms in RA are in regulatory sequences or coding sequences of PADI4 (10), SLC22A4 (11) and others. However, it may be deduced from the heterogeneity of the clinical phenotype and the fact that new pathogens identified in the last decades, such as *Borrelia burgdorferi*, may contribute to the multitude of arthritic conditions that rheumatic diseases may not be explained by a monogenetic trait or one polygenetic pattern. Therefore, multiparameter analyses will also develop for genetic screenings in rheumatic diseases.

Microarrays used for expression profiling may contribute to the identification of genetic markers. Interesting bioinformatic technologies were applied to expression candidates that were common to two genetically different murine arthritis models. Using a spatial autocorrelation function, a statistical technique in astrophysics, critical clusterings of selected genes in these two different genetic backgrounds were identified. These may contain relevant polymorphisms and are therefore possible hot spots for further investigation.

Microarrays and Autoreactivity

The role of B-cells and their products, antibodies, has been the first focus in RA research. It was initiated by the identification of RF *via* its complex formation by Waaler and Rose. RF is found in about 70% of all RA patients and is directed against the constant region of immunoglobulins. Formation of complexes and activation of complement in the joint significantly promote inflammation by constant recruitment of inflammatory cells. Furthermore, B-cells also contribute by their function as antigen presenting cells and their co-stimulatory capacities. As a consequence, B-cell depleting therapies by anti-CD20 targeting have been started and may provide new tools in the treatment strategies for RA (12). Serologic autoreactivity screening also has important practical consequences in connective tissue diseases like SLE for identification and classification of the disease. As current autoimmune screens already demand for many different autoantigens, protein microarrays were developed for application in eight different rheumatic diseases (13). Using the multiple spotting technique, combined spotting of antibodies on top of the immobilized antigen revealed extremely high sensitivity and specificity (14). These qualities of microarray based autoreactivity screening may support the development of criteria for sub-classification of rheumatic diseases. Furthermore, these microarrays may provide parameters of disease activity by quantification of selected seroreactivities.

Clinical Relevance

RA, which has the highest socioeconomic impact, serves as an example for other rheumatoid diseases. Within a decade after onset, this disease leads to work disability defined as a total cessation of employment in more than 50% of patients. Total costs of this disease are at least 2–3 \times higher than direct health care costs. These are consequences of the unknown etiology and the lack of appropriate treatment.

Current clinical classification criteria have not been able to predict responsiveness to defined therapy regimens. Different models of molecular pathology have been suggested in the past based on candidate

driven research, but none is unequivocal or has provided sufficient exclusiveness to improve classification. A common molecular diagnostic tool beyond screening for rheumatoid factor or anti-citrullinated peptide antibodies is not available. On the other hand, a predominantly TNF driven pathophysiology obviously occurs only in a subgroup of RA patients, suggesting that subgroups may be defined by molecular means. Microarray analyses as published up to now seem to confirm this assumption. Considering this complexity, it cannot be expected that a single gene or factor will molecularly describe the multiple characteristics of rheumatic diseases.

Diagnostic improvement is needed for many decisions in clinical care for these patients. The identification of the disease at the very early stage when treatment can be most effective to prevent joint destruction is important. Identification of subgroups may reveal different pathomechanisms and define criteria for therapeutic stratification, especially for the evolving targeted therapies with biologicals. Therapeutic effectiveness is difficult to estimate by solely clinical criteria, since these will not allow differentiation between insufficiency of dosage or failure because of targeting the wrong pathway(s). Finally, infections may complicate the disease but present clinically similarly to a flare of the rheumatic disease.

All these difficulties demonstrate the importance of new molecular screening technologies. Microarrays as we know them today will certainly have high impact on the development of these diagnostic tools. However, rapid technological progress in this field may find cost effective ways of integrating not only the currently dominating gene expression and autoreactivity screens but also protein detection and modification.

References

- Krenn V, Morawietz L, Häupl T et al (2002) Grading of chronic synovitis – a histopathological grading system for molecular and diagnostic pathology. *Pathol Res Pract* 198:317–325
- Heller RA, Schena M, Chai A et al (1997) Discovery and analysis of inflammatory disease-related genes using cDNA microarrays. *Proc Natl Acad Sci USA* 94:2150–2155
- Pouw Kraan TC van der, Gaalen FA van, Kasperkovitz PV et al (2003) Rheumatoid arthritis is a heterogeneous disease: evidence for differences in the activation of the STAT-1 pathway between rheumatoid tissues. *Arthritis Rheum* 48:2132–2145
- Häupl T, Stuhlmüller B, Janitz M et al (2003) Gene expression profiling in rheumatoid arthritis. *Z Rheumatol* 62:49–50
- Morawietz L, Gehrke T, Frommelt L et al (2003) Differential gene expression in the periprosthetic membrane: lubricin as a new possible pathogenetic factor in prosthesis loosening. *Virchows Arch* 443:57–66
- Judex M, Neumann E, Lechner S et al (2003) Laser-mediated microdissection facilitates analysis of area-specific gene expression in rheumatoid synovium. *Arthritis Rheum* 48:97–102
- Baechler EC, Batliwalla FM, Karypis G et al (2003) Interferon-inducible gene expression signature in peripheral blood cells of patients with severe lupus. *Proc Natl Acad Sci USA* 100:2610–2615
- Bennett L, Palucka AK, Arce E et al (2003) Interferon and granulopoiesis signatures in systemic lupus erythematosus blood. *J Exp Med* 197:711–723
- Gu J, Marker-Hermann E, Baeten D et al (2002) A 588-gene microarray analysis of the peripheral blood mononuclear cells of spondyloarthritis patients. *Rheumatology* 41:759–766
- Suzuki A, Yamada R, Chang X et al (2003) Functional haplotypes of PADI4, encoding citrullinating enzyme peptidylarginine deiminase 4, are associated with rheumatoid arthritis. *Nat Genet* 34:395–402
- Tokuhiro S, Yamada R, Chang X et al (2003) An intronic SNP in a RUNX1 binding site of SLC22A4, encoding an organic cation transporter, is associated with rheumatoid arthritis. *Nat Genet* 35:341–348
- Dörner T, Burmester GR (2003) The role of B cells in rheumatoid arthritis: mechanisms and therapeutic targets. *Curr Opin Rheumatol* 15:246–252
- Robinson WH, Di Gennaro C, Hueber W et al (2002) Autoantigen microarrays for multiplex characterization of autoantibody responses. *Nat Med* 8:295–301
- Angenendt P, Glöckler J, Konthür Z et al (2003) 3D protein microarrays: performing multiplex immunoassays on a single chip. *Anal Chem* 75:4368–4372

Microcalorimetry

Definition

Calorimetry is the science of heat, i.e. it is about how a given material responds to temperature changes on both the atomic and macroscopic level. Microcalorimetry is the calorimetry of small samples.

- Protein-Protein Interaction
- Two Hybrid System

Microcephalus

Definition

Microcephalus is characterized by a small head circumference.

- Prader Willi and Angelman Syndromes

Microchannel

Definition

A microchannel is a conduit that is capable of transporting very small volumes of fluid (L – nL).

► [Proteomics in Microfluidic Systems](#)

Microdeletion

Definition

Microdeletion (meaning: chromosomal microdeletion) describes the loss of several thousand to around four or six million bases of genomic DNA from one homologue of a specific chromosome pair. It is frequently caused by unequal crossing over at low copy repeats ► [LCRs](#) (low copy repeats). Microdeletions are typically detected by ► [fluorescent in situ hybridization](#) (FISH).

► [Microdeletion Syndromes](#)

Microdeletion Syndromes

OLIVER BARTSCH¹, EVA SEEMANOVA²

¹Institute of Human Genetics, Mainz University School of Medicine, Mainz, Germany

²Institute of Biology and Medical Genetics, 2nd Medical School, Charles University, Prague, Czech Republic

bartsch@humgen.klinik.uni-mainz.de

eva.seemanova@lfmotol.cuni.cz

Synonyms

Chromosomal microdeletion syndromes; Contiguous gene syndromes

Definition

Progress in diagnostic methods allows the identification of the etiology of increasing numbers of clinical syndromes showing that microdeletions or occasionally microduplications are the specific type of causative mutations. The cytogenetic term microdeletion can be misleading, in particular to molecular geneticists, who consider microdeletions as extremely large at the molecular level. Other expressions such as “cryptic”

or “submicroscopic” deletion are no better, as molecular cytogeneticists regard chromosomal microdeletions neither as cryptic nor as submicroscopic. The term “microdeletion” should be avoided in favor of chromosomal microdeletion or simply deletion.

Chromosomal microdeletions represent deletions below the cytogenetic detection limit. The smallest unit in cytogenetics is the chromosomal subband or band which, given the length of the haploid human genome (~3,000,000,000 base pairs) and the number of bands resolved by standard karyotyping (450–550), spans around 6 Mb of genomic DNA in length. However, these 6 Mb cannot be simply equated with the cytogenetic detection limit, which strongly depends on the resolved number of bands, chromosomal location and banding pattern of the area in question. Microdeletions typically span in the order of 2–6 Mb of genomic DNA. The “subtle” rearrangements at the cytogenetic detection limit cover around 4–12 Mb. Strictly speaking, there are no exact borders between the cryptic, near-cryptic and subtle chromosome rearrangements and any rigid classification is artificial.

Characteristics

In 1956 a new syndrome of severe neonatal hypotonia, hypogenitalism, small hands and feet, obesity, short stature and mild mental retardation in the older child was delineated, and named ► [Prader-Willi syndrome](#). In the 1980s it was found that a subset of Prader-Willi patients have deletions at chromosome 15q11-q13 visible only by high-resolution methods. The 15q11-q13 deletion was found to be a constitutional deletion, removing 3–4 Mb of genomic DNA from the paternal chromosome 15q11-q13, due to the presence of ► [low copy repeats](#) in this area. With the advent of molecular cytogenetic techniques using fluorescence labeled DNA probes on metaphase chromosomes (fluorescent *in situ* hybridization, FISH) in the 1990s, several familiar and less well known phenotypes were discovered to be microdeletion syndromes, and became accessible for laboratory diagnosis. Microdeletions have been found in numerous genes and virtually all human chromosomes. At least 50 different microdeletions (Table 1) recur and generate more or less recognizable microdeletion syndromes (Figs. 1, 2, 3, Fig. 4c, d, e, f, g) that may be diagnosed by the medical specialist (clinical dysmorphologist). There is no distinct line that can be drawn between the microdeletion syndromes and the private syndromes (Fig. 4a, b) that are occasionally diagnosed in unique patients showing exceptional clinical signs.

Chromosomal microdeletions are one of the multitude of mutational mechanisms causing Mendelian disorders in man. Mutations such as single nucleotide substitutions are well known in human medicine, but



Microdeletion Syndromes. Figure 1 Phenotypes in Wolf-Hirschhorn syndrome; (a–b) 1 year old boy – note severe failure to thrive, anisochromia of iris and hypertelorism; (c–d) 14 year old girl with Pitt-Rogers-Danks syndrome (mild Wolf-Hirschhorn variant).

chromosomal microdeletions removing a string of neighboring genes out of a chromosome or a number of exons out of a large gene such as *DMD* (►Duchenne muscular dystrophy) represent a less known concept. Chromosomal areas harboring microdeletions at a high frequency include 7q11.23, 15q11-q13, 17p11.2-p12 and 22q11 (Fig. 3a). These areas are also prone to the formation of other constitutional cytogenetic microarrangements such as inversion 7q11.2, supernumerary marker chromosomes 15 or 22 and duplication 17p11.2-p12.

The genetic content of the different chromosomal microdeletions varies greatly but can be determined if the size and location of the deletion have been defined. The number of genes included in a microdeletion can be roughly estimated based on the number of genes and the total length of the haploid human genome. Genes are spaced at an average distance of 100,000 base pairs within the human nuclear genome. Therefore, a typical microdeletion such as the common 22q11 deletion, which spans around 3 Mb, removes approximately 30 genes. The precise content of a deletion can be computed using human genome browsers (Ensembl: ►<http://www.ensembl.org>, UCSC Genome Browser: ►<http://genome.ucsc.edu>) that display known and predicted genes in molecularly defined areas.

Nearly all microdeletion syndromes show genetic heterogeneity, for example by different mechanisms

of mutation (e.g. Prader-Willi syndrome; microdeletion, ►uniparental disomy, imprinting center mutation or chromosomal translocation) or by mutations on different chromosomes (e.g. ►DiGeorge syndrome due to mutation at 22q11 or at chromosome 10p14).

Moreover, deletions and duplications in the same area may cause different phenotypes (clinical heterogeneity). Deletion of the *PMP22* gene at 17p12 causes HNPP, whereas the 17p11.2-p12 duplication including *PMP22* leads to the more severe ►Charcot-Marie-Tooth disease (Table 1). Due to the heterogeneous origin, the specification of the cause is necessary for genetic evaluation of reproduction risks in affected families. The clinical work-up should include parents of those affected being checked for chromosomal rearrangements and cytogenetic mosaicism, even if they appear healthy.

The incidence of many well-known microdeletion syndromes is high, ranging from 1:4,000 to 1:30,000 in the general population. The constant and relatively high rate of microdeletions results from the common etiological mechanism that includes unequal crossing over during meiosis.

Rearrangements at the chromosome ends, such as the 4p16 deletion causing ►Wolf-Hirschhorn syndrome (Fig. 1), the 7q36 deletion causing ►holoprosencephaly and related defects, or other subtelomeric deletions are known to arise in a subset of cases secondary to a balanced familial translocation. More recently, low copy repeats (LCRs) were identified as a major cause of microdeletions during meiosis, in particular, interstitial microdeletions. By definition LCRs are large rare DNA motifs at specific chromosomal sites and have been reported to cause the constitutional micro-rearrangements at chromosomes 7q11.23, 15q11-q13, 17p11.2 and 22q11. LCRs have been associated with different types of rearrangements showing constant molecular breakpoints such as deletions, duplications, inversions, supernumerary marker chromosomes and translocations. The presence of defined LCRs and molecular breakpoints also explains why certain microdeletions typically show constant chromosomal breakpoints and why these microdeletions sometimes coincide with a different micro-rearrangement present in one parent. For example, the 7q11.23 deletion (Williams syndrome) has been associated with micro-inversion on one homologue 7q in a parent.

Due to the diploid nature of the human genome, a microdeletion typically results in heterozygous deletion (monosomy), with a normal allele of the deleted gene remaining present on the intact homologous chromosome. Yet a few microdeletions occur at imprinted chromosomal areas and therefore cause genetic nullisomy, e.g. ►Prader-Willi syndrome represents nullisomy of the paternal chromosome 15q12 and Angelman

Microdeletion Syndromes. Table 1 Chromosomal microdeletion syndromes

Syndrome or disorder	OMIM	Molecular region	Chromosomal region
Monosomy 1p36 syndrome	607872	–	1p36.3
Charcot-Marie-Tooth disease, type IB, CMT1B; hereditary motor and sensory neuropathy, type IB; HMSN1B	118200	MPZ	1q22
Holoprosencephaly 6, HPE6	605934	HPE6	2q37.1-q37.3
Pitt-Rogers-Danks syndrome; PRDS	262350	LETM1 to HOX7	4p16.3
Wolf-Hirschhorn syndrome; WHS	194190	LETM1 to HOX7	4p16.3
Cri-du-chat syndrome	123450	TERT	5p15.3-p15.2
SOTOS syndrome, cerebral gigantism	117550	NSD1	5q35.2
Saethre-Chotzen syndrome; SCS	101400	TWIST	7p21
Williams-Beuren syndrome; WBS; Williams syndrome	194050	ELN to CYLN2	7q11.23
Monosomy 7q36; holoprosencephaly 3; HPE3	142945	SHH	7q36.3
Nijmegen breakage syndrome; NBS; ataxia teleangiectasia variant VI	251260	NBS1	8q21.3
Langer-Giedion syndrome; LGS; trichorhinophalangeal syndrome, type II; TRPS2	150230	TRPS1 to EXT1	8q24.11-q24.13
Gonadal dysgenesis, XY female type; doublesex- and MAB3-related transcription factor 1; DMRT1	602424	DMRT1	9p24.3
DiGeorge syndrome/velocardiofacial syndrome spectrum of malformation 2; DGS2	601362	GATA3 to D10S2190	10p14
Beckwith-Wiedemann syndrome; BWS; exomphalos-macrocephaly-gigantism-syndrome	130650	CDKN1C	11p15.5
Aniridia, type II; AN2	106210	PAX6	11p13
WAGR syndrome; Wilms tumor, aniridia, genitourinary anomalies, and mental retardation	194072	PAX6 to WT1	11p13
Wilms tumor 1; WT1, nephroblastoma	194070	WT1	11p13
Proximal 11p deletion syndrome; p11pDS; Potocki-Shaffer syndrome	601224	EXT2 to ALX4	11p11.2
Ataxia teleangiectasia; AT1; Louis-Bar syndrome	208900	ATM	11q22.3
Noonan syndrome 1; NS1; Turner phenotype with normal karyotype	163950	PTPN11	12q24.1
Retinoblastoma 1, RB1	180200	RB1	13q14.3
Prader-Willi syndrome, PWS	176270	SNRPN	15q12-q13
Angelman-Syndrome, AS; happy puppet syndrome	105830	UBE3A	15q12-q13
Tuberous sclerosis; TS; Bourneville-Pringle disease; adenoma sebaceum; TS, type II; TSC2	191100	TSC2	16p13.3
Polycystic kidneys; polycystic kidney disease, PKD; Potter type III adult PKD, included	173900	PKD1	16p13.3
Rubinstein syndrome; Rubinstein-Taybi syndrome; RSTS; RTS	180849	CREBBP	16p13.3

Microdeletion Syndromes. Table 1 Chromosomal microdeletion syndromes (Continued)

Syndrome or disorder	OMIM	Molecular region	Chromosomal region
Miller-Dieker lissencephaly syndrome; MDLS; LIS1	247200	PAFAH1B1	17p13.3
Lissencephaly 1; LIS1; classic lissencephaly	607432	PAFAH1B1	17p13.3
Neuropathy, hereditary, with liability to pressure palsies; HNPP; tomaculous neuropathy	162500	PMP22	17p12 deletion
Charcot-Marie-Tooth disease, type IA; CMT1A; hereditary motor and sensory neuropathy, type IA; HMSN1A	118220	PMP22	17p11.2-p12 duplication
Smith-Magenis syndrome; SMS	182290	RAI1	17p11.2
Neurofibromatosis, type I; NF1; von Recklinghausen disease	162200	NF1	17q11.2
Holoprosencephaly 4; HPE4	142946	TGIF	18p11.3
Diamond-Blackfan anemia; DBA; Blackfan-Diamond syndrome	105650	RPS19	19q13.32
Alagille syndrome; AGS; arteriohepatic dysplasia	118450	JAG1	20p12.2
Albright hereditary osteodystrophy, AHO	103580	GNAS1	20q13.33
Holoprosencephaly 1, familial alobar; HPE1	236100	–	21q22.3, 2q37
DiGeorge-syndrome; DGS; CATCH phenotype; deletion 22q11 syndrome	188400	TBX1	22q11.2
Velocardiofacial syndrome; VCFS; Shprintzen-Sedláčková VCF syndrome	192430	TBX1	22q11.2
Leri-Weill dyschondrosteosis; LWD	127300	SHOX	Xp22.32, Yp11.32
Langer mesomelic dysplasia; homozygous dyschondrosteosis	249700	SHOX	Xp22.32, Yp11.32
Short stature, X-linked; SHOXX haploinsufficiency	312865	SHOXX	Xp22.32
Ichthyosis, X-linked; steroid sulfatase deficiency	308100	ARSC1	Xp22.32
Kallmann syndrome 1; KAL1; anosmic hypogonadism	308700	KAL1	Xp22.31
Ocular albinism, type I; OA1	300500	OA1	Xp22.22
Microphthalmia and linear skin defects; MLS; MIDAS syndrome	309801	ARHGAP6	Xp22.22
Aicardi syndrome; AIC; agenesis of corpus callosum with chorioretinal abnormality	304050	–	Xp22
Muscular dystrophy, Duchenne type, with deletion of exons of the DMD gene	310200	DMD	Xp21.2-p21.1
Mental retardation; X-linked 1; MRX1	309530	–	Xp11.3-q21.1
Myotubular myopathy with abnormal genital development	300219	F18 to MTM1	Xq28
Incontinentia pigmenti; IP; Bloch-Sulzberger syndrome	308300	NEMO	Xq28
Short stature; Y-linked; SHOXY haploinsufficiency	400020	SHOXY	Yp11.32
Gonadal dysgenesis, XY female type; Swyer syndrome; sex-determining region Y; SRY	480000	SRY	Yp11.31

syndrome is due to nullisomy of the maternal chromosome 15q12. These syndromes can be caused by microdeletion or alternatively, in a subset of cases, by uniparental disomy, or by defective imprinting.

Some 50 microdeletion syndromes (Table 1) have been delineated. A few syndromes such as the Wolf-Hirschhorn (Fig. 1), Prader-Willi, and ►**Smith-Magenis syndromes** (Fig. 2) demonstrate a fairly constant phenotypic appearance, but many others display considerable phenotypic variation between affected individuals, with the 22q11 deletion syndrome (Fig. 3) at the extremely variable end of the spectrum. Variability in the extent and severity of features is observed within the family as well as inter-familially. In particular, the 22q11 deletion syndrome has been reported with normal stature, normal appearance and normal intelligence (Fig. 3l, m).

The clinical signs of a microdeletion syndrome are defined by the hemizygous deletion of a string of genes on the recombinant chromosome. Between different chromosomes, a general correlation between the size of the microdeletion and the phenotypic manifestation does not exist. The phenotypic variability reflects the fact that the microdeletion syndromes are distinguished from each other by the clinical signs and symptoms that arise mainly from the deletion of genes with dominant effects within the deletion interval (monogenic traits), and to a lesser extent by other genes.

The paucity of the genes with dominant effects in the human genome (~1,000) can help to explain some of the peculiar characteristics of the microdeletion syndromes. Relatively few microdeletion syndromes (Williams-Beuren syndrome, proximal 11p deletion syndrome) represent true contiguous gene syndromes, in which the haploinsufficiency of two or more

neighboring ‘dominant’ genes contributes significantly to the patient’s phenotype. In other microdeletion syndromes the phenotype largely reflects the haploinsufficiency of a single ‘dominant gene’ (*SHH* gene causing holoprosencephaly and related defects in the monosomy 7q36 syndrome, *UBE3A* causing neurological defects in Angelman syndrome, *TBX1* causing pharyngeal pouch defects in 22q11 deletion syndrome). In some syndromes there is a minor contribution from another ‘dominant’ gene in a subset of cases.

A considerable number of microdeletions do not generate specific traits at all but only unspecific clinical signs and symptoms representing polygenic traits; this has been referred to as the chromosomal phenotype. To facilitate the diagnosis of subtle chromosomal imbalances in individual patients with an unspecific phenotype (Fig. 4a, b, c), clinical checklists have been established (2, 4). These checklists characteristically include four key points; 1) developmental delay, 2) abnormal (compromised) growth, 3) congenital anomalies (compromised development of complex body parts such the face, heart or reproductive organs) and 4) the chromosomal pedigree. Premature birth it is another unspecific sign frequently occurring with chromosome aberrations and could also be integrated into such checklists.

The developmental retardation (brain and neurological dysfunction) in the chromosomal phenotype is polygenic in origin and highly variable including prenatal disorders and severe hypotonia in the newborn and, at the mild end of the spectrum, only mildly abnormal behavioral phenotypes, impaired learning or even normal development. Correspondingly, the compromised growth and development may already be recognizable by prenatal ultrasound (intrauterine



Microdeletion Syndromes. Figure 2 Phenotypes of two unrelated male individuals, aged 14 and 20 years, with Smith-Magenis syndrome – note broad forehead, flat midface, refraction disorder and short neck.



Microdeletion Syndromes. Figure 3 Partial metaphase by FISH – note deletion at 22q11.2 (left homolog of chromosome 22) and normal control signals at both chromosome 22q13.3 homologues; b-m: Phenotypes of four unrelated children and mother and daughter with Di George syndrome at different ages – note hypertelorism, full cheeks, bulbous nose, small mouth and dysplastic ears.

growth retardation), may appear later in life (short stature, microcephaly), or may not even become apparent at all (normal stature included). Moreover, characteristic short stature and microcephaly in the chromosomal phenotype can be camouflaged in certain situations. The 22q13.3 deletion has been reported with tall stature and macrocephaly in some individuals and with normal (Fig. 4c) or short stature in others, predicting a growth regulating gene within the deletion interval, as yet unknown, but in the hemizygous state results in overgrowth in some individuals and in normal or compromised growth in others. The genetic background such as familial tall stature or familial macrocephaly can also camouflage the chromosomal phenotype of short stature or microcephaly in some patients. Congenital anomalies such as unspecific craniofacial dysmorphism (abnormal ear shape/structure, small chin, etc.) and unspecific heart defects are other common signs non-randomly associated with

chromosome aberrations, indicating disturbed embryonic and fetal development.

The chromosomal pedigree (recurrent miscarriage, stillbirth, affected relatives, and non-Mendelian transmission) is predominantly found with the terminal deletion syndromes (examples 4p- or Wolf-Hirschhorn syndrome (Fig. 1), 5p- or [Cri-du-chat syndrome](#), 17p- or [Miller-Dieker syndrome](#)), because these can arise from a subtle or cryptic balanced translocation in one of the parents. Interstitial microdeletions may also occur secondarily to familial chromosome rearrangements in rare cases (Williams syndrome, proximal 11p deletion syndrome).

Microdeletions result in the loss of a number of neighboring genes, causing the impaired function (by haploinsufficiency) of a subset of the proteins that are encoded by these genes and thereby generating the corresponding phenotype in the patient (contiguous gene syndromes). The molecular cytogenetic techniques



Microdeletion Syndromes. Figure 4 Phenotypes of children with sub-microscopic or subtle deletions at the ends of the chromosomes. (top) Three unrelated boys – note mild unspecific facial dysmorphism and absence of seminal features; (middle, bottom) Two unrelated male infants with subtle terminal deletion at 13q – note unspecific facial dysmorphism and characteristic ano-genital malformations.

(FISH, fluorescent *in situ* hybridization) made possible the visualization of single genes, and even parts of genes, on metaphase chromosomes and in non-dividing cells (interphase nuclei). The detection of chromosomal microdeletions is a standard procedure in clinical medicine today. Standard FISH tests have become available for most well-known microdeletion syndromes, but many disorders (Table 1) still require “home-made” or research FISH tests. Microdeletion syndromes are increasingly defined by the ‘dominant’ genes causing the major clinical signs. BAC/PAC clones

representing these causal genes are ideal DNA probes in the diagnostic FISH tests.

Subtelomeric microdeletions have been identified as the cause of 5% of severe idiopathic mental retardation (Fig. 4, top part) and the population incidence of all subtelomeric deletions combined is apparently at higher than 1 in 4,000 newborns. Nearly 50% of subtelomeric microdeletions are inherited from one parent, who then is a carrier of a balanced translocation or another cryptic rearrangement. These couples have a high recurrence risk (up to 50%).

Unfortunately, the test for detecting deletions or duplications at the telomeres, the subtelomere screen, is laborious and demanding, and therefore subtelomeric rearrangements are certainly underdiagnosed. Several working groups presently attempt solving this problem by use of microarrays (“DNA chips”) typically representing 6,000 and more BAC/PAC clones.

Cellular and Molecular Regulation

The molecular basis of the microdeletion syndromes is becoming increasingly elucidated. The disease-associated ‘dominant’ genes (Table 1) within the deletion interval give rise to the specific clinical signs making up the phenotype and gestalt of the syndrome. Most human genes are not dosage-sensitive to the deletion of one of the two alleles. For these genes, the lack of function of the allele on the deleted homologue of the chromosome is compensated by the other allele, which is present on the normal non-deleted homologous chromosome. Therefore, it is mainly the small subset of autosomal ‘dominant’ genes within the deletion interval that causes the phenotype of the individuals with a microdeletion. For example, the malformations in the common large 22q11 deletion are thought to be caused by only one gene (*TBX1*) out of the ~30 deleted genes. In addition, other genes within the deletion interval may contribute to multifactorial traits such as intelligence or behavior (1). Compromised intelligence is a frequent sign in chromosomal microdeletion syndromes, with the individual IQ outcome ranging from normal to mental retardation depending on many factors. As a rule, the smaller the deletion or duplication of a given chromosomal region, the better mental outcome can be expected. For example, there is evidence that individuals with a 1p36 deletion of 2.2 Mb and less in size are much less retarded than patients with a 1p36 deletion encompassing 3.6 Mb and more; and patients with Angelman syndrome caused by uniparental disomy (imprinting defect) often can speak several words whereas Angelman patients with the typical 15q11-q13 deletion (haploinsufficiency of additional genes) are typically completely a verbal. A small subset of individuals with a chromosomal microdeletion syndrome may present exceptional clinical signs, usually of a recessive disorder. Such cases can be explained by the high frequency of disease-causing recessive alleles in the population.

Clinical Relevance

Molecular cytogenetic testing (►FISH) has bridged the gap between cytogenetic diagnosis (typically detecting congenital disorders caused by genomic imbalance involving more than 6 Mb of genomic DNA, or 60 genes) and molecular diagnosis (typically detecting disorders caused by a molecular defect within a single

gene), leading to the growing recognition of the chromosomal microdeletions as a significant cause of human congenital disorders. The most frequent microdeletion, the 22q11 deletion (Fig. 3a), occurs at 1 in 4,000 newborns and was found to be second most frequent autosomal aneusomy in humans after trisomy 21 (incidence 1 in 700 newborns).

Microdeletion or subtelomere screening (using different DNA probes that are specific to each of the human chromosome ends) may be extremely valuable in patients with congenital anomalies and/or developmental delay (4), and has also been utilized in hematologic malignancies, in recurrent miscarriages, for prenatal detection, and for preimplantation diagnosis, given that it is allowed in the given country. Chromosomal microdeletions make up for a significant proportion of congenital disease. Congenital disorders in man occur in 3–4% of live births in the general population; mild mental retardation (MR) occurs at 1–2% and severe MR at about 0.5% of adults. Microdeletions, including those that do not cause a recognizable syndrome but can be detected by the FISH subtelomere screen or DNA microarray techniques, occur at an estimated total incidence of more than 1 in 1,000, or >0.1% of newborns. Taken together they account for around 3% of all inborn human disorders in man, for 3–5% of mild MR and for 5–10% of severe idiopathic MR.

Prenatal prevention is feasible after gestational week 11 by chorion villus biopsy, and by amniocentesis.

Clinical Features

In the last decade the number of newly recognized chromosomal microdeletion syndromes has greatly increased (Table 1). Characteristics common for this group of disorders include

1. frequently a characteristic craniofacial phenotype or recognizable face (Figs. 1, 2, 3), growth retardation (often with prenatal onset) (Fig. 1), congenital heart defects, developmental delay and mental retardation (all patients shown, except Fig. 3l, m).
2. Many microdeletion syndromes (DiGeorge, Williams, Prader-Willi, Angelman, and Smith-Magenis syndromes) have defined chromosomal breakpoint areas, and the seminal clinical features result from the deletion of a few dominant gene(s). Thus, in many cases, the phenotypic variability between patients with a given syndrome does not reflect differences in size of the microdeletions, but most likely is due to unknown (environmental or genetic) factors.
3. New mutations are the common etiology; only 1–10% of chromosomal microdeletions are inherited due to the identical deletion, or a balanced translocation, point mutation in imprinting centers or gonadal mosaicism in one parent.

4. Recurrence risk is generally low (new mutations) but may be as high as 50% in cases of an inherited mutation and spontaneous abortions can occur.

References

1. Bartsch O, Nemecková M, Kocárek E et al (2003) DiGeorge/velocardiofacial syndrome: FISH studies of chromosomes 22q11 and 10p14, and clinical reports on the proximal 22q11 deletion. *Am J Med Genet* 117A:1–5
2. De Vries BB, White SM, Knight SJ et al (2001) Clinical studies on submicroscopic subtelomeric rearrangements: a checklist. *J Med Genet* 38:145–150
3. De Vries BBA, Winter R, Schinzel A et al (2003) Telomeres: a diagnosis at the end of the chromosomes. *J Med Genet* 40:385–398
4. Walter S, Sandig K, Hinkel GK et al (2004) Subtelomere FISH in 50 patients with mental retardation and dysmorphism, identified by a checklist, detects 10 rearrangements including a de novo balanced translocation of chromosome 17p13.3 and 20q13.33. *Am J Med Genet* 128A:364–373

Microfilaments

Definition

Microfilaments are thin actin filaments comprising of part of the cytoskeleton that function in structure and movement of eukaryotic cells.

- Actin Cytoskeleton
- Cell Polarity
- Cytoskeleton

Microfluidics

Definition

Microfluidics denotes manipulating and understanding transport of small volumes of fluid (μL – nL).

- Proteomics in Microfluidic Systems

Microglia

Definition

Microglia are a neuroglial subtype of mesodermal origin. They are the resident macrophage population of

the CNS, and are considered to be its major immunocompetent elements. Microglia are activated by any type of brain pathology. Their activation results in conversion to so-called brain macrophages. Activated microglia secrete proinflammatory cytokines and chemokines, which can migrate to the lesion site and start to phagocytose.

- *In Vivo* Imaging of Transgenic Mice with Fluorescent Protein Expression

Micrometastasis

- Genomic Analysis of Single Disseminated Cancer Cells

Micrometastatic Tumour Cells

- Genomic Analysis of Single Disseminated Cancer Cells

Micromolding

Definition

Micromolding means shaping plastics by curing their liquid (or monomeric) form over a master that has the required patterns. For example, music CDs are produced using this method.

- Proteomics in Microfluidic Systems

MicroRNA

V. NARRY KIM

School of Biological Sciences and Institute of Molecular Biology and Genetics, Seoul National University, Seoul, Korea
narrykim@snu.ac.kr

Synonyms

The first two ► *microRNAs*, *lin-4* RNA and *let-7* RNA, were collectively named small temporal RNAs

(stRNAs), because of their temporal expression patterns and their roles in temporal regulation (5). After the discovery of hundreds of small RNAs in 2001, members of this non-coding RNA family were renamed as microRNAs (miRNAs) (1).

Definition

MicroRNAs (miRNAs) are defined as single-stranded RNAs of ~22 nt in length (range 19~25 nt) generated from endogenous transcripts that can form local hairpin structures (1, 2) (Fig. 1). MiRNAs act as guide molecules in gene silencing, by base pairing with the target mRNAs leading to translational repression and/or mRNA cleavage. The target gene locus is distinct from the miRNA gene locus itself.

Most miRNAs were found by the cDNA cloning method but a significant number of miRNAs are also being discovered by computational prediction. To identify and distinguish miRNAs from siRNAs or fragments of other RNAs, a combination of the following criteria are used (1): (A) detection of a distinct ~22 nt RNA by Northern blotting, (B) identification of the ~22 nt sequence in a library of cDNAs made from size-fractionated RNA, (C) presence of the ~22 nt sequence in one stem of a potential hairpin precursor structure, (D) phylogenetic conservation of the ~22 nt sequence and (E) accumulation of precursor when Dicer function is reduced. Usually evidence for (A+D) is sufficient. An RNA sequence meeting the criteria (A+C) or (B+D) is also regarded as miRNA.

MiRNAs are named with the “miR” prefix and a sequential number. Identical mature sequences from distinct precursors get names like miR-16-1 and miR-16-2, while highly related ones get names like miR-15a and miR-15b. When two miRNAs are cloned from opposite strands of one precursor, the more abundant one used to be assigned a name like miR-142 and the other one has an added asterisk as in miR-142*. When the relative abundance is not known, they were named miR-142-s (from the 5' strand) and miR-142-as (from the 3' strand). The current convention, however, is that miRNAs from the 5' strand are named miR-142-5p while those from the 3' strand are referred to as miR-142-3p, regardless of their relative abundance. The nomenclature follows the conventions of the organism so that miR-1 is in *C. elegans* and *Drosophila* and MIR39 in plants. The genes encoding the miRNAs are named with italics and capitalization like *mir-1* in *C. elegans* and *Drosophila* and *MIR39* in plants. The Rfam database of RNA families provides a searchable database for published miRNAs and assigns the miRNA gene names before publication (►<http://www.sanger.ac.uk/Software/Rfam/mirna>).

Characteristics

The Gene

Hundreds of small RNA genes have been found in animals, plants and viruses (2, 6). Many miRNAs are conserved across species, indicating that they have evolutionarily conserved roles in gene regulation (Fig. 1A). Most of the miRNA genes are found in intergenic regions or in antisense orientation to annotated genes, suggesting that they constitute independent transcription units. MiRNAs that are located in the boundary of annotated genes are mostly found in intronic regions. These miRNAs are likely to be transcribed as a part of the hosting gene.

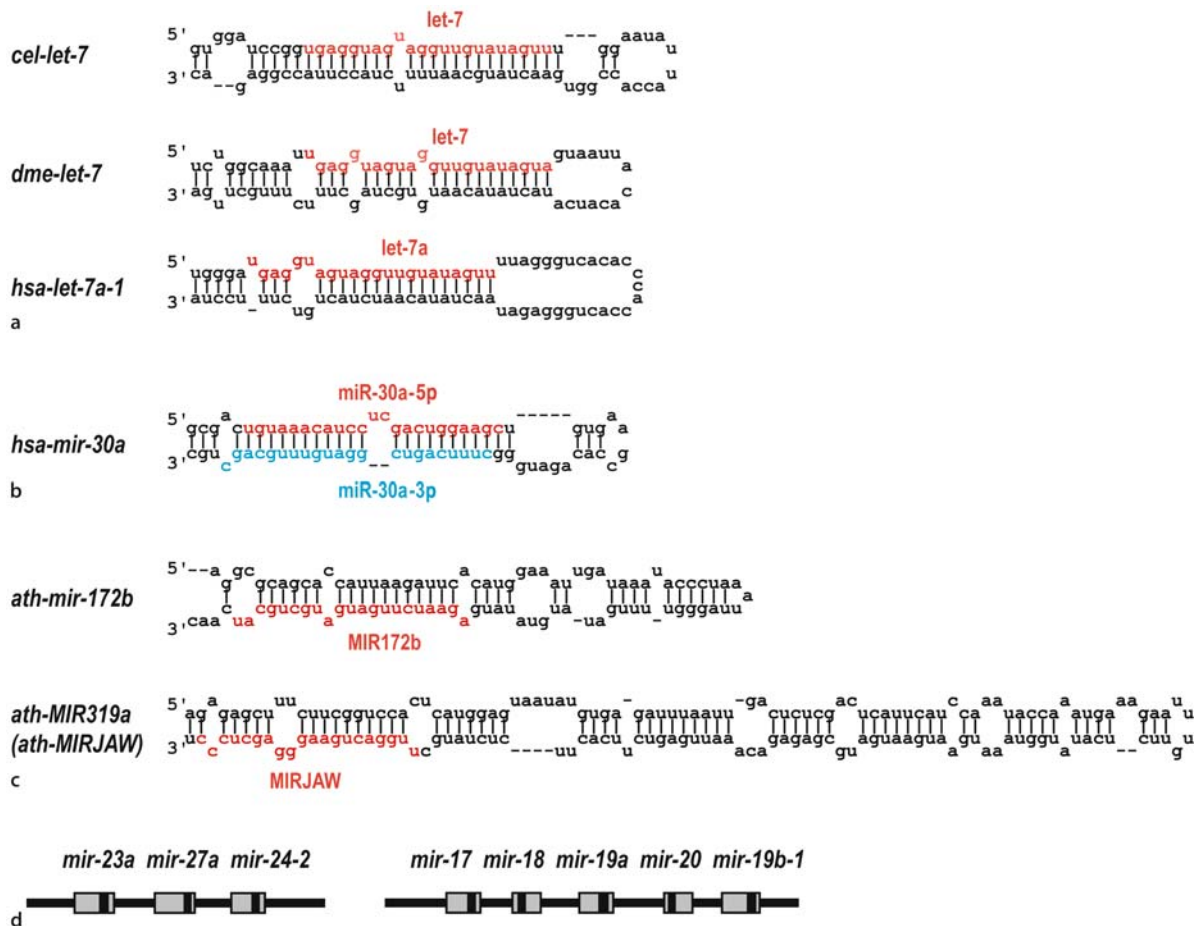
More than half of miRNA genes are in close vicinity to each other, forming operon-like gene clusters (Fig. 1D). They are expressed as polycistronic transcripts containing multiple hairpins. Because the miRNAs in a given cluster are transcribed together and often related to each other, this operon-like gene organization may provide a coordinated control over functionally related protein-coding genes.

The gene structure and the transcription mechanism of miRNA genes are still largely unknown. Detailed analysis of the gene structure (especially the promoter elements) and the transcription machinery will be important to elucidate the regulation mechanism of miRNA biogenesis.

Biogenesis

MiRNA genes are transcribed by RNA polymerase II (pol II) to generate primary transcripts (►*pri-miRNAs*) that are much longer than the 60~70 nt stem-loop (Fig. 2). Pri-miRNAs are first trimmed to release the hairpin intermediate (►*pre-miRNA*). This cleavage is executed by the RNase III type enzyme Drosha in the nucleus (4). Pre-miRNA then gets exported to the cytoplasm by exportin-5 (Exp5), which is a member of the Ran-dependent nuclear transport receptor family. Upon arrival in the cytoplasm, pre-miRNAs are subjected to the second processing by Dicer, the cytoplasmic RNase III type protein. Pre-miRNA is cleaved into the short-lived miRNA duplex, whose one strand is degraded by an unknown nuclease while the other strand remains as a mature miRNA.

MiRNA maturation appears to be well coordinated. Exportin 5 recognizes its cargo by specifically interacting with the short 3' overhang and the dsRNA (stem) structure, which are the characteristic features common in the products of Drosha-mediated processing. Dicer has a preference towards a terminus of dsRNA containing short 3' overhang so that it cleaves at ~22 nt away from the terminus of the pre-miRNA stem. Thus Drosha is responsible for producing optimal substrates for the downstream events, thereby pre-determining the



MicroRNA. Figure 1 Examples of miRNAs. Predicted stem-loops are shown here with the mature miRNA sequence (red or blue) and the flanking sequence. (a) *let-7* RNA genes in *C. elegans* (*cel-let-7*), *D. melanogaster* (*dme-let-7*) and human (*hsa-let-7a-1*) are highly conserved. The human genome has at least nine *let-7* subfamily genes while nematode worm and fruit fly have one each. (b) The human *mir-30a* gene expresses a pair of miRNAs from both arms, the more abundant miR-30a (red) and the less abundant miR-30a* (blue). (c) Examples of miRNAs from *Arabidopsis thaliana*, MIR172b and MIRJAW. (d) Examples of miRNA genes that are organized as clusters. Grey and black boxes indicate the regions corresponding to 60–70-nt precursors and mature miRNAs, respectively.

sequence of the mature miRNA and facilitating biogenesis.

Many miRNAs accumulate differentially in a spatially and/or temporally regulated manner. MiRNA gene regulation is likely to take place mainly at the transcriptional level like other genes, although some miRNAs may also be under post-transcriptional control such as in processing and export.

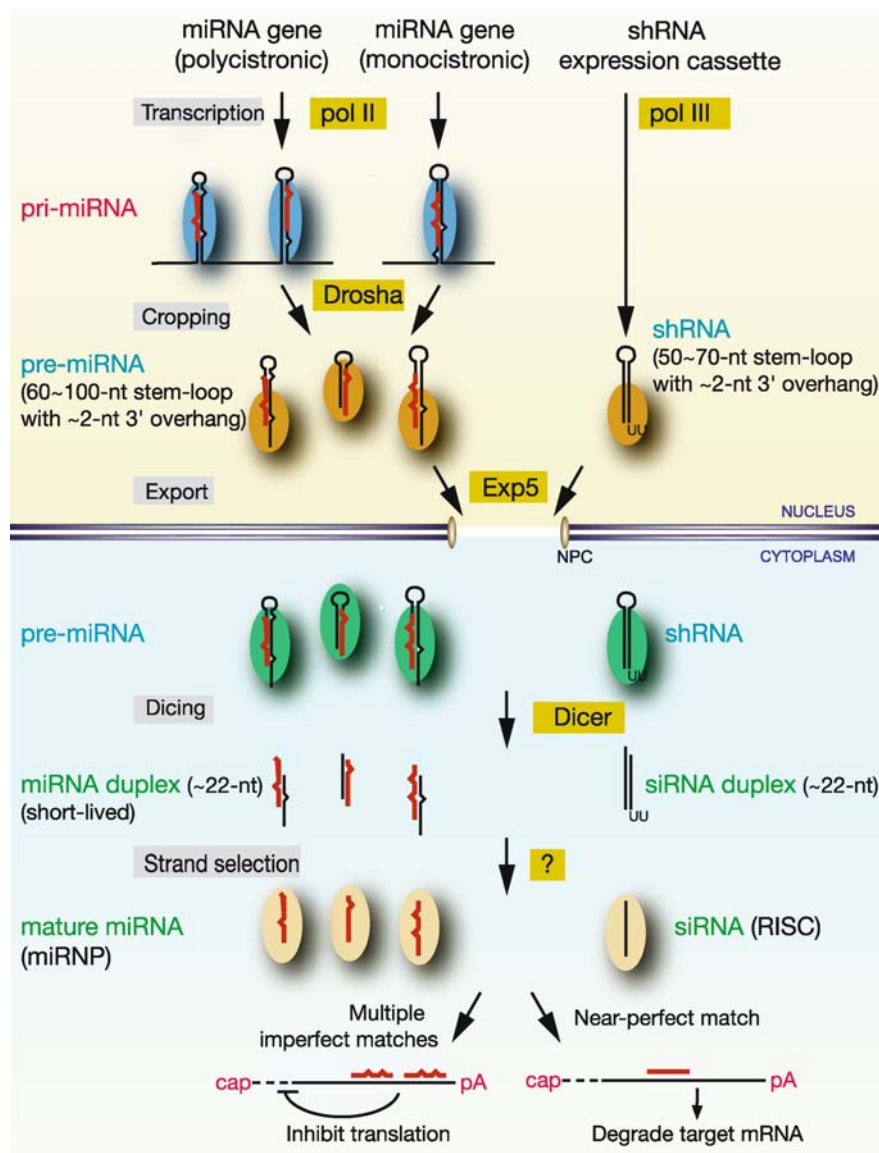
MiRNA biogenesis in plants is different from that in animals in many respects (2). MiRNA stem-loops in plants are highly variable in length and structure. The precursors of 60–70 nt are not found in plants. DCL1 (Dicer-like 1), HEN1 (a putative dsRNA methylase) and HYL1 (dsRNA RNA-binding protein) are required for miRNA accumulation. Plants have several Dicer homologues while no Drosha homologue has been

found. HASTY is a plant homologue of Exp5, and plays an important role in miRNA biogenesis.

Function

It is becoming clear that miRNAs have key roles in diverse regulatory pathways (2). The paradigm for the function of miRNAs has been provided by *lin-4* and *let-7* RNA. They act as post-transcriptional repressors of their target genes when bound to their specific sites in the 3' untranslated region (3'UTR) of the target mRNA. The level of target mRNA does not change, indicating that the inhibition occurs at the level of translation.

Out of hundreds of miRNAs, only a handful of miRNAs are known for their biological functions. For instance, *bantam* RNA from *Drosophila* suppresses



MicroRNA. Figure 2 Model for miRNA biogenesis and action mechanism. MiRNA genes are transcribed by RNA polymerase II to generate the primary transcripts, referred to as pri-miRNAs. The initiation step (cropping) by Drosha results in pre-miRNAs of ~70-nt, which are exported by Exp5. Upon export, Dicer participates in the second step (dicing) to produce miRNA duplexes. The duplex is separated and one strand is usually selected as mature miRNA while the other strand is degraded. Final products act as guide molecules in translational control or cleavage of certain mRNAs, depending on the degree of complementarity between the miRNA and the target mRNA. The question marks indicate unidentified biogenesis factors.

apoptosis and stimulates cell proliferation by inhibiting translation of *hid* mRNA. Being expressed in a temporal and tissue-specific manner, *bantam* RNA regulates tissue formation during development. Another nematode miRNA, *lsey-6* RNA, was recently identified in a gene screening for left/right asymmetry of neuronal chemoreceptor expression. *Lsey-6* RNA targets *cog-1* transcription factor. In mammals, miR-181 is involved in the control of hematopoiesis through

as yet unknown target(s). More recently, miR-196 miRNAs were shown to repress the expression of the *HOXB8* gene, which is a transcription factor important in developmental regulation. MiR-196 RNAs are the first examples of animal miRNAs that cause target mRNA cleavage rather than translational repression. Plant miRNAs generally show higher degrees of complementarity to the target mRNAs, resulting in target cleavage. For instance, MIRJAW binds and

guides the cleavage of TCP4 mRNA, which encodes a transcription factor regulating leaf development and embryonic patterning.

Interestingly, most of the known targets of plant miRNAs are transcription factors, particularly those involved in developmental regulation or cell differentiation. Functions of the targets of animal miRNAs seem to be more diverse than those of plant miRNAs. MiRNAs and their targets seem to constitute remarkably complex regulatory networks, since a single miRNA can bind to and regulate many different mRNA targets and, conversely, several different miRNAs can cooperatively control a single mRNA target.

Action Mechanism

The fate of the target mRNA when bound to miRNA – translational inhibition or mRNA cleavage – seemingly depends on the degree of complementarity between the miRNA and the target mRNA (Fig. 2). In either case, the highest complementarity (near perfect match) is found at positions 2–8 nt relative to the 5' end of miRNA. Base pairing at this position appears to be important for target recognition. For translational inhibition, multiple binding of miRNAs on a single mRNA induces a synergistic effect.

MiRNA is strongly associated with the Argonaute family protein and is a part of a ribonucleoprotein complex termed **miRNP** (also called miRISC). Argonaute family proteins are highly basic proteins of ~100 kD that contain two common domains, PAZ and PIWI domains. The PAZ domain, consisting of ~130 amino acids, is usually located at the center of the protein and interacts with the 3' overhang of dsRNA. The C-terminal PIWI domain containing ~300 amino acids is important for its interaction with Dicer. Although Argonaute proteins play a central role in various aspects of small RNA pathways, their biochemical functions are still unclear. The miRNP in humans also contains other proteins such as Gemin3 (a putative helicase), Gemin4, and **FMR1/FMRP** (the fragile X mental retardation protein).

The mechanism of translational repression remains elusive because polysome profiling on the target mRNA indicates that ribosomes still proceed on the mRNA as if they are being normally translated. The detailed mechanism of mRNA cleavage has been unravelled by the recent discovery of the nuclease that executes the cleavage of mRNA. The nuclease in humans is hAgo2, one of the Argonaute family proteins, whose PIWI domain serves as the catalytic domain.

Related Molecules and Pathways

There are other types of small RNAs that share common structural features with miRNA, the size (19~27 nt) and the end structure (the 5' phosphate

group and the 3' hydroxyl group). RNase III proteins play central roles in the generation of the small RNA family. Dicer homologues can be found in *S. pombe*, *C. elegans*, *Drosophila*, mammals and plants, suggesting that small RNA-mediated regulation is evolutionarily ancient and may have critical biological roles. Small RNAs regulate remarkably diverse pathways, including **RNA interference** (RNAi), heterochromatin formation and chromosome rearrangement.

RNAi refers to sequence-specific gene silencing induced by double-stranded RNA (dsRNA) (7). The small RNA acting in RNAi is called **small interfering RNA** (siRNA), which is defined as small RNA of ~22 nt generated from long dsRNAs of exogenous or endogenous origin. Dicer cleaves dsRNA to short (~22 nt) RNA duplexes that contain a 2 nt overhang at each 3' end. Each strand contains a 5' phosphate group and a 3' hydroxyl group. siRNA is incorporated into a nuclease complex called **RISC** (RNA-induced silencing complex, also called siRISC). The initial RISC containing a siRNA duplex is inactive until it is transformed into an active form, which involves loss of one strand of the duplex. The activated RISC then targets and cleaves mRNA that is complementary to the siRNA. Biochemical studies using *Drosophila* S2 cell extracts and human HeLa cell extracts revealed the presence of argonaute family proteins in the RISC. Argonaute-2 (AGO2) and eIF2c2 (hAGO2) were found in *Drosophila* and human, respectively. Depletion of the eIF2C proteins by RNAi showed that they are required for RNAi. Intriguingly, human let-7 was found to be a component of a RISC and can catalyze target cleavage if the mRNA has perfect complementarity to let-7. Thus, RISC and miRNP may simply be different names for the same complex, although they might represent distinct complexes with slightly different compositions and functions. At least in *Drosophila*, some of the protein factors are distinct in the two pathways. Dicer-1 is responsible for miRNA accumulation whereas Dicer-2 functions in siRNA production.

One of the most popular RNAi techniques is to express short hairpin RNA (shRNA) from which active siRNA is generated by Dicer (Fig. 2). ShRNAs are transcribed from an RNA polymerase III-dependent promoter such as the U6 snRNA gene promoter that allows termination at a stretch of four Us. Because the structure of shRNA resembles that of pre-miRNA, shRNA is exported and processed by miRNA maturation machinery. Therefore shRNAs can be regarded as artificial pre-miRNAs.

Endogenous siRNAs found in *S. pombe* are derived from centromeric repeat transcripts and direct heterochromatin formation in this region. The centromeric RNAs were found as part of a ribonucleoprotein complex called **RITS** (RNA-induced initiation of transcriptional gene silencing) that also contains Ago1,

the only argonaute homologue in fission yeast, together with Chp1 (a chromodomain-containing protein) and Tas3p. Mutations of these protein genes result in the loss of the heterochromatin state at the centromeric repeats. A similar mechanism seems to exist in *Drosophila*, where Aubergine and Piwi (Argonaute proteins) are required for heterochromatin formation.

Clinical Relevance

Some miRNAs have been implicated in tumorigenesis (3). Human miR-15a and miR-16-1 are located in the intronic region of a non-coding RNA gene named LEU2. This region at chromosome 13q14 is frequently deleted and/or the expression is down-regulated in the majority (~68%) of B cell chronic lymphocytic leukemia (B-CLL) cases. Some miRNA genes are located near breakpoint regions in cancer, including miR-142 whose promoter is thought to be placed upstream of the MYC gene leading to MYC over-expression. There may also be miRNAs acting as proto-oncogenes. For instance, miR-155 is part of a non-coding RNA gene, BIC, originally found as a target for retroviral insertion-induced B cell lymphomas in chicken. Human BIC RNA, which is a precursor of miR-155, is highly expressed in pediatric Burkitt lymphoma and in Reed-Sternberg cells of Hodgkin lymphoma. Identification of the molecular targets for these miRNAs will be necessary to provide direct links between miRNA genes and cancer.

FMR1 (also named FMRP) whose loss of function in human causes fragile X mental retardation syndrome was found associated with miRNA and Argonaute proteins in *Drosophila* and human (5), implying that the miRNA pathway may be involved in fragile X syndrome.

► *C. Elegans* as a Model Organism for Functional Genomics

References

1. Ambros V, Bartel B, Bartel DP et al (2003) A uniform system for microRNA annotation. *RNA* 9:277–279
2. Bartel DP (2004) MicroRNAs: genomics, biogenesis, mechanism, and function. *Cell* 116:281–297
3. Calin GA, Sevignani C, Dumitru CD et al (2004) Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. *Proc Natl Acad Sci USA* 101:2999–3004
4. Kim VN (2005) MicroRNA biogenesis: coordinated cropping and dicing. *Nat Rev Mol Cell Biol* 6:376–385
5. Lee RC, Feinbaum RL, Ambros V (1993) The *C. elegans* heterochronic gene *lin-4* encodes small RNAs with antisense complementarity to *lin-14*. *Cell* 75:843–854
6. Pfeffer S, Zavolan M, Grassl FA et al (2004) Identification of virus-encoded microRNAs. *Science* 304:734–736
7. Tijsterman M, Plasterk RH (2004) Dicers at RISC; the mechanism of RNAi. *Cell* 117:1–3

Microsatellite/Microsatellite Marker

Definition

Microsatellites are DNA sequences, typically shorter than 1000 bp, which are composed of tandemly repeated units 1–10 bp long (also 1–6 bp or 1–13 bp ranges are used). Microsatellites are also known as short tandem repeats (STR) or simple sequence repeats (SSRs). These repeat sequences occur frequently and randomly across the human genome. Microsatellites are highly polymorphic and make ideal markers for linkage analysis.

- Common (Multifactorial) Diseases
- COPD and Asthma Genetics
- Genomic Information and Cancer
- Hereditary Nonpolyposis Colorectal Cancer
- Large-Scale ENU Mutagenesis in Mice
- Mutagenesis Approaches in Medaka
- Mutagenesis Approaches in Yeast
- Repeat Expansion Diseases
- Repetitive DNA

Microscope Slide

M

Definition

A microscope slide is a piece of glass (25 mm x 75 mm) holding an object for a microscope.

- Protein Microarrays as Tool for Protein Profiling and Identification of Novel Protein-Protein Interactions

Microtiter Dishes

Definition

These ‘plates’ or ‘dishes’ with (originally) 8 rows of 12 wells labelled A-H and 1–12 were first popular in immunology applications (from there the term ‘titre’), but have more recently become the backbone of the logistics of maintaining gridded libraries. Variants with 4, 16 or even more times the number of wells, but still maintain the same 9 mm spacing of their every second or fourth well, have come into increasing use.

- YAC and PAC Maps

Microtubule

Definition

Microtubule refers to a hollow tube of 25 nm diameter formed by 13 protofilaments. Each protofilament consists of polymerised α and β tubulin heterodimers. Microtubules are polarized and have a plus and a minus end. They form a structural network within the cytoplasm of eukaryotic cells, build up cilia and flagella, and are involved in intracellular transport and migration.

- Cardiac Signaling: Cellular, Molecular and Clinical Aspects
- Cell Polarity
- Cytoskeleton
- Mitotic Spindle

Microtubule Associated Proteins

Definition

Microtubule associated proteins (MAPs) are molecules that modulate the properties of microtubules by interacting with microtubules and play a role in their nucleation, growth, shrinkage, stabilization and motion.

- Cytoskeleton
- Mitotic Spindle

Microtubule Motor Proteins

- Motor Proteins

Microtubule-Organizing Center

Definition

Microtubule-organizing center (MTOC) describes a region in a cell from which microtubules grow.

- Centrosome
- Rho, Rac, Cdc42

Microvilli

EVELYNE FRIEDERICH¹, DANIEL LOUVARD²

¹Laboratory of Molecular Biology, Genetic Analysis and Modelling, Public Research Center for Health, Luxembourg

²Subcellular structure and cellular dynamics, UMR 144 CNRS/IC, Institute Curie – Research Division, Paris, France

evelyne.friederich@crp-sante.healthnet.lu

daniel.louvard@curie.fr

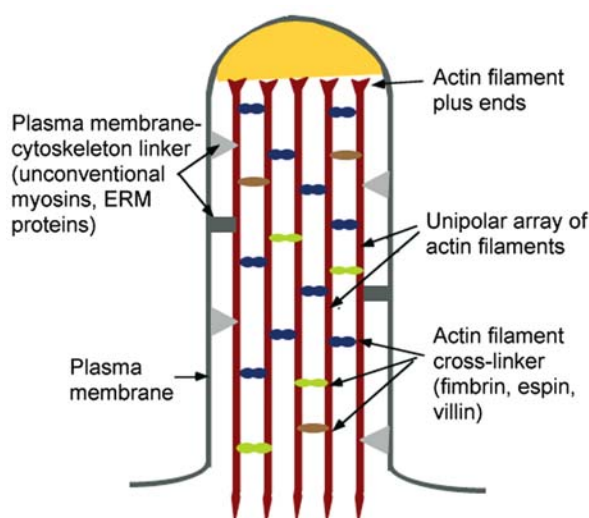
Definition

► Microvilli are finger-shaped ► plasma membrane protrusions that are found at the surface of a large variety of cell types but are most numerous and elaborated on simple epithelial, for example intestinal mucosa and the ► epithelium of the kidney proximal tubule. In contrast to other plasma membrane extensions, microvilli are not attached to the extra-cellular matrix but cover free cell surfaces that are, in many cases, facing a lumen where exchange occurs. These protrusions were termed microvilli by analogy with the intestinal villi because they increase, as do those, the absorptive surface of the gut. Microvilli of cells of the placental syncytiotrophoblast contribute to the bi-directional foeto-maternal exchanges. However, the function of microvilli is not restricted to absorption. Microvilli of the apical surface of hepatocytes participate in secretion. ► Stereocilia are surface projections present on the surface of the hair cells of the sensory neuroepithelium of the ► inner ear that participate in the transformation of a mechanical stimulation into an electrical auditory signal. Similarly, the photoreceptor cells of the eye also bear modified microvilli on their apical surface. Microvilli covering the surface of circulating ► immune cells, such as leucocytes, mediate adhesion to the endothelium, a key step in the activation of migration of immune cells to the sites of inflammation. Microvilli-like structures are also induced by pathogens and are linked to their propagation cycle within the infected organism.

Characteristics

Common and Function-Specific Features of Microvilli

Microvilli contribute to very diverse biological functions, yet they are all constructed following the same principle. The organisation of a prototype microvillus was established on the basis of the information gained from ultrastructural and biochemical analysis of the relatively stable microvilli of ► absorptive intestinal cells (1). The plasma membrane is supported by a



Microvilli. Figure 1 Schematic representation of a prototype microvillus. The major cytoskeleton components are indicated.

bundle of a variable number of cross-linked proteinaceous filaments, made of **actin**, a 43 kD monomeric protein (Fig. 1). Actin has an intrinsic asymmetry and self-assembles into a polar filament with a fast- and a slow-growing end, also termed the plus- and minus-ends. The fast growing ends of microvillar actin filaments are oriented towards the tip of the microvillus and addition of monomers at this end is required for membrane extension. In some cases, the cross-linked actin filaments exhibit a hexagonal, paracrystalline organisation, the structural constraint of which dictates the binding of other proteins. Linker proteins tethering the F-actin core bundle to the inner face of the plasma membrane increase the stiffness of the microvillus and play an important role in the maintenance of its structural and functional integrity and in the integration of signals. The microvillar plasma membrane is a specialised domain containing lipid rafts that are lipid-based sorting platforms for proteins involved in **endocytosis**, **signalling** and cell adhesion.

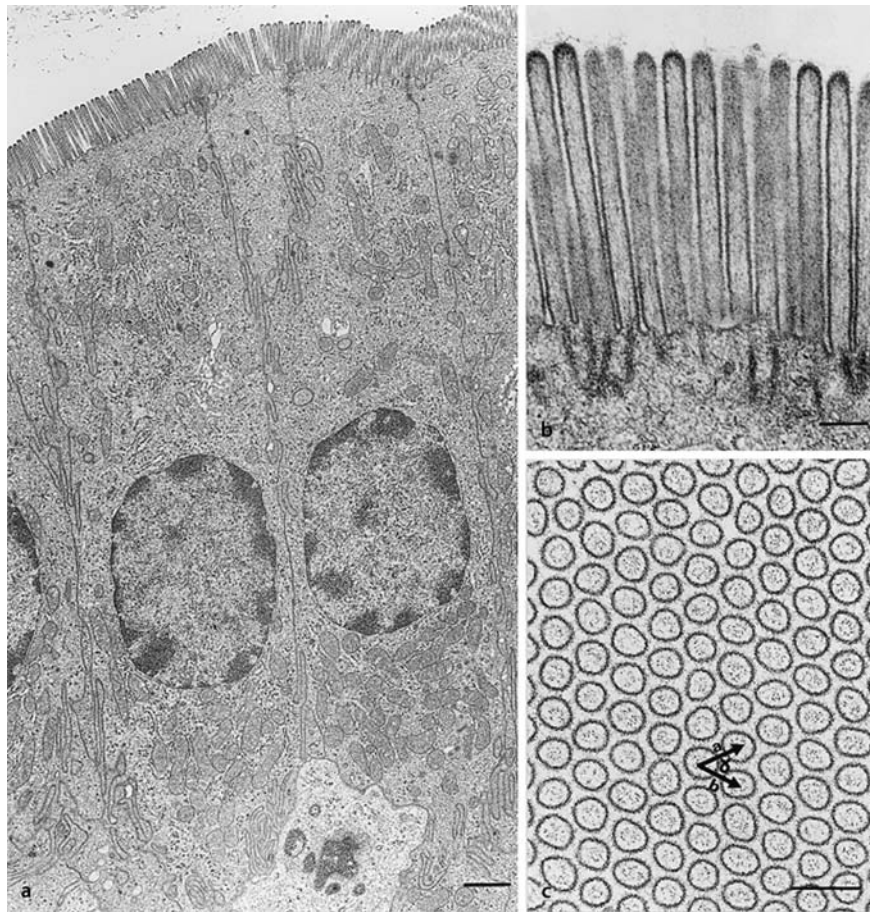
Comparison of different types of microvilli reveals that, in addition to the proteins involved in specific, cell type-dependent functions, they also contain a common repertoire of proteins. Highly conserved throughout evolution, these proteins frequently control the assembly of microvilli or are required for their structural integrity. Three representative types of microvilli with distinct biological functions will be considered here to illustrate the common and specific features of these structures: 1) microvilli of absorptive epithelial cells of the gut, 2) stereocilia of the inner ear hair cells and 3) microvilli of immune cells.

Molecular Interactions

Microvilli permanently covering the apical surface (**Apical Membrane**) of polarised absorptive epithelial cells or the stereocilia of the sensory neuroepithelium of the inner ear are the basic units of two highly specialised structures (Fig. 2, Table 1).

The well-characterised microvilli of the absorptive epithelial cells of the small intestine are densely packed and all of similar size (1–2 μm), forming a stable structure named the “**brush border**” that is specialised for nutrient up-take (2). The actin filament core bundle comprises 15–20 actin filaments plunging into a cap of dense material present at the tip of the microvillus. The major cytoskeletal components of the brush border microvilli are the actin filament bundling proteins **villin**, fimbrin and espin. These proteins harbour two closely spaced F-actin binding sites endowing them with the capacity to cross-link actin filaments into a tight, unipolar bundle. The non-conventional brush border myosin I, a Ca^{2+} -calmodulin-complexed actin-based ATPase **motor protein**, and ezrin, an **ERM (ezrin radixin moesin)** family member, link the actin core bundle to the plasma membrane. Ezrin and other family members bind to F-actin as well as to membrane proteins including adhesion and signalling molecules. Fodrin, a tetrameric actin cross-linking protein, anchors the actin core bundle “roots” to a dense meshwork of filaments, named the terminal web, further stabilising the microvillus. Related to its function in absorption, the plasma membrane of intestinal microvilli contains proteins involved in sugar transport and enzymes implicated in the degradation of nutrients like dipeptidyl peptidase IV or in ion transport. Intestinal microvilli also contain signalling molecules, including Ca^{2+} -buffering proteins like calmodulin, enzymes involved in phosphoinositide metabolism, serine/threonine or tyrosine protein kinases and **small GTPases** like ras. The microvillar plasma membrane is enriched in cholesterol-glycosphingolipid-containing microdomains, the “lipid rafts” that have been assigned a role in membrane trafficking. Playing an important role in the biology of brush border microvilli, these microdomains contain proteins involved in nutrient up-take, including ion channels, enzymes, immunoglobulin receptors and signalling molecules.

Stereocilia found at the apical surface of hair cells are modified microvilli that are arranged in ranks of increasing height (up to 120 μm) and connected to each other with extracellular linkages. Linked stereocilia, 30–300 in number, form a functional unit, the hair cell bundle, that is a primary detector of sound vibrations. Despite their distinct biological roles, intestinal microvilli and stereocilia share common structural and biochemical features. The stereociliary membrane is supported by a hexagonal, unipolar, actin



Microvilli. Figure 2 Transmission electron microscopy images of the intestinal cell brush border. (a) The apical surface of polarised epithelial cells of the small intestine exhibit finger-like plasma membrane extensions, the microvilli, forming the brush border. (b) Higher magnification of the brush border visualising the microvillar actin cytoskeleton. (c) Transverse sections of the microvilli. Bar, 1 μm (a) and 0.2 μm (b and c).

filament bundle with which the actin filament bundling proteins fimbrin and espin are associated. Radixin, an ERM protein, is detected at the base of the stereociliary bundle. The rootlets of the core bundle are tethered *via* the cross-linking proteins fodrin and alpha-actinin to a dense meshwork of filaments, named the cuticular plate. Comparison of brush border microvilli and stereocilia also reveals differences. Unlike brush border microvilli, stereocilia contain several hundreds of actin filaments and exhibit a staircase organisation, implying a specific control mechanism of actin filament length. The inter-stereocilia proteinaceous linkages play an important role in mechanotransduction. Mechanically gated channels that are part of the molecular mechanotransduction machinery are located at the tip of short and on the sides of long stereocilia, at interconnection sites. When stereocilia deflect during sound stimulation, the elastic linkages stretch, thereby opening the

transduction channels. In keeping with their function, stereocilia harbour a specific set of proteins, including adhesion molecules like cadherin 23 and proteins controlling stereocilium length or adjusting the position of the linkages. ► **Unconventional myosins** may contribute to these regulatory mechanisms by controlling membrane traffic and membrane protein distribution. Among these, myosins VI, VIIa and XVa are the best studied because of their implication in ► **genetic hearing diseases** (3). They distribute with the cuticular plaque (e.g. myosin VI), along the stereociliary core bundles (e.g. I beta, VIIa) or are restricted to the tip (myosin XVa), suggesting that they have distinct functions in hair cell assembly and maintenance.

Unlike microvilli of the brush border and stereocilia, those covering the entire surface of immune cells are non-permanent structures whose formation depends on the activation state of the cell. The microvilli of

Microvilli. Table 1 Comparison of the properties of representative types of microvilli

Major associated diseases						
Origin of tissue	Cell type	Major cytoskeletal proteins	Major microvilli-function-related plasma membrane proteins	Function(s) of microvilli	Pathology/ Syndrome	Causal factor(s)
Absorptive endoderm	Enterocyte	β and γ actin	Enzymes involved in nutrient degradation (e.g sucrose-isomaltase, dipeptidyl peptidase IV) Ion channels Ig receptors	Absorption Immuno- protection	Diarrhoea Breakdown of microvilli	Enteropathogenic E. coli bacteria (EPEC) Salmonella Entamoeba histolytica
		Villin I-fimbrin Unconventional myosin I-calmodulin Espin ERM protein (Ezrin)				
Endoderm	Hepatocyte	β and γ actin Villin ERM protein (Radixin)	Phospholipid transporter Bile-salt excretory pump	Secretion	Biliary atresia	Impaired villin expression
Sensory neuroepithelium	Hair cell	β and γ actin I-fimbrin Unconventional myosins Espin ERM protein (Radixin)	Mechanical gated transduction channels Adhesion molecules (e.g Cadherin 23)	Mechano-transduction of signals	Deafness/ blindness USHER DFNB3 DFN31	Myosin VIIa, Cadherin 23 Myosin XV Whirlin
Hematopoietic	T cell B cell Leucocyte	β and γ actin L-fimbrin ERM protein (moesin) WASP	Adhesion molecules (integrins, selectins) ion channels	Cell-cell adhesion, signalling	Impaired immune response Wiskott Aldrich syndrome	Wiskott Aldrich syndrome protein (WASP)

lymphocytes and leucocytes mediate contact with the endothelium under flow, also named rolling adhesion, the first event of a cascade leading to their emigration from circulation to tissue in response to inflammation (4). Microvilli also mediate homotypic and heterotypic interactions of lymphocytes. Although less information is available on the structural and biochemical properties of these transient structures when compared to brush border microvilli or to stereocilia, they contain, as do those, a polarised actin bundle that is linked *via* the ERM family member moesin to the plasma membrane. The plasma membrane of immune cells' microvilli is a domain specialised in adhesion and signalling. Segregation of specific adhesion molecules to microvilli or planar membrane areas is thought to depend on lipid rafts. In addition, the repertoire of adhesion molecules of the microvillar membrane is defined by cell type specific regulation of gene expression. While selectin type adhesion molecules are present at the tip of leucocyte microvilli, $\alpha 4 \beta 7$ (CD 49d) ►integrins mediate lymphocyte attachment under flow.

Regulatory Mechanisms

Assembly Mechanisms of Microvilli

The assembly of all microvilli is regulated in time and space by a specific set of associated proteins. However, the time course of the assembly process varies considerably when comparing different types of microvilli. For example, while microvilli of epithelial cells form progressively during embryonic development and undergo terminal differentiation in adults, activation of the immune response requires rapid growth of microvilli on the surface of circulating blood cells. Nevertheless, in both cases the important steps of the cascade of events leading to the morphogenesis of microvilli include

1. temporally and spatially regulated nucleation of a rudimentary actin ►cytoskeleton at the inner face of the plasma membrane,
2. elongation of actin filaments and assembly of the microvillar actin core bundle,
3. regulation of the length and width of microvilli
4. spatial restriction of plasma membrane proteins and lipids.

Although signals triggering the assembly of microvilli are diverse and depend on the cellular background, common molecular mechanisms underlie microvillar morphogenesis. Indirect evidence indicates that Rho GTPase family members, molecular switches regulating actin assembly and disassembly through effector proteins, may control the assembly of the microvillar cytoskeleton. In the same way, WASp (►Wiskott Aldrich syndrome protein), a cdc42 and rac GTPase effector protein, is a plasma membrane-cytoskeleton

linker. Initially identified as the gene causing the Wiskott Aldrich syndrome, a rare ►genetic X-linked disease associated with immune cell disorders, WASp appeared to be required for the morphogenesis of microvilli of immune cells (5). WASp deficient lymphocytes exhibit fewer and shorter microvilli when compared with normal cells. Upon inactivation of intramolecular autoinhibition of ►WASp by cdc42 or rac, this scaffolding protein localises to the plasma membrane and recruits the Arp2/3 complex (actin related protein). Made up of seven subunits, this complex binds to the sides of actin filaments and initiates the assembly of a new filament with a free barbed end from which it will grow, yielding a branched filament network. Whether WASp initiates the assembly of the microvillar cytoskeleton is still a matter of debate. In support of such a mechanism, WASp was shown to be required for the temporally correct assembly of rhabdomeres, microvilli-like structures of the insect eye. However, WASp deficient lymphocytes are not totally devoid of surface projections, suggesting that other WASp family members, like WAVE, or alternatively another so far unknown nucleating protein initiate the assembly of the microvillar actin cytoskeleton. Evidence has been given that ERM family members may also contribute significantly to microvillar morphogenesis. Activation of the EGF receptor results in tyrosine phosphorylation of ezrin and concomitant growth of microvilli in epithelial cells. Phosphorylation and PIP2 binding release the intramolecular autoinhibition of ERM proteins, exposing their plasma membrane and actin binding sites. Inhibition of the expression or function of ERM family members severely impairs the formation of cell surface projections. Consistent with a function in microvillar regulation, inactivation of ERM actin-membrane linkers by dephosphorylation precedes microvillar breakdown in immune cells. In addition to the stabilisation of actin filaments, ERM proteins may also control the spatial distribution of plasma membrane proteins. Bundling proteins arrange the elongating actin filaments into a unidirectional array that grow from their free, membrane-oriented ends, generating the force required for plasma membrane protrusion. Bundling proteins may not only align actin filaments into unipolar arrays, but also control elongation. Indeed, forced expression of the bundling proteins villin, fimbrin or espin causes the growth of microvilli. Several, non-exclusive mechanisms may account for this activity. Binding of bundling proteins to the sides of actin filaments may interfere with Arp2/3-dependent nucleation of a dendritic network and contribute to its conversion into a parallel array of actin filaments, as shown for fascin, a filopodia-associated bundling protein. Alternatively, they may control the length of

microvilli by regulating the turnover of actin in the microvillar core bundle, as suggested in a recent study with espin. In addition, non-conventional myosins may control the length of microvilli by regulating membrane traffic. However, the precise mechanism of action of these motor proteins remains to be elucidated. Recently, activation of the serine/threonine kinase LKB1, a human homologue of the *C. elegans* Par-4 polarity proteins, by the adaptor protein STRAD was shown to induce the polarisation of single intestinal epithelial cells and the assembly of a brush-border like structure (6). In the future, the identification of the downstream effectors of LKB1 will contribute to unravelling the assembly mechanism of brush border microvilli.

Dynamics of Microvilli and Remodelling

Mature microvilli are not static but dynamic structures, capable of shrinking or growing in response to specific stimuli. This is most obvious for immune cell microvilli that assemble or disassemble depending on the activation state. However, permanent microvilli of the intestinal brush border or of stereocilia are also dynamic structures. Intestinal microvilli shorten in physiological situations like fasting or in pathological situations, such as infections of intestinal cells by viral or bacterial pathogens. Similarly, exaggerated bending can damage terminally differentiated stereocilia so that they need to be repaired. Actin incorporates at the tip of mature microvilli, showing that their actin cytoskeleton is continuously renewed. Thus, the above-described mechanisms of molecular regulation are not only for microvillar assembly but also contribute to the maintenance of their integrity, their renewal or remodelling. This duality is illustrated by villin, a multifunctional, Ca^{+2} -regulated actin-binding protein that bundles actin filaments in the absence of Ca^{+2} and depolymerises actin filaments in a Ca^{2+} -dependent manner. Consistent with its *in vitro* properties, villin might have a role in both the assembly and breakdown of microvilli in epithelial cells.

Disorders of Microvilli and Pathology

Although exhibiting very different clinical features, microvilli-associated pathologies are frequently caused by mutations in genes encoding cytoskeleton components or adhesion molecules. The close relationship between the actin cytoskeleton, cell signalling and communication is illustrated by the X-linked Wiskott-Aldrich syndrome, a rare recessive disorder caused by mutations in the *WAS* gene yielding severe defects in platelets and immune cells (5). The ►[scaffolding protein](#) WASp not only regulates actin polymerisation in microvilli but also controls T-cell receptor complex formation. As a consequence, T-lymphocytes of WAS

patients exhibit abnormal short microvilli, impaired interaction with B-cells and poor response to protein antigens. Hearing impairment is frequently caused by mutations in genes encoding unconventional myosins (3) and actin-associated proteins like espin that regulate stereociliary length and shape. For example, myosin VIIA seems to play a role in the formation of the linkages connecting neighbouring stereocilia (USH-ER1B syndrome). Myosin VIIA was shown to interact with verzinin, a protein of cadherin-catenin ►[adherens junctions](#) and with signalling molecules. Notably, mutations in the myosin VIIA gene also lead to blindness, demonstrating the close structural and functional relationship of these microvilli-like structures. Recently, mutations in the whirlin gene were shown to cause human autosomal recessive deafness (DFNB31). This ►[PDZ domain](#) protein might organize a group of proteins into a functional complex required for stereociliary elongation. In addition, mutations in cell-cell contact molecules, like cadherin 23, also impair hair cell function by causing spraying of stereocilia in hair cell bundles. Pathologies linked to microvilli dysfunction are also found in endoderm-derived cells like hepatocytes and enterocytes. ►[Biliary atresia](#), a clinical disorder resulting in progressive liver cholestasis, correlates with the disorganisation of bile canalicular microvilli and the absence of the actin bundling protein villin (7).

References

1. Mooseker MS, Tilney LG (1975) Organization of an actin filament-membrane complex. Filament polarity and membrane attachment in the microvilli of intestinal epithelial cells. *J Cell Biol* 67:725–743
2. Bretscher A, Weber K (1978) Localization of actin and microfilament-associated proteins in the microvilli and terminal web of the intestinal brush border by immunofluorescence microscopy. *J Cell Biol* 79:839–845
3. Gibson F, Walsh J, Mburu P et al (1995) A type VII myosin encoded by the mouse deafness gene shaker-1. *Nature* 374:62–64
4. Bevilacqua MP, Stengelin S, Gimbrone MA Jr et al (1989) Endothelial leukocyte adhesion molecule 1: an inducible receptor for neutrophils related to complement regulatory proteins and lectins. *Science* 243:1160–1165
5. Kenney D, Cairns L, Remold-O'Donnell E et al (1986) Morphological abnormalities in the lymphocytes of patients with the Wiskott-Aldrich syndrome. *Blood* 68:1329–1332
6. Baas AF, Kuipers J, van der Wel NN et al (2004) Complete polarization of single intestinal epithelial cells upon activation of LKB1 by STRAD. *Cell* 116:457–466
7. Phillips MJ, Azuma T, Meredith SL et al (2003) Abnormalities in villin gene expression and canalicular microvillus structure in progressive cholestatic liver disease of childhood. *Lancet* 362:1112–1119

MIDAS

Definition

MIDAS stands for Metal Ion-Dependent Adhesion Site. It is formed by an array of key residues in some integrin head domains which coordinate a divalent cation during ligand binding.

► [Integrin Signalling](#)

MIF

Definition

This term has been used in several different ways:

- a) ► <http://www.copewithcytokines.de/cope.cgi> Mesoderm inducing factor: factors inducing the formation of the mesoderm of amphibian embryos. These are members of the TGF- β , FGF, and Wnt families of proteins.
- b) Macrophage migration inhibitory factor is a proinflammatory cytokine that was originally defined on the basis of inhibition of emigration of mononuclear cells from capillaries. MIF has been associated with various types of adenocarcinoma.
- c) Müllerian-duct inhibiting factor refers to a factor that suppresses the further development and maintenance of the Müllerian-ducts in embryos.

Migration

Definition

Migration corresponds to the movement of cells on a certain substrate, in response to a certain chemotactic or chemokinetic signal. During migration, cell surface receptors are continuously recycled from the rear to the leading edge of the cell as result of dynamic cytoskeleton rearrangements.

► [Cell Migration](#)

► [Cancer Invasion and Metastasis: Cellular, Molecular and Clinical Aspects](#)

Milk-Alkali Syndrome

Definition

Milk-alkali syndrome refers to a chronic disorder that is characterized by deposition of calcium in many sites, especially in the kidney. It is induced by ingestion of large quantities of calcium and alkali, formerly used to treat peptic ulcers.

► [Hyper- and Hypoparathyroidism](#)

Miller-Dieker Syndrome

Definition

Miller-Dieker syndrome (MDS) is a contiguous gene syndrome (1:20,000 newborns) that is caused by loss of one allele of PFAH1B1 (platelet activating factor acetylhydrolase, isoform Ib alpha subunit), and neighboring genes at chromosome 17p13.3. Features include lissencephaly and a very mild facial dysmorphism. PFAH1B1 point mutations result in a similar clinical outcome (isolated lissencephaly sequence) but without the facial dysmorphism of MDS.

► [Microdeletion Syndromes](#)

Miller Planes

Definition

Planes parallel to the faces of the unit cells and all other diagonal planes formed using the vertices of combinations of unit cells. These sets of equally spaced parallel planes correspond to the Miller indices of the individual diffracted spots. The planes parallel to the faces of the unit cell correspond to the basic Miller indices (100), (010) and (001); higher indices correspond to the more narrowly spaced diagonal planes. In general, the Miller indices (h,k,l) indicate the number of planes crossing the a, b and c axes of the unit cell, respectively. The reflection F(h,k,l) can be measured when the orientation of the Miller planes relative to the incoming X-ray beam fulfills Bragg's law.

► [X-Ray Crystallography—Basic Principles](#)

MIM

Definition

MIM stands for Mendelian Inheritance in Man, and is a catalogue of human genes and genetic disorders (authored and edited by Dr. Victor A. McKusick and colleagues at Johns Hopkins, Johns Hopkins University Press, Baltimore, 12th ed., 1998; Online accessible as OMIM: www.ncbi.nlm.nih.gov/Omim/).

- Bloom Syndrome
- OMIM

MIN/MSI

Definition

Mutations in genes that repair damaged DNA cause genomic regions called microsatellites to get longer or shorter, a phenomenon that is called microsatellite instability (MIN, also termed MSI). Microsatellites usually consist of repeated mono, di-, or trinucleotide sequences, which are prone to errors during DNA replication. MSI occurs in some but not all sporadic tumors, and can lead to frameshift mutations of genes when present in the coding sequence.

- Colorectal Cancer

Mineralocorticoid Receptors

- Glucocorticoid/Mineralocorticoid Receptors

Mineralocorticoid Resistance

- Glucocorticoid/Mineralocorticoid Resistance

Mineralocorticoids

Definition

Mineralocorticoids are steroid hormones that are synthesized in the adrenal gland (aldosterone, deoxycorticosterone). The main function is the raising of reabsorption of Na^+ -ions in the distal tubulus of kidneys.

- Mendelian Forms of Human Hypertension and Mechanisms of Disease

Minisatellites

Definition

Minisatellites are tandemly repeated arrays of DNA sequences, with lengths of the basic sequence motifs between 10 to (several) 100 base pairs, which are longer than in microsatellites. Minisatellites span from 1 kb to 100 kb. They are found dispersed throughout the genome and exhibit comparatively high to extremely high mutation rates in eukaryotes. Also known as variable number of tandem repeats (VNTRs)

- Repetitive DNA
- Repeat Expansion Diseases

Minor Groove

- Major and Minor Groove

miRNA

- MicroRNA

miRNP

MicroRNA-protein complex

- Micro RNAs

Misexpression

- Overexpression/Misexpression

Misfolded Proteins

Definition

In misfolded proteins, the protein conformation(s) is disturbed or lost and, therefore, the protein lacks functionality.

- Repeat Expansion Diseases

Mismatch (Nucleotide)

Definition

Mismatch (nucleotide) refers to a particular position in an oligonucleotide reporter molecule on the array, which differs from the nucleotide in a matching position (CG or AT) in the hybridising sample molecule.

- DNA Recombination
- Microarray Data Analysis

Mismatch Repair

Definition

Mismatch repair describes a DNA repair pathway that can correct single mismatches shortly after S-phase in *E. coli*. The mismatch repair pathway in human cells can additionally correct nucleotide expansion mutations.

- DNA-Repair Mechanisms
- Hereditary Nonpolyposis Colorectal Cancer
- Mutagenesis Approaches in Yeast

Missense

Definition

Missense mutation is a genetic mutation that results in the substitution or elimination of one amino acid for a different amino acid in the protein encoded by the affected gene.

- Hereditary Spastic Paraplegias
- Heritable Skin Disorders
- Hereditary Nonpolyposis Colorectal Cancer
- Limb Development
- Protein Interaction Analysis: Suppressor Hunting
- Splicing

Mitochondria – Biogenesis and Structural Organization

ANDREAS S. REICHERT, WALTER NEUPERT
Department of Physiological Chemistry, Ludwig-Maximilians-University Munich, Munich, Germany
andreas.reichert@med.uni-muenchen.de
neupert@med.uni-muenchen.de

Definition

► **Mitochondria** are cellular compartments of eukaryotes that are enclosed by two membranes. They are often called the powerhouses of the cell because of their central role in energy transduction. Fundamental metabolic processes such as oxidative phosphorylation, iron-sulfur cluster biogenesis, beta-oxidation of fatty acids, the citric acid cycle, part of the urea cycle and heme synthesis take place in this organelle. Mitochondria further play important roles in apoptosis, ageing, signal transduction and cancer.

They are bounded by two membranes, the outer membrane and the inner membrane. The aqueous compartment enclosed by the inner membrane is called the ► **matrix** and the space between the two membranes is called the ► **intermembrane space**. The inner membrane can be further divided into the ► **inner boundary membrane** and the ► **cristae** membrane. The former is in continuous and close apposition to the outer membrane; together with the outer membrane it forms a kind of envelope of the mitochondria. Stable and dynamic ► **contact sites** between these two

membranes are necessary for a number of central mitochondrial functions, such as the channeling of metabolites, coordinated fusion and fission of mitochondria and protein transport into mitochondria (1). Cristae are invaginations of the inner boundary membrane that protrude into the matrix space. They harbor the complexes required for respiration and ATP synthesis. Cristae are connected to the inner boundary membrane by so-called ►**cristae junctions** (2). Rather than being wide openings, as often shown in biochemistry textbooks, cristae junctions are narrow tubular structures of varying length, which potentially act as diffusion barriers between the inner boundary and the cristae membrane.

Characteristics

Biogenesis of Mitochondria

Mitochondria contain their own genome (mtDNA), which, however, encodes only 8–13 proteins (yeast and human, respectively). The majority of mtDNA encoded proteins are components of oxidative phosphorylation, in particular of respiratory complexes I, III and IV, and of the F_1F_0 -ATP synthase. Approximately 99% of all mitochondrial proteins are encoded in the nuclear genome. Consequently, most of the proteins are imported into mitochondria after being synthesized in the cytosol (3). The mitochondrial proteomes of various species and organs including the human heart, rice, baker's yeast and the inner membrane of mouse liver mitochondria have been analyzed. In yeast 750 proteins were identified, of which a large proportion (25%) is involved in the maintenance and expression of the mitochondrial genome (4). Thus, nearly 200 proteins are involved in the synthesis of only eight mitochondrial proteins. The function of another 200 proteins is still unknown. Mitochondria contain translocases responsible for import and sorting of preproteins into the different mitochondrial subcompartments. Translocases control the specificity of the sorting processes and facilitate the vectorial movement of polypeptides across and into mitochondrial membranes.

Morphology of Mitochondria

The macroscopic appearance of mitochondria in cells as observed by fluorescence microscopy depends on the organism and the cell type as well as the metabolic state. Mitochondria exhibit a great variety of shapes, ranging from spherical or threadlike organelles to branched tubular networks (5). The reticular tubular network of mitochondria observed in different cells from yeast to human is highly dynamic and requires a balance of constantly occurring fusion and fission processes (reviewed in 6). This dynamic nature of mitochondrial morphology is essential for the inheritance of mitochondrial DNA and for spermatogenesis. Furthermore,

fragmentation of mitochondria is required for cells to undergo apoptosis. In apoptotic cells, release of cytochrome c from the intermembrane space is further accompanied by remodeling of the cristae membrane. Moreover, complementation of mutations in mitochondrial DNA within cells depends on the capability of mitochondria to fuse with each other. Thus, fusion of mitochondria seems to be important as a defense mechanism against oxidative damage and ageing (6). Taken together, the biogenesis as well as the dynamic nature of the structural organization of mitochondria is crucial for maintaining their functionality within the cell.

Molecular Interactions

Protein Import into Mitochondria

Correct targeting of preproteins requires the presence of signal sequences. A typical mitochondrial targeting sequence is a positively charged stretch of about 15–80 amino acid residues located at the N-terminus which is able to form an amphipathic helix. Internal sorting signals also exist such as those found in members of the mitochondrial solute carrier family. The ►**TOM complex** (translocase of the outer membrane) recognizes preproteins on the cytosolic side of the outer membrane of mitochondria and mediates their transfer into the intermembrane space or into the outer membrane. A recently discovered complex in this membrane, the ►**TOB complex**, is essential for the integration of β -barrel proteins into the outer membrane after they have passed the TOM complex. The inner membrane contains three known translocases: the TIM23, the TIM22, and the ►**OXA1 complex**. The ►**TIM23 complex** transports preproteins containing typical N-terminal mitochondrial targeting sequences into the matrix or into the inner membrane. The ►**TIM22 complex** facilitates the import of hydrophobic inner membrane proteins that contain internal signal sequences such as members of the solute carrier family. Other proteins are first imported into the matrix *via* the TIM23 complex before they are exported and inserted into the inner membrane by the OXA1 complex. The latter complex also mediates the insertion of mitochondrially-encoded proteins. The mitochondrial translocases act in a coordinated fashion. This cooperation is instrumental in correct sorting and maintenance of mitochondrial identity and thereby in the functionality of this highly complex organelle.

Fusion and Fission of Mitochondria

Many of the molecular mechanisms underlying fusion and fission of mitochondria were initially understood from studies in baker's yeast. For example, deletion of Mgm1 leads to extensive fragmentation of mitochondria, loss of mitochondrial DNA and abolishment of mitochondrial fusion. Mgm1 is an intermembrane

space localized dynamin-like protein. Dynamins are large GTPases, which are normally known as key players in vesicle formation during endocytosis. Mgm1 was proposed to be involved in modulating the cristae membrane morphology and to be essential for the fusion of mitochondria (6). Fusion requires at least two more proteins in yeast, Fzo1 and Ugo1. Both are outer membrane proteins and were shown to interact with Mgm1. Thus, Mgm1, Fzo1 and Ugo1 are likely to be part of dynamic contact sites in mitochondria that mediate fusion processes between mitochondria. Fission of mitochondria depends on the dynamin-related GTPase Dnm1. This protein is mostly localized in the cytosol but part of it assembles into a ring-like complex on the cytosolic side of mitochondria in order to facilitate the division of mitochondria. This division apparatus contains two other proteins, Fis1/Mdv2 and Mdv1/Fis2/Gag3, which are equally important for mitochondrial fission. However, the molecular details of how fusion and fission of a double-membrane enclosed organelle occurs are still far from being understood, as are the mechanisms of the regulation of these processes.

Regulatory Mechanisms

Energy Dependency of Mitochondrial Biogenesis and Dynamics

Oxidative phosphorylation is usually the major source of ATP in mammalian cells. The glycolytic pathway, however, is the predominant origin of ATP in cancer cells. Baker's yeast is able to switch between these two pathways depending on the type of carbon source present and is therefore a highly suitable model organism for studying the energetic requirements of mitochondrial biogenesis. Import of proteins into or across the inner membrane as well as mitochondrial fusion and fission are energy dependent. Dissipation of the membrane potential of the inner membrane using uncoupling chemical compounds such as CCCP impairs import of preproteins into the matrix or into the inner membrane. In addition to a membrane potential, import of preproteins into the matrix using the TIM23 complex requires the presence of matrix ATP. This results from the ATP-dependency of the 70 kD mitochondrial heat shock protein (mtHsp70) which is an essential component of the import motor. However, in the case of preproteins using the TIM22 or the OXA1 translocase, the presence of a membrane potential is sufficient to complete translocation into the inner membrane. Insertion of proteins into the outer membrane depends on cytosolic ATP but not on a membrane potential nor on matrix ATP. Fusion of mitochondria is also a process that depends on the membrane potential; however, the underlying molecular mechanism is unknown.

Both, fusion and fission of mitochondria are processes that require the action of a dynamin-like GTPase,

Mgm1 and Dnm1, respectively. Therefore it is likely that at least one energy dependent step in both processes is the hydrolysis of GTP. Fragmentation of mitochondria and loss of mitochondrial DNA upon deleting *Mgm1*, *Fzo1*, or *Ugo1* can be prevented by blocking mitochondrial fission, e.g. by deletion of *Dnm1*. When fission is blocked, but fusion still going on, mitochondria adopt a characteristic net-like structure. Fusion and fission of mitochondria therefore are independent processes, which under normal conditions are balanced in order to maintain a tubular reticular structure. How this balance is regulated within cells is one of the major unresolved questions concerning mitochondrial dynamics.

Mitochondrial Diseases Caused by Nuclear Mutations

Not surprisingly, mitochondrial dysfunctions play an important role in the pathogenesis of a variety of diseases (reviewed in 7). They are caused either by alterations in the mitochondrial or the nuclear DNA and mostly manifest as a reduced respiratory function accompanied by reduced ATP synthesis. Defects in the biogenesis and the structural organization of mitochondria are also associated with diseases in humans. The molecular mechanisms related to two such disorders are briefly summarized.

Mohr-Tranebjaerg syndrome

The Mohr-Tranebjaerg syndrome is the only known mitochondrial disorder associated with a defect in the mitochondrial import machinery (reviewed in 8). It is a rare neurodegenerative disorder caused by mutations in the X-linked gene coding for the deafness-dystonia peptide (DDP1/TIMM8a). DDP1 is localized in the intermembrane space of mitochondria and, together with TIM13, facilitates the import of preproteins into the inner membrane. In general, this process depends on the presence of a membrane potential across the inner membrane. The corresponding Tim8/Tim13 complex from yeast was shown to be required for import of preproteins into the inner membrane, in particular under bioenergetically unfavorable conditions leading to a reduced membrane potential. One substrate that requires Tim8/Tim13 under these conditions is Tim23, which itself is an essential component of the TIM23 complex. Mutations in DDP1 are likely to exert pleiotropic effects on the import of the vast majority of mitochondrial matrix and inner membrane proteins. The exact molecular pathomechanisms leading to the Mohr-Tranebjaerg syndrome, however, are unknown.

Optic atrophy type I

Optic atrophy type I (OPA1) is the most prevalent hereditary optic neuropathy, occurring with a frequency of 1:12000 to 1:50000. OPA1 is an autosomal-dominant

inherited disease with onset in the first two decades of life. It is characterized by progressive loss of visual acuity, development of cecocentral scotoma, color vision disturbances, and bilateral optic atrophy. Several mutations in the nuclear encoded gene *opal* on chromosome 3q28-29 are linked to *OPAL* in different families. *Opal* is homologous to *Mgm1* from baker's yeast, which was shown to be essential for mitochondrial morphology (see above). Down-regulation of *Opal* by small interfering RNA in cell culture results in alterations of cristae morphology. In addition, this leads to release of cytochrome *c* from the intermembrane space and promotes apoptosis. Nevertheless, it is not understood at all how mutations in *opal* lead to degeneration of the optic nerve and it will be interesting to see to what degree mitochondrially induced apoptotic pathways are involved in the pathogenesis of optic atrophies.

References

1. Reichert AS, Neupert W (2002) Contact sites between the outer and inner membrane of mitochondria-role in protein transport. *Biochim Biophys Acta* 1592:41
2. Frey TG, Mannella CA (2000) The internal structure of mitochondria. *Trends Biochem Sci* 25:319-324
3. Neupert W, Brunner M (2002) The protein import motor of mitochondria. *Nat Rev Mol Cell Biol* 3:555-565
4. Sickmann A, Reinders J, Wagner Y et al (2003) The proteome of *Saccharomyces cerevisiae* mitochondria. *Proc Natl Acad Sci USA*
5. Scheffler IE (1999) Mitochondria. Wiley-Liss, New York
6. Mozdy AD, Shaw JM (2003) A fuzzy mitochondrial fusion apparatus comes into focus. *Nat Rev Mol Cell Biol* 4:468-478
7. Wallace DC (1999) Mitochondrial diseases in man and mouse. *Science* 283:1482-1488
8. Bauer MF, Neupert W (2001) Import of proteins into mitochondria: a novel pathomechanism for progressive neurodegeneration. *J Inher Metab Dis* 24:166-180

Mitochondrial Disease

► Mitochondrial Myopathy

Mitochondrial DNA

► Mendelian Forms of Human Hypertension and Mechanisms of Disease

► Mitochondrial Disease
► Mitochondrial Genome
► Mitochondrial Myopathies

Mitochondrial Genome

MICHIO HIRANO

Department of Neurology, Columbia University
College of Physicians and Surgeons, New York, NY,
USA

mh29@columbia.edu

Synonyms

Mitochondrial genome = Mitochondrial DNA (mtDNA)

Complex I = Nicotinamide adenine dinucleotide hydrogen (NADH) dehydrogenase

Complex II = Succinate dehydrogenase (SDH)

Complex III = Ubiquinone-cytochrome *c* oxidoreductase

Complex IV = Cytochrome *c* oxidase (COX)

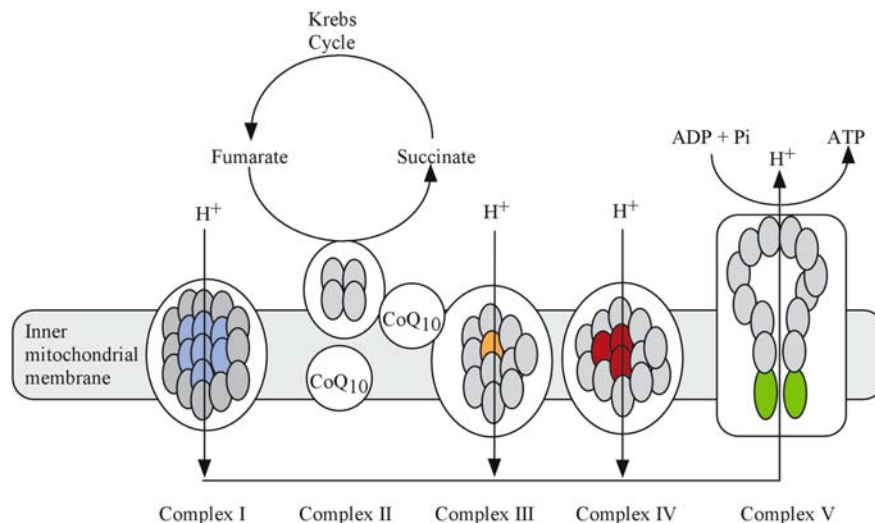
Complex V = ATP synthase

Definition

According to the endosymbiont hypothesis, mitochondria evolved from bacteria that were incorporated into protoeukaryotic cells more than a billion years ago. Those bacteria carried the capacity to generate cellular energy efficiently through aerobic metabolism, the oxidation of substrates to generate energy in the form of adenosine triphosphate (ATP). The pathway responsible for aerobic energy metabolism is composed of five enzymes complexes; four ►respiratory chain enzymes (complexes I-IV) generate a proton gradient across the inner mitochondrial membrane and complex V uses the proton gradient to generate ATP by ►oxidative phosphorylation (Fig. 1). In addition to their biochemical functions, the bacteria brought their own DNA, much of which has been incorporated during evolution into nuclear DNA. Nevertheless, mitochondria retained a small portion of their own genome. Consequently, mitochondria are the products of two genomes, one in the nucleus and the other within the organelle.

Characteristics

The ►mitochondrial genome is a small, 16,569 base-pair, circular DNA molecule (Fig. 2), efficiently packed with no introns and very few non-coding segments (1, 2). The genome contains 37 genes encoding: 22 transfer RNAs (tRNAs), two ribosomal RNAs (rRNAs)



Mitochondrial Genome. Figure 1 Schematic representation of the mitochondrial respiratory chain and oxidative phosphorylation enzyme complexes. Nuclear DNA encoded subunits are depicted as gray ovals and mitochondrial DNA encoded subunits are shown as colored ovals.

and 13 polypeptides required for respiratory chain function and oxidative phosphorylation. A mitochondrion has about five copies of mitochondrial DNA and each cell harbors hundreds of mitochondria; therefore, each cell contains hundreds to thousands of copies of mitochondrial DNA.

The two strands of the mitochondrial DNA differ in composition; one strand, disproportionately rich in guanine and thymine residues, is heavier in alkaline cesium chloride gradients and is therefore called the “heavy” or H-strand. Naturally, the other strand is identified as the “light” or L-strand.

Of the 13 polypeptides encoded in the mitochondrial genome, seven are subunits of complex I of the mitochondrial respiratory chain, one is a subunit of complex III, three are subunits of complex IV and two are subunits of complex V (Figs. 1, 2). In addition, about 70 structural polypeptides encoded in the nuclear DNA are required for mitochondrial respiratory chain activity. Only complex II is encoded entirely by nuclear DNA. In addition, 22 tRNA and 2 rRNA molecules encoded in the mitochondrial genome are required for mitochondrial protein synthesis.

Transcription of the Mitochondrial Genome

Unlike genes in nuclear DNA that are transcribed individually into RNA, the mitochondrial genes are transcribed as two large 16 kilobase polycistronic precursor molecules. The light strand and heavy strand promoters (LSP and HSP respectively) for transcription are located close to each other in the D-loop, a non-coding segment of the genome (Fig. 2). The

polycistronic transcripts are processed into individual messenger RNAs (mRNA) as well as tRNAs and rRNAs. Most of the mRNAs are separated by tRNAs that serve as “punctuation” signals for processing of the polycistronic transcripts into single gene RNA transcripts (Fig. 2).

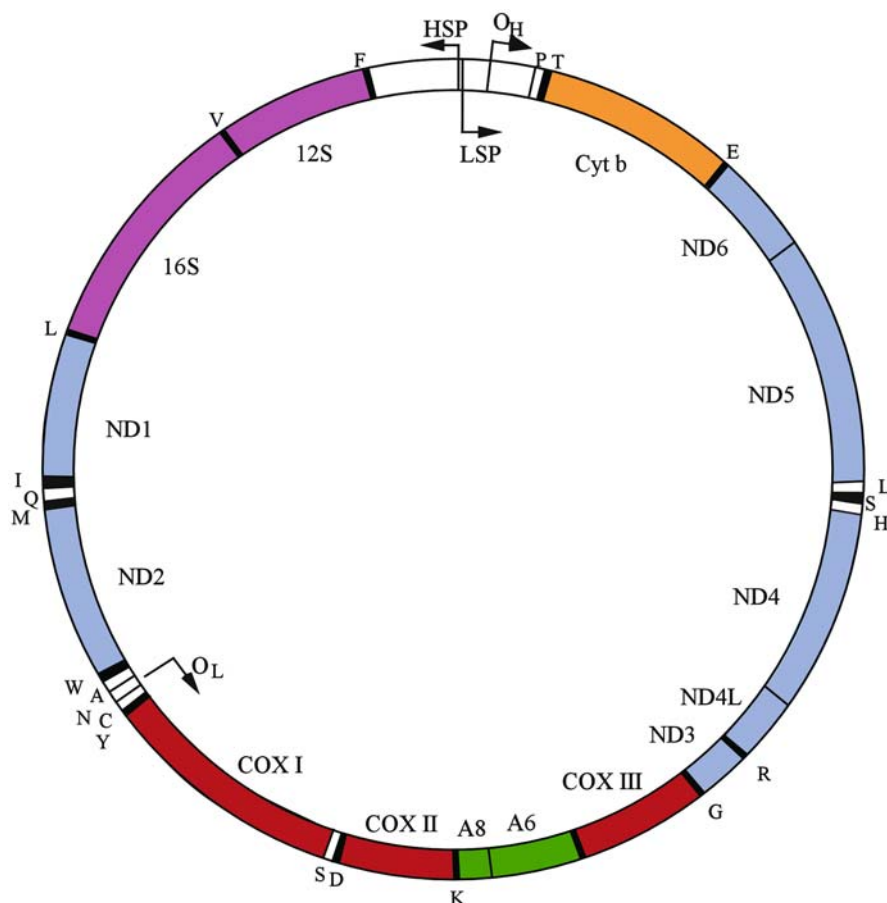
Translation of the Mitochondrial Genome

Within mitochondria, mRNAs are translated into proteins. Four of the mitochondrial codons (three nucleotide segments that encode either an amino acid or a termination signal) differ from the 64 “universal” triplets. For example, AUA encodes methionine in the mitochondrial genome but encodes isoleucine in nuclear DNA.

Replication of the Mitochondrial Genome

During evolution, transfer of mitochondrial genes into the nuclear genome has made the mitochondrial genome a “slave” of the nucleus because all the factors required for replication and repair of mitochondrial DNA are encoded in nuclear DNA.

The mechanism of mammalian mitochondrial DNA replication is a controversial subject. According to the strand-asymmetric model, replication of mitochondrial DNA originates at two sites, the origins of heavy-strand synthesis (O_H) and light-strand synthesis (O_L) (Fig. 3) (7). In this model, mtDNA replication begins with binding of mitochondrial transcription factors and mitochondrial RNA polymerase and initiation of RNA synthesis at the LSP. This light strand transcript may generate a polycistronic RNA molecule as



Mitochondrial Genome. Figure 2 The mitochondrial genome. Structural gene abbreviations: A, ATP synthase; COX, cytochrome c oxidase; Cyt *b*, cytochrome *b*; ND, NADH dehydrogenase; 12S and 16S, ribosomal RNA genes; 1-letter code, transfer RNA genes. Control regions: HSP, heavy-strand promoter; LSP, light strand promoter; O_H , origin of heavy-strand replication and O_L , origin of light-strand replication.

previously described or may produce a short RNA primer for polymerase γ to initiate DNA synthesis at O_H . Replication of the heavy strand proceeds about two-thirds of the way around the mitochondrial genomic circle and exposes O_L , which serves as the initiation site for light strand replication. In contrast, the strand-synchronous model for mitochondrial DNA replication asserts that H-strand synthesis is initiated at O_H and is closely coupled with synthesis of the L-strand (Fig. 3) (5).

Mitochondria maintain nucleotide pools that are distinct from those of the rest of the cell and rely more heavily on nucleotide salvage pathways than nuclei, which depend on *de novo* synthesis.

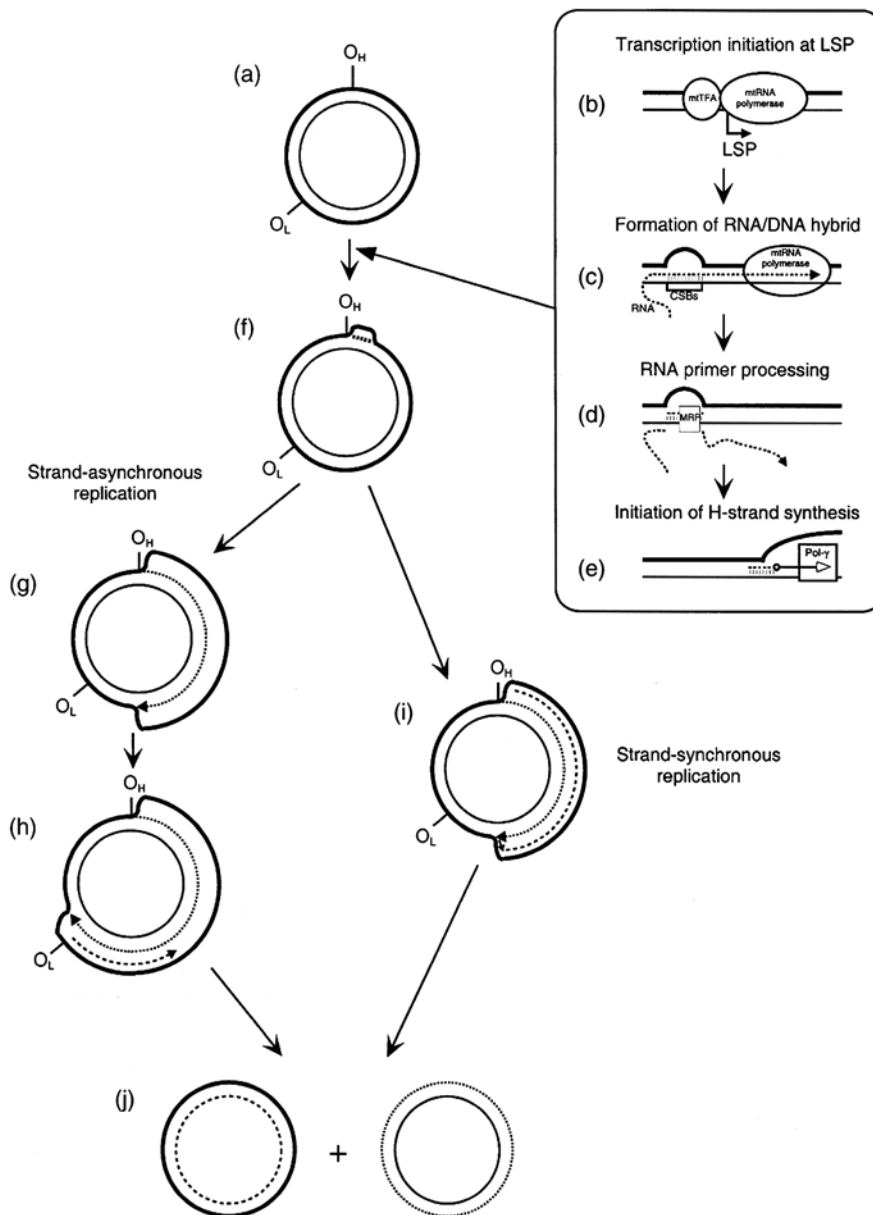
Repair of the Mitochondrial Genome

Because mitochondria generate large quantities of reactive oxygen species (ROS) that can damage DNA, lipids and proteins, the organelles have effective mechanisms to repair ROS-damaged DNA. In contrast,

mitochondria seem to lack repair systems for other forms of abnormal DNA. For example, mitochondria are unable to remove ultraviolet-induced pyrimidine dimers. The differences in repair mechanisms and exposure to mutagenic factors may contribute to the ten-fold higher mutation rate of the mitochondrial genome compared to that of nuclear DNA.

Mitochondrial Inheritance

Human mitochondria and mitochondrial DNA are maternally inherited; a mother transmits her DNA to all of her children. This non-Mendelian inheritance pattern may be in part due to the overwhelming number of mitochondrial genomes, about 100,000, in each oocyte compared to about 100 copies of mitochondrial DNA in each sperm. In addition, there is evidence that after entry into oocytes, sperm mitochondria are lost early in embryogenesis. The mechanism of the selective elimination of sperm mitochondria is not fully understood. Nevertheless, there is a single well-documented



Mitochondrial Genome. Figure 3 Replication of the mitochondrial genome (modified from (7)). See text for detail.

man who developed a mutation in paternally derived mitochondrial DNA (6). It is not certain whether this patient represents a unique medical curiosity or is part of a more generalized phenomenon.

Segregation and Heteroplasmy

As previously noted, each cell contains a large population of hundreds to thousands of copies of the mitochondrial genome, which is strikingly different from the situation in nuclear DNA, where somatic cells have only two copies of each autosome. In most individuals, the mitochondrial genomes are identical (homoplasmy); however, mitochondrial DNA

polymorphisms can be present in variable proportions, a situation known as **heteroplasmy**. Consequently, in an individual with a heteroplasmic mitochondrial DNA polymorphism, the composition of mitochondrial DNA can vary from tissue to tissue or even from cell to cell within a tissue.

In oocytes, although there are about 100,000 copies of the mitochondrial genome, there is a bottleneck of a relatively small number of segregating units, perhaps as few as 5–50 that determine the mitochondrial genomic population of the embryo. Furthermore, the distribution of the polymorphic mitochondrial DNA may be skewed in different tissues.

Clinical Relevance

Mitochondria are vital organelles required for energy production and can affect virtually any organ. Brain and skeletal muscle are particularly vulnerable to mitochondrial defects because of their high energy requirements. Due to their dual genetic origin, mitochondrial disorders can be caused by mutations in mitochondrial or nuclear DNA. More than 150 point mutations and hundreds of deletions of mitochondrial DNA have been associated with human diseases (3). Because of the unique features of mitochondrial DNA, these mutations show unusual clinical characteristics. For example, most disorders of mitochondrial DNA are maternally inherited. Because of heteroplasmy and skewed segregation of pathogenic mitochondrial DNA mutations, phenotypic variability is common even within families harboring an identical mitochondrial DNA mutation.

Autosomal diseases of the mitochondria are being identified with increasing frequency. A group of Mendelian disorders with secondary quantitative or qualitative defects in mitochondrial DNA is being defined at the molecular level.

The potential role of mitochondrial DNA in aging has generated much interest and even greater controversy. According to the mitochondrial theory of aging, ROS derived from mitochondria may cause somatic mutation of mitochondrial DNA that, in turn, leads to more ROS production and a vicious cycle of additional deleterious alterations in the mitochondrial genome (4). For further information about specific mitochondrial disorders, readers are referred to the article on mitochondrial disorders in this Encyclopedic Reference of Genomics and Proteomics in Molecular Medicine.

References

1. Anderson S, Bankier AT, Barrel BG et al (1981) Sequence and organization of the human mitochondrial genome. *Nature* 290:457–465
2. Andrews RM, Kubacka I, Chinnery PF et al (1999) Reanalysis and revision of the Cambridge reference sequence for human mitochondrial DNA. *Nat Genet* 23:147
3. Di Mauro S, Schon EA (2003) Mitochondrial Diseases. *N Engl J Med* 348:2656–2668
4. Di Mauro S, Tanji K, Bonilla E et al (2002) Mitochondrial abnormalities in muscle and other aging cells: classification, causes, and effects. *Muscle Nerve* 26: 597–607
5. Holt IJ, Lorimer HE, Jacobs JT (2000) Coupled leading and lagging-strand synthesis of mammalian mitochondrial DNA. *Cell* 100:515–524
6. Schwartz M, Vissing J (2002) Paternal inheritance of mitochondrial DNA. *N Engl J Med* 347:576–80
7. Shadel GS, Clayton DA (1997) Mitochondrial DNA maintenance in vertebrates. *Ann Rev Biochem* 66:409–435

Mitochondrial Membrane

Definition

There are two membranes for mitochondria that are small organelles found within eukaryotic cells. The outer membrane is semi-permeable, allowing small molecules like orotate to penetrate. The inner membrane is selectively permeable, and contains the electron transport chain required for respiration and reaction 4 of the pyrimidine pathway.

► [Mitochondria – Biogenesis and Structural Organization](#)

► [Nucleotide Biosynthesis](#)

Mitochondrial Myopathy

LENE SÖRENSEN, NILS-GÖRAN LARSSON
Karolinska Institute, Department of Laboratory
Medicine, NOVUM, Karolinska University Hospital-
Huddinge, Stockholm, Sweden
lene.sorensen@labmed.ki.se
nils-goran.larsson@labmed.ki.se

Synonyms

Mitochondrial disease; Oxidative phosphorylation disease

Definition

Mitochondrial myopathies are diseases of varying severity caused by a dysfunction of the mitochondrial respiratory chain. They are often accompanied by symptoms in other tissues, particularly those with a high energy demand such as brain, liver, kidney and heart muscle. Tissues that are ► [post-mitotic](#), meaning that most of their cells do not divide any more, are often affected in mitochondrial disease. The terms mitochondrial disease, mitochondrial cytopathy, mitochondrial myopathy, mitochondrial encephalomyopathy and oxidative phosphorylation disease all refer to a heterogeneous group of diseases caused by defective function of the respiratory chain. The symptoms are extremely variable and onset of disease may be from early infancy to late adulthood.

The mitochondrial network is located in the ► [cytoplasm](#) of the cell and contains the respiratory chain. The respiratory chain consists of five enzyme complexes that generate the common form of energy currency, ► [ATP](#), required for a large array of metabolic processes in the cell. The respiratory chain is dependent on two

separate genomes to function, the nuclear DNA and the mitochondrial DNA, ►**mtDNA**. The mtDNA contains 37 genes, encoding 13 proteins and 24 RNAs needed for mitochondrial protein synthesis. A typical human cell contains 1,000–10,000 mtDNA molecules. Mitochondrial myopathies are caused by a variety of genetic defects in the mtDNA or nucleus.

Characteristics

Rolf Luft described the first patient with mitochondrial dysfunction more than forty years ago, even before mtDNA was discovered (1). The clinical manifestations of mitochondrial myopathies can be of varying severity and symptoms include muscle weakness, exercise induced cramps and ►**myalgia**. Other frequent symptoms are ptosis and ►**progressive external ophthalmoplegia** (PEO). Ptosis is when the muscles surrounding the eyes are weak and the patient has droopy eyelids. PEO is weakness of the muscles that control eye movement. The severity of PEO can vary from mild to complete paralysis. Less common skeletal muscle symptoms are episodic weakness with ►**rhabdomyolysis** and ►**myoglobinuria** (2).

The primary morphological hallmark of mitochondrial myopathies is the occurrence of ragged red fibers, RRF. These are muscle fiber segments that have an abnormal accumulation of mitochondria. The term RRF refers to the irregular shape of the mitochondrial accumulations and the finding that these accumulations can be stained red with a specific chemical stain (modified Gomori trichrome).

Respiratory chain deficiency caused by nuclear mutations usually affects all cells in a tissue in a uniform way, whereas mutations of the mtDNA result in a characteristic mosaic pattern of enzyme deficiency. The explanation for this pattern is that patients harboring mutations of mtDNA often have a mixture of normal and mutated mtDNA.

The levels of mutated mtDNA often vary widely between different organs in the patient and even between different cells of the same organ. A certain minimal number of mutated mtDNA molecules, a threshold level, is needed to produce a respiratory chain deficiency. This threshold level is different for different mutations and also varies depending on the affected tissue. The occurrence of two or more different kinds of mtDNA in a single cell is called heteroplasmy.

The distribution of mutated mtDNA is an important determinant of symptoms in patients carrying heteroplasmic mtDNA mutations. One particularly striking example is children with an infantile anemia and pancreas insufficiency disorder (Pearson syndrome) caused by widespread distribution of high levels of deleted mtDNA. This syndrome is often lethal but some infants recover; however they always develop a

severe neuromuscular syndrome (Kearns-Sayre syndrome) years later. The likely explanation for this complete change of symptoms with time is that clones of rapidly dividing bone-marrow cells with high levels of deleted mtDNA may be lost due to selection, thus allowing a recovery from anemia. However, deleted mtDNA will accumulate with time in non-dividing tissues such as muscle and nerve cells and this will lead to later onset of the neuromuscular syndrome (3).

Nuclear mutations causing respiratory chain deficiency may also have varying clinical phenotypes. Patients with ►**autosomal recessive** depletion of mtDNA usually die in infancy with severe myopathy. Patients with autosomal dominant or recessive PEO and multiple mtDNA deletions often have a milder disease with adult onset of symptoms.

In spite of recent advances in the understanding of the pathogenesis of mitochondrial diseases, there is no efficient treatment. There are some pharmacological treatments but there is no consensus on the efficacy of these drugs. Some progress has been made using small nucleic acids that selectively inhibit the replication of mutated mtDNA, hindering it from reaching threshold levels within the muscle fiber. Also, there has been some success with a strategy where affected muscle fibers are selectively killed. In this way dysfunctional muscle cells are removed and new, functional muscle can take their places. The optimal cure would be gene therapy, introducing the missing gene(s) into the mitochondria or, in the case of a nuclear mutation, the nucleus.

Genetics

There are three main categories of genetic defects causing mitochondrial myopathy:

1. Known mtDNA mutations; point mutations and deletions
2. Known nuclear mutations
3. Unknown mutations

In addition to these genetic categories, there can also be an accumulation of sporadic mutations in aging muscle. This can then lead to a late-onset mitochondrial myopathy affecting elderly individuals.

Known mtDNA Mutations: mtDNA Point Mutations

Mitochondrial DNA is maternally inherited. This means that the mtDNA from an individual comes from the mother. During the production of the egg, there is a phenomenon called the mtDNA “bottleneck”. This is thought to be when only a few mtDNA molecules are transferred from the precursor germ cell and then replicated up to the 100,000 copies of mtDNA in the mature egg cell (3). The “bottleneck” explains how a mother who has no or only mild symptoms can have

severely affected children. The random shift of mutation load between generations also means that it is difficult to predict the probability of having an affected child.

Today there are over 100 point mutations of mtDNA that are known to cause mitochondrial dysfunction. Point mutations leading to mitochondrial myopathies are often in a ▶**tRNA** gene, but disease-causing mutations in protein-coding genes have also been found (4). A tRNA mutation often leads to an impairment of mitochondrial translation and therefore a loss of several mitochondrial encoded respiratory chain subunits.

Examples:

- a mutation in *tRNA*^{Lys} causes myoclonic epilepsy with ragged red fibers (MERRF)
- a mutation in *tRNA*^{Leu} causes mitochondrial myopathy, encephalopathy, lactic acidosis and stroke-like episodes (MELAS)
- a mutation in *tRNA*^{Leu} causes progressive external ophthalmoplegia (PEO)

Known mtDNA Mutations: mtDNA Deletions

Single large deletions of the mitochondrial DNA are usually sporadically occurring. There is no obvious explanation as to why mtDNA deletions are rarely transmitted from the mother whereas mtDNA point mutations are frequently transmitted. Single large mtDNA deletions are thus usually sporadic and often not maternally transmitted, i.e. they happen by chance. There are several known nuclear mutations that cause multiple mtDNA deletions. However, in these cases it is the predisposition to develop mtDNA deletions that is inherited, not the deletions themselves.

Large deletions in mtDNA that cause diseases always remove at least one tRNA gene, leading to impaired translation of mitochondrial proteins and respiratory chain dysfunction.

Examples:

- chronic progressive external ophthalmoplegia (CPEO)
- Kearns-Sayre syndrome

Known Nuclear Mutations

Around 85 of the protein subunits in the respiratory chain as well as all proteins needed for the maintenance of the mtDNA are encoded by the nucleus. A wide range of proteins involved in other aspects of mitochondrial function such as transport across the mitochondrial membranes or protein assembly are also nucleus-encoded. The number of identified mutations in nuclear genes known to cause mitochondrial disease has increased dramatically over the last few years, and so far more than 20 have been documented (3).

When a nuclear mutation is the cause of mitochondrial myopathy, it will be inherited in a ▶**Mendelian** fashion. There are different kinds of mitochondrial myopathies and most of them are inherited in an ▶**autosomal dominant** way. Nevertheless, diseases with autosomal recessive or ▶**X-linked inheritance** are also known.

The clinical manifestations of a nuclear mutation affecting mitochondrial function can be very severe if a crucial enzyme activity is totally abolished. On the other hand, symptoms may be less severe if there is only a partial decrease of function or the mutation affects a redundant protein.

Examples:

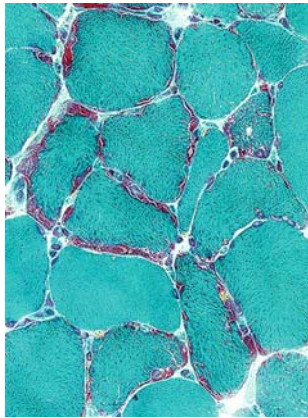
- a mutation in *adenosine nucleotide transferase 1* (▶**ANTI**) causes autosomal dominant progressive external ophthalmoplegia (adPEO)
- a mutation in *thymidine phosphorylase* (▶**TP**) causes mitochondrial neurogastrointestinal encephalomyopathy (MNGIE)
- mutations in *DNA polymerase gamma* (▶**POLG**) cause autosomal recessive or dominant progressive external ophthalmoplegia (arPEO/adPEO)
- a mutation in mitochondrial *thymidine kinase* (TK2) causes autosomal recessive mtDNA depletion and severe infantile myopathy

Unknown Mutations

There are many cases with clinically established mitochondrial disease where no causative mutations have been found. In fact, this group of patients is the largest. In these cases, it is prudent to believe that the cause is a nuclear mutation that has not been identified yet. The mitochondrial genome is rather small, only 16,569 base pairs, and is easily sequenced for mutations. The nuclear genome, on the other hand, is too large and contains too many polymorphisms to be routinely sequenced. Success in finding nuclear mutations causing mitochondrial myopathies largely depends on the presence of families where one can observe if a candidate mutation is inherited together with the disease.

Clinical Relevance

When mitochondrial disease is suspected, there are many ways to test whether this is indeed the case. Laboratory diagnosis of mitochondrial disorders often includes a combination of biochemical, morphological and genetic methods to diagnose respiratory chain dysfunction (2). The most widespread diagnostic method for mitochondrial myopathy is to look for the occurrence of ragged red fibers, RRF. These fibers can be visualized by the modified Gomori trichrome technique, which stains the irregular accumulations of mitochondria red (Fig. 1). When mtDNA is mutated,



Mitochondrial Myopathy. Figure 1 Ragged red fibers (RRFs), one of the hallmarks of mitochondrial disease, stained with the modified Gomori trichrome stain. A cross-section of a skeletal muscle specimen shows the individual muscle fibers as rounded structures. Normal muscle fibers are stained evenly blue, whereas RRFs show an accumulation of mitochondria around the edges, stained red. *Picture by Anders Oldfors, Gothenburg University.*

mitochondrial translation is often impaired. When this is the case, mitochondrially encoded respiratory chain enzymes will disappear. Another way to visualize mitochondrial deficient RRFs is to stain for the enzyme activities of succinate dehydrogenase (SDH) and cytochrome c oxidase (COX), the so-called COX/SDH stain. COX is a respiratory chain enzyme with critical components encoded by mtDNA, whereas SDH is entirely nuclear encoded. This means, that if the mtDNA is compromised there will be plenty of SDH activity but no COX activity. With this double staining method respiratory chain deficient muscle fibers will appear blue whereas normal fibers will appear brown. The activities of other respiratory chain enzymes can also be detected by enzyme histochemistry, although it is less common.

It is possible to measure the function of the different respiratory chain complexes directly on isolated mitochondria. In this case, one adds chemical substrates that enter the respiratory chain at different levels and monitors the oxygen consumption. If a particular enzyme complex is not working properly, the addition of that enzyme's substrate will not result in elevated oxygen consumption. It is also possible to measure the mitochondrial ATP production rate or the activities of the individual respiratory chain enzyme complexes.

In the case of mitochondrial myopathies, the mitochondria are often more abundant and display an abnormal morphology, with crystalline structures in the [▶matrix](#) and a general disorganized appearance. [▶Electron microscopy](#) is well suited to detect these abnormalities.

One can also use this method to detect the accumulation of mitochondria in dysfunctional muscle fibers and to calculate the percentage of the total cell volume occupied by mitochondria.

Molecular genetics methods are used to detect specific mtDNA and nuclear mutations. Point mutations in the mtDNA are easy to diagnose with the polymerase chain reaction, [▶PCR](#) or by sequencing candidate regions. [▶Southern blot](#) or [▶long PCR](#) analysis is used to detect mtDNA deletions and rearrangements. These analyses are sometimes performed on a sample of the patient's blood. There is one large pitfall with this approach, analyses of mtDNA mutations in blood are unreliable. Some classes of mtDNA mutations are absent in blood, even though they are present in high levels in skeletal muscle. If the cause of the disease is a nuclear mutation, one can with PCR analysis and sequencing easily test if one of the genes known to cause mitochondrial myopathies is mutated. However, a large number of nuclear mutations causing disease are still unknown and cannot be detected by this approach.

References

1. Luft R, Ikkos D, Palmieri G et al (1962) A case of severe hypermetabolism of nonthyroid origin with a defect in the maintenance of mitochondrial respiratory chain control: a correlated clinical, biochemical and morphological study. *J Clin Invest* 41:1776–1804
2. Larsson N-G, Oldfors A (2001) Mitochondrial myopathies. *Acta Physiol Scand* 171:385–393
3. Graff C, Bui T-H, Larsson N-G (2002) Mitochondrial diseases. *Best Pract Res Clin Obstet Gynaecol* 16: 715–728
4. Larsson N-G, Clayton DA (1995) Molecular genetic aspects of human mitochondrial disorders. *Annu Rev Genet* 29:151–178

Mitochondrial Permeability Transition Pore

Definition

The mitochondrial permeability transition pore (MTP) is a high conductance channel consisting of the adenine nucleotide translocator (ANT) in the mitochondrial inner membrane, the matrix protein Cyp22, as well as apoptosis-regulatory proteins from the Bax/Bcl-2 family which is involved in mitochondrial membrane permeabilization. Transient openings of the MTP play an important role in the pathways leading to cell death.

[▶Mitochondria – Biogenesis and Structural Organization](#)

[▶Peptidyl Prolyl Cis/Trans Isomerases](#)

Mitogen-Activated Protein Kinases

Mitogen-activated protein kinases (MAPKs) are an evolutionarily conserved family of serine/threonine protein kinases, involved in the regulation of a variety of cellular processes including proliferation, differentiation and apoptosis. MAPKs become activated by dual phosphorylation on threonine and tyrosine residues. The MAPK family includes the stress activated protein kinases (SAPKs), the Extracellular signal Regulated Kinases (ERKs) and p38-related kinases. MAPKs of all subgroups are activated by dual specificity kinases, the MAP kinases kinases (also designated as MEKs), which in turn are regulated by MAP kinase kinase kinases (also designated as MEKKs). Thus, defined groups of a MAPK, a MAPKK and a MAPKKK can act in a cascade as a functional unit. According to the identity of the MAPK involved, one discriminates between the ►JNK cascade, the p38 ►MAP kinase and the ►ERK cascade. The MAPK signalling modules can be triggered by a variety of stimuli including ligands of the TNF family.

- Cytokines
- Growth Factors
- Interleukins and Their Receptors, a Role in Regulating Hematopoietic Stem Cells
- Kinases
- Receptor Serine/Threonine Kinase
- TNF Receptor/Fas Signaling Pathways

Mitosis

Definition

Mitosis is the process of cell and nuclear division in somatic cells, which results in the production of two daughter cells from a single parent cell. The daughter cells are identical to one another and to the original parent cell (mitotic recombination).

- Cell Division
- Centromeres
- Chromosomal Instability Syndromes
- Mammalian Fertilization
- Meiosis and Meiotic Recombination
- Mitotic Recombination
- Mitotic Spindle

Mitotic Figure

Definition

Mitotic figure is a term that is used by pathologists. Cells in mitosis are discriminated by visible chromosomes (like dark-stained threads) and the absence of a nuclear membrane. The number of mitotic figures is sometimes used to distinguish a benign from a malignant tumor.

- Cell Division
- Centrosome

Mitotic Recombination

IAN TOMLINSON

London Research Institute, Cancer Research UK,
London, UK

ian.tomlinson@cancer.org.uk

Definition

Mitotic recombination (MR) results from somatic crossing over between homologous sequences during mitosis. MR is a term used inconsistently among geneticists, molecular biologists and biochemists. Herein, MR will be used to describe both exchange of genetic material between sister chromatids (SCEs) – which is an effective, common repair mechanism usually without genetic or phenotypic consequence – and exchange between homologous chromosomes. The latter process manifests either as so-called gene conversion (GC) in which a relatively short length of chromosome becomes identical to its homologue or as ‘mitotic recombination proper’ (which we shall term ‘mitotic crossing-over’, MXO) in which the entire distal chromosome arm becomes identical to its homologue. GC and MXO are rare compared with meiotic crossing-over because homologues do not normally pair at mitosis. GC and MXO lead to homozygosity in a daughter cell for markers distal to the crossover site for a variable length along the chromosome arm, up to and including the telomere. MXO is distinct from mitotic non-disjunction, in which a whole chromosome is lost or gained, but the definition of MR may include mitotic gene conversion, in which a relatively short, interstitial chromosome segment becomes identical to its homologue.

In its strict sense, MR may also be used to describe recombination between non-homologous chromosomes in mitosis, leading to translocation and associated rearrangements; I shall not, however, focus on non-homologous recombination below. Finally, I shall regard non-homologous end joining (NHEJ) repair in mitosis as distinct from MR.

Characteristics

History

Some of the first clues as to the existence of MR came from observations by Stern in *Drosophila* (7). He found that a few flies showed patches on their body, which could be best explained by somatic reduction to homozygosity of chromosomal segments between some location on the chromosome arm and that arm's telomere. MR appeared to occur largely at random in different tissues at different stages of development, thus producing flies with varying phenotypes. In principle, given enough heterozygous phenotypic markers, an MR map could be built up by identifying phenotypes that were revealed together when MR occurred. MR has since been shown to occur in a variety of organisms from *Aspergillus* to humans. In humans, MR by MXO or GC can sometimes be visualised as chiasmata in mitosis and these are distributed non-randomly, with hotspots hypothesised to correspond to gene-rich regions. MR crossovers can also, however, occur in heterochromatin.

The Normal Role of MR in DNA Repair

Mitotic recombination is envisaged to occur as the result of DNA repair processes. Specifically, MR is probably one of several possible outcomes of DSB (double strand break) repair. DSBs classically occur as a result of ionising radiation, but other genotoxic agents can cause them and 'spontaneous' DSBs can occur at stalled replication forks, perhaps at sites of existing single strand breaks. Non-homologous end joining is believed to repair most DSBs, but homologous recombinational mechanisms appear important in certain cells and at certain stages of the cell-cycle (S and G2). Figure 1 shows the model proposed whereby DSB repair leads to MR (5). In this model, DSB repair occurs as a result of invasion of one strand of the broken DNA molecule into the double helix of its homologue. Unless alternative repair mechanisms (such as synthesis-dependent strand annealing or break-induced replication) occur, repair proceeds through new DNA synthesis complementary to the unbroken strands. In this model, it is hypothesised that the form of the resolution of the resulting Holliday junctions determines whether or not a crossover occurs. In most cases, MR will occur in the form of SCE. If, however, homologous chromosomes happen to be in close physical alignment during mitosis, the crossover

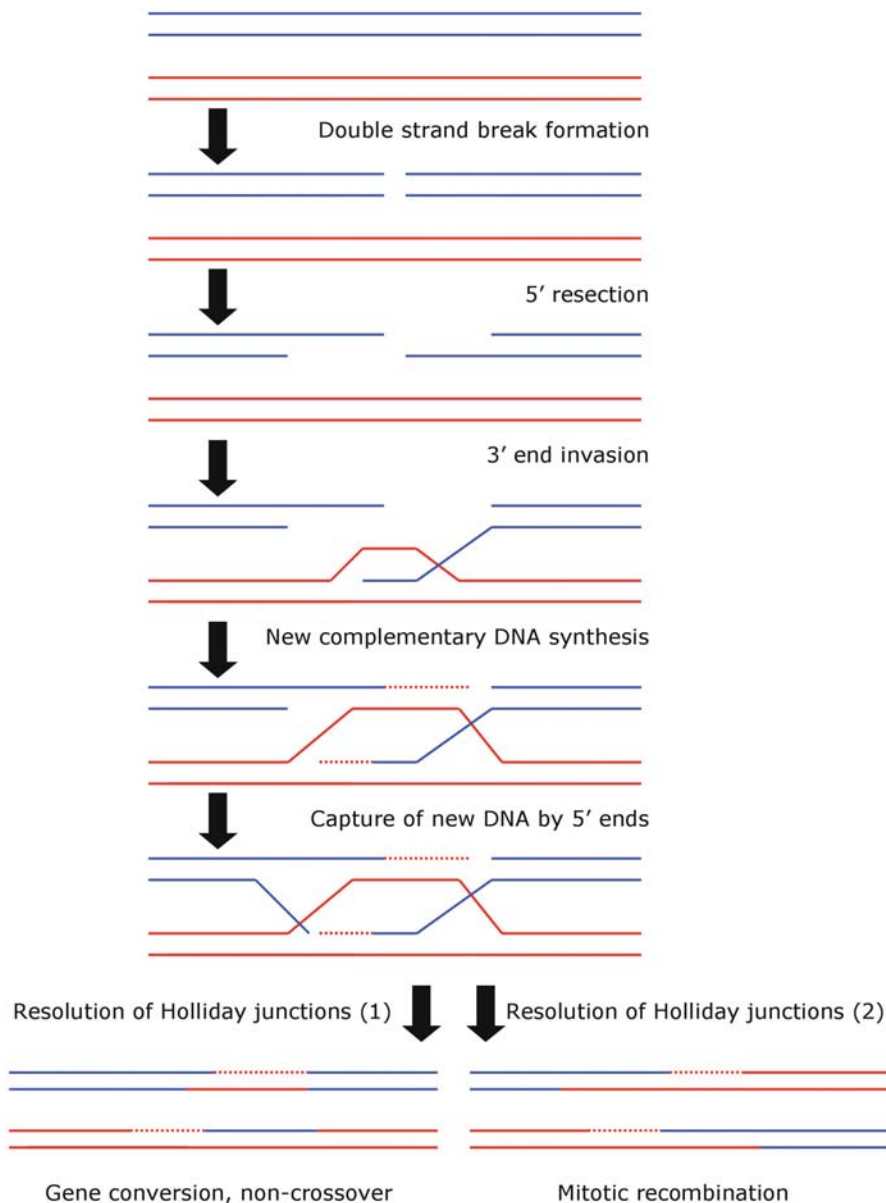
can occur between non-sister chromatids and GC or MXO results. It can readily be seen both that MXO is not the usual result of DSB repair and that MXO involves a small region of GC close to the DSB. If non-homologous chromosomes participate in MR – for example if orthologues or repeated sequences act as 'illegitimate' templates for repair – reciprocal or unbalanced translocations may result.

Molecules Involved in MR

The molecules involved in DSB repair and MR remain largely unknown in mammalian cells. Prado et al. (5) have reviewed the genes that are known to be important in yeast, which include *RAD51*, *RAD52* and components of the mismatch repair machinery. Mammalian homologues of *RAD51* and *RAD52* include *RAD51B/C/D*, *RAD54*, *RAD54B*, *RAD50*, *MRE11*, *XRCC2* and *XRCC3*. The mechanism of DSB repair by SCE or other forms of MR in higher organisms is not well characterised. It is suggested that ATM and NBS1, complexed with MRE11 and RAD50 may be involved in the initial signalling to the repair machinery that a DSB has occurred. RAD52 and other proteins may recognise the DNA ends and a single-strand region with a 3' overhang is generated by exonuclease digestion. RAD51 may polymerise onto the single stranded DNA with the help of RPA and RAD52, and the resulting nucleoprotein filament in some way searches for a homologous DNA duplex. RAD52 and RAD54 may then help with single strand invasion of the intact duplex. DNA polymerases re-synthesise DNA complementary to the intact strands, before strand ligation and resolution by unknown enzymes. Mouse *Rad54*^{-/-} cells show decreased SCE and *Xrcc2*^{-/-} and *Xrcc3*^{-/-} hamster cells show reduced levels of homologous recombination. Interestingly, RAD51 interacts with the breast cancer protein BRCA1, although – perhaps paradoxically for tumour promotion – *BRCA1*^{-/-} cells have been reported to show reduced homologous recombination (reviewed in 8). It has been suggested that one role of BRCA1 is to promote homologous recombination repair when DSB damage occurs in S- or G2-phase.

Diseases Resulting from Deficient MR

Humans with two mutant copies of the ► **Bloom syndrome** gene (*BLM*) have genetic instability, which is characterised by an increased frequency of SCEs (OMIM 210900). An ancestral mutation accounts for over-representation of the disease in Ashkenazi Jews. Bloom syndrome patients have proportionate pre- and post-natal growth deficiency, are sun-sensitive, suffer from telangiectasia and have hypo- and hyper-pigmented skin; patients also have an increased risk of a wide variety of malignancies, including both solid tumours and haematological malignancies. Men with Bloom



Mitotic Recombination. Figure 1

syndrome are sterile; women have reduced fertility and a shortened reproductive span. *BLM* encodes a DNA helicase that is homologous to RecQ in *E. coli*. The *Blm* 'knockout' mouse has been found to develop several types of tumour as a result of raised GC and *Blm*^{-/-} cells show raised frequencies of chiasma formation, homologous recombination (including SCE) and loss of heterozygosity (see below). Breaks, gaps and complex chromosomal rearrangements are also seen relatively often. The BLM protein promotes branch migration of Holliday junctions and may help to re-start stalled replication forks. Recently, BLM has been shown directly to suppress recombination.

Individuals with Werner syndrome (OMIM 277700), a recessively inherited condition of premature ageing, have scleroderma-like skin changes, cataract, subcutaneous calcification, premature arteriosclerosis, diabetes mellitus and a wizened facies. There is also an increased risk of sarcoma. Werner cells have normal SCE and appear to acquire chromosomal changes through increased non-homologous MR and consequent translocations/deletions. Like *BLM*, the Werner syndrome gene (*WRN*) is a RecQ homologue. BLM and WRN may interact (8), although the specific role of WRN is unknown. Rothmund-Thomson syndrome (OMIM 268400) is, in some cases, caused by mutations

in another helicase, *RECQL4*; this condition is characterised by hereditary dermatosis (comprising atrophy, pigmentation and telangiectasia) and is frequently also associated with juvenile cataract, saddle nose, congenital bone defects, disturbances of hair growth and hypogonadism. Sarcomas, largely arising from bone and soft tissue, are relatively common in Rothmund-Thomson syndrome.

Other molecules with putative roles in mammalian MR include the ataxia telangiectasia protein (ATM), Nijmegen breakage syndrome protein (NBS1), RAD50, BRCA2 (which interacts with RAD51) and MRE11. ATM may control many aspects of DSB repair. Recessively inherited germline *ATM* mutations cause a phenotype of chromosome breakage associated with cerebellar ataxia, telangiectases, immune defects and a predisposition to malignancy (OMIM 208900). *NBS1* (OMIM 251260) mutations produce a similar recessively inherited syndrome of microcephaly, growth retardation, immunodeficiency and predisposition to cancer. *BRCA1* and *BRCA2* mutations, by contrast, cause dominantly inherited, early-onset breast and ovarian cancer in which the genes act as classical tumour suppressors (OMIM 113705, OMIM 600185), although double *BRCA2* mutants have a phenotype of Fanconi anaemia (OMIM 227650), another recessive chromosomal breakage syndrome.

Factors Predisposing to MR in 'Normal' Individuals

The frequency of MR is increased by any factor which increases DNA damage that can be repaired by a mechanism in which homologous recombination occurs. DSBs are, for example, more prevalent in individuals exposed to X-rays or a variety of chemical carcinogens (9) or after oxidative damage.

There is also evidence that the frequency of MR varies in the normal population (2), although it is not known whether this occurs as a result of increased rates of DNA damage, a specific tendency to repair damage by homologous recombination or some other factor, such as an increased tendency for homologues to pair at mitosis. Holt et al. (2) estimated MR to occur at a rate of $\sim 1 \times 10^{-6}$ per locus per cell division in culture. Females showed higher MR levels than males, and rates increased with age. Variation in MR might therefore contribute to diseases such as cancer that depend on somatic mutations.

Shao et al. (6) showed that MR was decreased in crosses of distantly related strains of mice, whereas meiotic recombination was little affected. Similar effects have been found for the *Apc*^{Min} mutant mouse, which develops intestinal polyps when the wild type *Apc* allele is lost by MXO or GC; the frequency of allelic loss generally decreases when mouse strains are crossed and hence the number of polyps is reduced

commensurately (1). MR by MXO or GC therefore appears to depend greatly on the availability of closely homologous sequences, perhaps because this affects the probability of homologue pairing at mitosis.

How is MR Detected and Measured?

In mammalian cells, MR can be detected indirectly by observing reduction to homozygosity of polymorphic markers along a chromosome arm from the recombination site towards the telomere, together with the exclusion of alternative mechanisms, notably large deletions. Such a task is, in general, impractical as a means of screening for a rare event in large numbers of somatic cell clones (although for expansions of single clones, such as tumours or cases of UPD, it is feasible). Somatic cell clones that may have undergone MR are therefore generally identified initially in culture using a selectable marker which is heterozygous in the germline and for which homozygosity can be detected. Shao et al., for example, analysed cultured fibroblasts from *Aprt* +/- mice and selected for cells that had lost the wild type allele using 2,6-diaminopurine. Polymorphic markers showed about 80% of such losses to occur by MR. The estimated median frequency of cells with *Aprt* MR was 1.2×10^{-6} . MR breakpoints were not distributed randomly along the chromosome, although the reasons underlying breakpoint clustering were not clear. In a similar analysis in cultured human lymphocytes, Turner et al. found that oxidative damage by H₂O₂ predisposed to MR at the HLA locus. Other studies have, in general, found similar results and are reviewed by Wijnhoven et al. (9).

MR as a Mutagenic (or Sub-Optimal Repair) Process

In addition to inherited defects that may increase the rate of MR, 'spontaneous' GC or MXO (that is, at a normal frequency) may also contribute to disease processes as a side effect. In such a case, of course, it might well have been better for SCE to have occurred, or even for DSB repair to have failed and apoptosis to have been triggered. If GC or MXO happens to occur early in human development, it is one of several mechanisms whereby uniparental isodisomy may occur. Here, part or all of a chromosome is derived from a single parent. Various pathologies – or none – may occur, resulting either from unmasking of deleterious recessive alleles or from failure of controlled genetic imprinting. In the latter case, MXO has been implicated in mosaic forms of Beckwith-Wiedemann syndrome (UPD11p15.5) and in Prader-Willi and Angelman syndromes (UPD15q11-q13). The involvement of MR in the generation of UPD raises the question as to whether MXO, in particular, is a true mechanism of DNA repair, or a side effect of the preferred repair process, SCE.

MR and Tumorigenesis

Arguably, the most important occurrence of MR in humans is as a cause of loss of heterozygosity (LOH) in tumorigenesis. LOH is generally envisaged as the 'second hit' which knocks out a tumour suppressor gene (4). There is plentiful evidence to show that tumour suppressor genes often acquire a 'second hit' by LOH, but very few data to show the mechanism involved (MXO, GC, whole chromosome loss (with or without reduplication) or deletion (interstitial or terminal)). The reason for this lack of knowledge is that the underlying cause of LOH can only be determined by both knowing the pattern of apparent homozygosity (generally using polymorphic markers) and successfully making the distinction between true homozygosity and hemizyosity (using an *in situ* method such as FISH, or dosage quantitation, for example using real-time PCR or array-comparative genomic hybridisation). The type of 'second hit' is undoubtedly influenced by two factors, first the rate of occurrence of each mutation (small-scale change, large deletion, whole-chromosome loss, GC and MXO) and second the selective consequences of each change (which cannot readily be predicted and may depend on local sequence variation). Of the studies which have undertaken a comprehensive analysis of the mechanisms of LOH, there is good evidence that LOH at *RB1* in retinoblastoma almost always occurs by MXO, that MXO usually accounts for LOH at *NF1* in neurofibromatosis, but that MXO and GC are common at *APC* in human colorectal tumours. These examples are all probably early events in the pathogenesis of these tumour types. In contrast, there is evidence that many examples of LOH in malignant tumours result from chromosomal scale losses and gains, usually on a background of an aneuploid/polyploid karyotype. It is a moot point as to whether or not any human tumours gain an advantage by acquiring an increased tendency to MR as they do others forms of genetic instability, such as those resulting from silencing of DNA mismatch repair. It is notable that sporadic tumours very rarely harbour mutations in any of the known genes associated with MR, although this does not exclude other forms of aberrant expression or mutations in other, unknown genes in the same pathways.

Clinical Relevance

The role of mitotic recombination in the somatic inactivation of tumour suppressor genes and in generating parental isodisomy (UPD) has been described above. Genetic testing for mitotic recombination in clinical practice is generally related to the detection of isodisomy and is therefore indirect. Isodisomy analysis chiefly involves three diseases, Beckwith-Wiedemann, Prader-Willi and Angelman

syndromes. Other syndromes that are infrequently tested for isodisomy include transient neonatal diabetes (UPD6q), cystic fibrosis (UPD7q), Silver-Russell syndrome (UPD7p) and problems associated with UPD. Isodisomy testing generally uses multiple highly polymorphic markers within the region of the disease loci, searching for absence of heterozygosity. Distinguishing isodisomy from deletion therefore usually relies on a failure to demonstrate the latter using methods such as FISH. Isodisomy resulting from mitotic recombination is usually not distinguished from other mechanisms such as chromosomal loss and rescue.

Testing for mitotic recombination – or, more generally, for allelic loss – in cancers is rarely performed routinely in clinical practice. The value of testing for somatic genetic changes in tumours, for indicating outcome and response to therapy, is likely to become better recognised and routine in the coming years.

Conclusions

Mitotic recombination generally takes one of two forms; SCE, which is very difficult to detect and has no functional consequence or GC or MXO, which may have important effects on human disease. MR is used by the cell to repair DNA damage, in particular DSBs. SCE is perhaps best viewed as an efficient form of repair, generally likely to preserve genetic integrity, whereas GC or MXO are prone to exposing deleterious recessive mutations or, in some situations, causing defective imprinting. The existence of human diseases such as Bloom and Werner syndromes shows that excessive MR is pathogenic; whether the disease phenotype results primarily from increased MXO/GC rather than SCE is not known. Arguably, however, the most important "role" of MR in a pathological situation is as a cause of LOH in tumorigenesis. It appears that MXO and GC are not uncommon as the 'second hit' at tumour suppressor loci, but comprehensive studies are relatively rare and there is variation in the importance of MR among both genetic loci and tumour types. Future work will further characterise the biochemistry of MR in mammalian systems, define the roles of GC and MXO as both anti- and pro-mutagenic processes and search for germline variants which influence the rate of MR, both as a general process and in a site-specific fashion.

References

1. Haigis KM, Hoff PD, White A et al (2004) Tumor regionalism in the mouse intestine reflects the mechanism of loss of Apc function. *Proc Natl Acad Sci USA* 101:9769–9773
2. Holt D, Dreimanis M, Pfeiffer M et al (1999) Inter-individual variation in mitotic recombination. *Am J Hum Genet* 65:1423–1427

3. Johnson RD, Jasin M (2001) Double-strand-break-induced homologous recombination in mammalian cells. *Biochem Soc Trans* 29:196–201
4. Knudson AG (2001) Two genetic hits (more or less) to cancer. *Nat Rev Cancer* 1:157–162
5. Prado F, Cortes-Ledesma F, Huertas P et al (2003) Mitotic recombination in *Saccharomyces cerevisiae*. *Curr Genet* 42:185–198
6. Shao C, Stambrook PJ, Tischfield JA (2001) Mitotic recombination is suppressed by chromosomal divergence in hybrids of distantly related mouse strains. *Nat Genet* 28:169–172
7. Stern C (1960) A mosaic of *Drosophila* consisting of 1X, 2X and 3X tissue and its probable origin by mitotic non-disjunction. *Nature* 186:179–180
8. Kobbe C von, Karmakar P, Dawut L et al (2002) Colocalization, physical, and functional interaction between Werner and Bloom syndrome proteins. *J Biol Chem* 277:22035–22044
9. Wijnhoven SW, Van Sloun PP, Kool HJ et al (1998) Carcinogen-induced loss of heterozygosity at the *Aprt* locus in somatic cells of the mouse. *Proc Natl Acad Sci USA* 95:13759–13764

Mitotic Recombination

Definition

Mitotic recombination is a recombination event in a somatic cell that generates a diploid daughter cell with a combination of alleles different from that in the diploid parental cell.

► [Mouse Genomics](#)

Mitotic Spindle Poles

Definition

Mitotic Spindle poles are structures at opposite ends of the Mitotic Spindle.

► [Centromeres](#)

Mitotic Spindles

RYOKO KURIYAMA

Department of Genetics, Cell Biology, and Development, University of Minnesota, MN, USA
ryoko@lenti.med.umn.edu

Definition

The mitotic spindle is a special device essential for chromosome segregation (► [mitosis](#)) during ► [cell division](#). It consists of three parts, chromosomes, two asters and a spindle region (Fig. 1). The overall shape of the mitotic spindle is due to ► [microtubules](#), which are made of α and β -tubulin subunits. Microtubules are dynamic, changing their length all the time by repeating growing and shrinking cycles at both ends. This microtubule property is known as ► [dynamic instability](#). One end of the microtubule (minus end) is associated with the pole where the microtubule-organizing center (MTOC) called the ► [centrosome](#) is located. The other end (plus end) is attached to a special region of the chromosome called the ► [kinetochore](#) to form ► [kinetochore](#) microtubules. Some microtubules have a free plus end, ► [astral microtubules](#) splay out widely from each pole, other microtubules originate from opposite poles and overlap in the middle of the spindle to constitute ► [interpolar microtubules](#) (Fig. 1). The mitotic spindle is transient as it is assembled at the onset of cell division phase/M phase and disappears entirely after completion of mitosis. Because it is composed of dynamic microtubules, the mitotic spindle is also dynamic, undergoing continuous morphological alterations in mitosis. Mitosis is divided into five distinct stages, prophase, prometaphase, metaphase, anaphase and telophase (Figs. 1, Parts 2–8).

Characteristics

Macromolecular Composition

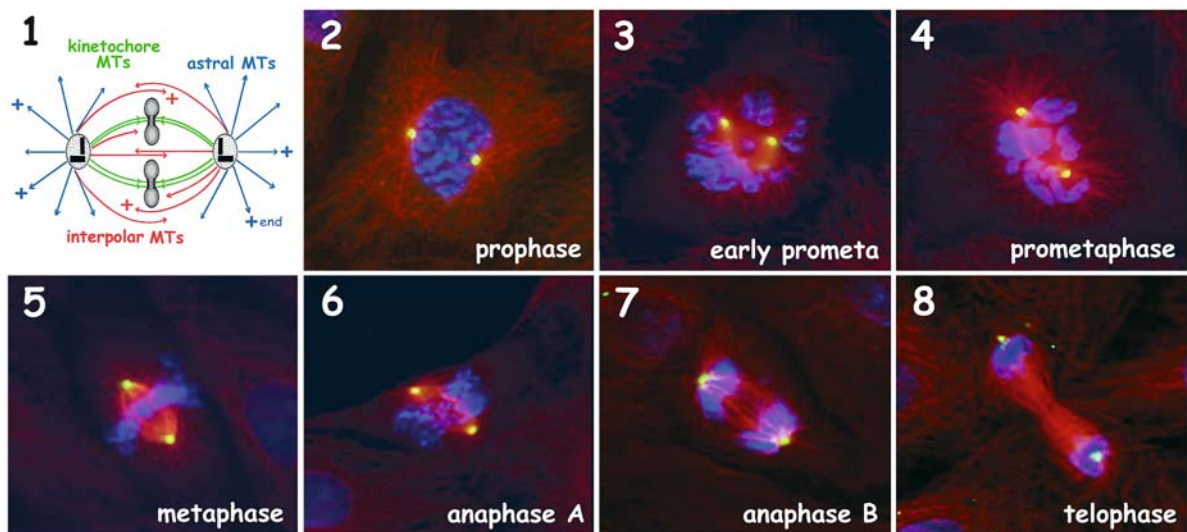
The spindle consists of a number of molecules, which have been identified using multiple approaches, including biochemical analysis of isolated spindles, genetic analysis of genes affecting spindle functions and immunological analysis of anti-spindle antibodies. A major component of the spindle is the microtubule, whose function is largely dependent on its associating molecules collectively called ► [microtubule-associated proteins](#), or MAPs. The spindle has an abundant variety of MAPs, which can be classified into two types, motor MAPs and non-motor MAPs. The mechanochemical motor proteins (cytoplasmic dynein and kinesin-like proteins) have been subjected to extensive analysis. Although less is known of the non-motor MAPs, evidence has been accumulated to indicate that they play an important role in modulating the dynamics of spindle microtubules, in particular at the kinetochore region. Phosphoproteins and kinase/phosphatase enzymes are numerous in the spindle, which is consistent with the fact that the M phase has been characterized as that with an increased level of protein phosphorylation and kinase activities. The mitotic spindle recruits signaling molecules important for controlling the mitotic progression and completion of cell division;

these components tend to be positioned at particular locations, such as poles, kinetochores and the spindle midzone. Although many other molecules have been detected in the spindle, their roles in the structure and function of the spindle have not yet been explored.

Spindle Poles

The pole has major three functions: 1) organizing spindle microtubules, 2) serving as the locus toward which the chromosomes move in anaphase and 3) specifying the axis that determines the orientation and position of the plane for cytoplasmic division. In animal cells, the morphological expression of the poles is the centrosome, which consists of a pair of centrioles and a surrounding dense pericentriolar material. By

serving as an MTOC, centrosomes are responsible for nucleation, orientation and anchorage of microtubules. Individual microtubules are nucleated from ~ 25 nm ring-shaped structures (γ -TuRC: γ -tubulin ring complex) composed of a ubiquitous centrosome protein, γ -tubulin, plus additional minor components. The microtubule-nucleating sites are held together and connected to coiled-coil proteins, which are abundant in the centrosome and presumably maintain the overall shape of the pericentriolar material by serving as the centrosome scaffold. To anchor the microtubule minus ends, molecules with an affinity for the end and/or wall of microtubules are embedded in the pericentriolar material. It has become apparent that the centrosome is involved not only in microtubule organization but also



Mitotic Spindles. Figure 1 (1) Diagram of the mitotic spindle consisting of chromosomes, two asters and a spindle part. The overall shape of the spindle is framed by three types of spindle microtubules: kinetochore microtubules (green), astral microtubules (blue), and interpolar microtubules (red). Microtubules are a polarized structure containing two distinct ends, the fast growing (plus) end and slow growing (minus) end. All minus ends are attached to the spindle pole where the microtubule-organizing center, the centrosome, is located. While astral and interpolar microtubules have a free plus end, kinetochore microtubules interact with the kinetochore on the chromosome at their plus end. The change in the shape of mitotic spindles in mammalian cells fluorescently stained with DAPI (DNA: blue), anti-tubulin (microtubules: red), and anti- γ -tubulin (centrosome: green) antibodies. At the onset of mitosis (prophase, 2), DNA-containing materials condense to form well-defined chromosomes. Each chromosome is composed of two genetically identical parts called chromatids and two sister chromatids are bound together back-to-back in a single unit. Toward the end of prophase, two mitotic poles separate and start to organize spindle microtubules. Prometaphase is defined as the stage of nuclear membrane breakdown (3 and 4). As a result of the disappearance of the nuclear membrane, the spindle microtubules, which have been lying outside the nucleus, are able to enter the nuclear region and interact with chromosomes through a special region called the kinetochore. Once linked to the microtubules, the chromosomes that are initially scattered throughout the nuclear region start to align in one plane midway between the poles. All chromosomes align at the metaphase plate in metaphase cells (5). Sister chromatids move toward the opposite poles in anaphase, which is traditionally divided into two phases, anaphase A and anaphase B. During anaphase A, the chromosomes move to the pole, causing the distance between the pole and chromosomes to decrease (6). During anaphase B, microtubules elongate and the distance between poles increases (7). In many cases, anaphase A and anaphase B occur simultaneously. After arriving at the pole, the separated chromatids, now called daughter chromosomes, decondense and a new nuclear envelope forms at the end of mitosis (telophase: 8).

in regulation of a wide range of cellular activities, such as signal transduction and cell-cycle control. A number of molecules with diverse functions are thus likely to be targeted to the centrosomes/poles in a stage specific manner.

Kinetochores

The kinetochore is a multiprotein complex assembled on the ►centromere located at the primary constriction of the chromosome. It consists of two distinct regions (inner kinetochore and outer kinetochore) and has three functions: 1) serving as the site where spindle microtubules attach to the chromosomes, 2) generating the forces necessary for chromosomes to align at the spindle midzone and move toward the pole and finally 3) determining the timing of anaphase onset. Because a kinetochore is included in each chromatid, two kinetochores are arranged back to back and the microtubule attachment sites face opposite directions in a single chromosome. To capture the microtubule end, both motor and non-motor MAPs reside in the kinetochore. They bind microtubules and alter the dynamic nature of microtubules, which ultimately leads to the changes in the relative position of the chromosome between the two poles in prometaphase cells and the distance between pole and chromosome in anaphase cells. Signaling molecules that control the time of anaphase onset become recruited to the outer region of the kinetochore.

Central Spindles

As chromosomes move toward the pole, the midzonal region of the spindle starts to organize the central spindle. It consists of highly bundled interpolar microtubules with a dense material surrounding antiparallel microtubules at the equator. With the progression of cytoplasmic constriction, the central spindle transforms to the midbody in the center of which is a coalesced dense dot. The central spindle/midbodies accommodate a number of components; some are shifting from the entire spindle region to the midzone during chromosome movement, whereas others are chromosomal passenger proteins that are centromere/kinetochore proteins but remain at the spindle equator after chromosome movement rather than shifting to the spindle pole along with chromosomes in anaphase. Recent evidence has suggested that the central spindle and midbodies play an important role in controlling cytoplasmic division, in particular completion of cell separation to liberate two daughter cells at the end of cell division.

Visualization of the Mitotic Spindle

Because it occupies a large part of the cell and changes its shape drastically during mitosis, the mitotic spindle

is easy to detect with a light microscope. Using microtubule-stabilizing reagents such as taxol/paclitaxel, hexylene glycol and glycerol, mitotic spindles have been isolated from naturally synchronized marine invertebrate eggs as well as synchronized cultured mammalian cells. Biochemical quantities of highly purified spindle isolates are useful for both morphological and biochemical analysis of the spindle. However, because no chromosome movement has yet been demonstrated, isolated spindles do not allow us to investigate the dynamic nature of the spindle and spindle microtubules *in vitro*. Individual microtubules have been successfully visualized by fluorescence microscopy using cells loaded with fluorescently tagged tubulin molecules. By monitoring the position of microtubules labeled by irradiation, it is possible to measure the dynamics of spindle and spindle microtubules by time-lapse video fluorescence microscopy.

Mechanism of Chromosome Movement

Because the mitotic spindle undergoes significant morphological alterations during mitosis, factors governing microtubule dynamics must be the key to the mechanism of chromosome movement. M phase onset triggers the global change in microtubule organization, which is caused by the increased level of both the microtubule-nucleating activity of the centrosome and the frequency of catastrophes defined as the transition from microtubule growth to shrinkage. MAPs and small molecular weight compounds (such as Kin I, XMAP215 and Op18) are likely to be responsible for the change in microtubule dynamics at the phase transition from interphase to M phase. This change results in the formation of many short microtubules from each spindle pole; while some keep growing at their plus end, others shrink rapidly. Once their plus end is captured by MAPs at the kinetochore, these kinetochore microtubules become stabilized by preventing shrinkage. Simultaneously, MAPs located at the spindle midzone crosslink antiparallel microtubules to stabilize interpolar microtubules. MAPs and motor activities are also indispensable for the assembly and maintenance of normal size bipolar spindles by separating spindle poles and generating forces acting at different parts of the spindle structure.

During prometaphase, chromosomes move back and forth between two poles until they align properly at the spindle equator, called the ►metaphase plate. Although no precise molecular mechanisms have yet been established, it is tempting to believe that a variety of molecules located at the kinetochores/centromeres participate in regulating the length of kinetochore microtubules by mediating addition (microtubule

growth) or removal (microtubule shrinkage) of tubulin subunits at the plus end. These molecules include motor proteins, such as the cytoplasmic dynein complex, CENP-E, Kin I and chromokinesin, as well as regulatory MAPs and kinases, including NuMA, EB1 and [▶chromosome passenger proteins](#).

The cell is ready to enter anaphase when all chromosomes are aligned at the spindle equator. A complex of signaling molecules, consisting of Mad1, Mad2, Bub1, Bub3, BubR1 and others, is recruited to the outer layer of the kinetochore. Upon attachment to the plus end of microtubules, these signaling molecules become silenced. This is the way in which cells sense that all chromosomes are associated with kinetochore microtubules and aligned at the metaphase plate. The presence of a single non-attached kinetochore causes the delay of anaphase onset. This surveillance mechanism is called the [▶spindle checkpoint](#). As long as this wait signal is on, Cdc20, a part of APC/C (anaphase promoting complex/cyclosome) required for protein degradation, is inactive and does not allow APC/C to recruit a substrate, securin. Securin binds to separase, resulting in the inhibition of separase activity in digesting SCC1. To activate the protease activity of separase, securin must be removed from separase. Because SCC1 is a part of the multiprotein complex called cohesin, which glues two sister chromatids together, SCC1 digestion causes cohesin cleavage and sister chromatid separation. Therefore, the activation of Cdc20 and APC/C is a prerequisite for chromosome segregation.

During anaphase A, chromosomes move toward the pole and the distance between pole and chromosome decreases. Two mechanisms have been proposed to account for the chromosome movement. One is that the minus-end-directed motor proteins present in the kinetochore generate the force necessary to pull the kinetochore towards the pole; the other is depolymerization of kinetochore microtubules at both plus and minus ends, which is probably mediated by microtubule destabilizing Kin I type motor proteins located at the kinetochores and spindle poles. As microtubule ends disappear, the kinetochore tends to slide toward the poles in order to retain the microtubule binding; this results in chromosome movement toward the pole. During anaphase B, the spindle elongates and the distance pole-to-pole increases. This is believed to be mediated by microtubule sliding at the spindle midzone. As new tubulin subunits are incorporated to the plus end, the length of interpolar microtubules increases. The motor proteins located at the area where microtubules originating from opposite poles overlap generate the forces necessary for antiparallel microtubules to push away. Thus, the degree of microtubule overlap decreases at the interzone as the spindle elongates.

Clinical Relevance

The mitotic spindle is responsible for faithful transmission of the genetic material from one cell to its progeny. Equal chromosome segregation is ensured by establishment of spindle bipolarity. Thus the proper number of spindle poles/centrosomes is strictly maintained in the cell. Any defects in the spindle and spindle poles/centrosomes induce cells with abnormal numbers of chromosomes, most notably cancer cells. In fact, genetic instability is a hallmark of all human cancers. It is easily imagined that molecules controlling the precise number of centrosomes and proper assembly and function of spindles play important roles in tumorigenesis. When over-expressed, the spindle/centrosome components malfunction, ultimately leading to genetic instability in cancer cells. This is consistent with the fact that many oncoproteins identified thus far have turned out to be spindle components, Aurora kinases, polo-like kinases, MAPs such as TACCs (transforming acidic coiled-coil proteins) and XMAP215/ch-TOG (colonic hepatic tumor-over-expressed gene) and spindle checkpoint proteins including Mad2 and Bub1. These components must be of great clinical potential as markers for molecular diagnosis of human cancers and as targets for therapeutic development.

[▶Cell Cycle – Overview](#)

[▶Centromere](#)

References

1. Cleveland DW, Mao Y, Sullivan KF (2003) Centromeres and kinetochores: from epigenetics to mitotic checkpoint signaling. *Cell* 112:407–421
2. Mitchison T, Kirschner M (1984) Dynamic instability of microtubule growth. *Nature* 21:237–242
3. Mitchison TJ, Salmon ED (2001) Mitosis: a history of division. *Nat Cell Biol* 3:E17–E21
4. Rieder CL, Khodjakov A (2003) Mitosis through the microscope: advances in seeing inside live dividing cells. *Science* 300:91–96
5. Wilson EB (1928) *The Cell in Development and Heredity*, 3rd edn. MacMillan, New York
6. Wittman T, Hyman A, Desai A (2001) The spindle: a dynamic assembly of microtubules and motors. *Nat Cell Biol* 3:E28–E34

Mitral Valve Prolapse

Definition

Mitral valve prolapse is a floppiness of the mitral valve, which separates the left atrium and left ventricle from one another.

[▶Marfan Syndrome](#)

Mixed Disulfide

Definition

Mixed disulfide describes a disulfide bond between a cysteine residue of a protein and a thiol-containing redox reagent such as glutathione or dithiothreitol.

► [Protein Disulfide Bonds](#)

MLH

Definition

MutL homologue (MLH) are human mismatch repair genes, all related to the bacterial mutL gene; mutations in that gene may account for hereditary non-polyposis colon cancer in humans.

► [Hereditary Non-Polyposis Colorectal Cancer](#)

Mixed Gonadal Dysgenesis

► [SRY – Sex Reversal](#)

Mixing Time

Definition

Mixing time is the time period in a multidimensional NMR experiment during which a transfer of magnetization from one spin to another is accomplished by some form of mixing sequence, consisting of several radio frequency pulses and delays.

► [Multidimensional NMR Spectroscopy](#)

MM-PBSA

Definition

MM-PBSA stands for the Molecular mechanics Poisson-Boltzmann surface area method. It obtains free energy estimates from the analysis of an MD trajectory using molecular mechanics force field terms, Poisson-Boltzmann electrostatic calculations and a term that depends on the solvent-accessible surface area, as well as an entropy term.

► [Molecular Dynamics Simulation in Drug Design](#)

MJD

► [Machado-Joseph Disease](#)

MMPs

► [Matrix Metalloproteinases](#)

MMR

► [Mismatch Repair](#)

MLC

Myosin Light Chain.

► [Signal Transduction: Integrin-Mediated Pathways](#)

MMTV

Mouse Mammary Tumor Virus

► [Retroviruses](#)

Mobile Fraction

Mobile fraction denotes the percentage of fluorescent molecules in a sample available to contribute to fluorescence recovery during the time course of a FRAP (fluorescence recovery after photobleaching) experiment. Note that this measure is used to assess the mobility of fluorophores within a continuous compartment.

►FRAP

Mobile Genetic Elements

►Transposons

Mode of Release

Definition

Neurotransmitters may be released from synaptic terminals in two different modes: the release may be uni-vesicular, i.e. one vesicle is released per synaptic contact per incoming action potential (all-or-none release), or the release may be multi-vesicular (more than one vesicle per incoming action potential).

►Neurons

Modified Antisense Oligonucleotide

Definition

Antisense reagents act through the complementarity between nucleic acid strands. Antisense oligonucleotides, however, are generally rapidly degraded. To increase their stability, and therefore biological efficacy, oligonucleotides can be modified. Typically, the phosphodiester bond between nucleotides is modified. In morpholinos, this linkage is replaced by a non-ionic phosphorodiamidate linkage. Another non-proprietary modification involves the use of phosphorothioate linkages.

►Morpholino Oligonucleotides, Functional Genomics by Gene ‘Knock-down’

►*Xenopus* as a Model Organism for Functional Genomics

Modifiers

Definition

Modifiers are alleles of genes that modify (increase or decrease) the severity of the phenotype caused by mutations in other genes.

►Mouse Genomics

MODY

Definition

Maturity onset diabetes of the young (MODY) is a less common form of diabetes, characterized by autosomal dominant transmission and by early age of onset of the disease, similar to type 2 diabetes.

►Diabetes Mellitus, Genetics

Moesin

►ERM Protein (Ezrin, Radixin, Moesin)

Molecular Aging Research

CHRISTIAN BEHL

Department of Pathobiochemistry, Johannes Gutenberg University, Mainz, Germany
cbehl@uni-mainz.de

Definition

At the beginning of the last century the average life expectancy was about 37 years in the developed

countries. Today life expectancy is approximately 70 years with a tendency to increase in the next decades. As early as the year 2050, approximately 25% of the population will have reached an age over 65. Consequently, as people are getting older, more and more will be affected by various age-associated disease including

1. cardiovascular disorders,
2. cerebrovascular diseases,
3. malignant tumors and cancer
4. chronic progressive neurodegenerative disorders.

In human ►aging an increase in the incidence of age-associated disorders is the cause of a decreased survival rate or an increased mortality. In addition to life threatening disorders that may lead to death in elderly people, a variety of age-associated changes lead to dysfunctions of different organs that include the brain. In particular, various neuropsychiatric disorders that frequently occur in aged people are, although not directly life-threatening, a major burden for the aged human society. Important examples of this are the increased occurrence of depression and impairments of cognition. In summary, the actual phenotype of aged humans includes various pathological changes and is investigated experimentally in different model systems. Here the focus is on the cellular and molecular level of the analysis of the aging phenotype.

Different model systems have been developed to study aging processes at the cellular and molecular level. These include

1. cultivated fibroblasts
2. cultures of intact tissues (biopsies)
3. *Saccharomyces cerevisiae*
4. ►*Caenorhabditis elegans*
5. ►*Drosophila melanogaster*
6. *Mus musculus*

Employing these models the molecular and biochemical changes in the cell as well as in tissue can be investigated. Animal models of aging help to validate cellular and molecular findings in an intact organism. Based on such models different theories of aging have been developed.

Theories of Aging

By looking at the age-associated changes in cells and organisms from different angles the following main theories of aging have been developed

1. cross-linking theory that focuses on the chemical reaction between cellular proteins
2. immune theory that points to a dysfunction of the immune system leading to chronic inflammatory changes
3. senescence gene theory which defines specific age-associated gene products

4. free radical theory that focuses on the detrimental activity of accumulating free oxygen and nitrogen radicals
5. ►telomere theory which summarizes the continuing shortening of the cellular DNA (telomere shortening)
6. evolution theory concluding that aging is not genetically determined but rather a consequence of a continuing accumulation of somatic changes.

Although all these different theories have a wide experimental basis, the discussion will be continued with those research approaches that are at the moment followed most intensively.

Characteristics and Clinical References

Most of the actual research approaches that investigate age-associated changes in cells and organisms are based on three major theories, the senescence gene theory, the free radical theory and the telomere theory.

Senescence Gene Theory: Syndromes of Accelerated Aging in Humans

In the search for the genetic control of aging and correspondingly for the cause of long life syndromes of accelerated aging are mainly investigated. For example, ►Werner's syndrome is characterized through the early onset of different signs of human aging including the graying of the hair, skin atrophy, cataracts, arteriosclerosis, osteoporosis, benign and malign neoplasias and others. The patients mostly die early at approximately 47 years, either in consequence of a heart attack or due to cancer. Werner's syndrome is an autosomal recessive disease and the incidence of the homozygous occurrence is approximately 1–22 per million. Werner's syndrome is frequently also named progeria of the adult. The Werner's syndrome gene (WRN) has been identified and is in a mutated form responsible for the syndrome. WRN codes for a protein that belongs to the family of RecQ ►helicases and is localized on the short arm of chromosome 8. Further prominent examples of syndromes of accelerated aging are ►Down syndrome and ►Bloom syndrome.

Free Radical Theory

The free radical theory of aging is based on the detrimental effects of accumulated free oxygen and nitrogen radicals that attack macromolecules of the cell. Oxidation of functional cellular molecules leads to dysfunction and to detrimental changes. During evolution our life has developed under the continuous influence of the regulatory and destructive activities of molecular oxygen and oxygen radicals. Of course, a variety of fine tuned antioxidant systems have been developed for defence against free radical attacks. Such antioxidant systems include enzymes (e.g. catalase,

glutathione peroxidase, superoxide dismutase) as well as antioxidant compounds (e.g. vitamin C, vitamin E). The production of ►free radicals in the cell is a normal physiological process and such radicals also exert regulatory functions with respect to gene transcription. Their production is controlled and regulated and an accumulation of radicals is prevented by the activity of the antioxidant systems. Nevertheless, as a consequence of various exogenous and environmental changes or due to toxic stimuli, this fine tuned balance can be disturbed, leading to the accumulation of radicals that destroy vital functions in the cell. Such an unbalance is called oxidative stress. Interestingly, oxidative stress is not only a continuous companion of human aging but has also been nominated as one major player in a variety of non-neuronal and neuronal degenerative disorders including Parkinson's disease, Alzheimer's disease, amyotrophic lateral sclerosis, and stroke. The actual triggers for oxidative stress have been defined for a variety of neurodegenerative disorders and a common hypothesis of protein aggregation as one cause of an increased oxidative stress in neuronal tissue has been proposed. In consequence of the high impact of oxidations on nerve cell survival and on the development of chronic age associated neurodegenerative disorders, the concept of the use of antioxidants for the prevention of age-related dysfunctions has been intensively studied.

The most important antioxidants that contribute to cell protection in the human are vitamin C and vitamin E and these antioxidants have been successfully used in model organisms such as *Drosophila melanogaster* to increase life span. In addition, the combined over-expression of the gene Cu/Zn-SOD and of catalase in transgenic *Drosophila* flies also increased the average and maximal life of those animals. Furthermore, this model has been extended to mammalian organisms and synthetic SOD and catalase mimetica indeed increase the survival of transgenic mice that represent a model of amyotrophic lateral sclerosis. SOD and catalase mimetica also increase the life span of the nematode *Caenorhabditis elegans*. In summary, the following experimental findings strongly support the importance of the free radical theory of aging:

1. The over-expression of antioxidant enzymes delays age-associated damage and increases the maximal life span of model organisms.
2. The life expectancies of different species are indirectly proportional to the rate of increase of superoxide radicals and of hydrogen peroxide in mitochondria.
3. Caloric restriction and/or controlled fat and cholesterol metabolism decreases oxidative stress due to decreased metabolism in mitochondria and leads to decreased oxidative damage that consequently

delays age associated changes. Moreover, caloric restriction has been shown to increase the maximal life span in mammals.

Telomere Theory

The chromosomes of all eukaryotes are linear structures with telomeres at their end. The replication of cellular DNA during the S-phase of the cell cycle starts with the synthetic extension of small RNA stretches (RNA-primers) through the activity of DNA polymerase. RNA-primers are displaced after the synthesis of the DNA and gaps are filled up with DNA. The replacement of RNA-primer stretches with DNA can occur along the whole replication fork with extension of the five prime ends of the DNA stretch. In consequence, the newly synthesized DNA stretch is shortened by the corresponding length of the RNA-primers at the end of the stretch. This process is the molecular basis of the replication problem. These end stretches are called the telomeres. The DNA sequences of the telomeres do not carry coding sequences and can be missed on the basis of their genetic information. Nevertheless, they have important functions during replication. For instance, they prevent the chromosome end being recognized at double-strand break and influence the exact pairing of the nucleotide stretches, as well as the movement of the chromosomes through mitosis and meiosis. The structure of telomeres and different telomere binding proteins that can affect telomere function are known. The constant shortening of the telomeres is the biochemical basis of the ►Hayflick-limit. Among the telomere-associated and binding proteins ►telomerase is centrally involved in the regulation of telomere structure and function. The modulation of telomerase activity has been investigated as one possibility to prevent the replication problem and therefore the Hayflick-limit and to extend life span through the prevention of telomere shortening.

Mouse Models of Human Aging

As mentioned before, the phenotype of human aging involves a variety of different pathological changes. Consequently, a variety of mouse models that reflect single aspects of this phenotype of human aging have been developed. Some major models are mentioned here:

Klotho-Mouse: The Klotho gene codes for a transmembrane protein and its defect leads to a decreased life span, skin atrophy, infertility and osteoporosis as seen in Klotho knock-out mice.

Senescence accelerated-mice (SAM): mouse colonies that suffer from accelerated aging have been selected by breeding and experimental selection. The genetic changes in these mouse lines obviously appear to be of multigenetic origin and are under investigation. Some

of these colonies are associated with increased neoplasias, cataracts or brain atrophy.

Werner syndrome mouse: Mice that carry a mutated WRN gene were developed and the fibroblasts of these mice have a strongly reduced proliferative capacity. Otherwise these mice have no obvious phenotype.

Telomerase knock-out mice: Mice with a targeted disruption of the gene that codes for the RNA component of telomerase show infertility, reduced wound healing and increased incidence of spontaneous malign neoplasia among other changes. Despite the fact that these mice show a phenotype that reflects important aspects of human aging, these animals do not show any further signs of the age-associated diseases that can frequently be found in humans, such as arteriosclerosis, osteoporosis and brain atrophy. Therefore, it has been proposed that the shortening of the telomeres and the induction of cellular senescence *in vivo* do not directly lead to pathological changes characteristic of human aging.

Growth hormone overexpression: Chronic excess of growth hormone is associated with symptoms that are typical for mammalian aging. Transgenic mice that over-express growth hormone show signs of accelerated aging, including a decreased life span, loss of weight and reduced fertility.

In studying the different mouse models of human aging and the genotype *versus* phenotype it becomes clear that aging is a biological phenomenon that involves multiple genetic as well as environmental signals.

Frontiers of Aging—an Outlook

The next steps towards state of the art molecular and biochemical age research are the integration and combination of various expertises, including biochemistry, endocrinology, neuroscience, genetics, physiology, population biology, epidemiology and, last but not least, clinical gerontology. Novel technologies in functional genomics and proteomics will help to clarify functions of age-related genes and proteins. One example should show how powerful technology such as expression profiling could help to identify age-relevant gene expression by using gene chip technology; the gene expression changes in aging brain have been analyzed in mice. Interestingly, many expression changes have been found that reflect changes that have also been described for human neurodegenerative diseases (8).

What becomes of increasing importance is to not only try to identify genes that might be involved in the aging processes (if such genes really do exist), but rather to look at functions of genes and their gene products which are of vital importance under aged conditions. For instance, we need to investigate the activity of molecular signals in detail in young and aged cell model systems and organisms in parallel. Our own lab

is involved in the analysis of changes in intracellular signaling by comparing different young and aged cell types. Moreover, we are looking into functions of estrogen receptors and possible changes in estrogen receptor signaling depending on the age and proliferation states of cultured cells. Gene expression does change with age and the exact mechanisms that mediate these alterations need to be identified. Consequently, molecular aging research is not restricted to the study of cell free systems or cultured cells or even genetic models such as knockout and transgenic mice. From early in the investigations the human situation also needs to be included, which can be done by using human material, for instance cultivated primary human fibroblasts from human donors. Since the increased life span of the human society in industrial countries is a fact and the increased incidence of age-associated disorders is the direct consequence of this fact, the effort and the interdisciplinary approaches to studying the basis of human aging need to be implemented in order to circumvent and prevent age-associated diseases and to provide the possibility of a long and healthy life.

References

1. Collins K (2000) Mammalian telomeres and telomerase. *Curr Opin Cell Biol* 12:378–383
2. Finkel T, Holbrook NJ (2000) Oxidants, oxidative stress and the biology of ageing. *Nature* 408:239–247
3. Harley CB, Futcher AB, Greider CW (1990) Telomeres shorten during ageing of human fibroblasts. *Nature* 345:458–460
4. Harman D (1956) Aging: a theory based on free radical and radiation chemistry. *J Gerontol* 2:298–300
5. Hayflick L (1965) The limited *in vitro* lifetime of human diploid cell strains. *Exp Cell Res* 37:614–636
6. Melov S, Ravenscroft J, Malik S, Gill MS, Walker DW, Clayton PE, Wallace DC et al (2000) Extension of lifespan with superoxide dismutase/catalase mimetics. *Science* 289:1567–1569
7. Miki T, Morishima A, Nakura J (2000) The genes responsible for human progeroid syndromes. *Intern Med* 39:327–328
8. Lee CK, Weindruch R, Prolla TA (2000) Gene-expression profile of the ageing brain in mice. *Nat Genet* 25:294–297

Molecular Chaperones and Cochaperones

Definition

Molecular chaperones and cochaperones comprise of a large group of unrelated protein families whose

function is to stabilize unfolded proteins, unfold them for translocation across membranes or for degradation, and/or to assist in their correct folding and assembly and prevent undesirable oligomerization. Chaperones are not themselves components of the final functional structure.

► [Glycosylation of Proteins](#)

► [Nuclear Hormone Receptors: Regulators of Gene Transcription and Cellular Signaling](#)

► [Protein Folding](#)

Molecular Clock

Definition

The molecular clock approximation models evolutionary changes in proteins by a constant rate, i.e. the same number of substitution always occurs in the same time interval.

► [Sequence Annotation in Evolution](#)

Molecular Complementarity

Definition

The specific recognition of one molecule by another is very important for biochemical and cellular function. The molecular complementarity describes the geometrical fit, as well as the understanding of enthalpic and entropic contribution, to the stability of the protein-ligand complexes.

► [Classification of Active Centers](#)

Molecular Docking

KARIN SCHLEINKOFER¹, TING WANG²,
REBECCA C. WADE²

¹Department of Bioinformatics, Biocenter, University of Würzburg, Würzburg, Germany

²Molecular and Cellular Modeling Group, EML Research gGmbH, Heidelberg, Germany
rebecca.wade@eml-r.villa-bosch.de

Definition

Molecular docking is the computer-aided prediction of the bound geometry of two or more molecules. Molecules may be docked manually with the aid of computer graphics or automatically by using computer algorithms.

Manual docking is done using molecular visualization software, often in conjunction with hardware for stereo viewing and sometimes with haptic devices that provide force feedback to the user's hand.

There are a variety of algorithms for automatic docking (for reviews of methodology and applications see 1, 2, 3, 4). Rigid-body and ► [flexible docking](#) procedures are distinguished by their treatment of molecular flexibility and conformational changes upon binding. Docking procedures are developed and chosen according to the types and numbers of molecules to be docked. Different algorithms are generally used for docking two macromolecules as opposed to a macromolecule and a low molecular weight compound. More computationally demanding and accurate procedures can be applied to the docking of a single ligand to a receptor than to the screening of a large library of compounds for those that bind to a macromolecular target.

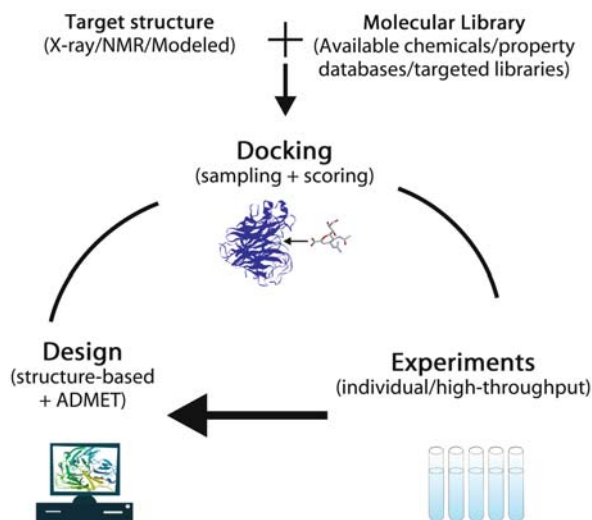
Docking is an established methodology in the structure-based drug design cycle (Fig. 1) and will play an increasingly important role in the exploitation of results from structural genomics and proteomics projects.

Description

The docking process can be divided into three steps: (i) identification of the ligand-binding site on the target macromolecular receptor, (ii) sampling of possible geometric arrangements of ligand and receptor and (iii) scoring of the sampled arrangements. These three steps are considered in turn here.

In most docking procedures, the ligand binding site on the target macromolecule should be defined prior to sampling. The binding site can often be identified by comparison with structures of the target protein co-crystallized with different ligands or with structures of homologues of the target protein. Putative docking sites may also be identified by automatic binding site detection algorithms. Potential small molecule binding sites can be found by searching for concave surfaces on the target protein structure. Protein-protein binding sites can be identified on the basis of a combination of properties including amino acid sequence conservation, amino acid type and surface curvature.

Sampling procedures can be separated into two classes, rigid-body docking and flexible-body docking. The rigid-body docking model is historically the first docking approach. The ligand and receptor molecules are treated as rigid bodies and the flexibility of the interacting partners is not considered explicitly.



Molecular Docking. Figure 1 Scheme showing the role of molecular docking in the structure-based drug design process.

However, molecular flexibility and conformational changes upon binding may be implicitly accounted for by, for example, reducing atomic radii or removing the side-chains of high mobility residues. In addition, rigid-body docking may be performed for an ensemble of rigid ligand or receptor conformations that represent the conformational variability of the molecules. The advantage of rigid-body docking is that the docking problem for two molecules is reduced to a search through three translational and three rotational degrees of freedom. There are many different techniques for performing this sampling. Major classes are complete enumeration, geometric hashing, and gradient-based techniques.

Complete enumeration

- Many rigid-body docking methods, particularly those applicable to protein-protein docking, make use of complete enumeration for shape-based docking. Typically, the molecules are discretized onto grids with values assigned at each grid point to show whether the grid point is buried on or near the molecular surface or outside in solution. Values to describe electrostatic or chemical properties can also be assigned to the grid points. For each relative orientation of the two molecules, all possible translations of the molecules are scanned by superposition of the grid points. The computational efficiency of this approach can be improved by applying a convolution strategy in which fast Fourier transforms are used to compute a correlation function for each relative orientation of the two molecules. Reverse Fourier transforming then permits scoring of geometric and electrostatic matching

of all possible translations at this particular orientation. After searching all orientations, a set of potential docked arrangements is ranked.

Geometric hashing

- As alternatives to complete enumeration, a number of techniques perform docking by matching important features of ligand and receptor. This strategy was adopted in one of the earliest approaches for using docking in structure-based drug design that was introduced in the DOCK program by Kuntz and colleagues. In this method, a binding site on a target receptor is filled with spheres that together represent the binding site's shape. A ligand is then docked by optimizing the superposition of its atoms onto the spheres in the binding site. Many docking programs make use of a computer-vision-based method called geometric hashing. Matching is performed for defined critical points in different parts of the ligand in local coordinate systems. These matches are ranked. Then full docking of the complete ligand is performed for the highest scoring local coordinate systems. Finally, the docked structures are filtered and re-ranked.

Gradient-based techniques

- Gradient-based techniques for docking make use of forces between the molecules that are defined by a [molecular mechanics force field](#) or by a simplified function. Sampling may be performed by energy minimization, Monte Carlo simulation, simulated annealing or Brownian dynamics.

Currently, most flexible docking procedures incorporate explicit treatment of the conformational flexibility of low molecular weight ligands only. Receptor conformational flexibility is usually neglected because of the large number of degrees of freedom of macromolecules. Flexible docking is computationally more demanding than rigid-body docking because more conformational degrees of freedom have to be taken into account. There are various algorithms that address the challenge of flexible docking. A few examples are:

- Incremental ligand construction. The ligand is assembled in the binding pocket by incremental connection of fragments (building blocks). Only energetically favorable additions enter the next iterative step. As a prerequisite, an initial base fragment should be positioned in the binding pocket; this is usually a rigid part of the molecule, such as a ring system.
- Distance geometry. In this formalism, the ligand conformation and location are described by a distance matrix containing intra- and inter-molecular distances; these distances are optimized to match the receptor binding properties.
- Genetic algorithms. These mimic the concept of Darwinian evolution and require the generation of an

initial population of ligand-receptor configurations that evolve by application of a fitness function aimed at improving the binding affinity.

- Monte Carlo simulation. At each step of a series of iterations, the ligand is randomly translated, rotated or distorted in the binding site of the receptor. The movement of the ligand is accepted if it results in a lower energy. If it results in a higher energy, the movement of the ligand is accepted or rejected according to a Boltzmann probability criterion.
- **►Molecular dynamics simulation.** Docking is based on molecular mechanics force fields that are used to calculate the forces on the atoms. Newton's equations of motion are solved at a series of time steps to generate the motions of the atoms resulting from the molecular mechanics forces. Simulated annealing, in which high temperature is used to sample many configurations and then the system is gradually cooled to identify energetically favorable configurations, is often used in docking. All classical degrees of freedom of the molecule can be considered but this can be computationally costly.

During a docking calculation, a large number of possible ligand orientations in the receptor-binding site are generated. Rigorous computation of the binding affinity of each arrangement using molecular dynamics simulation techniques is generally computationally infeasible. Consequently, approximate **►scoring functions** are used to rank the possible solutions by providing an estimate of their binding affinity. There are four major types of scoring function, shape complementarity, force field, knowledge-based and empirical regression.

- Shape complementarity is the first property that has been applied to score docking and is still widely used. It can be estimated in a variety of ways including burial of surface area, grid point matching of discretized molecules and minimization of inter-facial cavities.
- In the force field-based scoring functions, the potential energy (Lennard-Jones, Coulombic) is calculated using standard molecular mechanics force fields. Entropic and solvation effects are usually neglected but can be estimated empirically as terms dependent on the change in buried surface area upon binding. Good results have been obtained recently by using the MM/PBSA or MM/GBSA procedure in which the average binding energy, including Poisson-Boltzmann (PB) or generalized Born (GB) electrostatic solvation energy and surface-area (SA) dependent hydrophobic energy terms, is computed from a series of snapshots from a molecular dynamics simulation. However, this approach is

too computationally intensive for large-scale **►virtual screening**.

- Knowledge-based scoring functions exploit the large number of available 3-dimensional structures of protein-ligand complexes. The probability distribution functions of inter-atomic distances for different combinations of atom-types may be converted into energy functions by inverse formulation of Boltzmann's law. Preferred interactions score favorably, repulsive interactions are penalized.
- Empirical regression functions can be constructed as generally applicable scoring functions or as scoring functions for specific target systems. They are derived using a training set of experimentally determining structures of ligand-receptor complexes for which binding affinities have also been measured. A function is constructed that consists of the weighted sum of a number of terms that are chosen to describe the important physico-chemical properties contributing to binding affinity, such as hydrogen-bond formation and ionic and hydrophobic interactions. The weights of the various terms are derived by regression using the training set. Most functions derived for scoring give good results for some types of complexes and poor results for others. For lead optimization problems, when a number of similar ligand-receptor complexes should be docked, tailored scoring functions can be used. These can be trained on a small training set of related complexes. These functions can be derived by classical regression for a small number of terms or by partial least squares projection to latent variables (PLS) for models based on terms in a molecular mechanics force field.

Clinical Applications

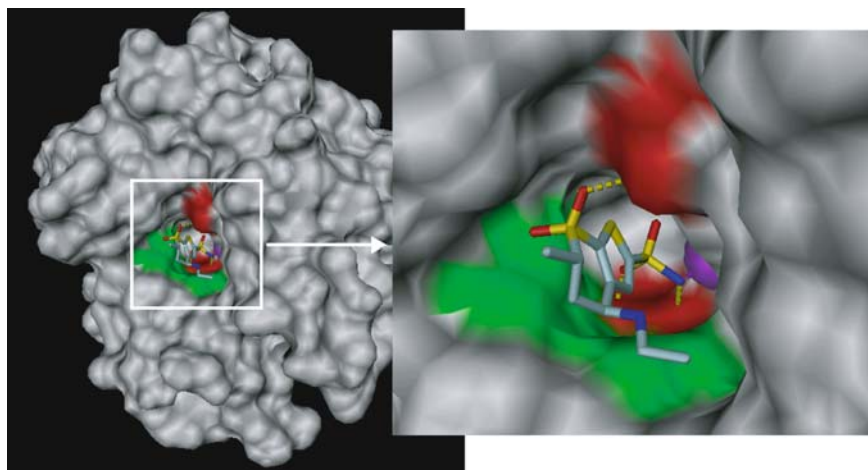
The early applications of molecular docking techniques in drug design were as an aid to lead optimization and this is still an important application of molecular docking. A designed molecule is docked into the target protein and then the docked orientation is compared with that of other ligands. If the new molecule fits well in the binding pocket, further experiments (**►X-ray crystallography** and bioassay) and optimization of the biophysical properties (solubility, pharmacokinetic profile) can be carried out to obtain a more potent compound. Docking can also provide insights important for further optimization of the compound and can be used in conjunction with molecular design software. The anti-AIDS drug Indinavir, an HIV-1 protease inhibitor, provides an excellent example of the use of docking in lead optimization (5). Analogues of Indinavir were first designed based on the transition state mimic concept. Then a basic amine in a rigid ring

system was incorporated into the backbone of the known peptidomimetic inhibitor to increase the aqueous solubility and decrease the entropy lost upon binding. The proper fit of the modified inhibitor in the binding site of HIV protease was predicted by docking studies using energy minimization and confirmed by X-ray crystallography. Indinavir emerged from further optimization in which the amine ring was replaced by a more soluble pyridine moiety.

With the increased understanding of the rules governing receptor-ligand binding and the development of molecular docking algorithms, the docking technique has been used for discovering novel leads by virtual screening of molecular databases. Docking has been applied to many disease targets and successful results have been obtained (2). Here, we describe three illustrative examples.

The first example is the identification of non-peptidic CD4 protein inhibitors (6) for the treatment of autoimmune diseases. The DOCK program was used to automatically dock the 150,000 compounds in the Available Chemicals Directory (Molecular Design Ltd, San Leandro, USA) into the crystal structure of the human CD4 protein D1 domain. From this virtual screening, the 1000 compounds with the best shape complementarity scores and the 1000 compounds with the best force field scores were selected for visual inspection with molecular graphics software. 41 of these compounds, representing the diversity of chemical structure and receptor interactions found amongst the screening hits, were finally chosen for testing in an

inhibitory activity assay. Out of these, four potent compounds were identified. This 10% success rate is a significant improvement over the typical yield of ca. 0.01% from conventional random screening procedures. The second example is the discovery of the first active non-peptidic HIV entry inhibitors targeting gp41 (7). The DOCK program was again used but this time to screen a database of 20,000 compounds (ComGenex Inc., Budapest, Hungary) for binding to a hydrophobic pocket on the protein. 200 screening hits were inspected visually and 16 of these were selected for biological assays. Of these, two polysulfonic acid compounds displayed inhibitory activity at micromolar concentrations on the formation of the gp41 core structure and on HIV-1 infection. In the above two cases, docking was applied directly to the original diverse molecular databases and resulted in a large quantity of hits, most of which had to be filtered out manually. Screening a more focused targeted molecular database, generated by applying filters prior to docking, should increase the hit rate. This was the approach taken in the search for novel inhibitors of human carbonic anhydrase, a target for anti-glaucoma and anti-tumor drug design (8). The program FlexX was used to flexibly dock 100 compounds. These compounds were preselected by a 3D [pharmacophore](#) search of 90,000 compounds in the Maybridge and LeadQuest databases and a subsequent superimposition scoring. 13 hits with high docking rankings were experimentally tested and four of them showed comparable potency to the known drugs, Dorzolamide (Fig. 2) and Brinzolamide, with



Molecular Docking. Figure 2 Steric and chemical complementarity between a drug and its protein target are illustrated by the crystal structure of the complex of the anti-glaucoma drug Dorzolamide (stick representation) bound to human carbonic anhydrase II in the catalytic pocket. The yellow dashed lines show hydrogen bonds between the drug and the protein. The hydrophobic and the polar regions on the protein-binding site are shown in green and red, respectively. The Zn ion in the active site is shown by a purple ball. See text.

subnanomolar or nanomolar IC₅₀ values. A further seven compounds were micromolar inhibitors. Moreover, the predicted docking modes of two of the compounds were in good agreement with subsequent crystallographic structure determinations.

Therapeutic Consequences

A therapeutic drug needs not only to bind tightly and specifically to its macromolecular receptor. It also needs to have the appropriate properties to reach its site of action in appropriate amounts without causing undue side effects and toxicity problems. It is now recognized that the ►**ADMET** properties of a compound should be considered in parallel, rather than after, the structure-based drug design phase. Virtual compound libraries can be filtered before screening by docking so that compounds with known toxicity problems or those that may have poor physicochemical properties for oral availability are filtered out. Further, the metabolism of compounds identified by docking as ligands of a macromolecular target can be predicted by, along with other calculations, docking them to structurally characterized drug metabolism enzymes such as cytochromes P450.

Docking technology can also be used to investigate the potential for specific action and side effects. Hits identified by screening a compound library against a given macromolecular target can then be docked against a library of other macromolecular structures. These may be of related proteins, homologues or orthologues from different species or from a diverse set of protein targets.

While current docking techniques can be usefully applied in medical and agrochemical areas, they also have many shortcomings and, consequently, improving docking methodology is an area of active research. Nevertheless, it is now possible, using grid computing with PC workstations that can be widely distributed, to routinely screen molecular databases containing millions of compounds in short periods of time. This has been demonstrated for cancer and anthrax targets amongst others.

References

1. Taylor RD, Jewsbury PJ, Essex JW (2002) A review of protein-small molecule docking methods. *J Comput Aided Mol Des* 16:151–166
2. Shoichet BK, McGovern SL, Wei B et al (2002) Lead discovery using molecular docking. *Curr Opin Chem Bio* 6:439–446
3. Halperin I, Ma B, Wolfson H et al (2002) Principles of docking: An overview of search algorithms and a guide to scoring functions. *Proteins* 47:409–443
4. Ehrlich LP, Wade RC (2001) Protein-protein docking. *Rev Comput Chem* 17:61–97

5. Dorsey BD, Levin RB, McDaniel SL et al (1994) L-735,524: The design of a potent and orally bioavailable HIV protease inhibitor. *J Med Chem* 37:3443–3451
6. Li S, Gao J, Satoh T et al (1997) A computer screening approach to immunoglobulin superfamily structures and interactions: Discovery of small non-peptidic CD4 inhibitors as novel immunotherapeutics. *PNAS* 94:73–78
7. Debnath AK, Radigan L, Jiang S (1999) Structure-based identification of small molecule antiviral compounds targeted to the gp41 core structure of the human immunodeficiency virus type 1. *J Med Chem* 42:3203–3209
8. Gruneberg S, Stubbs MT, Klebe G (2002) Successful virtual screening for novel inhibitors of human carbonic anhydrase: strategy and experimental confirmation. *J Med Chem* 45:3588–3602

Molecular Dynamics Simulations

Definition

The motion of a set of atoms is simulated using a classical model with interatomic forces described by a molecular mechanics force field. Newton's equations of motion are solved at a series of time steps (typically at 1 fs time intervals) to generate a movie of the dynamics of the molecule(s).

► **Molecular Docking**

► **Protein Structure Prediction**

Molecular Dynamics Simulations in Drug Design

CHRISTOPH A. SOTRIFFER

Department of Pharmaceutical Chemistry,
Philipps-University, Marburg, Germany
sotriffer@staff.uni-marburg.de

Definition

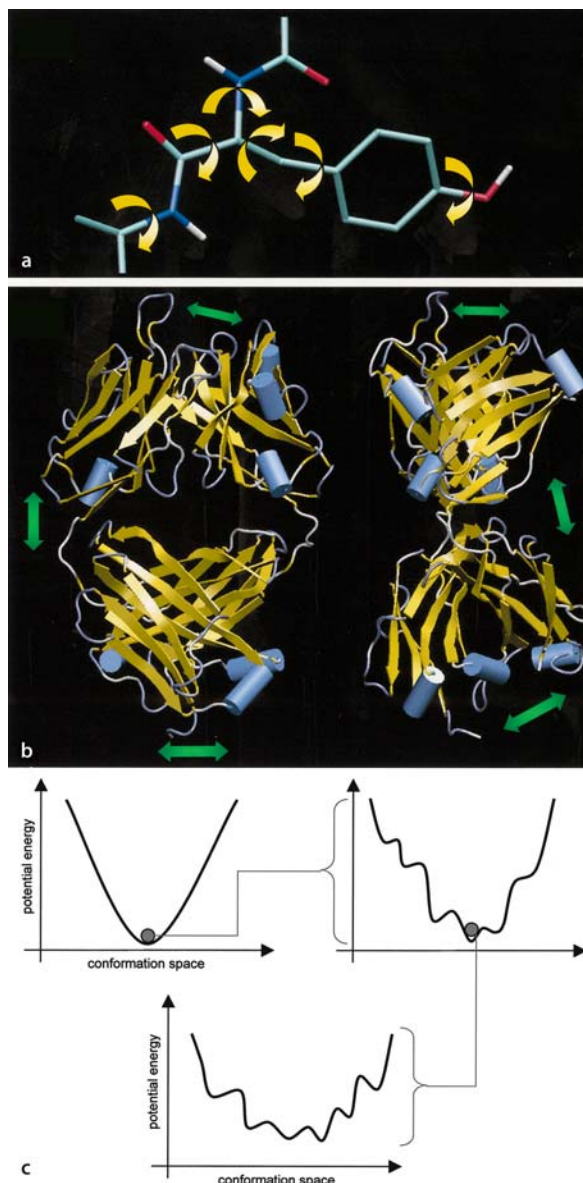
Molecular dynamics (MD) is a computational tool to simulate the motions of a molecular system. The method requires an interaction potential from which interatomic forces can be calculated and equations of motion that govern the dynamics of the system. MD simulations sample the ► **configuration space** and generate a ► **trajectory** that follows molecular movements as a function of time. Originally developed to

study properties of the liquid state, MD simulations are nowadays routinely applied to macromolecular systems of biological and pharmaceutical interest. Applications include the refinement of experimentally determined structures, ►conformational analysis and protein ►homology modeling, the elucidation of biochemical and biophysical mechanisms at the atomic level, ►docking of biomolecular complexes and a variety of approaches for calculating ►free energy changes. Most of these tasks are of utmost importance in structure-based drug design.

Characteristics

Physical Principles and Computational Methods

At ambient temperatures, biological macromolecules are never at rest. Proteins, for example, are inherently dynamic systems (Fig. 1). They are characterized by a hierarchy of motions that range from small fluctuations of hundredths of an Ångström to large domain movements of tens of Ångströms and cover time scales from sub-picoseconds to seconds and more (1). Many of these motions are essential for function; ligand binding, enzyme catalysis and signal transduction would frequently be impossible without significant flexibility of the biological macromolecules involved. Since interfering with these processes is a common task in structure-based drug design, any insight into the dynamics of the investigated system comes as a significant support for rational design approaches (2). MD simulations can provide such insights, often at a level of detail that would be difficult or impossible to reach by experimental studies. Apart from following the time course of a dynamic process, however, it is often the ability to sample ►configuration space that makes MD simulations attractive for certain applications; examples include the exploration of thermally accessible ►conformations and the evaluation of thermodynamic quantities, most notably free energy changes. Classical MD simulations are based on Newton's equation of motion, $F_i = m_i a_i$, according to which the acceleration a_i of a particle (atom) i of mass m_i is directly proportional to the acting force F_i . Since the acceleration is the first derivative of the velocity with respect to time and the second derivative of the atomic positions with respect to time, expressions for the coordinates and the velocities of the particles as a function of time can be derived. Thus, given the interatomic forces as well as a set of starting coordinates $r_i(t_0)$ and velocities $v_i(t_0)$, the positions $r_i(t_1)$ and velocities $v_i(t_1)$ of the atoms at a shortly following time $t_1 = t_0 + \Delta t$ can be evaluated. The values calculated at t_1 will then serve to calculate positions and velocities at a subsequent time t_2 and so forth (Fig. 2). Various numerical integration algorithms are available to accomplish this task efficiently for appropriately chosen time steps Δt (3).



Molecular Dynamics Simulations in Drug Design.

Figure 1 Protein molecules are dynamic systems characterized by a hierarchy of different motions. Torsion angles of the backbone (ϕ, ψ) and the side chains (χ_n) show significant fluctuations at room temperature (a). If they are concerted over larger units of the protein structure, higher-order movements may arise, which are often of functional relevance, such as loop reorientations or hinge-bending domain motions (b). The potential energy landscape governing these motions is very complex and frequently characterized by a multitude of nearly isoenergetic conformational sub-states (c).

To carry out this integration, the forces between the particles must be known. Consequently, a suitable interaction potential that describes the energy of the

system as a function of the atomic coordinates must be assigned (the forces are then accessible through the first derivatives of the potential with respect to the atom positions). For biological macromolecules, this potential is usually provided by a ►**molecular mechanics force field**, which typically consists of empirically parameterized terms for bond lengths, bond angles and torsion angles, as well as terms for the non-bonded ►**van der Waals interactions** and ►**Coulomb electrostatics**. As the approach is entirely based on classical potentials, electronic degrees of freedom are not taken into account. Simulations of chemical reactions involving bond breaking or bond formation require quantum mechanical potentials.

Beyond the equation of motion and the interaction potential, practical MD simulations of biomolecules depend on a variety of further important issues. These include the definition of appropriate ►**boundary conditions** for the system, the handling of the solvent, the regulation of temperature and pressure and the treatment of long-range electrostatic forces. Detailed discussions of these topics can be found in the literature (3).

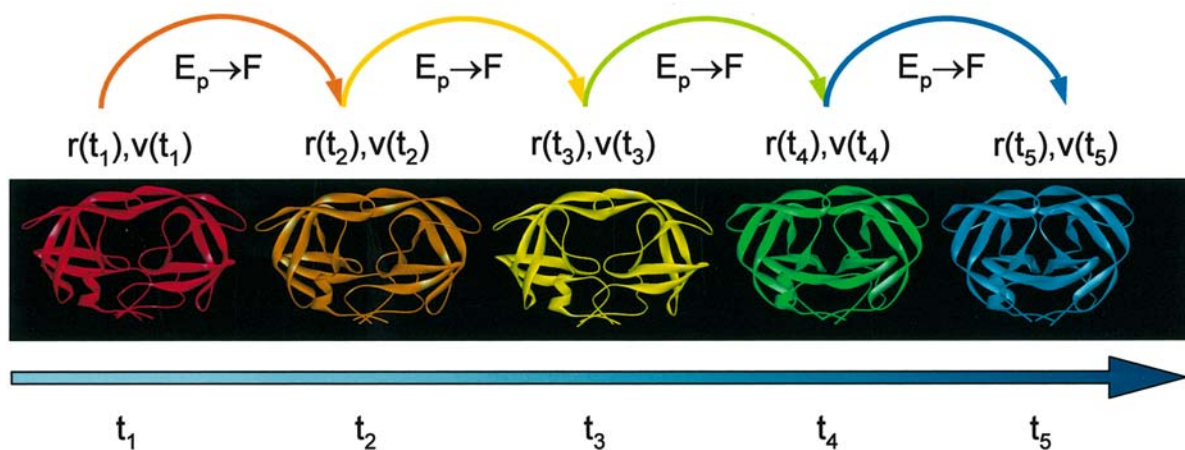
The primary result of an MD simulation is a trajectory of the simulated system, i.e. a series of atomic coordinates (and velocities) as a function of time. A trajectory can be analyzed with respect to a large variety of structural and energetic parameters (4). For any quantity Q of interest, the time course $Q(t)$ may be followed, averages $\langle Q(t) \rangle$ and fluctuations $\langle [Q(t) - \langle Q(t) \rangle]^2 \rangle^{1/2}$ can be calculated or an ►**autocorrelation** function $\langle Q(t')Q(t' + t) \rangle$ may be evaluated (angular

brackets denote average quantities). The results should always be complemented with and compared to available experimental data. In addition, any interpretation should be aware of the ►**convergence** and the relaxation properties of the analyzed quantity, as well as of the inherent approximations of the entire approach (4).

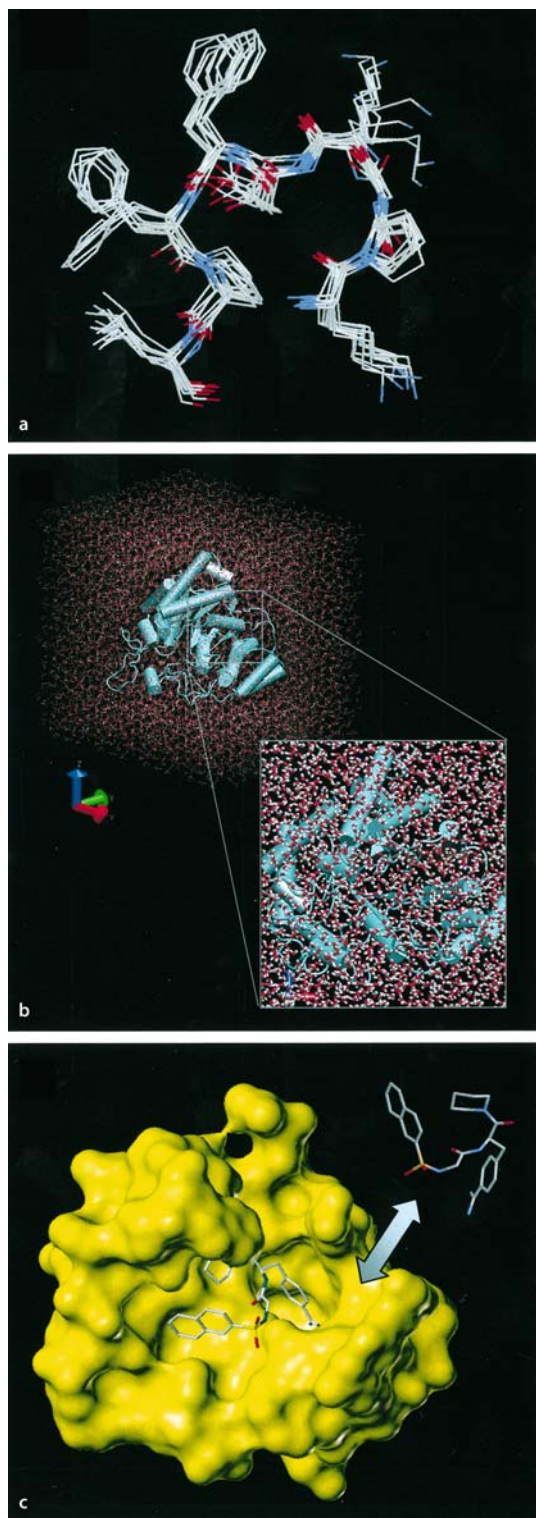
The main limitations of current MD simulations are inaccuracies in the potential energy function and insufficient ►**sampling** time. With respect to the first aspect, current research is focused on the reparameterization and optimization of ►**force fields** for simulations, as well as on the inclusion of ►**polarizabilities**. As for the simulation lengths, most is gained from the tremendous improvements in computer hardware. It is nowadays possible to routinely obtain multi-nanosecond trajectories for large macromolecules of many thousand atoms under the most realistic conditions (i.e. including explicit solvent molecules, counterions, periodic boundary conditions and long range electrostatic forces). The continuing increase in computing power and improvements in existing algorithms can be expected to progressively extend the current limits, eventually bringing within reach dynamic processes that take place on much longer time scales (5).

Applications in Drug Design

MD simulations are applied in three major areas of structure-based drug design: (i) structure determination and modeling, (ii) docking and ligand design and (iii) free energy calculations (some common applications are illustrated in Fig. 3).



Molecular Dynamics Simulations in Drug Design. Figure 2 MD allows simulation of molecular motion and generation of configurational ensembles. This is achieved by iteratively solving Newton's equation of motion for short, discrete time steps Δt to update the positions r_i and velocities v_i for every atom i in the molecular system. An appropriate potential energy function E_p is used to evaluate the forces F_i required for this purpose.



Molecular Dynamics Simulations in Drug Design.

Figure 3 Common applications of MD simulations are the conformational sampling and the refinement of macromolecular structures determined by X-ray crystallography or NMR (a), mechanistic studies on catalytic reactions or folding/unfolding processes in aqueous

Determination, Modeling and Analysis of Target Structures

Structure-based drug design depends on the availability of a 3D structure of the investigated target. Such structures are usually obtained *via* X-ray crystallography or NMR spectroscopy. Both methods frequently rely on MD simulations to refine the initial models obtained from the primary experimental data. Based on its ability to explore conformation space, MD is used here as a tool to optimize the simultaneous agreement of a structural model with observed data and with *a priori* chemical information (6). This is normally done in such way that the experimental data are transformed into energy penalties or constraints that are combined with the standard interaction potential. MD simulations are then carried out subject to these constraints, usually following a ►[simulated annealing](#) protocol.

If no experimental structure of the target is available, it may be theoretically predicted. For proteins, this is reasonable only if structures of other proteins with significant sequence homology are known. Homology modeling techniques may then be applied to obtain structural models of the target. MD simulations in the form of simulated annealing are used to refine these models. Conformational sampling is performed primarily for loop regions and areas of low sequence similarity in order to identify conformations corresponding to low energy minima.

Obviously, MD simulations are also of use if a high-resolution experimental structure of the target is already known. A single, fixed conformation, even the average provided by a crystal structure, may not be an adequate representation of a macromolecule, unless the system is very rigid. Instead, even under standard equilibrium conditions, the native folded state of a protein is best characterized by a collection or ensemble of energetically nearly equivalent conformations. MD simulations may be run to study the accessible conformations and reveal the inherent flexibility of the system. The dynamic results can be compared with experimental structural information and data from ligand binding or enzyme catalysis studies, providing clues for their interpretation and insights about the underlying structural mechanism. Thus, functionally relevant motions may be revealed that would not be evident from the crystal structure alone. Examples of such MD studies on pharmaceutically relevant proteins are included in Table 1.

solution (b) and research in the area of protein-ligand interactions and molecular recognition, including docking and free energy calculations (c).

Molecular Dynamics Simulations in Drug Design. Table 1 Selection of MD-based studies on biologically important structures and targets of pharmaceutical interest

Target structure	Investigated features	First author	Reference
acetyl cholinesterase	activity regulation, role of water	Shen T	Acc Chem Res 35: 332 (2002)
catechol O-methyltransferase	enzyme energetics and regioselectivity	Kuhn B	J Am Chem Soc 122: 2586 (2000)
cell cycle control phosphatase Cdc25A	ligand induced conformational changes	Kolmodin K	FEBS Lett 465: 8 (2000)
cytochrome P450 3A4	regioselectivity of drug metabolism	Kuhn B	J Med Chem 44: 2027 (2001)
dihydrofolate reductase	selectivity of ligand binding	Graffner-Nordberg M	J Med Chem 44: 2391 (2001)
fatty acid binding protein	dynamics of internal water molecules	Likić VA	Proteins 43: 65 (2001)
fructose-1,6-bisphosphatase	relative free energies of inhibitor binding	Reddy MR	J Am Chem Soc 123: 6246 (2001)
HIV integrase	active site conformations, role of metal ion	Lins RD	Biophys J 76: 2999 (1999)
HIV integrase	ligand mobility, interactions with water	Ni HH	J Med Chem 44: 3043 (2001)
HIV protease	sequence variation and inhibitor binding	Zoete V	J Mol Biol 315: 21 (2002)
HIV protease	enzymatic reactivity	Piana S	J Mol Biol 319: 567 (2002)
HIV protease	role of water, catalytic mechanism	Okimoto N	J Am Chem Soc 122: 5613 (2000)
HIV protease	dimer stability	Wang W	J Mol Biol 303: 567 (2000)
HIV reverse transcriptase	inhibitor binding modes and energetics	Wang J	J Am Chem Soc 123: 5221 (2001)
human glutathione transferase P1-1	induced-fit mechanism	Stella L	Proteins 37: 1 (1999)
human water channel aquaporin-1	diffusional permeation of water molecules	De Groot BL	Science 294: 2353 (2001)
KcsA potassium channel	energetics of ion conduction	Bernèche S	Nature 414: 73 (2001)
metallo- β -lactamase	effects of inhibitor binding	Salsbury FR	Proteins 44: 448 (2001)
p53 - oncoprotein Mdm2 complex	energetics of protein-protein interaction	Massova I	J Am Chem Soc 121: 8133 (1999)
prion protein	pH-induced conformational changes	Alonso DOV	Proc Natl Acad Sci USA 98: 2985 (2001)
protein kinase C	binding mode prediction, docking	Pak Y	J Med Chem 44: 1690 (2001)
thymidylate synthase	ligand binding, inhibitor design	Lee TS	J Am Chem Soc 122: 4385 (2000)
tyrosine kinases c-Src, Hck	phosphorylation effects	Young MA	Cell 105: 115 (2001)

Docking and Ligand Design

In the context of drug design, docking is mainly concerned with the prediction of protein-ligand interactions and the identification of small molecules binding tightly to a given macromolecular receptor. MD simulations could in principle be used to simulate the actual binding process, hence providing a “realistic” view of how the docking process proceeds, but this is computationally not yet feasible. Although pure MD docking approaches have been reported (e.g. Table 2:1,2), more often MD simulations are used to analyze and refine already docked complexes. This can be valuable to overcome simplifications of fast docking tools, most importantly the neglect of binding-site flexibility. As comparisons of experimental protein structures in the ligand-free and the complexed state show, ligand binding is, in fact, frequently associated with significant conformational changes in the protein. While most docking tools can not handle this appropriately, a few algorithms have been described where at some stage of the process, short MD runs are used to explore the local conformation space, often restricted to the vicinity of the binding site (e.g. Table 2:3). Alternatively, rigid docking runs may be performed with different conformations of a given protein. The required ensemble of conformations could be assembled from multiple crystal structures of the protein, from NMR studies or from MD trajectories. Apart from the protein, conformational sampling by MD may also be useful for the free ligand, either to generate low energy conformations as input for fast, but purely rigid docking or ►[virtual screening](#) routines or to perform a thorough conformational analysis of the ligand in order to estimate important conformational contributions to the overall affinity.

As mentioned above, most frequently MD simulations are applied as a post-docking tool to refine and analyze the generated complexes. This can help to identify more stable conformational substates and binding modes or, conversely, to reveal instabilities under dynamic conditions. Different ligands and ligand binding modes can be compared by analyzing, for example, the dynamics of contact distances, hydrogen bonds and solvent accessible surface areas, as well as the residual mobility and fluctuations of the ligands in the binding site (2). In addition, the important role of water molecules, which is rarely considered explicitly in docking, may be investigated.

Besides the identification of ligands by docking and database screening, an important field of drug design is concerned with the rational modification of known binders or the *de novo* design of new molecules fitting to a binding site. Even here, MD simulations can be of use. Current computational design methods usually work with a set of atoms, functional groups or

fragments that are first placed at favorable positions in the binding site and then connected in suitable ways to form new potential ligands. Among the large variety of approaches, some methods use MD simulations to identify low-energy configurations and localize favorable binding positions. Multiple-copy simultaneous search (MCSS) methods have been developed to speed up such searches by using numerous ligand copies, which are transparent to each other but subject to the full force of the protein. These and other MD applications in ligand design and docking are briefly listed in Table 2.

Free-Energy Calculations

The calculation of binding affinities is a common problem in drug design. Docking methods, for example, rely on estimates of the binding free energy provided by fast ►[scoring functions](#). The theoretically most rigorous and – in principle – most accurate approach to free energy differences is based on the evaluation of certain ►[ensemble averages](#) according to the principles of ►[statistical thermodynamics](#) (7). ►[Free energy perturbation](#) (FEP) and ►[thermodynamic integration](#) (TI) are the most prominent methods in this context (8). The required ensemble averages are typically obtained by MD. Sampling problems, however, preclude direct access to absolute free energies of binding. Instead, relative free energies are obtained by exploiting the fact that free energy is a ►[state function](#), which allows setting up a thermodynamic cycle to evaluate free energy changes by following unphysical pathways more amenable to simulation. Typically, a series of ligands binding to a given protein or the interaction of a ligand with a set of protein mutants is analyzed in this way. Besides the considerable degree of accuracy that can be achieved, a further advantage is the possibility for in-depth analysis of the results, which may provide clues with respect to the origin of the free energy changes. However, due to the slow convergence, both FEP and TI are suitable only for cases where the structural difference between the investigated ligands (or mutants) is small. Even then, the required computational resources remain significant.

Due to these restrictions, some less rigorous, but faster and often sufficiently accurate methods have been developed that are suitable to handle larger numbers of ligands. An example is the “one window free energy grid” (OWFEG) method. Here, an MD simulation is carried out with the ligand-free, solvated binding site of the protein, generating a free energy grid for a probe atom based on the FEP equation. Once such grids are calculated for different probes, ligand positions can be scored as fast as with any other grid-based scoring function. Other methods have been developed that try to obtain free energy estimates directly (without

Molecular Dynamics Simulations in Drug Design. Table 2 Selection of docking, design, and free-energy calculation methods making use of MD simulations

Method	Representative reference
Docking	
1 MDD: MD docking, fully flexible, in vacuo	Proteins 19: 174 (1994)
2 extension of MDD to fully flexible docking in solution	Proteins 35: 153 (1999)
3 grid-based docking with partial flexibility through local MD sampling	J Comput Chem 16: 454 (1995)
4 MD-based simulated annealing	J Comput Chem 19: 1623 (1998)
5 hierarchical MD-based method for exploration of low-energy conformations	Proteins 33: 475 (1998)
6 grid-based docking using protein ensemble generated by MD	J Mol Graphics Model 18: 247 (2000)
7 MD docking using a smoothed generalized effective potential	J Med Chem 44: 1690 (2001)
8 MCSA-PCR: multi-conformation simulated annealing pseudo-crystallographic refinement	J Mol Biol 314: 607 (2001)
Ligand Design	
9 MCSS: multiple copy simultaneous search	Proteins 11: 29 (1991)
10 MCSS and CCLD: Computational Combinatorial Ligand Design	J Comp Aided Mol Design 14: 161 (2000)
11 CONCEPTS: Creation Of Novel Compounds by Evaluation of Particles at Target Sites	J Comput Chem 14: 1184 (1993)
12 CONCERTS: extension of CONCEPTS to fragments	J Med Chem 39: 1651 (1996)
13 dynamic pharmacophore modeling with MD-generated protein conformations	J Med Chem 43: 2100 (2000)
Free Energy Calculations	
14 FEP & TI: Free Energy Perturbation and Thermodynamic Integration	Ann Rev Biophys Biophys Chem 18: 431 (1989)
15 PROFEC: Pictorial Representation of Free Energy Components	J Comp Aided Mol Design 12: 215 (1998)
16 CMC/MD: chemical Monte Carlo/Molecular Dynamics	J Med Chem 42: 868 (1999)
17 OWFEG: One Window Free Energy Grid	J Med Chem 42: 4313 (1999)
18 MM-PBSA: Molecular Mechanics Poisson-Boltzmann Surface Area	Acc Chem Res 33: 889 (2000)
19 MM-GBSA: Molecular Mechanics Generalized Born Solvent Accessibility	J Med Chem 44: 4325 (2001)
20 LIE: Linear Interaction Energy approach	Acc Chem Res 35: 358 (2002))

FEP-like procedures) from the trajectories of the interacting molecules in the complexed and the free state. The ►**MM-PBSA** method calculates trajectory averages based mainly on force field energies combined with ►**Poisson-Boltzmann Electrostatics** and surface area terms. The ►**LIE** method, instead, evaluates simulation averages of the electrostatic and van der Waals interaction energies and arrives at binding free energy estimates by scaling with empirically determined factors. References and further examples of MD

applications in free energy calculations are provided in Table 2.

References

1. Karplus M, Petsko GA (1990) Molecular dynamics simulations in biology. *Nature* 347:631–639
2. Rognan D (1998) Molecular dynamics simulations: a tool for drug design. *Persp Drug Discov Design* 9/10/11:181–209

3. Gunsteren WF van, Berendsen HJC (1990) Computer simulation of molecular dynamics: methodology, applications, and perspectives in chemistry. *Angew Chem Int. Ed. Engl* 29:992–1023
4. Gunsteren WF van, Mark AE (1998) Validation of molecular dynamics simulation. *J Chem Phys* 108:6109–6116
5. McCammon JA, Wolynes PG (eds) (2002) Theory and simulation. Enlarging the landscape. *Curr Opin Struct Biol* 12:143–213
6. Karplus M (ed) (2002) Molecular dynamics simulations of biomolecules. *Acc Chem Res* 35:321–489
7. Kollman PA (1993) Free energy calculations: applications to chemical and biochemical phenomena. *Chem Rev* 93:2395–2417
8. Reddy MR, Erion MD, Agarwal A (2000) Free energy calculations: use and limitations in predicting ligand binding affinities. *Rev Comp Chem* 16:217–304

Molecular Imaging

Definition

With the aid of molecular imaging, biological processes can be measured and characterized on a cellular level. In the context of diagnostic medicine, conventional imaging methods can only be used to detect anatomical developments, however, such molecular techniques may be used to access the causes of illness at a cellular level. Molecular imaging also differs from conventional techniques because it identifies specific gene products and intracellular processes. Above all, in the field of clinical diagnostics, the ►positron emission tomography (PET), the ►single photon emission computer tomography (SPECT), and the magnetic resonance tomography (►Nuclear Magnetic Resonance) (MRT) provide representative images at a cellular or molecular level in the living organism.

In the context of the basic cellular mechanisms (cellular signal transduction, protein-protein interactions, mobility and functionality of molecules, etc.), molecular imaging such as fluorescence microscopy is described as ►functional imaging.

Molecular Imprints

Definition

Molecular imprints are synthetic polymers synthesized in the presence of potential binding partners so that they can bind these molecules.

►Monoclonal Antibodies

Molecular Mechanics Force Field

Definition

Molecular mechanics force field refers to a classical, heuristic approach for the description of molecular structures and energetics. Molecules are handled as if they consist of charged spheres (atoms) connected by springs (bonds) of different elasticity (force constant). The energy is typically evaluated using expressions for bond stretching, angle bending, and torsional changes, as well as for van der Waals and electrostatic interactions. The terms must be empirically parameterized for all types of molecules that the force field should describe.

►Molecular Docking

►Molecular Dynamics Simulation in Drug Design

►Structure-Based Drug Design

Molecular Mimicry

Definition

Molecular mimicry describes the phenomenon that structurally related molecules bind mostly to the same binding site. They can fulfil the same function (agonists) or can block the active site without any regular function (antagonist, competitive inhibitor)

►Classification of Active Centers

Molecular Motors

MANFRED SCHLIWA

Institute for Cell Biology, University of Munich, Munich, Germany
schliwa@med.uni-muenchen.de

Definition

Molecular motors are enzymes that transform chemical energy into mechanical work. In the cytoplasm of eukaryotic cells, three different classes of motors that generate linear movement are known to exist – myosin, kinesin and dynein. Figure 1 shows prototypical representatives of these motor classes.

Characteristics

The three types of motors share several features (1). Firstly, they require a polar track along which they

Molecular Medicine: Genomics Encounters Classical Medicine

JÖRG T. EPPLEN

Department of Human Genetics, Ruhr University,
Bochum, Germany
joerg.t.epplen@rub.de

The present part of the encyclopedic reference for molecular medicine and genomics comprises 77 mostly disease-centered entries as composed by more than one hundred authors. The subjects range from chromosomes (►[Chromosome 21, Disorders](#)), many monogenic disorders (reviewed in ►[Hereditary Disease, Genetic Basis](#)) and common multifactorial diseases (see overview ►[Common Diseases, Genetics](#)) to ethical issues in the general context of medical genetics (►[Ethical Issues in Medical Genetics](#)). Thus a wide field of genome orientated basic and applied biomedicine is covered by exemplary articles illustrating that particularly active and competitive research efforts have yielded substantial progress, with one or the other even bearing on clinical medicine.

Molecular Medicine and Genomics

Molecular medicine and genomics are the virtual partners in this modern biomedical research scenario that drives progressive innovation. Molecular medicine gathers information from the study of the molecular, cellular and whole organism biology of man as well as of model organisms (►[Transgenic and Knock-Out Animals](#)). This assembled information, together with the knowledge obtained from the traditional fields of medicine from e.g. (molecular) cytogenetics (►[Microdeletion Syndromes](#)) to biochemistry (see e.g. ►[Acute Intermittent Porphyria](#); ►[Defective Protein Folding Disorders](#)) leads to a better understanding of a living organism in health and disease. The term genomics is obviously derived from ‘genome’, which comprises all the genetic material of an organism. Roughly speaking, genomics encompasses the research into the composition and function of the genetic material of all living organisms. Genomics covers a broad area that cannot be placed under the umbrella of one conventional scientific field. It is more than just biology, just chemistry or just medicine. Genomics is a new scientific field that covers a wealth of novel topics, innovative technologies and original approaches. Genomics is the science that studies genes and their function, but also the previously disrespected genomic desert (28), i.e. DNA sequences representing merely evolutionary junk. It is

not a great surprise to biological scholars that the ‘junky DNA desert’ is partly redefined by identifying therein a wealth of sequence elements with crucial regulatory functions for gene expression (►[Hypoxia Inducible Factors](#); ►[Transcription Factors and Regulation of Gene Expression](#)). Thus more specifically, genomics is the technology that sets out the map of all the genetic information of humans, animals, plants and microorganisms. It unravels their blueprint, for instance by means of DNA mapping and sequencing as well as by wide-scale research into the functions of genes and the manner in which hereditary characteristics are translated into the function of a cell – and furthermore ultimately into the function of an entire organism. Genomics also includes other high throughput technologies (as supplied by other fields like proteomics, metabolomics and bioinformatics among other ‘-omics’). Thus genomics comprises a set of technologies that has become an indispensable tool in the current field of life science research.

Undoubtedly, genomics is revolutionizing biomedical science and its applications to clinical medicine are starting to expand. Large-scale genomic sequencing has only been the basis of genomics in the genome projects. Translation of sequence information into an understanding of how the genome actually works in health and disease will dramatically alter the ways in which scientists and doctors think and work. With final regard to patient care, this accumulated knowledge has also to be transformed and applied integrally on higher organizational levels such as the immune system and the central nervous system (see e.g. ►[Autoimmune Diseases](#); ►[Addiction, Molecular Biology](#)).

Already ~150 years ago, Virchow in his famous book, *Cellular Pathology* wrote that ‘All diseases are reducible to active or passive disturbances in cells’ (26). Thus the key to every biological problem must finally be sought in the molecular changes in the cell. Before there is a cure for a disease, before drugs and/or life-saving procedures can be developed, even before clinical trials are commenced, there is always a crucial first step – research and discovery. Already today molecular medicine has become a central basic scientific discipline in biomedical research and in medical practice. It bridges biochemistry and structural biology to clinical medicine. Modern day scientists have the unprecedented opportunity to understand basic cellular processes as well as their derangements in diseases in molecular terms – often referred to as molecular cell biology and molecular medicine. Today researchers utilize information obtained from biochemical, genetic, molecular and sophisticated morphological techniques, such as deconvolution, confocal and video microscopies, in the study of basic problems in cell biology and molecular medicine. They also take

advantage of insights gained from ultramodern proteomic, genomic and bioinformatic analyses. After completing the human genome project (16, 34), the challenge of understanding the roles of all individual proteins and genes in coordinated cell function and disease remains. Thus the tempting frontier for the post-genomic era is centered in modern molecular medicine. A story of success in molecular medicine has been evidenced primarily by the progress in the diagnostics of monogenic diseases during the past two decades. Nearly all the 3000 human maladies inherited according to the Mendelian laws (see OMIM, Online Mendelian Inheritance in Man, a database of human genes and genetic disorders) can be diagnosed directly today, since they depend on a mutation in a given gene. Those few hundred disorders that still need to be tackled by the indirect approach should be prone to direct DNA analysis within the next year or two (►Hereditary Disease, Genetic Basis).

Monogenic vs Common Multifactorial Disorders plus Epigenetics

The conceptual gap between classical monogenic and common multifactorial diseases is bridged or even beginning to be closed by deeper insights into molecular basics as well as their pathogenetic consequences on the cellular and organismal levels. The 'monogenic disorder' *sensu stricto* is extremely rare, whereas 'common multifactorial disease' (►Common Diseases, Genetics) appears to be the general rule – and the latter ailments strike frequently. 'Multifactorial' comprises virtually all common diseases, although it is not yet known to what extent the individual disease entity is caused by very few (oligogenic) or many genes (multigenic). In addition, the critical environmental triggers especially but also other contributions remain largely elusive, so far. Thus, it appears that there rather exists a continuum between apparently strictly monogenic (Huntington disease, HD) and truly multifactorial disorders (multiple sclerosis, MS; inflammatory bowel disease, IBD). As early as 1983 the genomic region responsible for HD was defined on the tip of the short arm of human chromosome 4 (14). Only after a decade was the *huntingtin* gene and the relevant mutation described as a critical trinucleotide (CAG) expansion in an 'interesting transcript IT15' (30). Since then direct DNA and even prenatal diagnostics have been possible. Psychosocial counseling in combination with DNA testing for HD served as a model for other monogenic diseases. More and more of the causal pathogenesis can be defined, including the so far elusive genetic background (►Huntington's Disease).

With regard to the causal pathogenesis, eight other neurodegenerative diseases, including the spinocerebellar ataxias (►Repeat Expansion Diseases, Dynamic

Mutations Cause (Neurological) Model Disorders), are caused, like HD, by the expansion of a CAG repeat in the coding region of a transcribed gene. All these CAG expansion diseases are characterized by progressive neuronal dysfunction, starting usually during adult life and resulting in severe neurodegeneration. One of the hallmarks of the pathology of polyglutamine diseases is protein aggregation and formation of neuronal intranuclear inclusions in the affected areas of the brain (9). In addition to the repeat expansion, other factors, such as proteasomal activity and expression or recruitment of cellular chaperones, probably influence polyglutamine toxicity and disease development (35). Several findings point towards an involvement of the ubiquitin proteasome system (UPS) in the pathogenesis of polyglutamine diseases and to an enhancement of neurodegeneration by further impairment of the UPS (7). Thus inhibition of protein aggregation (►Acute Intermittent Porphyria; ►Defective Protein Folding Disorders) and proteasomal activation may turn out to be promising avenues for the treatment of polyglutamine diseases.

With regard to the genetic contributions to HD, the length of the expanded polyglutamine tract is clearly by far the most important factor in determining the age at onset (AO) of HD, although substantial variability remains after controlling for repeat length, particularly in cases where CAG repeat numbers range in the high 30s or low 40s (1, 2). These insights have prompted the search for additional genes influencing AO in HD. Defining such additional predisposition factors is important, since they may provide further clues pertaining to HD pathology arising from the expanded repeats. Neuronal death is obvious in various regions of brains of HD patients, including the neocortex and lateral hypothalamus accompanied by dramatic loss of GABAergic medium spiny neurons in the striatum. These neurons receive glutamatergic input from the cortex and dopaminergic input from the substantia nigra. A mechanism for the loss of these neurons in HD is excitotoxic cell death mediated by the release of glutamate from cortical afferents and activation of glutamate receptors (29). For example, excitotoxicity mediated by a N-methyl-D-aspartate (NMDA) receptor, in particular the GRIN2B subunit of the NMDA-type glutamate receptor, has implications for the pathogenesis of the neurodegenerative process in HD as a disturbance of neurotransmission. At least 10% of the genetic effect on the AO is attributable to combinations of the *GRIN2B* genotypes (2). Consequently, neuroprotective strategies for HD patients and persons at risk should be reconsidered.

In contrast to HD and other monogenic neurological and psychiatric disorders (►Neurofibromatosis Type 1 (NF1), Genetics; ►Hereditary Spastic Paraplegia), MS is a chronic inflammatory disease of the central

nervous system depending on a complex genetic background as do other multifactorial entities (see Baptista et al., this volume; Finckh, this volume; Nöthen, this volume; Norton and Owen, this volume; Pollmächer, this volume). Despite apparent genetic components, the responsible predisposing genes remained largely undefined. Only the *HLA-DRB1*15/16* alleles are well-established factors associated with susceptibility, conferring a nearly 4-fold relative risk for development of MS in northern European individuals (► [Genetic Predisposition to Multiple Sclerosis](#)). Disease manifestation and age at onset appear to depend on evasive environmental factors as well as on disease-modulating genes acting in genetically susceptible individuals. Although the etiology of MS remains to be understood, a plethora of putative candidate genes have already been screened with respect to mutations and polymorphisms associated with the disease itself or with clinical characteristics of the disease, such as severity of symptoms and disease course. The results of such candidate gene approaches have been unspectacular, as in most other autoimmune diseases, suggesting that MS is genetically complex. So far, no genes of major attributable risk have been found, but several genes probably confer a low or moderate effect (see *HLA-DRB1*15/16* alleles).

The genetic contributions to another group of multifactorial disorders, inflammatory bowel diseases (IBD), have already been clarified to a much greater extent. Hence, and also because of their model character for other multifactorial diseases, IBD are covered here in more detail. Both animal models and clinical investigations indicate that luminal bacteria play a pivotal role in the pathogenesis of morbus Crohn (► [Crohn disease](#), CD). This hypothesis of bacterial relevance is supported by the identification of the first susceptibility gene for CD. Single nucleotide polymorphisms (SNPs) have been identified in the *caspase recruitment domain 15* gene (*CARD15*, formerly known as *NOD2*) (reviewed in 15). Certain alleles of these SNPs increase the risk of developing CD. *CARD15* protein is located in the cytoplasm of mononuclear cells, where it is involved in the recognition of bacterial components, leading to subsequent activation of nuclear factor κ of B lymphocytes (NF κ B) by inactivation of its inhibitor NF κ BIA (see ► [NF- \$\kappa\$ B Pathway](#)). Disturbed activation of the innate immune system by bacterial antigens seems to be crucial in a subgroup of patients with CD. As rare alleles of the SNPs are not found in all patients with CD and as these alleles are found in healthy persons as well, variations in the *CARD15* gene are neither necessary nor sufficient for genetic predisposition to CD. Genome-wide linkage based screens with highly informative markers succeeded in identifying regions on chromosomes 1, 3, 5, 6, 7, 12, 14 and 16

linked to IBD (reviewed in 15). Yet the susceptibility genes located within the chromosomal regions still remain to be identified. Linkage of IBD to chromosome 12p13.2-q24.1 has been confirmed in several genome-wide screens. The *STAT6* gene is located within this chromosomal region. The transcription factor STAT 6 is involved in the regulation of the TH1/TH2 immune response and autoimmune reactions (► [Autoimmune Diseases](#)). Increased production of TH1 cytokines is crucial in the pathogenesis of IBD. In comparison to controls, certain *STAT6* and *NFKBIA* alleles are more frequent exclusively in CD patients without any predisposing variation in the *CARD15* gene (18, 19).

The study of human disease has focused on genetic mechanisms, but disruption of the balance of epigenetic networks can cause several major pathologies including cancer (► [Hereditary Nonpolyposis Colorectal Cancer](#)), syndromes involving chromosomal instabilities (► [Chromosomal Instability Syndromes](#); ► [Bloom Syndrome](#)) and mental retardation (► [Rett Syndrome](#); ► [Prader-Willi and Angelmann Syndromes](#); ► [Fragile X Mental Retardation Syndrome](#); ► [Inherited Mental Retardation Syndromes](#)). Epigenetic mechanisms, which involve DNA and histone modifications, result in the heritable silencing of genes without a change in their coding sequence (10). The development of new diagnostic tools might reveal other diseases that are caused by epigenetic alterations. Great potential lies in the development of ‘epigenetic therapies’ – several inhibitors of enzymes controlling epigenetic modifications, specifically DNA methyltransferases and histone deacetylases (► [Polyglutamine Disease, the Emerging Role of Transcription Interference](#)), have shown promising anti-tumorigenic effects in some malignancies.

In summary, major research efforts in genomics and molecular medicine will have to be concerned from now on with the identification of the complex genetic bases for the common multifactorial diseases. The so-called genetic background is critical or even decisive – not only for multifactorially caused diseases but also for the disorders due to ‘simple’ monogenic traits. Nevertheless the non-genetic contributions to disease manifestation and development (epigenetics and environment) will have to be simultaneously in the focus of major future research efforts.

Novel Developments in Genomics and Molecular Medicine – from Biobanks to Genome-wide Association Studies and Beyond

In addition to all modern methodological developments in the classical disciplines of biomedical research, specific new approaches and strategies are being established that virtually straddle genomics and molecular medicine. The definition of genetic and

environmental factors that contribute to health, disease and response to treatment is essential for the reduction of morbidity and mortality. The sequence of the human genome (16, 35) and increasing information about the genome's variation (31) and function provide a robust platform for the investigation of human health and disease. This information will be especially useful when combined with reliable and cost-effective (high throughput) methods that can be used to genotype these variations in large population samples. In parallel with the expansion of genomic tools and knowledge, methods for measuring non-genetic factors and environmental exposure have improved. Development of robust analytical methods for assessing disease risk relationships and interactions is beginning to allow the disentangling of such complex effects on a population scale (33). Together, these developments present opportunities to address unanswered questions related to the complex contributions of genes, the environment and gene/gene and gene/environment interactions. Large-scale national programmes to develop genomic epidemiological studies are already in place (UK (BioBank), Iceland (deCODE), Estonia) or planned (Canada, China, Japan, USA); international projects need to be launched in the near future (8).

On this basis of well-characterized patient and control cohorts, so-called disease association studies can be performed. In association studies, allele frequencies of polymorphic markers are compared in unrelated patients and healthy controls in order to identify markers that differ significantly between the cohorts (see also Vilmundur, this volume). Many studies used only a small number of polymorphisms, but most of the identified associations have been difficult to replicate (21) for a number of different reasons: First, there may be important differences in allele frequency and/or linkage disequilibrium structure across different populations. The situation may be further complicated by the presence of hidden stratification in some populations, producing spurious association or altering the pattern of a true association (12). Other sources of complexity are allelic heterogeneity, in which different alleles at the same locus are responsible for increased disease risk in different populations, and locus heterogeneity, in which alleles at different loci are responsible for increased disease risk in different populations. Both of these scenarios are likely when there are multiple rare variants that are fairly recent in origin. Currently, little is known about the true nature of allelic heterogeneity with respect to disease. Aside from these genetically driven phenomena, study design and publication bias may also lead to complex patterns (25).

Because it seems likely that common modest risk variants exist, developing tools such as high-density

genetic maps for whole genome association analyses is an important goal for analyzing the human genome. Comprehensive analysis of candidate regions and the whole genome should produce robust results. Meanwhile, the technical, informatic and statistical foundations have been laid for whole genome association analyses (6). It is important to understand that association analysis using dense maps, such as the HapMap, is fundamentally different from the family linkage studies used to identify rare high-risk alleles. In linkage analysis, any polymorphism between a pair of linked markers will be associated with both markers. As a result, linkage maps have a logical hierarchical structure wherein an initial genome scan can be performed at low density (<500 markers), with additional markers used to fine-map the boundaries of linked regions.

After decades of study, at least some facts have been learned about the genetics of common diseases, but most of this knowledge relates to rare families segregating high-risk alleles (5). Such alleles are generally very rare in the population and therefore explain relatively little of overall disease prevalence. Identification of the genetic polymorphisms that contribute to susceptibility for common diseases such as MS and inflammatory bowel disease will aid in the development of diagnostics and therapeutics. New (chip based) technologies and experimental resources make whole genome association studies more feasible (6, 13). Association studies of this type have good prospects for dissecting the genetics of common diseases, provided that the challenges, including multiple testing and study design, definition of intermediate phenotypes and interaction between polymorphisms are tackled successfully.

As a further development gene based approaches to association analysis are feasible and have many advantages. In contrast to SNP and haplotype approaches, a gene-based approach is less susceptible to erroneous findings due to genetic differences between populations. By capturing all or most of the potential risk conferring variations, a gene-based approach is capable of excluding association, subject only to the issue of power. Use of a gene-wide significance level should reduce the problem of chance findings due to multiple testing. Gene-based approaches lend themselves to meta-analysis of combined data from multiple studies. As the knowledge of the variation in genes grows, a gene-based approach will become the natural end point for association analysis and will provide pointers for genetic analysis at the functional level.

Gene-based approaches require detailed knowledge of genetic variation in coding sequences as well as regulatory and other regions affecting gene function. This level of knowledge is not generally available yet.

In the future, when knowledge of genetic variation allows a gene-based approach to be routinely employed, gene-based studies may become the preferred option for the genetic dissection of complex traits. At present, a gene-based approach should ideally be used for the replication of previous positive association findings or for the study of very strong candidate genes. The gene-based approach may not be necessary when there is very detailed knowledge about functional variation in the gene and would not be efficient for screening a very large number of genes (27).

Screening, Predicting and Preventing Diseases via Genomics and Molecular Medicine

For most diseases, medical treatment is offered only when the patient has visited the doctor with symptoms and complaints. By this time it is usually late in the natural history of the disease, and therapeutic intervention may be limited to alleviating symptoms and slowing disease progression. In the era of genomic medicine, physicians will have the opportunity to move from crisis-driven intervention to predictive medicine (17). It seems likely that entire populations or specific subgroups will be screened for genetic information in order to target interventions to individual patients that will improve their health and prevent disease. Until now, population screening involving genetics has focused on the identification of persons with certain Mendelian disorders before the appearance of symptoms and thus on the prevention of illness (screening of newborns for phenylketonuria etc.) and the testing of selected populations for carrier status and the use of prenatal diagnosis to reduce the frequency of disease in subsequent generations (Tay–Sachs disease among Ashkenazi Jews etc.). In the future, genetic information might increasingly be used in population screening to determine individual susceptibility to common disorders such as heart disease, diabetes and cancer. Such screening will identify groups at risk so that primary prevention efforts (diet, exercise etc.) or secondary prevention efforts (early detection, pharmacologic intervention) can be initiated. Such information could lead to the modification of screening recommendations, which are currently based on population averages (screening of people >50 years of age for early detection of colorectal cancer etc.).

In a few areas of medicine, standard approaches have progressed to incorporate multiple diagnostic tests that precisely identify disease mechanisms, indicate the most appropriate type of therapeutic intervention and evaluate therapeutic response as well as disease outcome. This is particularly evident in the management of infectious and metabolic diseases as well as cancer. Prediction, prevention and counseling of

individuals at risk of genetic diseases have been aimed largely at single gene disorders that have Mendelian patterns of inheritance. These disorders offer the advantage that the identification of genetic variations responsible for disease can lead directly to clinically helpful and reasonably accurate prediction or diagnosis of disease. Clinical genetics, however, will soon have to move on to aspects of genetics that are less ‘deterministic’ for a particular disease. Even apparently simple Mendelian disorders prove to have widely variable clinical phenotypes (see e.g. the genetic background of HD above). As a result of secondary genetic and environmental factors, individuals with exactly the same mutation may suffer either from a severe and early onset disorder or be affected relatively late in life and only modestly. Further examples pertain to situations when – on closer evaluation – genetic variants are not proved to have consistent levels of penetrance. Initially the genes such as the mismatch repair genes in hereditary non-polyposis colorectal cancer (HNPCC, ►[Hereditary Nonpolyposis Colorectal Cancer](#)) and *CFTR* in cystic fibrosis (see ►[Epistasis in Cystic Fibrosis](#)) have been identified because of their relatively high penetrance. Thus even single gene disorders have a significant level of heterogeneity (see ►[Hereditary Disease, Genetic Basis](#)) that limits any prognoses as asked for by the patient and/or his family.

Information valuable but also critical to patients and general practitioners is emerging from genomics (see Sass, this volume), especially in the form of future comprehensive genetic profiling. At present, this science is being accumulated in specialized university centers and it will need to be transferred more competently to a (routine) clinical setting, creating many different challenges for health professionals, both technically and educationally. Much of the recently generated genomic data that are clinically relevant are in formats that are inappropriate for diagnostic testing. The technology has not yet been established for rapid, inexpensive typing of most genomic biomarkers, with the exception of SNPs and too little is known about which SNPs to type for non-Mendelian disease. As stated above, very large epidemiological population samples followed prospectively for years and characterized for their biomarker and genetic variation will be necessary to demonstrate the clinical utility of these tools. Obstacles to the routine application of these data in clinical practice include an educational gap between the approach to clinical practice that is currently employed and that which is possible with these new tools. Diagnostic medicine that includes predisposition testing (►[Predictive Testing and Genetic Counseling](#)), early detection, individualized therapy (►[DNA-based](#)

Vaccination) and gene therapy (►Clinical Gene Transfer) as well as therapeutic monitoring is neither systematically applied nor well taught in the current health care system. Its implementation will require not just clear data demonstrating its benefits, but also demand by patients and acceptance by health care professionals. Cost effectiveness will have to be demonstrated. The opportunities for clinical genetics to become a mainstream component of clinical medicine are now apparent. This move to the clinic appears to be inevitable (3), but the transition will certainly take longer than another decade.

Partial Merger of Genomics and Molecular Medicine

Until recently, genomics and clinical research were independent areas, each with distinct intellectual traditions and communities of investigators. With the completion of the Human Genome Project, there is the need to combine these disciplines to study disease mechanisms and improve patient care, an endeavor that might be termed 'clinical genomics'. Genome sequence data are enabling clinical genomic investigation, in which the characteristics of human patients are explored using comprehensive inventories of biomolecules. Successful investigators must navigate through rapid technological change, collect and analyze large volumes of data and engage systems of clinical care. Such projects will increasingly rely on fully integrated multidisciplinary teams, demanding new organizational models in academic biomedical research. These facts apply also to pharmacogenetics and pharmacogenomics. Individuals respond differentially to drugs and sometimes the effects are unpredictable. Differences in DNA sequences that alter the expression or function of proteins that are targeted by drugs can contribute significantly to variation in the responses of individuals. Many of the genes examined in early studies were linked to highly penetrant, single gene traits, but future advances hinge on the more difficult challenge of elucidating multigene determinants of drug response (24). This intersection of genomics and medicine has the potential to yield a new set of molecular diagnostic tools that can be used to individualize and optimize drug therapy (►Idiosyncratic Drug Reactions).

If the use of medication is to evolve from a 'trial and error' approach to individualized therapy using genetics, at least two aspects of health care need to change. We must introduce protection against the misuse of genetic information and accept the added costs that may be incurred during the transition to genetically guided decisions about drug therapy. In the long run, decreasing the frequency of adverse drug effects and increasing the probability of successful

therapy will probably lower the cost of health care. Pharmacogenomics has the potential to facilitate this process by translating knowledge of human genome variability into more individualized therapeutics (11).

The Next Decade

While unimaginable only a decade ago, molecular medicine and genomics are now reality because of recent rapid progress and an enormous burst of new scientific opportunities. Thus, we are already beginning to gain ground in the fight against many dreaded diseases, including cancer and cardiovascular disorders and introduce more effective preventive measures, e.g. in the form of neuroprotection (32). One can realistically hope to develop increasingly effective treatments and preventive measures for these diseases as well as for the banes of diabetes (►Diabetes Mellitus, Genetics), obesity (►High-HDL Syndrome), Alzheimer disease (22), degenerative diseases of aging (►Molecular Aging Research) and emerging infectious agents.

The future is hard to predict, especially the longer it extends beyond a reasonably foreseeable span of 3–5 years. What will medical practice and patient care be like in 2014? Genomics and molecular medicine could grow as the basis for human health care in the next decade, especially if investments in biomedical research are increased substantially. The practice of molecular medicine will consist of new prevention, diagnosis and treatment methods that directly target the molecular, cellular or physiological defects causing disease. These medical methods will be based on precise, non-invasive imaging and other diagnostic techniques. They will be implemented with directed, rationally designed molecular and pharmaceutical therapies and they will be rooted in a deep understanding of normal human cellular and molecular physiology and genetics.

Most important in this context are new and far-reaching initiatives in functional and physiological genomics. Such new 'mega-projects' also involving the consolidated bio-medical industry sector would have the goal of understanding the normal functions of the many genes discovered in the complete genetic blueprints of humans and diverse model organisms by the Human Genome Project. These efforts will lead to a detailed understanding of normal cellular, molecular and integrative organismal physiology, which in turn will allow us to create therapies targeted directly to the cellular, genetic and physiological defects that cause disease and organ dysfunction. New achievements will also allow the defence of people against the ever-present and increasing dangers of emerging pathogens and viruses, by developing the next generations of vaccines and antibiotic drugs (see e.g. (23)).

This encyclopedic reference brings together several aspects of medicine that have significance at the molecular level with the advances in molecular biology that are relevant to medicine. With 77 separate primary entries, this comprehensive treatise covers the rapidly expanding field to a substantial extent, with well-studied examples. The coverage in this encyclopedic reference is multilayered, ranging from the organ to the cell to the molecular aspects and includes extensive cross-referencing. Molecular medicine covers diagnostics, epidemiology, pathogenesis, genomics, genetic testing, animal models, vaccines, gene therapy and drug design. The chapters in the Encyclopedic Reference of molecular medicine might be of interest to all those working in the fields of clinical medicine and medical research and to researchers in the pharmaceutical and medical biotechnology industries. Thus geneticists, biochemists, molecular biologists, medical researchers and doctors with either an academic or industrial background should find this an essential up-dated resource.

References

- Arning L, Jagiello P, Wiczorek S et al (2004) Glutathione S-Transferase Omega 1 variation does not influence age at onset of Huntington's disease. *BMC Med Genet* 5:7
- Arning L, Kraus PH, Valentin S et al (2005) *NR2A* and *NR2B* receptor gene variations modify age at onset in Huntington disease. *Neurogenetics* 6:25–28
- Bell J (2004) Predicting disease using genomics. *Nature* 429:453–456
- Bonizzi G, Karin M (2004) The two NF-kappaB activation pathways and their role in innate and adaptive immunity. *Trends Immunol* 25:280–288
- Botstein D, Risch N (2003) Discovering genotypes underlying human phenotypes: past successes for mendelian disease, future approaches for complex disease. *Nat Genet* 33:228–237
- Carlson CS, Eberle MA, Kruglyak L et al (2004) Mapping complex disease loci in whole-genome association studies. *Nature* 429:446–452
- Ciechanover A, Brundin P (2003) The ubiquitin proteasome system in neurodegenerative diseases: sometimes the chicken, sometimes the egg. *Neuron* 2003 40: 427–446
- Collins FS (2004) The case for a US prospective cohort study of genes and environment. *Nature* 429: 475–477
- Di Figlia M, Sapp E, Chase KO et al (1977) Aggregation of huntingtin in neuronal intranuclear inclusions and dystrophic neurites in brain. *Science* 277:1990–1993
- Egger G, Liang G, Aparicio A et al (2004) Epigenetics in human disease and prospects for epigenetic therapy. *Nature* 429:457–463
- Evans WE, Relling MV (2004) Moving towards individualized medicine with pharmacogenomics. *Nature* 429:464–468
- Freedman ML, Reich D, Penney KL et al (2004) Assessing the impact of population stratification on genetic association studies. *Nat Genet* 36:388–393
- Gödde R, Rohde K, Becker C et al (2005) Association of the HLA region with multiple sclerosis as confirmed by a genome screen using >10,000 SNPs on DNA chips. *J Mol Med* 83:486–494
- Gusella JF, Wexler NS, Conneally PM et al (1983) A polymorphic DNA marker genetically linked to Huntington's disease. *Nature* 306:234–238
- Hugot JP (2004) Inflammatory bowel disease: a complex group of genetic disorders. *Best Pract Res Clin Gastroenterol* 18:451–462
- International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860–921
- Khoury MJ, McCabe LL, McCabe ER (2003) Population screening in the age of genomic medicine. *N Engl J Med* 348:50–58
- Klein W, Tromm A, Folwaczny C et al (2004) A polymorphism of the *NFKB1A* gene is associated with Crohn's disease patients lacking a predisposing allele of the *CARD15* gene. *Int J Colorectal Dis* 19:153–156
- Klein W et al (2004) The G2964A polymorphism of the *STAT6* gene in inflammatory bowel disease. *Digest Liver Dis* 37:159–161
- Knight JC (2005) Regulatory polymorphisms underlying complex disease traits. *J Mol Med* 83:97–109
- Lohmueller KE, Pearce CL, Pike M et al (2003) Meta-analysis of genetic association studies supports a contribution of common variants to susceptibility to common disease. *Nat Genet* 33:177–182
- Mattson MP (2004) Pathways towards and away from Alzheimer's disease. *Nature* 430:631–639
- MM 2020: Vision for the Future of Medical Research and Human Health (1998) *FASEB Newsletter* 31:2, 6
- Muller M (2003) Pharmacogenomics and drug response. *Int J Clin Pharmacol Ther* 41:231–240
- Neale BM, Sham PC (2004) The future of association studies: gene-based analysis and replication. *Am J Hum Genet* 75:353–362
- Virchow R (1858) *Die CELLULARPATHOLOGIE in ihrer Begründung auf physiologische und pathologische Gewebelehre*. Zwanzig Vorlesungen gehalten während der Monate Februar, März und April 1858 im pathologischen Institute zu Berlin von Rudolf Virchow, o.ö. Prof. der pathologischen Anatomie, der allgemeinen Pathologie u. Therapie an der Universität, Direktor des patholog. Instituts u. dirigirendem Arzte a. d. Charité. Verlag August Hirschwald, Berlin
- Neale BM, Sham PC (2004) The Future of Association Studies: Gene-Based Analysis and Replication. *Am J Hum Genet* 75:353–356
- Ohno S (1982) Evolution is condemned to rely upon variations of the same theme: the one ancestral sequence for genes and spacers. *Perspect Biol Med* 25:559–572
- Schwarcz R, Bennett Jr, Coyle JT 1977 Inhibitors of GABA metabolism: implications for Huntington's disease. *Ann Neurol* 2:299–303

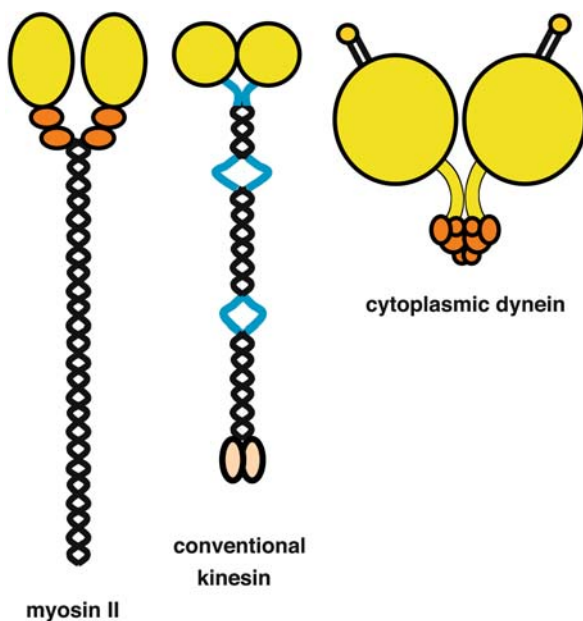
30. The Huntington's Disease Collaborative Research Group (1993) A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* 72:971–983
31. The International HapMap Consortium (2003) The International HapMap Project. *Nature* 426:789–796
32. Vajda FJ (2002) Neuroprotection and neurodegenerative disease. *J Clin Neurosci* 9:4–8
33. van den Oord EJ (1999) Method to detect genotype-environment interactions for quantitative trait loci in association studies. *Am J Epidemiol* 150:1179–1187
34. Venter JC, Adams MD, Myers EW et al (2001) The sequence of the human genome. *Science* 291:1304–1351
35. Wexler NS (2004) Venezuelan kindreds reveal that genetic and environmental factors modulate Huntington's disease age of onset. *Proc Natl Acad Sci USA* 101:3498–3503

move unidirectionally – actin filaments for myosins and microtubules for dyneins and kinesins. The information on polarity is provided by the uniform molecular orientation of the subunit proteins, actin and tubulin respectively, within the track. Intermediate filaments, the third major structural component of eukaryotic cells, are non-polar and are not known to support oriented movement. Secondly, all motors move unidirectionally either towards the plus end or the minus end of the track. Thirdly, all motors use the energy derived from ATP hydrolysis to undergo conformational changes that result in directional movement. The initial conformational changes in this process are similar in kinesins and myosins but totally different in dyneins. Fourthly, the binding sites for both ATP and the track are located in a roughly globular

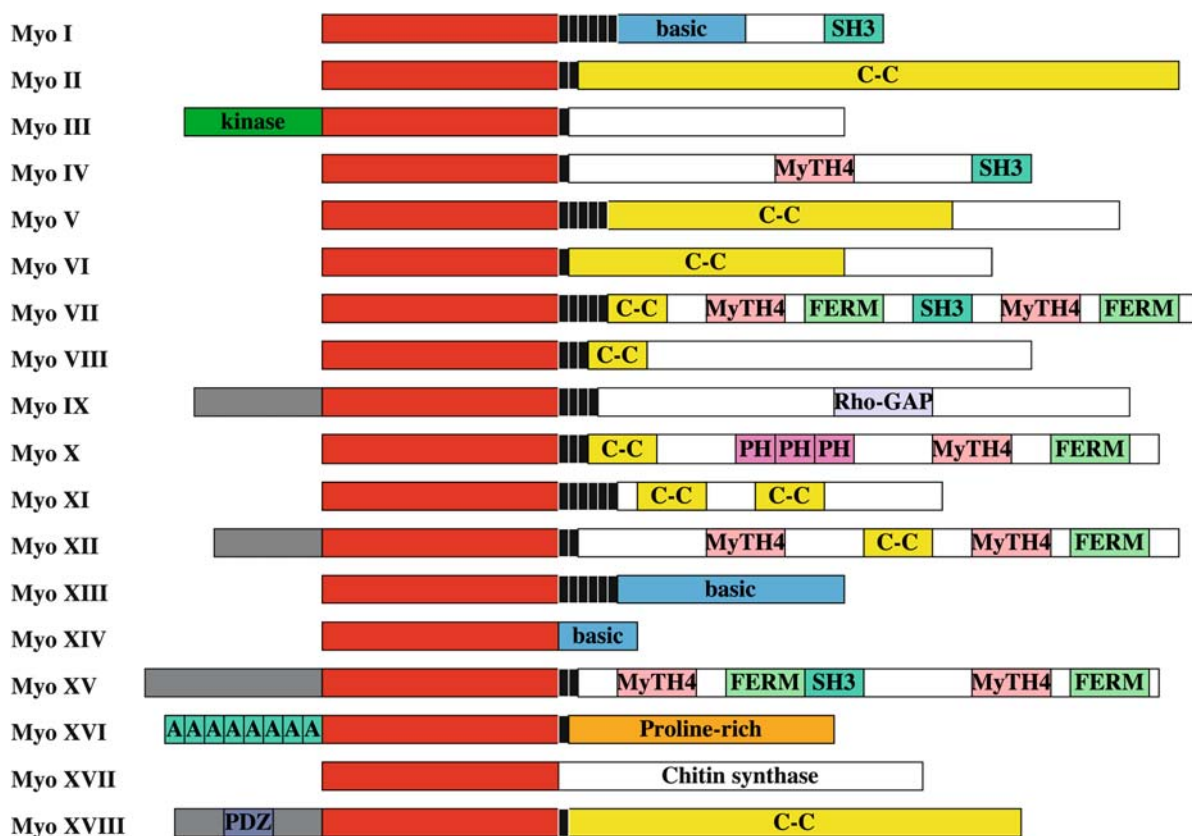
catalytic (or motor) **▶domain**, also referred to as the “head”. The two sites need to communicate with each other in the process of movement. In addition, the non-motor domains can harbor secondary structural elements important for motor function, including **▶coiled-coil** segments involved in dimerization and domains that function in mechanical amplification. Fifthly, most motors studied so far in some detail can generate a force that is sufficient to move even large objects through viscous cytoplasm. Finally, motors possess domains and/or associated proteins that confer specificity of function or are involved in regulation.

A combination of biochemical, molecular genetic and genomic approaches revealed that each of the three motor classes comprises superfamilies of motors of strikingly varied makeup and function outside the motor domain. Today, we can distinguish at least 18 different classes of myosins (Fig. 2), 10 different families of kinesins (Fig. 3) and two groups of dyneins, axonemal and cytoplasmic (Fig. 4). The complement of motors varies widely in different organisms (Table 1), ranging from 12 in yeast to over 100 in mammals. However, posttranslational modifications or various associated proteins may well increase the number of motors in a given cell several-fold. Many motors have not yet been characterized and clear functions are assigned to only a small subset. For a compilation of reviews, see (2).

The motor domain is the center of action in force generation and movement. Its size differs markedly between the three types of motors, being surprisingly small in kinesins (45 kD), about twice that size in myosins (100 kD) and exceptionally large in dyneins (500 kD). Structural information at atomic resolution has uncovered an unexpected similarity between kinesin and myosin. The central portion of the myosin head, which harbors the nucleotide-binding pocket, is virtually identical in structure to the core of the kinesin motor domain, despite a lack of significant sequence homology. In addition, both structures show similarities to the G-proteins in the region surrounding the nucleotide, the so-called switch I and switch II motifs. In all three proteins the switch regions shift position in

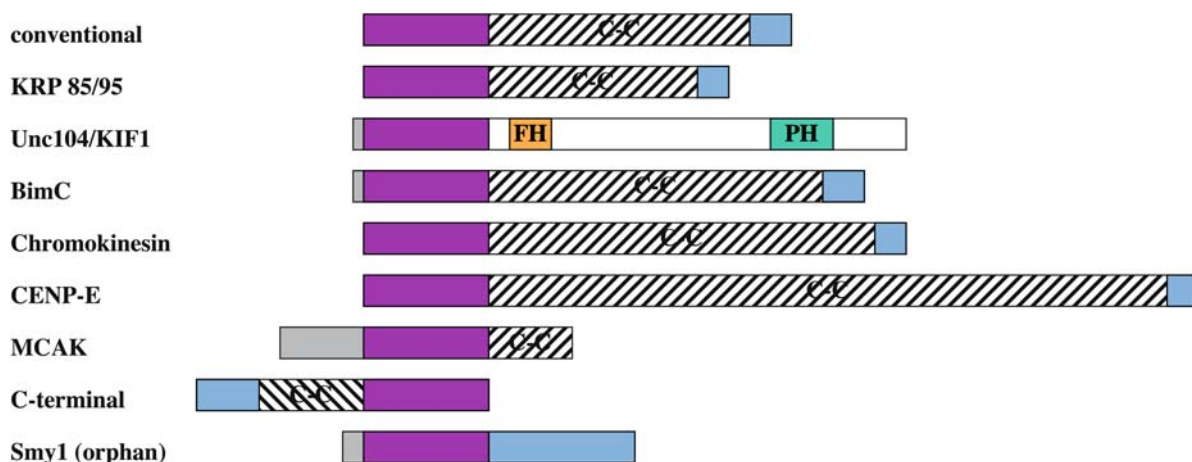


Molecular Motors. Figure 1 Schematic representation of prototypical molecular motors. Yellow: motor domains of the heavy chains. Brown: accessory light and intermediate chains. Blue: flexible regions. Black: coiled-coil domains.



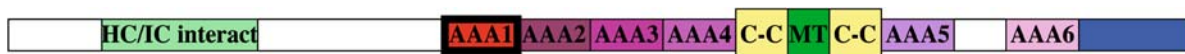
Molecular Motors. Figure 2 Overview of the domain organization of myosin motors. The motor domain is shown in red. Light chain binding regions (IQ-motifs) are shown in black. N-terminal extensions in grey. For more information, see the myosin home page at <http://www.mrc-lmb.cam.ac.uk/myosin/myosin.html>

A, ankyrin repeats; Basic, positively-charged region; C-C, coiled-coil domain; FERM, band 4.1; ezrin, radixin, moesin - family domain; kinase, kinase domain; MyTH4, myosin tail homology domain 4; PDZ, PSD-95, discs-large, ZO-1 -domain; PH, pleckstrin homology domain; Rho-GAP, rho GTPase-activating protein.



Molecular Motors. Figure 3 Overview of the domain organization of kinesin motors. The motor domain is shown in magenta. Tail domains are shown in blue. Smy 1 cannot be clearly assigned to one of the families and therefore is often designated as "orphan". For more information, visit the kinesin home page at <http://www.proweb.org/kinesin/>

C-C, coiled-coil domain (often interrupted); FH, forkhead-associated domain; PH, pleckstrin homology domain.



Molecular Motors. Figure 4 Domain organization of dyneins. The CC/MT domains form the microtubule-binding stalk extending from the globular motor domain (enlarged boxes). The AAA1-domain (boxed in bold) carries the catalytic site. AAA, triple-A domains 1-6 (ATPases Associated with various cellular Activities); C-C, coiled-coil regions; HC/IC interact, domain involved in the interaction with intermediate chains; MT, microtubule binding region.

Molecular Motors. Table 1 Complement of motors in selected model organisms

	Kinesins	Myosins	Dyneins
Saccharomyces cerevisiae	6 (4)	5 (3)	1
Schizosaccharomyces pombe	9 (7)	5 (3)	1
Dictyostelium discoideum	13 (7?)	12 (5)	1
Caenorhabditis elegans	21 (9)	17 (8)	1
Drosophila melanogaster	24 (10)	13 (10)	12
Homo sapiens	45 (10)	40 (12)	15
Arabidopsis thaliana	61 (7)	17 (2)	0

Total number of different motors (number of families in parenthesis)

Data compiled from various sources

response to nucleotide hydrolysis and thus are instrumental in the catalytic process. These findings are consistent with a common evolutionary origin of myosin, kinesin and G-proteins.

Dyneins are different. The dynein heavy chain is a member of the ►AAA⁺ ATPases, a highly diverse superfamily of proteins involved in a wide spectrum of cellular activities. In its C-terminal portion it contains six AAA⁺ modules apparently arranged in a ring. Only the first highly conserved AAA domain is believed to hydrolyze ATP and thus represents the catalytic site.

Initial Events in Force Generation

The key to an understanding of motor action is intradomain movement within the motor domain. In kinesins and myosins, small, initial structural changes triggered by ATP hydrolysis are translated into larger conformational changes associated with movement. In this process, nucleotide-dependent changes at the ►active site and the filament binding are tightly coupled. In myosin, the conformational change is communicated to the C-terminal region of the motor domain, an α -helix stabilized by associated light

chains. This domain swings through an angle of $\sim 70^\circ$ and thus acts as a mechanical amplifier of the small conformational change initiated in the ATP-binding site. According to a widely held model this swing of the lever arm is the ultimate cause for a “step” of the myosin motor along an actin filament.

Kinesin does not appear to possess a lever arm, but it has an analogous element, the neck linker, a motif of ~ 10 amino acids at the C-terminus of the motor domain. According to the most popular model for the kinesin mechanism, the neck linker undergoes a positional shift in response to nucleotide hydrolysis, being docked along the motor domain in the ATP state or mobile in the ADP state (3). Thus the structural elements involved in sensing and transmitting hydrolysis-dependent changes are similar in kinesin and myosin, but the last step of translation into a large-scale conformational change differs.

In dynein, the microtubule-binding site is located at the end of a unique ~ 10 nm stalk on the opposite side of the ring of six AAA⁺ modules. How dyneins transmit a conformational change in the first AAA unit to a microtubule-binding site located 20 nm away is presently unclear.

Taking a Step

A spectrum of behaviors is observed in the translation into movement of the structural changes in the motor domain. For example, some motors, such as conventional kinesin or mammalian myosin V, can move along their respective tracks for long distances without dissociating, a property referred to as processivity. In these dimeric motors the two motor domains are strictly coupled and cooperate. However, certain other motors that are also dimeric are non-processive (e.g., myosin II, members of the KAR3 family of kinesins). Here the motor domains presumably operate independently of each other.

A basic concept to explain these differences is the concept of the duty ratio, which describes the fraction of the time a motor domain spends attached to its track during one hydrolysis cycle (4). A single step can be envisioned to consist of a working stroke (where the head is bound to the track) and a recovery stroke (where it is detached). The hydrolysis cycles of the two heads must be coordinated in such a way that one of the heads is bound to the microtubule while the other is moving

forward to find a new binding site, and the two heads have to be kept out of phase. For this “hand-over-hand” cycle to work, the duty ratio must be at least 0.5 and a phase must exist in which both heads are bound to the microtubule. This enables a single molecule of conventional kinesin to take several hundred steps of 8 nm, which corresponds to the spacing of tubulin dimers, without dissociating from the microtubule. Motors that have a low duty ratio are non-processive. The two best-studied non-processive motors are myosin II and the kinesin-like protein *ncd*. Both are dimeric, but the two heads apparently do not cooperate. In muscle, for example, cooperation of the heads of myosin II is unnecessary because they operate as part of a large ensemble of motors within the sarcomere. Processivity probably evolved as a result of a physiological requirement for long-range movement driven by only one or a few motor molecules. Indeed, several processive motors function in organelle transport.

Directionality

Though most myosins and kinesins move to the plus end of their respective tracks, both superfamilies also have motors that move in the opposite direction, whereas dyneins are exclusively minus end-directed. Motors that can switch the direction of movement apparently do not exist.

Analysis of minus end-directed motors has offered important insights not only into the basis of reversed directionality, but also motor mechanochemistry. Minus end kinesins have the motor domain at the C-terminus, as opposed to the N-terminus for plus end motors. This placement alters head-neck interactions, which, in turn, are the key determining factor for directionality. When motor domains of forward and reverse motors are swapped, the resulting chimaeras adopt the direction of movement specified by the neck. The reversed polarity of the minus end-directed myosin VI motor is attributed to a unique 53-amino acid insertion in the converter domain that is proposed to reverse the direction of the lever arm swing.

Force and Velocity

Force measurements have been made on only a subset of motors in each superfamily. To measure these forces, ingenious micro-devices were developed that operate with unprecedented precision and sensitivity. They show that the forces developed by kinesin, myosin and dynein motors – around 1–10 pN – are extremely minute by our macroscopic standards. For example, to lift a 5 kg weight about 10,000,000,000,000 myosin motors are required. However, in the realm of the cell, a single motor can move an object many times its own size through viscous cytoplasm at near maximum speed.

Molecular Motors. Table 2 Speed and ATPase activity of selected motors

	Species	Speed (µm/sec)	K _{cat} (ATP/head/sec)
Conv. kinesin	animals	0.6	80
Conv. kinesin	fungi	2.6	260
Unc104/KIF1	mouse	1.2	110
<i>ncd</i>	<i>Drosophila</i>	– 0.1	1
NOD	<i>Drosophila</i>	0	10
Myosin I	<i>Acanthamoeba</i>	0.2	6
Myosin II (muscle)	mammals	8	20
Myosin V	birds	0.5	13
Myosin VI	mammals	– 0.06	1
Myosin XI	<i>Nitella</i>	60	–
Ciliary dynein	sea urchin	10	20
Cytoplasmic dynein	mammals	1.2	2

ncd and myosin VI are minus end-directed motors
Data compiled from various sources

Not surprisingly, the different families of motors exhibit a wide spectrum of motile properties, ranging from extremely slow to exceedingly fast movement (Table 2). Some motors are even unable to generate movement and instead have adopted a function in regulating microtubule dynamics (e.g. kinesins of the MCAK family).

Molecular Interactions

Molecular motors are involved in a wide variety of cellular activities, including contraction, organelle transport, cell movement, cell division, signaling, RNA localization, sensory transduction and basic developmental processes (Table 3). These activities are mediated by either direct or indirect interactions of the motor with its cellular targets (5). In both processes, non-motor domains and associated proteins play key roles and a spectrum of attachment mechanisms is observed.

Members of all three superfamilies of motors are implicated in organelle transport and membrane association, and various interaction mechanisms are known. Some motors, such as monomeric myosins, exhibit direct linkage to the phospholipid bilayer, notably acidic phospholipids. Likewise, a member of

Molecular Motors. Table 3 Involvement of molecular motors in various cellular activities

	Kinesins	Myosins	Dyneins
Organelle transport and maintenance	unc 104/KIF1, conv. kinesin, KRP85/95, C-terminal motor, chromokinesin/KIF4, rabkinesin6 (ungrouped)	myo I, myo V, myo VI, myo VII,	Cytoplasmic dynein
Endocytosis		myo I, myo VI	
Mitosis/meiosis	BimC, C-terminal motor, chromokinesin/KIF4, CENP-E, MKLP, MCAK/KIF2	myo V	Cytoplasmic dynein
Cytokinesis	MKLP, CENP-E; rabkinesin6, DmKLP3A (ungrouped)	myo II	
Transport of RNA, macromolecular complexes	KRP85/95, conv. kinesin, chromokinesin/KIF4	myo V	Cytoplasmic dynein
Flagellar movement			Axonemal dyneins
Axonemal biogenesis	conv. kinesin, KRP85/95		Cytoplasmic dynein
Microtubule dynamics	MCAK/KIF2		Cytoplasmic dynein
Microtubule interactions	BimC, MKLP, CENP-E		Cytoplasmic dynein
IF interactions	conv. kinesin		Cytoplasmic dynein
Actin interactions	MKLP		
Left/right asymmetry	KRP85/95		Axonemal dynein(s)
Cell movement/shape		myo I, myo II, myo III, myo V, myo VI, myo VII, myo X, myo XIV	
Microvilli and filopodia		myo I, myo V, myo VII, myo X	
Signaling	COSTAL2 (ungrouped)	myo I, myo III, myo IX	
Binding to other motors	conv. kinesin, Smy1 (to myo V)	myo V (to Smy1 and conv. kinesin), myo V (to dynein light chain 8)	Dynein light chain 8 (to myo V)
Sensory functions	KRP85/95	myo IIa, myo III, myo VI, myo VII, myo XV	LC Tctex1, LIS1 (interactor of DHC1)
Virus transport	conv. kinesin		Cytoplasmic dynein

Data compiled from various sources

the unc104/KIF1 family of kinesins interacts directly with lipids. Alternatively, motors such as conventional kinesin may bind directly to integral membrane proteins. Conventional kinesins from animal species co-purify in a 1:1 complex with a 70 kD protein termed kinesin light chain. *Via* these light chains, kinesin binds to amyloid precursor protein (APP), a transmembrane protein of certain neuronal vesicles. This link is of potential medical significance since APP has gained fame as the precursor of a proteolytic fragment that gives rise to amyloid plaques in patients with Alzheimer's disease.

In most cases, however, motors link with their target molecules *via* intervening proteins, often in the form of large assemblies. Thus conventional kinesin, again *via* its light chains, interacts with a class of scaffolding proteins that can bind certain components of signaling pathways. These scaffolding proteins in turn, may bind a transmembrane receptor of the LDL receptor family. Linkage *via* a subdomain in the non-motor region of kinesin heavy chain on the other hand, couples kinesin to the glutamate receptor. Two myosins involved in membrane transport likewise make use of linker proteins. Thus myosin VI forms a complex

Molecular Motors. Table 4 Molecular motors involved in disease aspects

Disease or defect	Motor involved
Myosin myopathies	Myosin II
User syndrome (USH1B)	Myosin VIIa
Griscelli syndrome (pigmentation disorder)	Myosin V
Nonsyndromic hearing loss (DFNA22)	Myosin VI
Hearing loss (DFNB3)	Myosin XVa
Hearing loss (DFNB30)	Myosin IIIa (ninaC)
Retinitis pigmentosa (photoreceptor degeneration)	Cytoplasmic dynein Kinesin Krp85/95
Primary ciliary dyskinesia	Axonemal dynein
Kartagener syndrome (situs inversus)	Axonemal dynein Kinesin Krp85/95
Polycystic kidney disease	Dyneins and kinesins
Lissencephaly	L1S1 (dynein/dynactin interactor)
Charcot-Marie-Tooth disease type 2A	K1F1B (unc104 kinesin family)
Virus transport	Conventional kinesin and cytoplasmic dynein
Anthrax susceptibility	K1F1C (unc 104 kinesin family)
Neurodegenerative disease	Kinesin and cytoplasmic dynein

with adaptor protein AP-2 and clathrin, promoting endocytosis. Myosin V, on the other hand, requires the small GTPase Rab27a and a recently identified Rab-binding protein, melanophilin, to bind to **melanosomes** in pigment cells. Here the GTPase binds to membranes first and recruits melanophilin, which then binds myosin V. Melanophilin binding is GTP-dependent, thus offering a means of regulating motor-cargo binding. A large protein assembly also seems to be involved in linking dynein to many of its sites of action. Through its intermediate chains, dynein interacts with a unique activator complex, dynactin. The dynactin complex seems to affect not only cargo binding but also motor activity *via* an as yet poorly defined mechanism.

Motors of all three superfamilies are also involved in mRNA localization. In yeast, certain mRNAs are transported in a complex with myosin V, whereas in neurons or insect oocytes microtubule motors are

required. The RNA is integrated into a ribonucleoprotein complex by RNA-binding and adaptor proteins, though the precise molecular interactions within this complex have yet to be determined.

Motors are increasingly being implicated in a number of human diseases (Table 4). These implications are firmly established in some cases (e.g., myopathies or hearing loss) and more tenuous in others (e.g., neurodegenerative diseases), but they suggest that motors may turn out to be useful targets for therapeutic applications (6).

Regulatory Mechanisms

The question of motor regulation is largely uncharted territory, though four different mechanisms have been identified to date, changes in calcium levels, post-translational modification, intramolecular inhibition, and modulation by accessory proteins (7).

The interaction of myosin with actin in skeletal muscle cells is regulated by calcium ions in cooperation with the accessory proteins troponin and tropomyosin. In response to a calcium spike, the troponin-tropomyosin complex associated with actin filaments is shifted, exposing the binding sites for myosin on actin and thus activating contraction. In mollusks, calcium binds to the light chains, turning on the ATPase activity and contraction. The light chains of vertebrate skeletal muscle on the other hand, do not bind calcium but can be phosphorylated, thus increasing force production. Other motors, or the proteins they interact with, may also be the substrate for kinases and phosphatases, affecting either the activity of the motors or interactions with other cell components. Thus phosphorylation of the intermediate chain of axonemal dynein inhibits its activity, whereas cytoplasmic dynein may be activated by phosphorylation *via* an as yet unspecified pathway. Phosphorylation affects the cargo binding of several motors. For example, phosphorylation inhibits docking of the globular tail of myosin V to melanosomes and releases dynein from membranes. There are hints that kinesin-based organelle movement can also be regulated by phosphorylation, but it is unclear whether cargo binding is affected directly.

A different mechanism that couples motor regulation and cargo binding is used by kinesin, where the tail folds up like a pocketknife to inhibit the motor domain. Binding to cargo de-represses tail inhibition and allows the motor to unfold, a process critically dependent on a flexible domain in the stalk. How the tail inhibits the motor is not known, but a crucial role is played by a tail motif conserved in all conventional kinesins.

Some organelles can either switch the direction of movement on the same track or even change tracks and move on both microtubule and actin filaments. This behavior requires the presence of opposite-polarity motors and/or motors from different superfamilies. Thus

in amphibian melanophores heterotrimeric kinesin and myosin V cooperate in the dispersion of pigment granules, while during aggregation myosin V is switched off, presumably by phosphorylation-dependent release from the granules. Myosin V and conventional kinesin have been shown to interact directly in their tail domains, but it is unclear whether motor coordination requires physical interaction in other instances as well. Small regulatory G-proteins are involved in several aspects of motor function. G-proteins may mediate motor-cargo interactions, as in the case of myosin V outlined above. Other studies likewise hint at a link between G-proteins or their effectors and several types of myosin motors. Since kinesin and dynein can also pair up with Rab effectors, a GTPase-dependent link between membrane traffic and motors is emerging.

References

1. Schliwa M, Woehlke G (2003) Molecular Motors. Nature 422:759–765
2. Schliwa M (ed) (2003) Molecular Motors. Wiley-VCH, Weinheim
3. Rice S, Lin AW, Safer D et al (1999) A structural change in the kinesin motor protein that drives motility. Nature 402:778–784
4. Howard J (2001) Mechanics of motor proteins and the cytoskeleton. Sinauer Associates, Sunderland
5. Vale RD (2003) The molecular motor toolbox for intracellular transport. Cell 112:467–480
6. Hirokawa N, Takemura R. (2003) Biochemical and molecular characterization of diseases linked to motor proteins. Trends Biochem Sci 28:558–565
7. Reilein AR, Rogers SL, Tuma MC et al (2001) Regulation of molecular motor proteins. Int Rev Cytol 204: 179–238

Molecular Symptom

Definition

Molecular symptom describes genes that are characteristic for a certain group of patients; they allow for molecular stratification of patient cohorts.

► [Computational Diagnostics](#)

Molten Globule

Definition

The molten globule was originally defined as a structural state that has a significant secondary structure, but with no fixed tertiary interactions. The

definition is used nowadays in a more relaxed way, to describe an ensemble of conformations that are highly dynamic, with a limited number of native-like contacts and only a partial native-like secondary structure.

► [Protein Folding](#)

Monilethrix

Definition

Monilethrix refers to an autosomal dominant transmitted hair disorder that is characterized by a beaded appearance of the hair shafts. Mutations in the hair cortex keratin hHb6 cause monilethrix.

► [Heritable Skin Disorders](#)

Monoclonal Antibodies

BURKHARD MICHEEL

Institute of Biochemistry and Biology, Department of Biotechnology, University of Potsdam, Germany
bmicheel@rz.uni-potsdam.de

Definition

Antibodies are molecules that specifically bind to those substances (antigens) against which an immune defence has been raised. Antibodies are synthesized by fully differentiated B lymphocytes (plasma cells). Each B lymphocyte and its descendants after clonal expansion produce a single antibody (a monoclonal antibody) that is identical in its amino acid composition and in its binding specificity to all the cells of that clone.

Characteristics

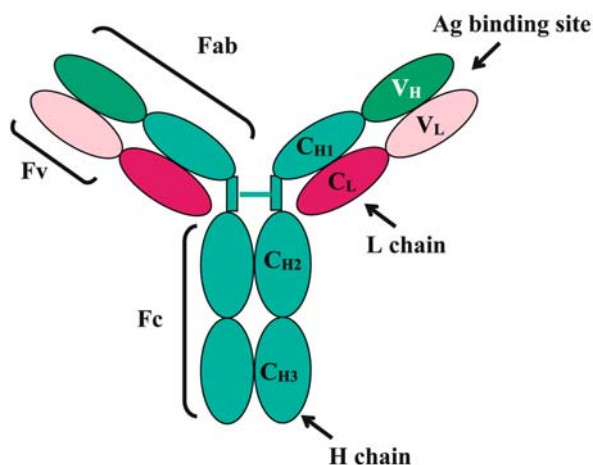
Production of Monoclonal Antibodies

Structure and Function of Antibodies

Antibodies or immunoglobulins (Ig) are globular glycoproteins that are produced by the immune system of vertebrates as a defence against foreign, non-self molecules, so-called antigens (antibody generators; 1). The antigens as targets of antibodies may be proteins, carbohydrates, lipids, nucleic acids etc. or completely synthetic compounds.

Under natural conditions antigens are in most cases parts or products of infectious agents. Antibodies are produced by fully differentiated B-lymphocytes that are called plasma cells.

Antibodies circulate in the body, bind to those antigens against which they have been produced and mediate the



Monoclonal Antibodies. Figure 1 Structure of antibodies of the IgG class. IgG molecules are composed of two identical heavy (H) and two identical light (L) chains which are linked by several disulfide bridges. The L chains contain one variable (v) and one constant (c) domain; the H chains contain one v domain and three c domains. A flexible region – the hinge region – between the first (CH1) and the second (CH2) domain of the H chain allows the free movement of the antigen-binding “arms” of an antibody molecule to different angles. Enzymatic cleavage may result in two antigen-binding fragments (Fab) and one crystallizable fragment (Fc). The Fc fragment mediates effector functions. A fragment containing only the v domains is called Fv.

neutralization and removal of foreign and potentially dangerous substances.

There is a very large number of different antibodies in the body (about 10^{11} different antibody specificities in one human individual) so that practically every potentially infectious agent can be controlled.

One B lymphocyte synthesizes only one antibody type, i.e. antibody molecules of unique structure and unique binding specificity. Daughter cells derived from such a lymphocyte (i.e. the corresponding B lymphocyte clone) therefore produce antibodies of the same specificity. The corresponding antibodies are called monoclonal antibodies.

Antibodies are relatively stable heterodimeric molecules built up by two identical heavy (H) and two identical light (L) chains. These chains associate in symmetric molecules with a “Y”-shaped structure (Fig. 1). This is the typical structure of IgG – immunoglobulin gamma – molecules, the most abundant antibody class in mammals.

Antibodies of different binding specificities have different amino acid sequences in the variable amino acid regions (V regions) in the N terminal part of both the heavy and light chains. Three hypervariable regions in the V regions of the heavy and the light chains (also called

CDRs, complementarity determining regions) constitute the antigen-combining site. The region that is detected by the antigen-binding site on an antigen is called the antigenic determinant or epitope. The binding specificity of an antibody is therefore determined by its ability to bind to a special epitope by means of its unique antigen-combining site. Each antibody of the IgG class has two identical antigen-combining sites, e.g. it is bivalent. Those regions of the antibodies not involved in antigen binding have identical amino acid sequences in any one Ig class or subclass (the so-called constant regions, C regions) and mediate binding to effector molecules and effector cells. These effectors mediate the elimination of the foreign substances from the organism. Antibodies are therefore bifunctional adapters between foreign antigens and the body’s own effector molecules and/or cells.

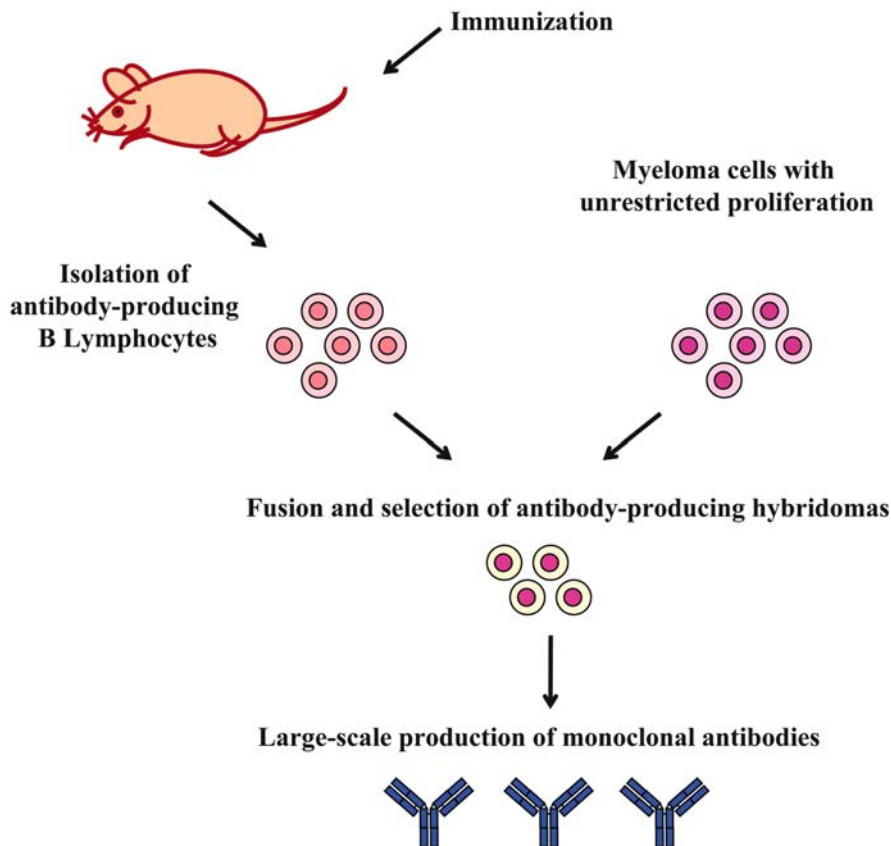
Since the immune system does not differentiate *a priori* between “dangerous” and “non-dangerous” substances, antibodies will also be produced after the injection of “harmless” antigens provided the antigens are foreign to the immunized vertebrate.

Since antibodies also bind their corresponding antigens *in vitro* or after transfer into other organisms, a large spectrum of useful applications for antibodies can be envisaged. ABO blood group determination, pregnancy assays and [passive immunization](#) against toxins during acute infections and after snakebites are well-known examples of antibody use.

Hybridoma Technique for the Production of Monoclonal Antibodies

Wide practical use of antibodies became possible after the hybridoma technique allowed the production of theoretically any antibody in high purity and large quantities (2). In the hybridoma technique, B-lymphocytes from immunized donors are fused *in vitro* with immortal myeloma cells (malignant B lymphocytes; Fig. 2). The resultant hybridoma cells contain the information for antibody production and unlimited multiplication. Those hybridoma cells producing the desired antibodies (because of their clonal origin called monoclonal antibodies) have to be selected and can then be propagated in large quantities for large-scale antibody production and also frozen for storage in liquid nitrogen. An uncountable number of different monoclonal antibodies against a multitude of antigens (including human tumor antigens) have been produced since the invention of the hybridoma technique. Most of these antibodies have been produced in mice because the immunization procedure is relatively easy and optimum myeloma cells for fusion are available. The production of human hybridomas by using human B-lymphocytes has turned out to be relatively complicated.

A breakthrough in solving this problem was the “creation” of the [XenoMouse](#), a transgenic mouse line that contains the complete human antibody gene



Monoclonal Antibodies. Figure 2 Hybridoma technique to produce monoclonal antibodies. After immunization of a mouse, spleen cells containing antibody-producing B lymphocytes are isolated and fused *in vitro* with “immortal” myeloma cells (malignant plasma cells). Fused hybridoma cells are selected by special selection media that kill non-fused cells. Hybridoma cells contain the information from both parental cells – for antibody production and unlimited cell growth. By a tedious screening process those hybridoma cells that produce the desired monoclonal antibody are identified. The cells can be mass cultivated to produce large amounts of antibodies and/or stored frozen in liquid nitrogen.

locus instead of the corresponding mouse locus (3). These mice produce fully human antibodies, even after the injection of antigens of human origin.

Production of Recombinant Antibodies

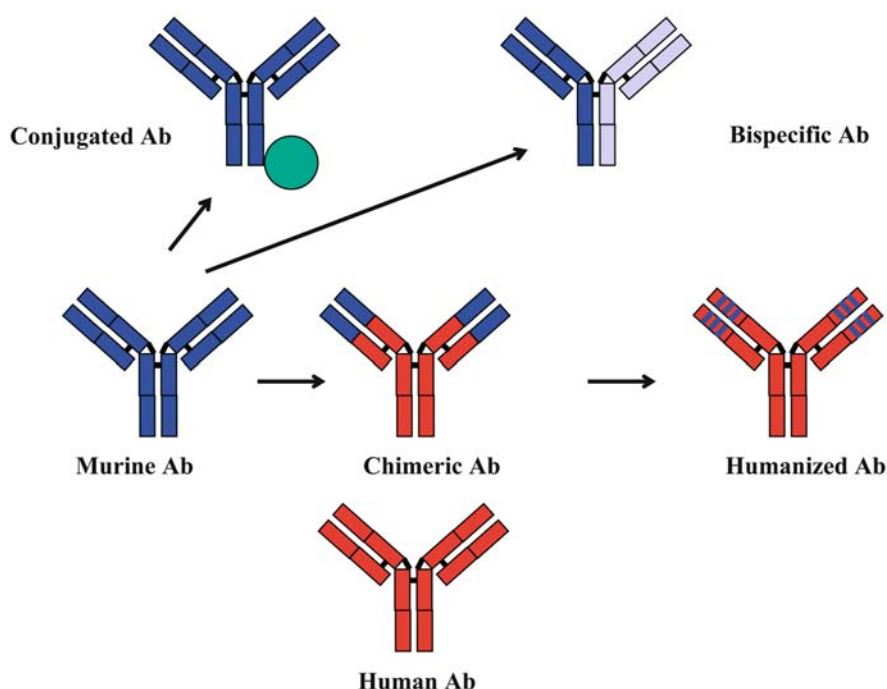
Antibody molecules can be conjugated with other proteins (e.g. enzymes) or small molecules (e.g. fluorescent markers and radioisotopes), which enhances the potential of use of antibodies as tools (Fig. 3).

In addition, DNA recombination techniques allow the cloning of antibody genes and the manipulation of antibodies at the gene level, which is a further step in enhancing the potential of the application of antibodies (Fig. 4).

A mixture of randomly generated primers (so-called degenerated primers) have to be used to clone and amplify by polymerase chain reaction (PCR) the variable region gene fragments of the heavy (V_H) and light chains (V_L) of antibodies – the unique and most important parts of antibodies. (Since all antibodies are different, one primer pair would not be enough for

amplification as in the case of genes coding for “normal” proteins and even when using a set of degenerated primers they do not “fit” for all antibody genes.) The V_H and V_L genes can be cloned in special expression vectors to allow the production of antibody fragments (e.g. single chain antigen-binding fragments – scFv) in prokaryotic (bacteria) or eukaryotic cells (yeasts, insect cells, mammalian cells, plant cells). Single chain antibody fragments (which are encoded by one gene in contrast to intact natural antibodies which are encoded by two genes) can be used as small binding molecules or can be linked on the level of the genes to other genes of interest.

The most important aspect of the cloning of antibody genes is the possibility of obtaining a large multitude of V_H and V_L antibody genes from normal and immunized human donors. These genes can be cloned as V_H - V_L -linked scAb genes in special display vectors. Thus one can create large antibody libraries, which allow, comparable to a normal immune system, the isolation of specifically binding antibody fragments (4). The



Monoclonal Antibodies. Figure 3 Modification of monoclonal antibodies. Monoclonal antibodies (in most cases of murine origin) can be modified in many ways depending on their intended use. They may be conjugated with specific markers (enzymes, radioisotopes, toxins, chemotherapeutics etc.) but it is also possible to produce bispecific antibodies with two different antigen binding sites by biochemical or cell fusion techniques (another possibility for producing bispecific antibodies is shown in Fig. 5). Two strategies are used to “transform” a murine antibody into a human or partly human antibody. The first is to fuse the variable regions of a murine antibody on the gene level with the constant regions of a human antibody. The results of this fusion are so-called chimeric antibodies. The second strategy is to insert the gene segments coding for the 6 complementary determining regions (CDR) into a human antibody framework; such antibodies are called humanized antibodies. By this method those parts of a human antibody that are responsible for antigen binding are replaced at the gene level by a murine sequence. Fully human monoclonal antibodies may be produced by hybridoma technology (using human B lymphocytes or B lymphocytes from transgenic mice containing a full set of human antibody genes) or by using human antibody gene libraries (as shown in Fig. 5).

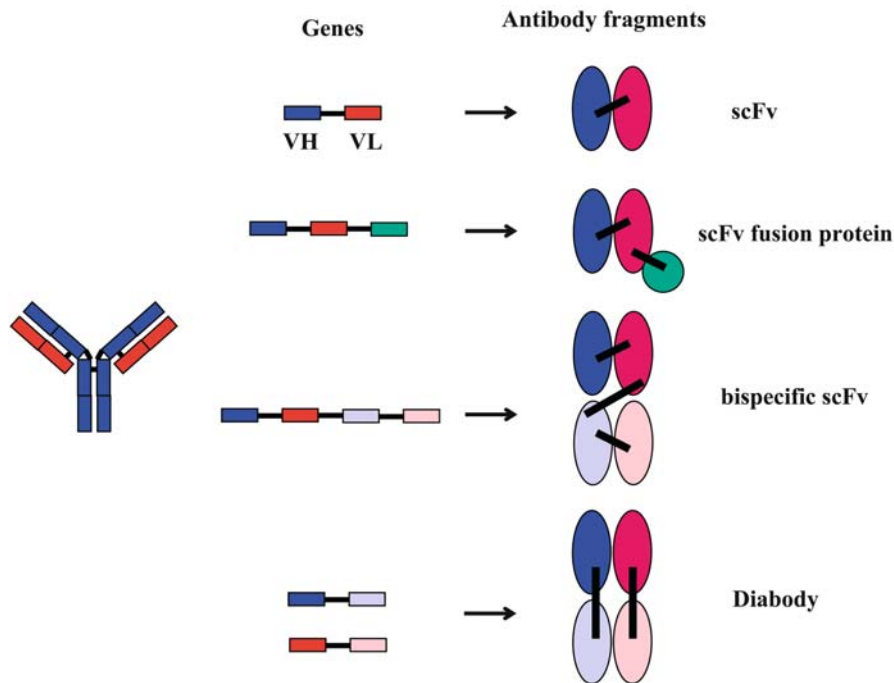
binding variability of such libraries can be enlarged by site-directed mutagenesis of the CDRs or by inserting synthetic oligonucleotides (semisynthetic or synthetic antibody libraries).

Different display systems (using phages, bacteria, yeast, ribosomes) were developed to select a specific antibody from such a library. The most common one is the ►phage display where the antibody genes are cloned into vectors based on filamentous fd phages. *Escherichia coli* infected and transformed by such phage vectors (phagemids) produce phages, which contain antibody genes and display the corresponding antibody fragments on the phage’s surface (Fig. 5). Those phages binding to a specific antigen can be selected and enriched by multiplying in *E. coli*. The corresponding antibodies can then be expressed and produced as scFvs. One can also produce other recombinant antibodies (Fig. 3). ►Chimeric antibodies can be produced after recombining the V_H and V_L antibody genes of murine origin with the corresponding

constant human light and heavy chain genes and ►humanized antibodies can be created by transferring the gene segments coding for the 6 complementary determining regions (CDR) into a human antibody framework (5).

Two further possibilities regarding antibody production and manipulation are worth mentioning. ►Bispecific antibodies, which, in contrast to natural antibodies, have two different binding sites, can be produced by biochemical manipulation of the molecules, by fusion of two hybridoma cells or by connecting two different scFv genes (Figs. 3, 4). Recombinant techniques also allow the production of antibodies with even more binding specificities.

The other possibility is the selection of antibodies with catalytic properties (e.g. esterolytic antibodies). Such ►catalytic antibodies (also called abzymes-antibody enzymes) may be especially valuable for catalyzing reactions for which no natural enzymes are available.



Monoclonal Antibodies. Figure 4 Antibody gene cloning and production of recombinant antibodies. Antibody genes coding for the variable heavy and light chain regions (v_H and v_L) may be amplified by the polymerase chain reaction (PCR) and cloned into expression vectors to produce antigen-binding single-chain Fv fragments in *E. coli* or in eukaryotic cells. The antibody genes may be fused with other genes (shown in green; coding for a toxin, an enzyme, green fluorescent protein, other antibody constant regions etc) allowing the production of fusion proteins, which may be applied for an assay or for diagnostic or therapeutic purposes. Antibody genes of different origin may also be combined to result in antibodies that bind two antigens (bispecific antibodies resulting from one gene construct or bispecific diabodies resulting from two constructs). Other constructs may allow the expression of antibodies with more than two binding specificities (not shown in the scheme).

Clinical Relevance

Monoclonal Antibodies in Diagnostics

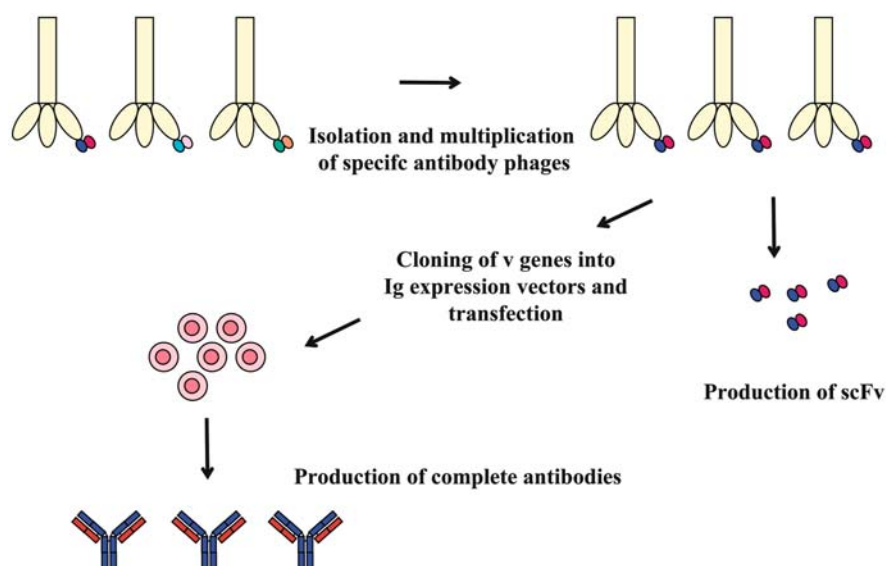
For diagnostic purposes only the binding specificities of antibodies are important. Monoclonal antibodies of murine origin are therefore sufficient.

In general, antibodies are applied for diagnostic purposes in immunohistology (or immunocytology), immunoassays and immunolocalization.

In immunohistology or immunocytology, specific antibodies are used to demonstrate specific foreign or self antigens in cells or on cell surfaces. The antigens may be normal cellular constituents or structures appearing because of pathological changes (tumor antigens, antigens of infectious organisms etc.). To visualize the antigens, the corresponding antibodies are labeled with enzymes (which convert a substrate into a visible product at the site of the antibody reaction) or fluorescent markers. The reactions can be visualized in a normal or fluorescence microscope. Especially valuable for hematological investigations is **cyto-fluorometry** where blood cells can be tested for their antigen profile under normal and pathological conditions. In some assays antibodies conjugated with

electron dense markers (colloidal gold, ferritin) can be used in electron microscopy, especially for the identification of virus infections.

In immunoassays, antibodies are used to detect soluble antigens (e.g. tumor markers, antigens of infectious organisms etc.) in body fluids or tissue extracts. The most commonly used test is the sandwich **ELISA** (enzyme-linked immunosorbent assay) where one antibody is linked to a solid phase (e.g. a plastic surface) as catcher and after incubating with the solution to be tested, the antigen bound to the first antibody is detected by a second indicator antibody. This second antibody is labeled with an enzyme that converts a nonvisible substrate into a colored (in some cases a fluorescent) product that can be measured. The intensity of the enzyme reaction is therefore a measure of the quantity of antigen in the solution investigated. There are several other test principles in use and in addition to enzymes other markers are used, including radioisotopes which were most popular in the past. In future **biosensors** and **microarrays** using antibodies will dominate many of the diagnostic tests at present still performed by conventional principles.



Monoclonal Antibodies. Figure 5 Antibody selection by phage display. Antibody genes coding for the variable heavy and light chain regions (v_H and v_L) are cloned in phage vectors (phagemids) which in *E. coli* allow the production of recombinant filamentous phages which display antibody fragments on their surfaces. Since a multitude of different v_H and v_L genes of natural and synthetic or mixed origin can be cloned into the phagemids, very large antibody libraries can be constructed. Phages displaying a specific antibody fragment can be isolated (e.g. by “panning” on the corresponding antigen adsorbed on plastic surfaces) and then multiplied by infecting *E. coli*. The genes coding for the specific antibody fragment now available from the corresponding phages may be used to produce single chain antibodies (scFv) in bacteria. They may also be cloned into vector cassettes containing the genetic information for the production of complete antibody molecules. These vectors can be used to transfect eukaryotic cells that may then synthesize intact whole antibody molecules.

Immunolocalization of diseased tissues *in vivo*, e.g. tumors, is based on the idea that, after injection of radiolabelled antibodies into tumor patients, the antibodies will be transported, as under normal conditions, to the tumor tissue and will then bind the tumor cells so that this site can be visualized scintigraphically. But under *in vivo* conditions, binding specificity is not the only prerequisite for the antibodies in use. Pharmacokinetics and especially the origin of the antibodies are important properties that have to be taken into consideration. These problems are especially obvious when using antibodies for therapy.

Monoclonal Antibodies in Therapy

The natural function of antibodies *in vivo* is to prevent dangerous substances from entering cells (neutralization of toxins and viruses), to mediate the removal of foreign substances (by binding to the foreign antigens and to the body's own phagocytic cells) and to destroy or to inhibit the growth of foreign, infected or abnormal cells (by inhibiting growth receptors, activating apoptotic signals and binding to foreign antigens and to the body's own effector molecules and/or cells). Antibodies are therefore regarded as “magic bullets” and were already used for passive immunization against acute infections (e.g. diphtheria) at the begin-

ning of the 20th century (6). The main obstacle in such a treatment is the very fast clearance of antibodies of non-human origin in the human body (production of HAMAs – human anti-mouse antibodies – after injection of murine monoclonal antibodies). Chimeric, humanized or fully human antibodies are therefore preferably used for immunotherapy (Fig. 3, Table 1). Immunotherapy by passively administering monoclonal antibodies is used for treating intoxications, infectious diseases, immunologically caused diseases and cancer (7). In the first examples, antibodies are injected which neutralize a toxin or a virus or destroy invading agents. To treat pathological immune reactions (transplant rejection, autoimmunity, allergies) antibodies are injected which bind to structures (molecules on immune cells, soluble mediators of immune reactions etc.) involved in overshooting immune reactions and thus diminish the effect. Antibody therapy of cancer uses in most cases the cytotoxic effects of anti-tumor antibodies and in a few cases the blocking of growth factors (Fig. 6). The antibodies are used as “naked”, conjugated or bispecific antibodies in a more or less “humanized” form. There are more than 10 monoclonal antibodies approved for clinical use in various countries at present, and about 200 antibodies in clinical trials.

Monoclonal Antibodies. Table 1 Therapeutically applied monoclonal antibodies¹⁾

Antibody ²⁾	Name	Used since ³⁾	Indication	Producer ⁴⁾
Humanized anti-F-protein	Synagis	1998	Viral disease ⁵⁾	MedImmune
Murine anti-CD3	OKT-3	1986	Organ transplants	Ortho Biotech
Chimeric anti-TNF- α	Remicade	1998	Chron's disease	Centocor
Humanized anti-IgE Fc	Xolair	2002	Allergy	Tanox
Murine anti-EpCAM	Panorex	1995	Colon cancer	Centocor
Humanized anti-ERBB2	Herceptin	1998	Mammary cancer	Genentech
Humanized anti-CD52	Campath	2001	B-cell leukemia	Mill. Pharmaceut.

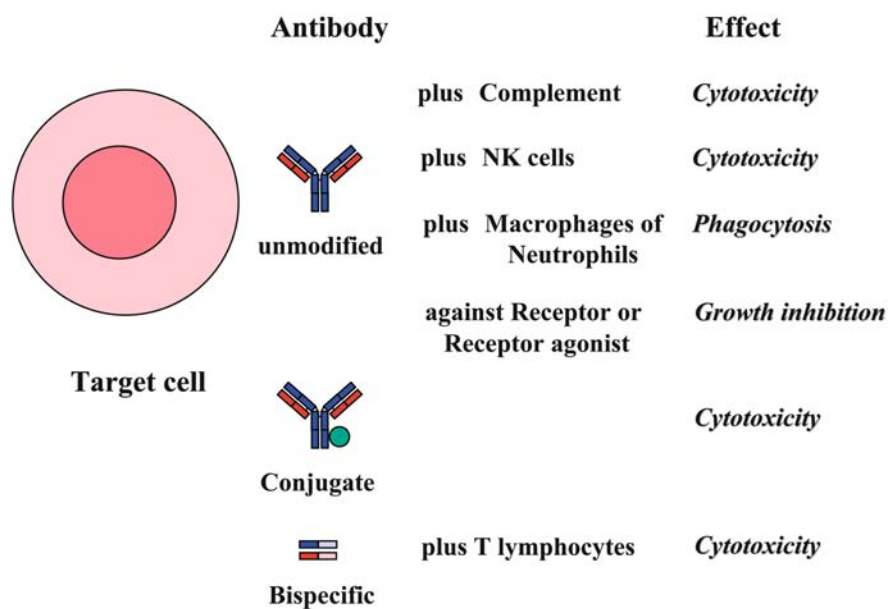
¹⁾ Modified after Brekke and Sandlie (2003)

²⁾ Indication of antibody type and target antigen

³⁾ Indication of approval date; in some cases only for some countries

⁴⁾ Company producing the antibody; full name not given in all cases, in some cases only the first producer is indicated

⁵⁾ Only against infection by respiratory syncytial virus



Monoclonal Antibodies. Figure 6 Efficiency of passively administered antibodies to inhibit the growth of target cells *in vivo*. Intact unmodified antibodies may lyse target cells with the help of complement or natural killer (NK) cells. The Fc part of the antibodies mediates the binding of the effector molecules or effector cells. Phagocytic cells may also take part in such reactions either by phagocytosing smaller cells or the remains of destroyed cells. Growth inhibition of tumor cells may be initiated by antibodies that specifically block the binding of a growth factor to its receptor at the tumor cell's surface (either by reacting with the growth factor or with the receptor). Killing of tumor target cells may also be possible by specific antibodies conjugated (directly or *via* genetic engineering) to a toxic molecule (a small toxic molecule or a toxic protein, a radioisotope, an enzyme which converts a harmless substance into a toxic one – the so-called ADEPT – antibody-dependent prodrug therapy etc.). Destruction of tumor target cells is also possible with the help of bispecific antibodies (this figure shows recombinant bispecific antibodies) that react *via* one “arm” with the tumor cells and *via* the other “arm” with T cells. The tumor cells are lysed by the activation of the cytotoxic potential of the body's own T cells after antibody binding. Other cytotoxic effector cells (e.g. NK cells) can also be activated by bispecific antibodies. In a few cases, bispecific antibodies have also been used in an ADEPT format with one antibody “arm” derived from a catalytic antibody that converts a harmless substance into a toxic one. All these possibilities have proven useful for many experimental models *in vitro*. Which of these mechanisms is most effective *in vivo* has yet to be clarified for most human diseases.

Concluding Remarks

Antibodies have proven to be valuable tools in many fields of biomedicine. The use of antibodies in therapy is promising but we are still far from having resolved all problems. For many diseases (e.g. infections, many forms of cancer) the antigens most suitable for attack still have to be identified and the corresponding antibodies still have to be isolated. It is obvious that in many cases one antibody is not enough for treatment (under natural conditions a polyclonal antibody response is initiated in all cases). It is not yet clear either for which purposes antibody fragments may be more convenient than intact antibodies and which of the many different antibody conjugates or fusion proteins may lead to better therapy.

For practical uses antibodies have to be produced in large quantities. Large fermenters used for the cultivation of bacteria and yeast can be applied for hybridoma cultivation in only a few cases. Because of their higher sensitivity to mechanical stress, cultivation in hollow fiber systems is a more gentle method for hybridoma mass cultivation. Since the production costs are very high, we have to look for cheaper alternatives, but it is not yet clear which of the methods developed so far (other expression systems as e.g. bacteria, yeast, insect cells, transgenic plants, larger transgenic animals) will be used in future. Efforts will also have to be intensified to develop quicker and easier selection methods for human monoclonal antibodies.

In addition to the applications mentioned, antibodies will also be used for other purposes, e.g. as mediators for specific [▶active immunization](#) or as intracellular binding molecules ([▶intrabodies](#) which may influence the metabolism of specific cells). It should also be mentioned that alternatives for antibodies might be possible in special cases. Other binding protein molecules have been selected and antibody mimetics with similar binding properties (peptides, [▶aptamers](#), [▶molecular imprints](#)) have been produced. The future will bring a mosaic of different and highly specific binders among which antibodies will surely play a very important but not the only role.

[▶High-Throughput Approaches to the Analysis of Gene Function in Mammalian Cells](#)

References

1. Janeway CA, Travers P, Walport M et al (2001) Immunobiology – The Immune System in Health and Disease, 5th edn. Garland Science, New York
2. Köhler G, Milstein C (1975) Continuous cultures of fused cells secreting antibody of predefined specificity. *Nature* 256:495–497
3. Green LL (1999) Antibody engineering via genetic engineering of the mouse: XenoMouse strains are a vehicle for the facile generation of therapeutic human monoclonal antibodies. *J Immunol Methods* 231:11–23

4. Vaughan TJ, Williams AJ, Pritchard K et al (1996) Human antibodies with sub-nanomolar affinities isolated from a large non-immunized phage display library. *Nat Biotechnol* 14:309–314
5. Hudson PJ, Souriau C (2003) Engineered antibodies. *Nat Med* 9:129–134
6. Gura T (2002) Therapeutic antibodies: magic bullets hit the target. *Nature* 417:584–586
7. Brekke OH, Sandlie I (2003) Therapeutic antibodies for human diseases at the dawn of the twenty-first century. *Nat Rev Drug Discov* 2:52–62

Monoclonality/Monoclonal

Definition

Monoclonality/monoclonal refers to a cell population derived from a single progenitor. The term is also used to describe antibodies that are produced by the identical offspring of a single, cloned antibody producing cell.

- ▶Autosomal Dominant (Inherited Disorder)
- ▶Hyper- and Hypoparathyroidism
- ▶Monoclonal Antibodies
- ▶Polycystic Kidney Disease, Autosomal Dominant

Monogenic Disorder/Monogenic

Definition

Monogenic disorder/monogenic describes a disorder that is caused by mutations in just one gene (such as hemophilia), as opposed to polygenic disorders (such as hypertension) that involve the interaction of several genes. Monogenic disorders are also known as “single-gene disorders” or as “Mendelian disorders”, because their inheritance patterns tend to follow the genetic laws first described by Gregor Mendel.

- ▶Chromosome 21 Disorders
- ▶Familial Dilated Cardiomyopathy
- ▶Genetic Screening in Populations
- ▶Hereditary Spastic Paraplegias
- ▶Mendelian Forms of Human Hypertension and Mechanisms of Disease
- ▶Parkinson’s Disease: Insights from Genetic Causes

Monogenic Hypertension

- ▶Mendelian Forms of Human Hypertension and Mechanisms of Disease

Monogenic Inheritance

Definition

Monogenic inheritance refers to the inheritance that is controlled by the alleles for one particular locus, as opposed to di- tri- or polygenic control exerted by two three or many non-allelic genes.

► [Mendelian Forms of Human Hypertension and Mechanisms of Disease](#)

Monolayer

Definition

Animal tissue cells growing on solid surfaces have the tendency to cover the surface with a complete layer only one cell thick, before growing on top of each other.

► [High-Throughput Approaches to the Analysis of Gene Function in Mammalian Cells](#)

Monooxygenases

► [Dioxygenases, Monooxygenases and Oxidases](#)

Monosaccharide

Definition

Monosaccharide defines a carbohydrate consisting of a single sugar unit with the general formula $C_nH_{2n}O_n$, such as tetraoses, pentoses, hexoses, and heptoses.

► [Biochemical Engineering of Glycoproteins](#)

► [Glycosylation of Proteins](#)

Monosomy 1p36 Syndrome

Definition

Monosomy 1p36 (Slavotinek syndrome) is a comparatively frequent (1:10,000 newborns), sometimes recognizable, subtelomeric deletion syndrome caused by monosomy 1p36. It is characterized by short stature, microcephaly, large fontanelle, and mild facial dysmorphism. The extent of mental retardation, mild to severe, depends on the actual deleted chromosomal portion.

► [Microdeletion Syndromes](#)

Mono-Ubiquitination

Definition

Ubiquitin is a polypeptide consisting of 76 amino acids which serves to tag proteins for degradation or intracellular trafficking. Ubiquitin is attached to ϵ -amino groups of lysine residues in the target protein by iso-peptide bond formation with its carboxyl-terminal glycine residue. Mono-ubiquitination can be extended by connecting the ϵ -amino group of lysine 48 of one ubiquitin molecule to the terminal glycine of another. Whereas mono-ubiquitin residues tag membrane proteins during endocytic and late endosomal sorting processes, poly-ubiquitinated proteins are degraded by the proteasome.

► [Protein and Membrane Transport in Eukaryotic Cells](#)

Monozygotic Twins

Definition

Monozygotic twins are “identical” twins that develop from one zygote. They are genetic duplicates of each other and identical at the genetic level.

► [Diabetes Mellitus, Genetics](#)

► [Schizophrenia Genetics](#)

Morbus Wegener

JULIE M. WILLIAMS, CAROLINE O.S. SAVAGE
Renal Immunobiology, The Medical School,
University of Birmingham, Birmingham, UK
j.m.williams.med@bham.ac.uk

Synonyms

Small vessel vasculitis; Wegener's Granulomatosis

Definition

Morbus Wegener (hereafter called Wegener's Granulomatosis, WG) is an autoimmune disorder of incompletely understood pathophysiology. It is a member of a family of diseases, characterised by chronic inflammation of the blood vessels, which are classified by the term ► [vasculitis](#). Histopathological features include perturbations to the vessel wall, an infiltration of leukocytes into the perivascular space and fibrin deposition into the subendothelial area. WG comprises a triad of granulomatous inflammation of the respiratory tract, a systemic necrotising vasculitis and a

segmental necrotising ► **glomerulonephritis** (although forms can be limited to one organ). Aetiology of the disease is probably multifactorial including genetic and environmental causes.

Characteristics

The incidence of WG is estimated at 20/million adults with no sex or race distribution but increased frequency in people in their 50s and 60s. The pathogenesis has yet to be elucidated but the majority of patients present with autoantibodies against their ► **neutrophils** (anti-neutrophil cytoplasm antibodies, ANCA). In >70% these antibodies are raised against an intracellular protein termed ► **proteinase 3** (PR3) and in <25% they are against another protein, ► **myeloperoxidase** (MPO) (1). The presence of these ANCA correlates with disease activity and is a good indicator of the onset of relapse. Interestingly high ANCA titres have been observed in patients without active disease and additionally active disease can exist without the presence of ANCA. This would therefore imply that other ANCA-independent mechanisms are also operating.

The chief autoantigen PR3 is a serine protease found mainly in the azurophil granules of neutrophils (but also in monocytes). It is predominantly involved in the host response to bacteria and exhibits lytic activity. There are a number of naturally occurring inhibitors *in vivo*, including alpha1-antitrypsin (α 1AT). PR3 can be constitutively expressed on the surface of neutrophils but predominantly is translocated there after an inflammatory stimulus. This leads to binding of ANCA via its F(ab')₂ (fraction antigen binding) fragment. However this is inadequate to lead to neutrophil activation, there must also be engagement of the Fc (fraction crystallisable) region of the antibody with Fc γ receptors present on the neutrophil surface. This results in downstream intracellular signalling, ► **reactive oxygen species** production, degranulation and production of proinflammatory ► **cytokines**.

Once activated, neutrophils also show increased adherence to the ► **endothelium** and transendothelial migration. *In vitro* the neutrophils exhibit cytolytic activity towards endothelial cells leading to cell damage. Additionally, the neutrophils undergo an accelerated rate of ► **apoptosis** but without a concomitant clearance by macrophages, resulting in secondary necrosis and release of intracellular contents.

Animal models of vasculitis have provided only limited information and there are no useful models of WG induced by PR3 directed ANCA.

Cellular and Molecular Regulation

It is now widely recognised that genetic factors can play a major role in autoimmune diseases and a number of familial cases have been recognised for WG. No

particular genes have been identified to date but there are reported associations with ► **polymorphisms** in the genes for PR3, α 1-AT and other genes involved in the immune response.

Polymorphisms

The genes for the constitutively expressed Fc receptors found on neutrophils (Fc γ RIIa and Fc γ RIIb) and monocytes (Fc γ RIIIa) have been extensively examined (2). For Fc γ RIIa both a high affinity (H/H131) and a low affinity (R/R131) form are known, as well as a heterozygous intermediate. Fc γ RIIIa also has an interchange at 158 being either V/V V/F or F/F. Fc γ RIIb is known to be either NA1 or NA2 (neutrophil antigen 1 or 2). No significant associations of any of the polymorphisms with WG have been detected. However, combinations of these mutations may be important. One group showed that if the Fc γ RIIa H/H phenotype was combined with the Fc γ RIIIa V/V then this was an increased risk for WG. They postulated that these lead to a stronger interaction of ANCA with the receptors and therefore more activation. Another study looked at risk of relapse and detected a correlation between this and combined Fc γ RIIa R/R and Fc γ RIIIa F/F. The reasoning this time was that these patients were weak IgG binders and therefore less able to mount an inflammatory response, leaving them susceptible to infection by known risk factors such as *Staphylococcus aureus*.

Associations have also been shown with the complement system but with no clear functional significance. A relationship of WG with both the C3F and the C4a3 allele have been observed.

Other polymorphisms have been examined pertaining to the immune system such as CD18, tumour necrosis factor- α (TNF α), interleukin-2 and interleukin-5Ra but with no connection to WG.

Polymorphisms have been observed in the PR3 gene itself, with an amino acid change at position 119, an 84 base pair deletion/insertion, seven individual nucleotide polymorphisms and a microsatellite. The polymorphism in the promoter region has been found to be prevalent in WG but no downstream functional effects are noted. Neutrophils have been shown to have PR3 in the secretory vesicles as well as the azurophil granules and this can be easily translocated to the plasma membrane. The percentage of these neutrophils with membrane expression of PR3 on resting cells was constitutively determined for each individual and appears to have a genetic background based on familial studies. Neutrophils can exhibit three different patterns of membrane PR3 expression, low, high or bimodal whereby only a subset of cells are positive. The proportions of positive cells remain stable over time. High expression on resting neutrophils was found to be increased in vasculitis but the functional significance

was not clear, as it was not restricted to patients with WG. There was however an increased incidence of relapse in this cohort (3).

The inhibitor of PR3 α 1-AT is also polymorphic at its PI locus with its phenotypes classified as normal (MM), slightly deficient (MZ) or severely deficient (ZZ). In WG patients, an increased incidence of the Z in either form was determined although there was no increased risk factor associated. A linkage disequilibrium has also been exhibited on chromosome 14q32.1 that contains the serpin gene cluster.

Antibody

Increases in ANCA titre are a good indication of active disease or onset of relapse. Data on subclass distribution for the antibodies are controversial with some data showing an increase in IgG3 whilst others IgG1 and 4. There is, however, an overall predominance of aglycosylated IgG. This is purported to lead to activation of the mannose binding lectin pathway and complement activation. Anti-PR3 antibodies are polyclonal but have been shown to recognise a restricted number of epitopes that are closely related. Using computer simulations these are reported to be intimately linked with PR3's catalytic site. Epitope switching has also been observed between active and remission phases of WG.

Cellular Immunity

Disease activity has been shown to correlate with monocyte activation. Activated monocytes and macrophages participate in ►granuloma formation by synthesising and secreting a variety of chemokines, cytokines and growth factors. PR3-ANCA applied to monocytes can induce the release of TNF α , interleukin-1 β , interleukin-6, interleukin-8, monocyte chemoattractant protein-1 and thromboxane A $_2$. Additionally they may exacerbate vascular damage by the release of reactive oxygen species and lytic enzymes.

There is now increasing evidence for a role of T cells in systemic vasculitis and ANCA IgG subclass switching indicates T cell help. Active CD4 $^{+}$ cells are seen in patient sera that show enhanced proliferative responses and secrete cytokines. Additionally high soluble interleukin-2 receptor levels are seen in active disease. Autoreactive T cells have been observed in some patients but there is no evidence to date that they initiate damage. CD4 $^{+}$ /CD28 $^{-}$ T cells are found in granulomas and produce a Th1-type cytokine pattern of interferon γ and TNF α .

HLA is involved in immune recognition by T cells, with class I presenting to CD8 $^{+}$ cells and class II to CD4 $^{+}$ cells. Associations, however, are hard to demonstrate in WG. There are reports of positive associations with DQw7 and weakly with DR1 and

negative associations with DR3, DR13 and DR6 but none have been observed to be linked with progression and no significant and consistent pattern has emerged. Polymorphisms in the CTLA-4 gene have been strongly associated with WG and may lead to an increased activation of T cells. CTLA-4 (AT)n microsatellites were strongly associated in affected patients, with the shortest allele showing a decreased prevalence over healthy controls.

WG has also been positively associated with nasal carriage of *Staphylococcus aureus* but there is no link to involvement of superantigens.

Gene Activation

Real time PCR and DNA micro array on leukocytes from patients and leukocytes treated with whole and F(ab') $_2$ ANCA IgG, showed activation of a subset of genes (4). Some of these genes were common to the whole IgG, some only to F(ab') $_2$ and some responded to either. These included DIF-2 (differentiation dependent gene-2 or radiation inducible immediate early gene, IEX-1, gly96 in mice, pituitary adenylate cyclase activating polypeptide responsive gene PRG1 in rat) and IL-8 (interleukin-8) along with a number of molecules associated with inflammation and cell signalling. DIF-2 was upregulated in patients and strongly correlated with disease activity and ANCA titre but was not present in remission or controls. IL-8 was increased only in remission.

Clinical References

WG can be fatal if left untreated. ►Immunosuppressants play a large role in preserving organ function but toxicities can contribute to mortality. Treatments are tailored so as to induce a rapid remission from acute episodes by administering high doses short-term. This is then tapered or altered to less toxic alternatives for maintenance. Survival rates are now estimated to be above 75%. Adverse prognostic factors are known to be age, renal impairment and pulmonary involvement (5).

References

1. Savage COS, Harper L, Holland M (2002) New findings in pathogenesis of ANCA-associated vasculitis. *Curr Opin Rheum* 14:15–22
2. Kallenberg CGM, Rarok A, Stegeman CA (2002) Genetics of ANCA-associated vasculitides. *Cleveland Clin J Med* 69 S2:SII-61–63
3. Rarok AA, Limburg PC, Kallenberg CGM (2003) Neutrophil-activating potential of antineutrophil cytoplasm autoantibodies. *J Leuk Biol* 74:3–15
4. Yang JJ, Preston GA, Alcorta DA et al (2002) Expression profile of leukocyte genes activated by anti-neutrophil cytoplasmic autoantibodies (ANCA). *Kid Int* 62:1638–1649

5. Jayne DR (2002) Conventional treatment and outcome of Wegener's Granulomatosis and microscopic polyangiitis. *Cleveland Clin J Med* 69 S2:SII-110–115

during the embryonic development of an organism, which results in the various tissues, organs and anatomy. Secreted factors, which are instructive in these processes, are referred to as morphogenes or are described as ►[morphogenic](#).

►[Bone and Cartilage](#)
►[Lung](#)

Morgan Unit(s)

Definition

Morgan unit(s) is a measure of genetic recombination by crossing over (on average one crossing over per meiosis). It defines the distance by which two genes are separated. One centimorgan (cM) refers to 1% of recombination.

►[Map Distances](#)

Morphant

Definition

Morphant refers to a mutant phenotype that is caused by gene inactivation by antisense morpholino oligonucleotides.

►[Morpholino Oligonucleotides, Functional Genomics by Gene 'Knock-down'](#)

Morphogen/Morphogenic Factor

Definition

A morphogen or a morphogenetic factor is an inductive developmental signal that is diffusible, and acts at a distance to regulate pattern formation in a concentration dependent manner.

►[Limb Development](#)
►[Neural Development](#)

Morphogenesis

Definition

Morphogenesis describes the process of cellular differentiation, distribution and growth that takes place

Morpholino Oligonucleotides, Functional Genomics by Gene 'Knock-down'

SALIM ABDELILAH-SEYFRIED

Max Delbrück Center (MDC) for Molecular Medicine, Berlin, Germany
salim@mdc-berlin.de

Definition

Morpholino oligonucleotides are chemically modified ►[antisense oligonucleotides](#) with a high affinity for binding to selected ►[messenger RNAs](#), thereby preventing ►[translation](#) or correct mRNA splicing (1). They are primarily used in a variety of species for the '►[knock-down](#)' of gene function (2, 3). Morpholino oligonucleotides consist of the four genetic bases (adenine, cytosine, guanine, and thymine) linked to a morpholine ring, from which the term morpholino oligonucleotide is derived. Typically between 18 and 25 bases are coupled by non-ionic phosphorodiamidate intersubunit linkages. This particular backbone moiety ensures resistance to ►[nucleases](#) and often has negligible toxicity that, unlike ►[nucleic acids](#), allows applications of high concentrations without reverse effects.

Characteristics

As a result of the tremendous pace of diverse genome sequencing projects, large sets of gene sequences that currently lack ascribed biological functions have been identified. Morpholino oligonucleotides present a tool for ►[reverse genetics](#) in model organisms that either lack ►[gene targeting by homologous recombination](#) ('►[knock-out](#)' technology) or in which gene inactivation through RNA interference or ►[siRNAs](#) are not reliably established. Moreover, morpholino oligonucleotides are a tool for *in vitro* and tissue culture systems. The term 'morphant' has been proposed to describe those animals in which a specific gene function has been inactivated by morpholino targeting

(4). The combination of genome sequencing projects and systematic 'knock-down' studies based on morpholino oligonucleotides in suitable model organisms allows for the cataloging of 'morphant' phenotypes of entire [▶transcriptomes](#). However, these studies are largely dependent upon injecting the morpholino oligonucleotides into early and accessible developmental stages of animals, thus limiting the use of morpholino oligonucleotides to genes that act early during development. Alternatively, square pulse [▶electroporation](#) and other modified delivery methods allow for the direct application of morpholino oligonucleotides for tissue-directed 'knock-downs' in chicken, mouse and rat (3, 5, 6). Titering the morpholino oligonucleotide concentrations applied for gene 'knock-down' experiments can provide a series of 'morphant' phenotypes ranging from mild to severe, the equivalent of an [▶allelic series](#).

Morpholino Oligonucleotide Design

The standard design is to generate antisense morpholino oligonucleotides that target between 18 to 25 bases of the mRNA [▶5' untranslated region \(5'UTR\)](#) prior to or overlapping with the [▶AUG translational start site](#) (AUG start). Efficiency of the [▶translational inhibition](#) is highly dependent upon the target position, enabling an almost complete translational inhibition provided that the target sequence is selected within about 85 base pairs 5' and 25 base pairs 3' of the AUG start. Target sequences that are selected more 3' to the AUG start show significantly decreased translational inhibition. To test the efficiency of inhibition, protein levels or activity levels of the target protein must be measured. Another application of antisense morpholino oligonucleotides is their use in blocking [▶pre-mRNA splicing](#) by targeting potential [▶exon/intron junctions](#), thereby forcing the splicing machinery to skip entire [▶exons](#) or to generate incorrect splice variants with unpredictable sequence alterations. Work by Schmajuk and colleagues (7) and by Draper and colleagues (8) has demonstrated the effectiveness of this method for gene inactivation in mammalian tissue-culture cell lines and in zebrafish, respectively. Advantages of this application are that splice alterations can easily be monitored by [▶polymerase chain reaction](#) techniques and that they can be assessed without the need for an antibody. Some applications may benefit from using splice variations for gene 'knock-down'. This approach can be applied if target gene sequences are incomplete, for instance when the 5'UTRs of gene sequences are unknown, precluding the standard morpholino oligonucleotide design for translational inhibition.

The cellular uptake of nonionic morpholino oligonucleotides is not feasible using the conventional lipid-based delivery systems. In contrast, pairing the morpholino oligonucleotide to a complementary

DNA and binding this compound electrostatically to partially ionized and weakly basic ethoxylated poly-ethylenimine (EPEI) provides a delivery system that is readily endocytosed and, upon reduction of pH levels within acidic [▶endosomes](#), results in the release of the morpholino oligonucleotides into the [▶cytosol](#) (9). This delivery system has been designed for efficient morpholino oligonucleotide uptake into tissue culture cells and may also result in more efficient cellular uptake during electroporation in chicken (3).

Morpholino Oligonucleotides Based Genetics

Gene inactivation through morpholino oligonucleotides targeting differs significantly in its mode of action from gene alteration through mutation. Importantly, the 'knock-down' by translational block causes the efficient inactivation of both maternal and zygotic transcripts, which, in zebrafish, can be fully penetrant through the first two days of development. Therefore, 'morphant' phenotypes have the potential to be both more severe and to have an earlier onset than zygotic mutant phenotypes. Moreover, fine-tuning the concentration of morpholino oligonucleotide applications may modulate the strength of 'morphant' phenotypes, providing the equivalent of an allelic series.

Morpholino oligonucleotides are highly amenable to genetic [▶redundancy](#) tests. 'Double-knock-downs' involving two morpholino oligonucleotides are feasible and can be used to determine genetic hierarchies or redundancy between genes based on the 'double-morphant' phenotypes. Morpholino oligonucleotide gene inactivation in true genetic mutants also provides a tool to determine potential genetic interactions. Moreover, morpholino oligonucleotide 'knock-downs' will become an increasingly important tool for testing candidate genes in [▶positional cloning](#) projects (4). In contrast to the standard method of translational inhibition, targeting pre-mRNA splicing specifically targets [▶zygotic transcripts](#) without affecting [▶maternal transcripts](#). Therefore, 'morphant' phenotypes generated by the latter method have the potential to be more closely related to zygotic mutant phenotypes. The inhibition of pre-mRNA splicing can also be deployed to specifically block splice variants and to study the function of different protein [▶isoforms](#) of a gene (8).

In developmental model systems, such as zebrafish, amphibians and sea urchin, the direct injection of morpholino oligonucleotides into the early [▶blastomeres](#) results in a general 'knock-down' of gene function throughout all tissues. In contrast, in chicken, square pulse electroporation has been successfully used to target morpholino oligonucleotides to only a subset of cells. Plans of both biotechnology companies and publicly founded consortia to employ large-scale

to genome-wide morpholino oligonucleotide ‘knock-down’ screens with the aim of inactivating the majority of genes within the respective model organisms will complement classic ‘►forwards genetics’ approaches such as those previously performed in zebrafish.

Limitations of Morpholino Oligonucleotide-based Gene ‘Knock-down’

Morpholino oligonucleotide mediated gene targeting has been reported for many genes with known mutant phenotypes. A special edition of the journal *Genesis* presented ‘morphant’ phenotypes of more than 20 known zebrafish mutants. The ‘morphants’ displayed phenotypic ranges but mostly had strong similarities to the true genetic mutant phenotypes. Nevertheless, a number of limitations have been reported in the usage of morpholino oligonucleotides. In some cases, gene inactivation is not achieved, possibly due to ►sequence polymorphisms in the targeting sequence. Strong polymorphic variations in the 5′ UTR sequences can occur in some genetic strains and may cause a reduced or highly variable efficiency of gene inactivation varying from animal to animal. The zebrafish *one-eyed-pinhead* locus appears to be an example of such a locus with, depending on the genetic background, 50%–100% of treated animals failing to respond to the morpholino oligonucleotide (2). Moreover, mistargeting of morpholino oligonucleotides toward unrelated gene sequences may be the cause for what appear to be unrelated ‘morphant’ phenotypes. One such example is the morpholino oligonucleotide against the zebrafish gene *bozozok/dharma* that reportedly resulted in additional phenotypes not observed in the mutant. These phenotypes may be due to the inactivation of another gene or other unknown morpholino oligonucleotide effects (2).

The targeting of pre-mRNA splicing has several limitations as well. Most importantly, aberrant splice variants generated by this method must be characterized in order to determine the type of mutation(s) introduced. Moreover, analysis can be clouded as targeting pre-mRNA splicing frequently results in several splice variants that may occur at different relative levels.

Due to these limitations, essential controls are routinely introduced when working with morpholino oligonucleotides. They include the design of a second morpholino oligonucleotide with an unrelated target sequence to ensure the specificity of the ‘morphant’ phenotype and to avoid the accidental loss of two distinct genes. In addition, rescue experiments, such as the injection of mRNA lacking the morpholino oligonucleotide target sequence, are essential and may reveal unrelated gene inactivation or non-specific defects.

Clinical Relevance

Morpholino oligonucleotides present the potential for ►functional genomics in tractable model organisms such as zebrafish, amphibians or chicken and in *in vivo* and tissue culture systems. The systematic analysis of medically relevant candidate genes or of genes with unassigned biological functions will result in a catalog of their early developmental functions as well as provide insight into general biological mechanisms. *In vivo* studies have indicated the feasibility of morpholino oligonucleotides for therapeutic applications. Examples are the application of morpholino oligonucleotides in mice by intranasal insufflation (5) or, in rat, by dermal absorption (6).

►Xenopus as a Model Organism for Functional Genomics: Rich History, Promising Future

References

- Summerton J (1999) Morpholino antisense oligomers: the case for an RNase H-independent structural type. *Biochimica et Biophysica Acta* 1489:141–158
- Nasevicius A, Ekker SC (2000) Effective targeted gene ‘knockdown’ in zebrafish. *Nature Genetics* 26:216–220
- Kos R, Tucker RP, Hall R et al (2003) Methods for introducing morpholinos into the chicken embryo. *Developmental Dynamics* 226:470–477
- Ekker SC (2000) Morphants: a new systematic vertebrate functional genomics approach. *Yeast* 17:302–306
- Qin G, Taylor M, Ning YY et al (2000) *In vivo* evaluation of a morpholino antisense oligomer directed against tumor necrosis factor- α . *Antisense Nucleic Acid Drug Development* 10:11–16
- Arora V, Hannah TL, Iversen PL et al (2002) Transdermal use of phosphorodiamidate morpholino oligomer AVI-4472 inhibits cytochrome P450 3A2 activity in male rats. *Pharmacological Research* 19:1465–1470
- Schmajuk G, Sierakowska H, Kole R (1999) Antisense oligonucleotides with different backbones. Modification of splicing pathways and efficacy of uptake. *Journal of Biological Chemistry* 274:21783–21789
- Draper BW, Morcos PA, Kimmel CB (2001) Inhibition of zebrafish *fgf8* pre-mRNA splicing with morpholino oligos: a quantifiable method for gene knockdown. *Genesis* 30:154–156
- Morcos PA (2001) Achieving efficient delivery of morpholino oligos in cultured cells. *Genesis* 30:94–102

Mosaic/Mosaicism

Definition

Mosaic/mosaicism describes the occurrence of cells in an organism or a tissue that are derived from a single zygote, but with distinct genetic constitutions due to a post-zygotic event. For example, a patient may be

mosaic for trisomy 21, with the trisomy present in 40% of cells, and the other 60% of cells having a normal set of chromosomes.

- ▶ [Huntington's Disease](#)
- ▶ [Mutagenesis Approaches in the Zebrafish](#)
- ▶ [Neurofibromatosis Type 1 \(NF1\), Genetics](#)
- ▶ [Prader Willi and Angelman Syndromes](#)
- ▶ [Repeat Expansion Diseases](#)
- ▶ [Tuberous Sclerosis](#)
- ▶ [X-Chromosome Inactivation](#)

Motif

Definition

Motif refers to a pattern of DNA sequence that is similar for genes of similar function. It also refers to a pattern for protein primary structure (sequence motifs) and tertiary structure that is the same across proteins of similar families. A motif can provide a basis for prediction of structural and functional characteristics within a family or group of proteins.

- ▶ [Protein Databases](#)

Motif Databases

Definition

A motif database contains representations of conserved sequences, for annotation of unknown sequences by comparison with individual, known sequences.

- ▶ [Protein Databases](#)

Motor Proteins

Definition

Motor proteins are eukaryotic proteins that convert chemical energy derived from the hydrolysis of nucleotides (ATP) to mechanical energy for movement of cytoskeletal tracks. The respective motor domains are linked to their cargoes via adaptor proteins. Accordingly, motor molecules are classified into two groups: one group includes the myosins that move along the actin filaments, the other group includes the

kinesins and dyneins that move along the microtubules. The kinesin motors only move to the plus and dynein only to the minus end of microtubules. Motor proteins of the centromere attach to spindle microtubules and provide energy and force to move chromosomes during mitosis. When motors are immobilized at their cargo binding area, they can move microtubules or F-actin, respectively.

- ▶ [Actin Cytoskeleton](#)
- ▶ [Cell Polarity](#)
- ▶ [Cytoskeleton](#)
- ▶ [Microvilli](#)
- ▶ [Mitotic Spindle](#)
- ▶ [Molecular Motors](#)
- ▶ [Centromeres](#)

Mouse Genomics

LOUISE VAN DER WEYDEN, DAVID J. ADAMS,
ALLAN BRADLEY
Wellcome Trust Genome Campus Hinxton, The Sanger
Institute, Cambridge, UK
abradley@sanger.ac.uk

Definition

'Mouse Genomics' is the study of the genetics of the laboratory mouse. An understanding of mouse genetics is essential to our understanding of human biology and disease, as the anatomical, physiological and genome similarities between mice and humans make the mouse an excellent experimental model for determining the function of human [genes](#). In addition, mice are small, easy to maintain and have a relatively short generation time compared to humans. The popularity of the mouse as a model organism also reflects the ease with which the mouse genome can be altered to generate models of human diseases.

Characteristics

Introduction to the Mouse Genome

▶ [Sequencing](#) of the Mouse Genome Since it began in 1990, the aim of the Human Genome Project was to sequence the genome of five key model organisms, one of which was the mouse. In 2000 the international Mouse Genome Sequencing Consortium began in earnest to sequence the genome of the [C57BL/6J](#) strain. Two years later, a high-quality draft sequence and annotation of the mouse genome was published in *Nature*. The mouse genome is estimated to be

2.6 billion base pairs, of which 97% can be aligned to the human genome sequence. This draft assembly of the mouse genome is freely available to the public and can be searched using sequence alignment tools such as ►http://www.ensembl.org/Mus_musculus/ssahaview or ►<http://www.ncbi.nlm.nih.gov/genome/seq/MmBlast.html>.

Manipulating the Mouse Genome

Genetic manipulation of the mouse genome can be performed using a random (phenotype-driven) approach or a sequence (►[genotype-driven](#)) approach (reviewed in (1)). In phenotype-driven approaches, the mice are exposed to a mutagen that generates mutations at random. The mutant mice are then screened with a battery of tests to identify any phenotypes. The responsible gene can then be localized by genetic mapping followed by detailed sequence analysis of genes in the region. Methodologies for generating random mutations in the mouse include exposure to radiation, chemicals or viruses (detailed below). In genotype-driven approaches, a gene hypothesised to be responsible for a particular function or trait is specifically mutated or altered and the resulting phenotype of the mouse can be characterised. Genotype-driven methodologies can be broadly divided into two categories, namely ►[transgenic](#) and ►[gene-targeting](#) strategies (detailed below). More recently, ►[chromosome-engineering](#) technology has allowed the generation of mice carrying specific chromosomal rearrangements (detailed below) and represents an additional method for manipulating the mouse genome.

Generating Random Mutations in the Mouse

Mutations occur spontaneously, however, they are very rare (occurring at a frequency of 5×10^{-6} mutations per gene per generation). Thus, mutagenesis strategies have been used to increase the mutation rate. Random mutations generated by these methods differ in their molecular nature and occur at varying frequencies.

Irradiation

X-rays induce mutations ranging from simple deletions to complex chromosomal rearrangements (such as translocations, inversions and duplications). Whilst chromosomal rearrangements have the advantage that they are relatively easy to identify/locate (using ►[FISH](#)), they often affect several genes on the chromosome, so it can be difficult to determine which gene is responsible for the phenotype of the mouse.

Chemical Mutagenesis

Two commonly used potent chemical mutagens are chlorambucil and N-ethyl-N-nitrosourea (►[ENU](#)). Whilst chlorambucil induces chromosomal rearrangements (frequently involving more than one gene), ENU

induces ►[point mutations](#), particularly in sperm precursor cells. Point mutations of DNA can result in ►[loss-of-function mutations](#), ►[gain-of-function mutations](#) or ►[hypomorphic alleles](#) (partial loss of function mutations). Thus ENU mutagenesis is a powerful genetic reagent for systematically analyzing the mouse genome (reviewed in (2)). ENU-mutagenesis experiments can be used in genome-wide screens for dominant or recessive mutations. In these experiments, male mice exposed to ENU are mated with numerous wild type females to produce large numbers of offspring, each of which carries a unique set of altered/mutated genes. These offspring will exhibit phenotypes caused by dominant mutations. Alternatively, these mice can be inter-crossed to generate mice with phenotypes caused by recessive mutations. Dominant screens are easier to conduct than recessive screens, however genetic reagents generated by chromosomal engineering have greatly facilitated recessive genetic screens (detailed below). ENU mutagenesis is efficient in generating mutations but does not provide molecular landmarks to identify the mutated genes. Thus these mutations must be mapped and ultimately confirmed by DNA sequencing.

Gene-Trapping

Gene-trap mutations are generated by ►[electroporation](#) or virus-mediated random insertion of vectors into the genes of mouse ►[embryonic stem \(ES\) cells](#), which are then used to generate mice that carry the mutated gene (reviewed in (3)). Gene-trap vectors consist of a ►[reporter gene](#) without a ►[promoter](#), such as *LacZ* or GFP, and a ►[selection marker](#), such as Neo or Puro. Mutagenesis and subsequent reporter expression occur when the vector inserts into an exon or intron of a gene to generate a fusion transcript between the upstream endogenous exonic sequence and the reporter gene. The fusion transcript generated by the trapping of a gene, serves as a template for techniques such as ►[RACE](#), which allows the trapped gene to be identified.

Generating Gene-Specific Mutations in the Mouse Transgenics

The ability to insert 'foreign' DNA (a transgene) into the mouse was first acquired in 1980 using the technique of '►[pronuclear microinjection](#)'. Mice carrying a transgene are termed "transgenic" (Fig. 1). This technique is robust and simple and relies on over-expression of the newly inserted transgene to produce a phenotype, which may not be physiologically relevant. However, the random integration site of the injected DNA will often result in variable expression of the transgene causing a variable phenotype. To generate a 'negative' effect, a mutant form of the gene (such as a ►[dominant negative](#)) can be used as the transgene.

Gene-Targeting

The ES cell genome can be modified by ►homologous recombination, also known as ‘gene targeting’, first achieved in mammalian cells in 1985 (Fig. 1). Using different strategies, it is possible to generate essentially any desired mutation and establish this genetic change in the germ-line of the mouse (reviewed in (1)). The different types of mutations and the vectors needed to generate such mutations are discussed below in more detail.

‘Knock-out’ Mice

1. One of the most common uses of gene targeting is to destroy (‘knockout’) a gene’s function and observe the effect this has on the phenotype of the mouse. This is usually achieved by introducing a selection cassette (such as Neo or Puro) into the target ►locus. Two basic types of targeting vector are used for knocking-out genes. These are ‘replacement’ vectors, which replace a portion of the targeted locus with a selectable marker and ‘insertion’ vectors, which mutate the target gene by ‘inserting’ into it and duplicating part of the locus. Both strategies disrupt the target gene and usually result in total loss of function of the gene (a ►null allele).

‘Knock-in’ Mice

2. Knock-in experiments are used to replace a targeted gene with a transgene, bringing the transgene under the transcriptional control of the targeted gene. The most commonly used transgenes include homologues of the target gene (to assess whether members of the same gene family have similar functions when expressed in the same spatial/temporal pattern) or reporter genes such as LacZ or GFP (to monitor expression patterns of the gene during development and in the adult mouse).

Conditional Mutations

3. Null mutations often cause ►embryonic lethality, precluding analysis of a gene’s function in adults. Conditional mutations can be generated which allow function to be examined in specific cell types *in vivo* (reviewed in (1, 4)). By regulating gene activity embryonic lethality is avoided, allowing gene function to be examined later in development or in adult life. Control over gene activity *in vivo* can be achieved using site-specific DNA recombination. The most widely used site-specific DNA recombination system is the enzyme ►Cre, which cleaves and ligates DNA at distinct target sequences known as ►loxP sites. LoxP sites can be incorporated into the targeting vector to flank (flox) the portion of DNA to be deleted, thus after homologous recombination these loxP sites will be incorporated into

the targeted locus. The ‘conditional’ nature of such a locus means that the ►floxed DNA will function normally until Cre is expressed and the floxed DNA is deleted (Fig. 2i). Deletion can be achieved in ES cells (using vectors expressing Cre) or in certain tissues in a mouse (by mating with a transgenic mouse expressing Cre). There are many established transgenic lines expressing Cre in different tissues (see ►<http://www.mshri.on.ca/nagy>). More recently, inducible forms of Cre have also been developed, such as ►Cre-ER^T, such that Cre-mediated recombination is additionally under the control of a drug such as tamoxifen and thus can be regulated temporarily as well as spatially.

Chromosomal Engineering in the Mouse

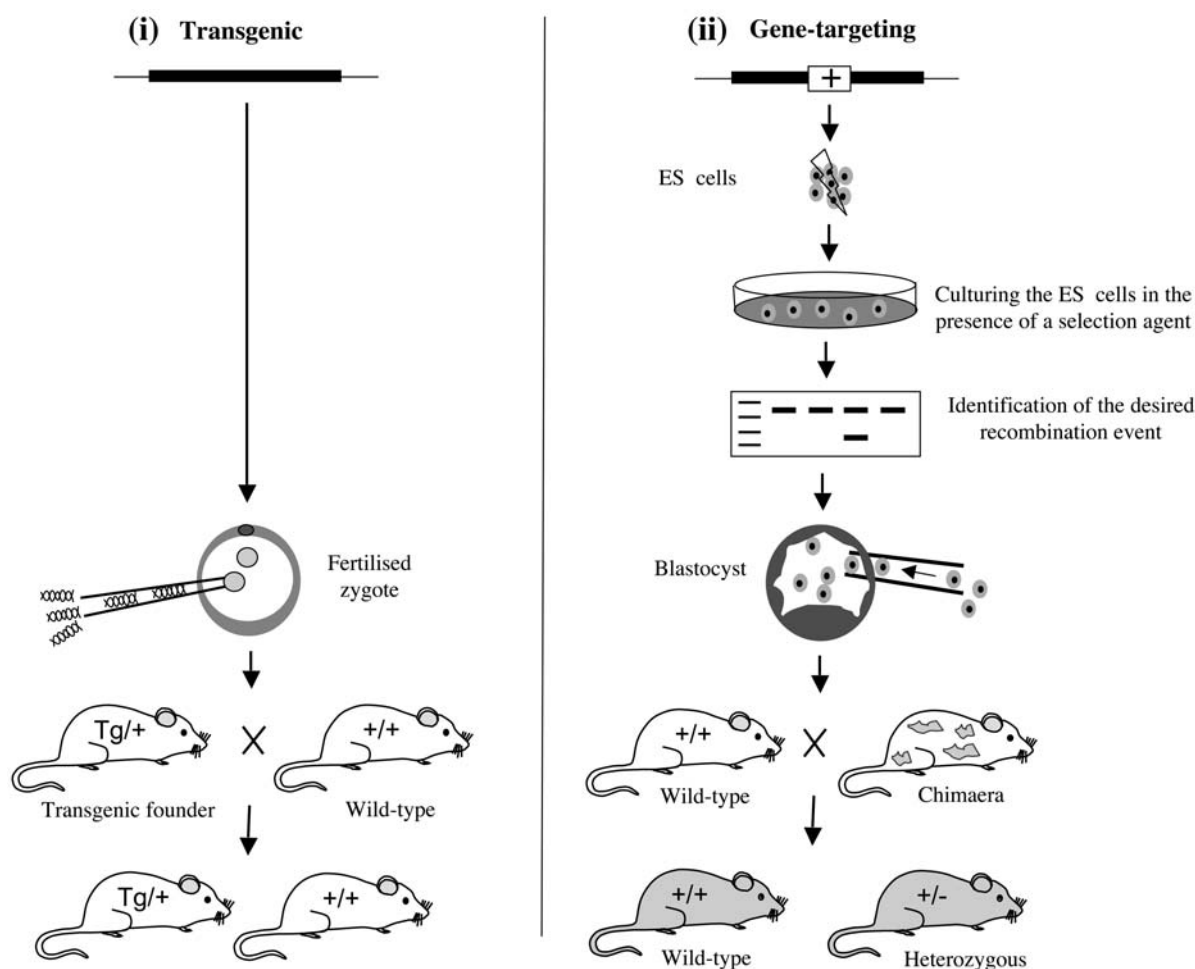
The combination of gene-targeting with the Cre-loxP recombination system resulted in 1995 in the emergence of chromosome engineering technology, which allows for modelling human diseases associated with chromosomal rearrangements, such as deletions, inversions, duplications and translocations (reviewed in (5)). Chromosome engineering requires the targeting of two ►HPRT vectors to defined sites within the genome (Fig. 3). The location of the 5’HPRT and 3’HPRT cassettes relative to each other (and hence the orientation of the loxP sites found on these vectors) dictates the structure of the mutant chromosome (Fig. 2).

Deletions

Together with duplications, deletions are the most common form of chromosomal abnormality in humans, and cause developmental syndromes such as DiGeorge and ►William’s syndromes (hereditary deletions) as well as tumorigenesis (►somatic deletions). Deletions can be engineered when the 5’HPRT and 3’HPRT vectors are targeted to the same chromosome and the loxP sites in the HPRT cassettes are facing in the same direction along the chromosome (Fig. 2i). Expression of Cre in these cells recombines the distant loxP sites, deleting the intervening segment of DNA. A conditional deletion can be engineered by expressing Cre in certain tissues in the mouse so that the deletion product is only generated in specific lineages. Deletions of many sizes have been engineered in ES cells. Although there is no limit on the size of deletion that can be engineered, larger deletions have often have biological problems such as ►haploinsufficiency and embryonic lethality that limit their use.

Duplications

Duplications are more commonly observed in human populations than deletions because the biological affects are often less severe. Duplications can be

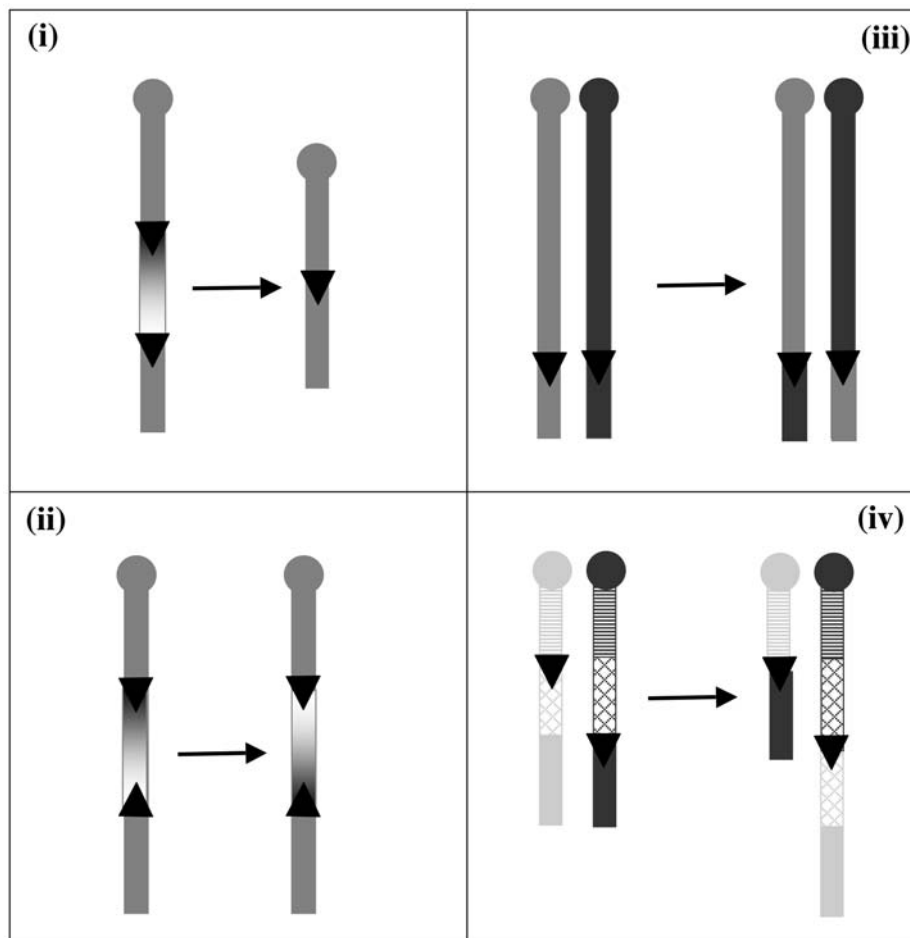


Mouse Genomics. Figure 1 Generation of genetically modified mice by (i) pronuclear injection or (ii) gene-targeting methods. (i) To generate transgenic mice, a vector containing the gene of interest is generated. The vector is linearized and then injected into the pronucleus of a fertilised zygote. The injected zygote is then transferred into the oviducts of a pseudo-pregnant foster mother. The resulting pups are screened for the presence of the transgene and those containing the transgene are termed the 'founders'. The founders are mated with wild type (+/+) mice to check for germline transmission of the transgene to the F₁ offspring (Tg/+ mice). (ii) To generate knockout mice, a targeting vector containing a selection cassette (+), and sequences of homology with the target locus (thick black line) is generated. The vector is then linearized and electroporated into mouse ES cells. These cells are then cultured in the presence of a selection agent to remove any cells that have not stably integrated the construct into their genome. ES cell clones that carry the desired recombination event are identified (using molecular methods). ES cells with the targeted allele are injected into blastocysts, which are then transferred into the uterus of a pseudo-pregnant foster mother. Male chimaeras generated from the blastocyst injections showing high percentage chimaerism are then mated with wild type female mice to check for germline transmission (indicated by the coat colour) of the targeted allele to the F₁ offspring. For both techniques, the heterozygotes may then be inter-crossed to breed to homozygosity. The phenotypes of the heterozygous and homozygous progeny are characterised.

engineered when the 5'HPRT and 3'HPRT vectors are targeted to different homologous chromosomes as shown in Fig. 2 (iv). After expression of Cre, recombination generates one chromosome with a duplication and another with a deletion.

Inversions

Inversions cause disease less frequently because the loss or gain of genetic material is limited to the ends of the rearrangement. Inversions can be engineered by targeting the 5'HPRT and 3'HPRT vectors to the same



Mouse Genomics. Figure 2 The actions catalysed by Cre. (i) Deletion: when two *loxP* sites (black triangles) are facing in the same direction on a chromosome (*cis*), Cre-mediated recombination will result in excision of the floxed DNA (shaded region). (ii) Inversion: if the *loxP* sites are orientated in opposite directions in *cis*, the sequence between the *loxP* sites will be inverted. (iii) Translocation: recombination between two *loxP* sites on different chromosomes (*trans*) will lead to the reciprocal exchange of the regions that flank the *loxP* sites. (iv) Duplication: if the *loxP* sites are targeted to different chromosomes in the same orientation, Cre-mediated recombination between sister chromatids will result in a duplication of the floxed region on one chromatid and deletion of this region on the other chromatid.

chromosome but ensuring the HPRT cassettes and their associated *loxP* sites are placed in an inverted orientation (Fig. 2ii). *LoxP* sites in the reverse orientation will invert ('flip') the intervening DNA segment upon exposure to Cre. Experimentally, inversions are very useful reagents for genetics because a chromosome with an inversion cannot recombine with its homologue and produce a viable gamete. Chromosomes with inversions are known as 'balancer chromosomes'.

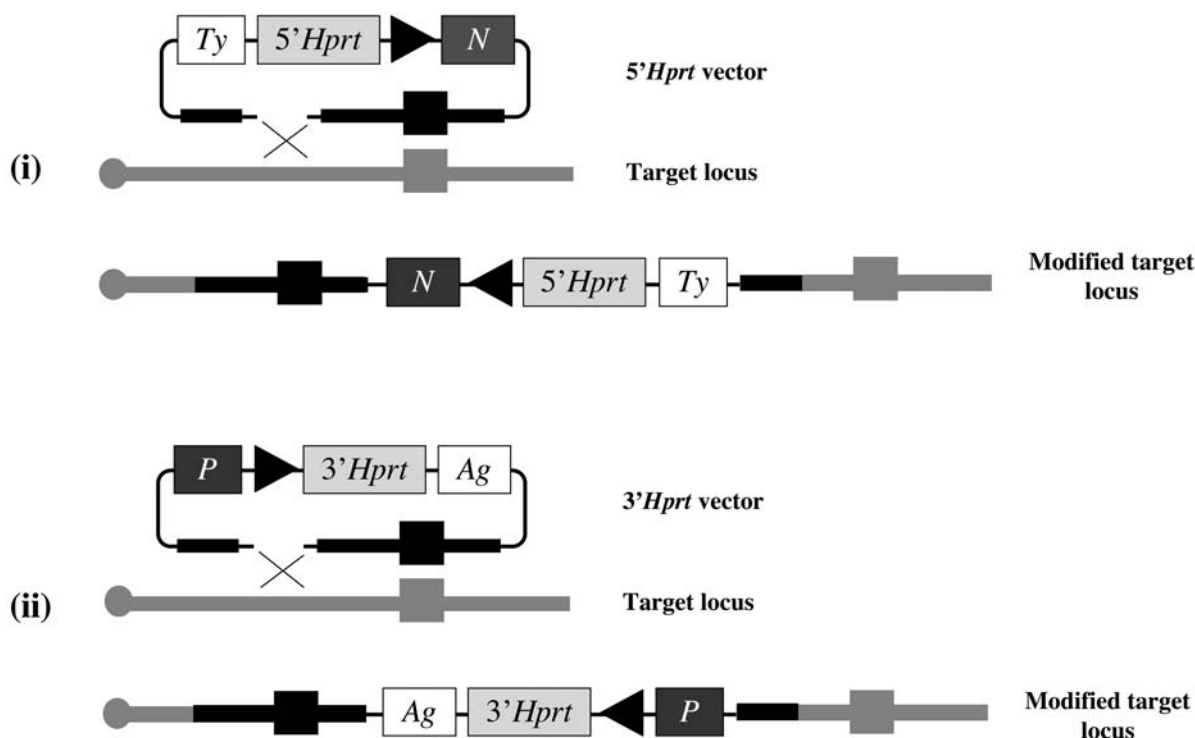
Translocations

Translocations are frequently observed in humans with infertility. Translocations are also important in malignant diseases, such as leukaemia. Translocations

usually have little or no net loss of genetic material, but genes that are normally tightly regulated are taken out of their normal context and novel fusion products can be generated which exert pathological effects. Translocations are generated when the 5'HPRT and 3' HPRT vectors are targeted to different chromosomes (Fig. 2iii). After the expression of Cre, a reciprocal exchange of the regions that flank the *loxP* sites will occur between the chromosomes.

References

1. van der Weyden L, Adams D, Bradley A (2002) Tools for targeted manipulation of the mouse genome. *Physiol Genomics* 11:133–164



Mouse Genomics. Figure 3 Chromosome engineering in embryonic stem cells. Targeting vectors can be used to insert *loxP* sites, positive selectable markers, *Hprt* gene fragments and coat-colour markers into a targeted locus in ES cells. The vector (thin black line) is linearized in the region of homology (thick black line) to stimulate targeted insertion into the locus (grey line). (i) The 5' vector contains the neomycin selectable marker (*N*), the 5' end of *Hprt*, a *loxP* site (black triangle) and the Tyrosinase gene (*Ty*) coat-colour marker. (ii) The 3' vector contains the puromycin selectable marker (*P*), the 3' end of *Hprt*, a *loxP* site and the K14-Agouti transgene (*Ag*) coat-colour marker. Using this system, the *Hprt* gene is divided into two complementary, but non-functional, fragments (*5'Hprt* contains exons 1–2 and *3'Hprt* contains the remaining exons 3–9) that are each linked to a *loxP* site. Cre-mediated recombination unites the 5' and 3' cassettes and restores *Hprt* activity (which is required for purine biosynthesis) thereby allowing the desired recombination events to be selected for in HAT (hypoxanthine, aminopterin and thymidine) medium.

- Justice MJ, Noveroske JK, Weber JS et al (1999) Mouse ENU mutagenesis. *Hum Mol Genet* 8:1955–1963
- Stanford WL, Cohn JB, Cordes SP (2001) Gene-trap mutagenesis: past, present and beyond. *Nat Rev Genet* 2:756–768
- Lewandoski M (2001) Conditional control of gene expression in the mouse. *Nat Rev Genet* 2:743–755
- Yu Y, Bradley A (2001) Engineering chromosomal rearrangements in mice. *Nat Rev Genet* 2:780–790

can move along the DNA to find a junction from double-strand to single-strand DNA.

► [Replication Fork](#)

MPNSTs

Definition

Malignant peripheral nerve sheath tumours (MPNST) are rare neoplasms, usually arising from somatic soft tissues or peripheral nerves. Approximately 10% of neurofibromatosis patients develop MPNSTs.

► [Neurofibromatosis Type 1 \(NF1\), Genetics](#)

Moving Platform

Definition

Moving platform is a ring shaped proliferating cell nuclear antigen molecule that encircles the DNA, and

MPTP

Definition

MPTP is an abbreviation for 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine, a contaminant of the addictive abused drug heroin, which causes dopamine responsive parkinsonism in heroin addicts, and which has served as a useful toxin in creating animal models of Parkinson's disease.

► [Parkinson's Disease: Insights from Genetic Cause](#)

MR Imaging

MATTHIAS G. FRIEDRICH

Department of Cardiac Sciences, University of Calgary, Calgary, AB, Canada
matthias.friedrich@ucalgary.ca

Definition

Understanding the impact of genomics and proteomics on the human biological system requires more than a simple description of experimental findings on a molecular level. Experience in molecular medicine teaches us that we need more knowledge on the nature and interaction of molecules in their natural physiological environment. Many knock-out animal models did not match the expected pathology due to the multiple compensation mechanisms of the biological system. Thus, medical research asks for highly accurate and reproducible data *in vivo*. The value in reflecting molecular physiology *in vivo* is limited for every sample either bred in an isolated *ex vivo* system or removed from the body. Therefore, we encounter a tremendous interest in powerful imaging techniques.

Among them, magnetic resonance imaging (MRI) may be of special importance. Unlike other imaging techniques, MRI provides information on molecules without affecting their physiological actions *in vivo*. Furthermore, the technical approach can be focused on certain tissue types or even certain molecules. Finally, the safety profile is an ideal setting for imaging volunteers and follow-up studies. Thus, MRI provides unique features for molecular imaging.

This chapter will describe the utility of MRI in molecular imaging.

Description

Image formation is based on the fact that in a magnetic field, the magnetic axis of nuclei with a magnetic

moment (such as ^1H , ^3He , ^{13}C , ^{19}F , ^{23}Na , ^{31}P , ^{129}Xe) spins, like a gyroscope. Each of these nuclei has a specific spinning frequency (the so-called Larmor frequency), which is a product of a nucleus-specific constant and the external magnetic field. When an electromagnetic excitation pulse of their frequency is applied, the nuclei show a resonant behavior and the orientation of their spins is changed. After the pulse, the nuclei will return to their original position in a specific time (relaxation time). This is associated with the emission of a very small signal, which can be detected by a receiver and further processed into an image. Since surrounding atoms alter the frequency ("chemical shift"), the emitted signal of a given nucleus depends on its molecular environment.

Thus, the specific resonant behavior of nuclei within different molecules can be exploited to generate molecule- or even tissue-specific signals. As an example, in a given MRI technique, water-bound ^1H nuclei would reveal high signal intensity, whereas fat-bound protons would hardly be detectable.

Since modern MRI systems provide numerous options for modifying the excitation-relaxation sequence, there are many ways to acquire images that provide information on the molecular pattern of a given tissue.

A Brief History of Molecular MRI

After the first report of Gerlach and Stern on the magnetic momentum of silver atoms in 1921, it needed many years to explore this phenomenon further, before an assessment of the molecular composition of compounds became possible. The introduction of the Fourier transformation by Ernst and Anderson in 1966, a mainstay of modern MR techniques, required computer power which was not available until the 70's, when it eventually enabled modern MR imaging. Damadian was the first to describe a difference in relaxation times between normal tissue and tumors and stimulated the clinical application of MR imaging. The following years encountered many advances in experimental MR research, enabling the definition of molecular structures and complicated 3-dimensional conformations of proteins, making MR an integral part of drug development, since the development curve of MR imaging was dependent on hardware power and thus showed a more exponential shape. In 1973, Nobel Prize laureate Paul Lauterbur described a technical approach to using magnetic field gradients for imaging ^1H in water and, in 1977, Damadian published the first image of a human chest *in vivo*. But in the following years, uncoupled from spectroscopy, the development of MR imaging in medicine was hampered by the limited field strength and computer power. In the 80's, eventually, the development of surface coils and time-dependent gradients enabled volume-selective studies and paved the road for higher resolution imaging.

Industry's interest in a clinical application of MR imaging fostered the development of hardware, software and a variety of technical approaches and after a tremendous development during recent decades, MRI has become a cornerstone of modern diagnostic imaging.

Nowadays, experimental studies deal with a pixel size of $20,000 \mu\text{m}^3$ (or 0.00002 mm^3) and *in vivo* studies in humans allow for pixels as small as 0.35 mm^3 .

Potential and Limitations of MRI

The outstanding potential of MRI is based on the unique possibilities of obtaining molecule-specific information, in many cases without interfering with the targeted molecular processes. The small proportion of nuclei available for signal generation however, limits the sensitivity of MRI. Thus, imaging approaches are still mainly confined to ^1H (which is highly abundant in tissue compounds), whereas the absolute number of other nuclei is often insufficient. Numerous approaches have addressed an increase in the sensitivity and, although the theoretical borders of sensitivity are not reached yet, reliable techniques have been developed to visualize tissue composition and molecular mechanisms.

MRI Systems

There are mainly two classes of MRI systems available, experimental, small-bore ($<20 \text{ cm}$) systems for probe and animal imaging and clinical, large-bore (around 60 cm) systems for whole body imaging of humans. Small systems generally offer a higher spatial resolution, but a smaller field of view.

The magnetic field strength depends on the purpose of the system. Experimental scanners often have high field strengths of $4.7\text{--}9.4 \text{ T}$; for *in vivo* imaging of humans, mostly 1.5 T and 3 T systems are currently in use.

Signal Contrast Mechanisms and Applications in Molecular Imaging

The specific changes in the nuclear relaxation properties offer numerous options for visualizing and quantitatively assessing molecules.

Nucleus-Specific Imaging

First of all, the study can be targeted on certain molecules by the frequency of the excitation pulse, e.g. in a 1.5 T field, a frequency of 63.87 MHz will excite ^1H nuclei, whereas 25.85 MHz would give ^{31}P images. Whereas ^1H nuclei have an unmatched abundance in biological tissue, other nuclei do not and thus are subject to limited sensitivity in the corresponding MR approaches.

Since the mere quantity of protons often does not provide specific information on molecules, the number

of applications is limited. Examples of MRI using nuclei other than ^1H include ^{31}P , ^{13}C and ^{23}Na .

The use of stronger magnetic fields such as 3 T or 7 T increases the sensitivity and eventually may lead to a more successful application *in vivo*.

Sodium MRI

^{23}Na ions have been subject to MR imaging for more than 20 years and were used to investigate multiple body areas such as the brain, muscular tissue, heart, and kidneys (1).

^{23}Na MRI has been used to quantify static sodium concentrations *in vivo*, but also during functional studies. It was applied to assess tissue viability in the heart. It may be of special importance in the follow-up of tumors.

Although the abundance of ^{23}Na ions is 100%, their inherent MRI sensitivity is low (due to a quadrupole electric moment), which leads to long acquisition times, limited spatial resolution and difficulties in accessing areas distant from the coils. Reliable separation of extracellular from intracellular sodium remains a challenge. Furthermore, as with MRI of other ions, broadband capability of the MRI hardware is required. More recent developments such as multiple quantum filters, shift reagents, new transmit-receiver systems and novel coil concepts are likely to facilitate ^{23}Na use in research and future clinical application.

MRI of Other Nuclei

Besides studies on rubidium (^{87}Rb) MRI in myocardial ischemia (2), there are reports on using potassium (^{39}K) to generate MR images related to the assessment of intracellular K^+ concentrations in myocardial infarction. The inherent sensitivity, however, limits its usefulness.

Relaxation-Time Based MRI

Making use of the difference in relaxation times (in the range of ms to seconds), it is possible to delay signal reception until a chosen time point, e.g. fat has shorter relaxation times, whereas components rich in water have longer relaxation times. Thus, a later temporal window will exhibit a strong signal coming from water and *vice versa*. Selecting specific imaging parameters, allows for acquiring images with known signal intensity properties. Basic types of images are called T1-weighted, T2-weighted or proton density-weighted. Additional high frequency pulses (e.g. inversion or saturation prepulses) allow for a more specific selection of image contrast.

BOLD (blood-oxygen-level-dependent) MRI has become an important tool to assess small changes in tissue blood supply. The underlying principal is based on the fact that the oxygenation status of hemoglobin defines its magnetic properties. In contrast to its

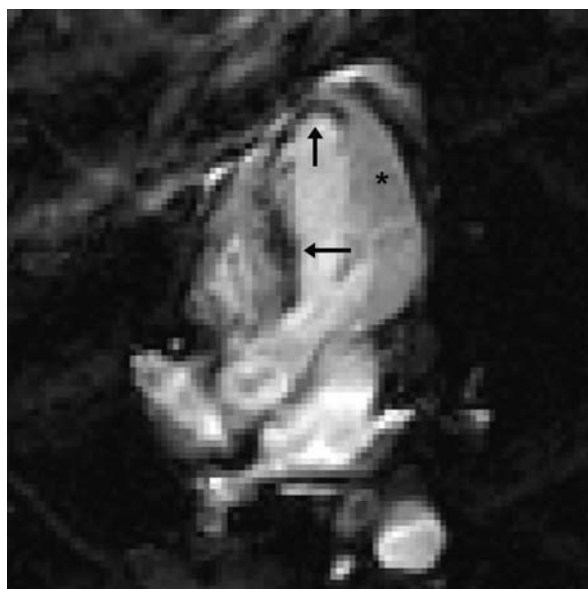
oxygenated counterpart, deoxygenated hemoglobin destabilizes the surrounding magnetic micro field and thus can be used as an internal contrast medium. The so-called BOLD effect has been exploited to visualize brain activity in the field of functional MRI as well in the diagnosis of stroke. More recently, BOLD MRI has been successfully applied to visualize myocardial ischemia (3) (Fig. 1).

Diffusion-weighted MRI is based on the phenomenon that the directional arrangement of molecules in ordered tissue structures leads to an alteration in relaxation times. This change of relaxation time is a source of error, but can yield information on the extent and direction of proton diffusion. The most important diagnostic use is in the detection of acute cerebral stroke.

Magnetization transfer MRI can be used to visualize pathological leakage from membranes, exploiting altered relaxation properties of protons bound to immobile tissue. The selective saturation of these protons is transferred to adjacent, mobile protons, acting as contrast agent, which indicates cellular injury. This approach has been applied to numerous diseases such as optic neuritis.

Contrast-Media Enhanced Approaches

The relatively low spatial resolution of molecular MRI represents a serious limitation and is faced with a much lower concentration of the metabolites to be studied.



MR Imaging. Figure 1 Hemorrhagic myocardial infarction as visualized by signal loss (arrows) as compared to normal tissue (asterisk) in this BOLD-sensitive image (gradient echo/echo planar imaging hybrid sequence). The signal loss reflects the local presence of degradation compounds of hemoglobin.

The current sensitivity limit of MRI is 0.5 mmol/kg, whereas the molecular concentration of the addressed targets or reporters is in the low nanomolar or even picomolar range.

Contrast agents may be used to increase the sensitivity and specificity of molecular MR. MR contrast media change the magnetic relaxation properties of the surrounding microenvironment. Gadolinium-III (Gd-III) complexes have a strong magnetic moment and have been the most frequently used compounds (4). A concentration of about 10–100 μmol is required to be detectable by present MRI systems *in vivo*, but recent reports indicate that MRI can even be applied in cell culture models.

More specific contrast agent complexes may be used to identify pathology on a molecular level. Gd-labeled antibodies were used to detect CA 19-9 and carcinoembryonic antigen in colorectal carcinoma. Gd-labeled mesoporphyrins have shown a selective affinity for necrotic cells and were applied for the detection of tumors and scarred tissue (5). Other studies addressed tumor angiogenesis using polymerized liposomes, which were targeted to the endothelial $\alpha\beta_3$ receptor by specific antibodies.

However, the sensitivity of these approaches is still limited in many situations firstly by the impact of most agents on magnetic relaxation, i.e. the relaxivity of the agent and secondly by the actual concentration of contrast-generating complexes in the region of interest which is still very low, especially when considering the small proportion of bound molecules within the equilibrium state of *in vivo* systems.

Consequently, current research is focused on the development of agents with high relaxivity and target affinity.

A novel fibrin-binding Gd complex with a high relaxivity was successfully used to visualize intravascular thrombi *in vivo* (6).

The use of nanoparticles containing specific antibodies and a large number of MR contrast agent molecules may provide an adequate sensitivity and enable detection of targets in the picomolar range (7). Liquid perfluorocarbon emulsions, containing nanoparticles with several tens of thousands of Gd molecules, have been used to image neovasculature and vulnerable plaques.

Newer developments have raised interest in so-called “smart molecules”, which are activated *in vivo* by certain pathology-related or reporter gene-related enzymes, but further results have to be awaited.

Perspectives

Due to ongoing technical advances, the role of MRI in molecular imaging is likely to become more and more important. Considering the limited spatial resolution and sensitivity of current MRI techniques *in vivo*, further development is to be expected. In several areas

of research, the current status may suggest a combination of MRI with other modalities such as optical imaging.

References

1. Maril N, Margalit R, Mispelter J et al (2004) Functional sodium magnetic resonance imaging of the intact rat kidney. *Kidney Int* 65:927–935
2. Kupriyanov VV, Xiang B, Sun J et al (2003) Imaging of ischemia and infarction in blood-perfused pig hearts using ⁸⁷Rb MRI. *Magn Reson Med* 49:99–107
3. Friedrich MG, Niendorf T, Schulz-Menger J et al (2003) Blood-Oxygen-Level-Dependent Magnetic Resonance Imaging in Patients with Stress-Induced Angina. *Circulation* 108:2219–2223
4. Aime S, Cabella C, Colombatto S et al (2002) Insights into the use of paramagnetic Gd(III) complexes in MR-molecular imaging investigations. *J Magn Reson Imaging* 16:394–406
5. Lee JW, Moon WK, Weinmann HJ et al (2003) Contrast-enhanced MR imaging of postoperative scars and VX2 carcinoma in rabbits: comparison of macromolecular contrast agent and gadopentetate dimeglumine. *Radiology* 229:132–139
6. Botnar RM, Buecker A, Wiethoff AJ et al (2004) *In vivo* magnetic resonance imaging of coronary thrombosis using a fibrin-binding molecular magnetic resonance contrast agent. *Circulation* 110:1463–1466
7. Wickline SA, Lanza GM (2003) Nanotechnology for molecular imaging and targeted therapy. *Circulation* 107:1092–1095

mRNA

► messenger RNA

mRNA Capping

YASUHIRO FURUICHI¹, AARON J. SHATKIN²

¹GeneCare Research Institute, Kamakura, Japan

²Center for Advanced Biotechnology and Medicine, Piscataway, NJ, USA

furuichi@genecare.co.jp

shatkin@cabm.rutgers.edu

Synonyms

5' end processing

Definition

Gene transcripts synthesized by RNA polymerase II, as well as other nascent pre-mRNAs including most viral mRNAs, are processed at the 5' end at an early stage of formation. The enzymatic addition of a m⁷GpppN ►cap structure marks the ►transcription initiation site, enhances RNA stability and splicing and promotes other downstream events, notably translation initiation through the action of cap binding proteins.

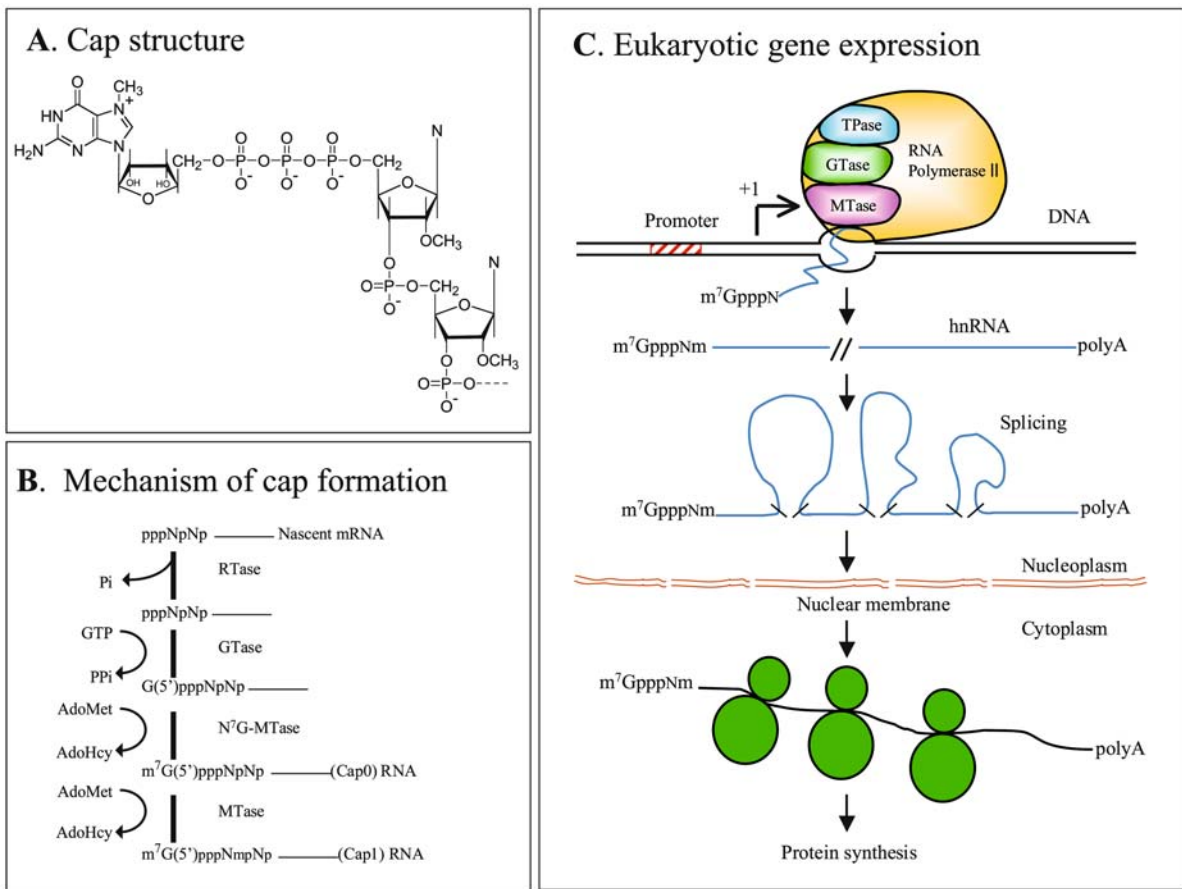
Characteristics

Discovery of the mRNA Cap from Virus Research

A blocked and methylated m⁷G(5')ppp(5')N structure, dubbed the cap, is present at the 5' ends of all eukaryotic cellular mRNAs and most but not all viral mRNAs (Fig. 1A). The 5' cap is a hallmark of eukaryotic mRNAs and is not present in rRNA or tRNAs or in prokaryotes. It has been 30 years since the discovery of this unexpected structure (1). This short review focuses on the impacts of this novel finding on the others that followed in the mid-70s and 80s. The story behind the discovery of the cap and how such a so-called “bizarre” structure was identified is described in our previous review “Viral and Cellular mRNA Capping: Past and Prospects” (2).

Although it is unclear why and how eukaryotes acquired the mRNA 5'-terminal element during evolution, studies from many laboratories have documented its crucial roles in several steps of gene expression. Elucidation of the cap also contributed significantly to the progress of molecular biology research and led to many other important findings. Indeed, molecular biology in the 1970s was filled with exciting surprises by a flood of discoveries on the fundamental structures of cellular components and the basic mechanisms by which genetic information is elicited from DNA *via* mRNA to generate the proteins needed for life.

Historically, new findings in science are related not only to those past but also mutually to each other, one-by-one and one-after-another. Accordingly, finding the cap was a consequence of a series of other discoveries, which then opened new avenues for exploration. Here it is important to note that virus systems played a pivotal role in defining the structure of the cap, its mechanism of synthesis and the biological functions of the cap in cells. Viruses such as insect cytoplasmic polyhedrosis virus and human reovirus, that contain as integral parts both genomic double-stranded RNA and a mRNA synthesizing system, played highly significant roles in elucidating the cap structure because one strand of the duplex genome RNA is identical to the viral messenger and the purified viruses produced capped mRNAs in the test tube when incubated with nucleoside triphosphates and the methyl donor S-adenosyl-methionine (Fig. 1B). Studies by Moss'



mRNA Capping. Figure 1 (A) Cap structure. N represents the nucleoside base in mRNA. (B) Enzymatic processes to form cap structure (AdoMet, S-adenosyl-methionine; AdoHcy, S-adenosyl-homocysteine). (C) mRNA formation and function in eukaryotic cells.

group on vaccinia virus, a DNA virus that also contains its own RNA polymerase and capping enzymes, similarly resulted in the identification of the enzymes and reactions involved in cap formation.

Fundamental Structure and Biological Functions of the Cap

We and others including the laboratories of Adams and Cory, Darnell, Miura, Moss, and Perry found that a cap is added to mammalian cell mRNA precursors synthesized by RNA polymerase II *via* a series of reactions catalyzed by a capping enzyme complex consisting of guanylyl transferase and methyltransferases in association with the polymerase (Fig. 1C). The reactions occur during the initial phases of transcription and before other processing events that include internal N6A methylation, 3'poly(A) addition and exon splicing. Most capped mRNAs have a single methyl group on the terminal G residue at the N7 position, but the adjacent nucleotides can also be 2'-O-methylated to different extents. Because capping

appears to be completed at a very early stage of transcription, co-transcriptional rather than post-transcriptional capping is likely for cellular mRNA synthesis *in vivo*. Biochemically, post-transcriptional capping can occur *in vitro* on 5' triphosphorylated RNA or oligonucleotides, or even on pppNpN dinucleotide structures, by the action of purified RNA 5' triphosphatase (RTase), guanylyl transferase (GTase) and methyltransferases (MTases) (Fig. 1B). However, this may not occur *in vivo* because a strong 5'-exonuclease activity that hydrolyzes uncapped RNA exists in the nucleus. Consequently, nascent mRNAs would be readily digested before completion of transcription unless their 5'-ends were protected by a m7G cap. Methylation of G at the N7 position prevents decapping, the GTase reverse reaction. In this context, it is interesting to mention that uncapped bacterial mRNAs have a short half-life (~ a few minutes) while capped eukaryotic mRNAs generally have half-lives of several hours to days, depending on the species and cell type. Thus one important function of cap is to help

stabilize mRNA. Also, in contrast to bacterial messengers, most eukaryotic mRNAs are produced from large nuclear precursor RNAs by complex, time-consuming splicing reactions that remove intronic sequences but with retention of the 5' cap as a transcription start site marker (Fig. 1C). Thus, the 5' cap in eukaryotic mRNAs might have arisen with the acquisition of nuclear compartmentalization and post-transcriptional processing systems including splicing.

In many lower eukaryotes including yeast, mRNAs contain a simple m⁷GpppN (cap 0) while higher organisms usually have more extensively methylated caps such as m⁷GpppNmpNpN (cap 1) or m⁷GpppNmpNmpN (cap 2). Similar structures with multiple methyl groups on the terminal G, i.e., 2,2,7m³GpppNm, are also present on small nuclear (sn)RNAs U1, U2, U3, U4, U5 and U8 which are synthesized by RNA polymerase II and act as scaffolds for RNA splicing. Despite the variations in the degree of methylation, the important biological consequences of a cap structure correlate with the N⁷-methyl group on the terminal G and the two pyrophosphoryl bonds that connect m⁷G in a 5'-5' configuration to the first nucleotide of mRNA (Fig. 1A). As noted above, caps increase [▶ mRNA stability](#) by protecting against 5' to 3' exonucleolytic degradation. Unmethylated caps can also stabilize mRNA to a certain extent, but unless the 5'-terminal GpppN is converted to m⁷GpppN, it can be hydrolyzed to (p)pN with the RNA becoming susceptible to subsequent degradation. Splicing accuracy and efficiency are both increased in the nucleus by the presence of the 5'-terminal m⁷GpppN and the mature mRNA enters the cytoplasm "head-first", all facilitated by nuclear cap-binding proteins analogous to the cytoplasmic cap-binding initiation factor eIF4E that is required for eukaryotic protein synthesis (Fig. 2A).

Reovirus with its associated RNA polymerase and capping enzymes proved extremely useful for deciphering how eukaryotic mRNAs initiate protein synthesis. Comparisons of reovirus RNAs containing various types of 5'-ends indicated that molecules with 5'-terminal m⁷GpppGm were preferentially translated in extracts of either plant or animal cells. Virus mRNAs with 5'-GpppG or 5'-ppG were not efficiently translated. Selection of capped over uncapped RNAs occurred during mRNA binding to the small ribosomal subunit. These data clearly demonstrated that cap recognition plays an important role in translation and, as shown later, in selection of the cap-proximal AUG as the initiator codon (3).

Impacts of Cap Discovery on Other Major Findings Caps in Virus Replication

Viruses reproduce in host cells by redirecting cellular systems that are lacking in the virus particle but needed for replication. Thus there exist a variety of virus

replication strategies. While the cap was discovered in studies on viruses, subsequent cap-related research in turn helped elucidate virus replication and pathogenesis mechanisms, information essential for developing antiviral drugs (4). Described below are some examples involving several well-studied viruses.

Poliovirus

Unlike many viruses that infect eukaryotes, Wimmer and others found that the picornaviruses including human poliovirus and mouse encephalomyocarditis virus do not have capped mRNA. Further studies led to the finding that infection by poliovirus and other picornaviruses converts the cap-dependent cellular translation system to cap-independence by inactivating the cap-binding protein important for host protein synthesis. These findings provided new insights on how pathogenic viruses redirect cellular systems for viral replication. Hepatitis C virus (HCV) notorious for causing chronic hepatitis, cirrhosis and finally a high frequency of human hepatocarcinomas also does not have capped mRNA and may follow a poliovirus-like strategy for proliferation and chronic infection. Further studies are needed as a basis for designing antivirals targeted to intrinsic replicative features of HCV and other life-threatening viruses.

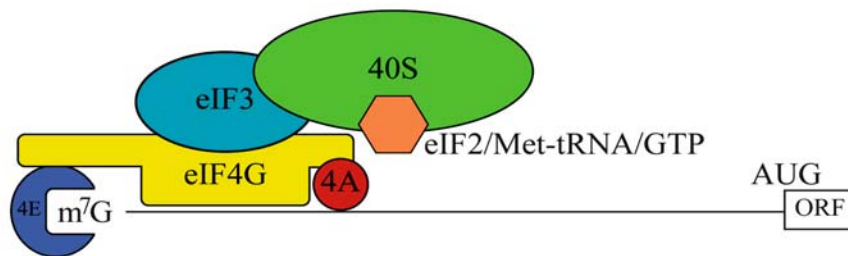
Influenza Virus

The virus core particles containing the single-stranded genome RNAs (negative strands) and viral RNA polymerase penetrate into the nucleus of infected cells. There the influenza complex that includes a cap-binding protein, endonuclease and polymerase binds to the capped ends of nascent cellular mRNAs, cleaves them ~10–20 residues downstream of the cap and utilizes the resulting capped fragments as primers of viral mRNA synthesis. This "[▶ cap-snatching](#)" mechanism results in viral messages containing short, diverse 5' sequences of cellular origin covalently linked to virus coding sequences. These findings showed clearly that influenza virus compensates for the absence of virus-associated capping enzymes by snatching capped oligonucleotides from the host, which also allows use of the [▶ cap-dependent translation](#) system for production of virus proteins required for replication. This unusual mechanism provides a possibility for drug intervention to prevent a major cause of human morbidity and mortality as well as worldwide economic damage.

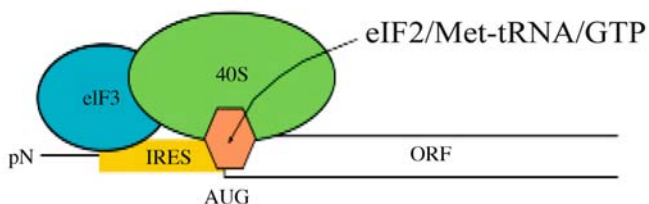
Corona Virus

The recent SARS pandemic in Southeast Asia was caused by a new corona virus, the severe acute respiratory syndrome virus. It contains a positive-sense, single-stranded RNA genome that functions as mRNA and is infectious. The virus replicates in the

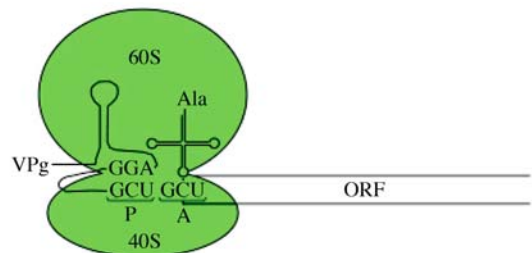
A. Cap-dependent ribosome recruitment



B. IRES-mediated translation



C. Structure of IRES



mRNA Capping. Figure 2 (A) Cap-dependent ribosome recruitment in translation initiation. eIF3, eIF4G, eIF4A, eIF2 and 4E are factors required for initiation of cap-dependent translation. 40S denotes the small ribosomal subunit and ORF the mRNA open reading frame. (B) IRES-mediated, cap-independent ribosome recruitment in protein synthesis. (C) Structure of IRES. Poliovirus IRES is shown as a typical IRES structure, together with tRNA-Ala and an 80S ribosome bound at the initiation site of protein synthesis. VPg is the virus protein covalently bound to poliovirus genomic RNA.

cytoplasm of infected cells and the genome RNA is capped. Capped viral RNAs of various sizes are synthesized by a viral polymerase, the individual mRNAs sharing a common 5' leader sequence of ~60–90 nucleotides corresponding to the 5' end of genomic RNA. These observations suggest that capped 5' leader fragments prime initiation of individual corona virus mRNAs, although other possibilities include post-transcriptional trans-splicing between leader RNA and pre-mRNAs or a polymerase jumping mechanism. In any case, SARS corona virus appears to contain a specific transcription system that could be a target for new drugs.

Retroviruses

The AIDS virus, HIV and many RNA tumor viruses contain enzymes, including reverse transcriptase (RT), packaged with the single-stranded RNA genome (positive strands). Upon infection, RT transcribes the genome and the resulting DNA copy gets integrated into the host genome to be transcribed by the host transcription machinery consisting of RNA polymerase

II and capping enzymes. Thus, like cellular mRNAs, the viral mRNAs and genome RNAs are capped as shown first for B77 avian sarcoma virus 35S RNA in our collaborative studies with Bishop's laboratory. The presence of cap similarly facilitates viral [mRNA translation](#), and protects the genome RNAs from exonucleolytic degradation until encapsidated in progeny viral particles. These and other findings on how retrovirus replication strategies resemble or differ from cellular processes helped to provide basic information leading to the development of HIV RT and protease inhibitors.

Adenovirus and Discovery of mRNA Splicing

Large, precursor mRNAs in the nucleus of mammalian cells were found by us in collaboration with Darnell's laboratory and by Perry's group, as well as by others in later studies on viral and cellular transcripts, to be capped at the initiation of transcription. Although these findings confirmed a hypothesis that mRNAs in the cytoplasm are processed from large nuclear RNAs (at the time called hn or heterogeneous nuclear RNA), the

pathway remained obscure, due to the complexity of cellular systems including the large numbers of mRNA species as well as the nuclear membrane separating the nuclear and cytoplasmic compartments (Fig. 1C). This challenging problem was approached successfully using adenovirus systems to reduce the complexity issues. Roberts' Cold Spring Harbor (CSH) group digested adenovirus-specific mRNAs, purified from infected cells by hybridization to viral genomic DNA, with RNaseT1 and separated the resulting fragments by two-dimensional DEAE-paper electrophoresis. They found a prominent spot corresponding to capped 5'-terminal oligomer that was apparently shared by multiple virus mRNAs transcribed from the same promoter but containing a variety of downstream (exon) sequences. Sharp's group, also then at CSH, obtained direct visual evidence for RNA splicing by electron microscopy. They demonstrated the presence of conserved (exon) and excised (intron) regions by examination of adenovirus DNA hybridized to viral mRNAs. These landmark virus findings dramatically changed our understanding of all eukaryotic gene expression.

Discovery of Two Types of Initiation in Protein Synthesis

Most eukaryotic proteins are produced by cap-dependent translation in which only capped mRNAs are recruited to ribosomes (like "beads on a string"). mRNA binding assisted by cap-binding protein initiates protein synthesis by accurate selection of AUG near the capped terminus (Fig. 2A). On the other hand, the absence of cap in some viral mRNAs prompted researchers to look for alternative ways to initiate protein synthesis. A cap-independent ribosome recruitment mechanism was soon reported by Sonenberg, Wimmer and others in which uncapped viral mRNAs could attach to ribosomes at an internal ribosomal entry site (►IRES, Fig. 2B, C). For example, HCV protein synthesis is initiated on a ~330-nucleotide IRES structure located downstream of the 5' uncapped end of the viral genome RNA. The protein initiation factors required for these two distinct modes of translation differ (Fig. 2A, B). IRES elements can also be present in capped mRNA and such mRNAs can presumably function in both translation systems.

Identification of Transcription Start Points Using the Cap to Prepare ►Full-Size cDNAs

Since transcription starts from the ►capping site downstream of the promoter region in a gene (Fig. 1C), the cap can be used as a "handle" to produce full-size cDNA containing the 5' terminal sequences often missing in the reverse-transcribed cDNA copy of the

mRNA or in cDNA libraries. An oligo-capping method by Sugano's group used tobacco nucleotide pyrophosphatase to remove 7mGp(p) from the 5' end of mRNA. The resulting 5' monophosphorylated end allowed fusion of the decapped RNA to a ribo-oligomer of defined sequence by RNA ligase. By using the "oligo-capped RNA" as template for reverse transcription, the cDNA produced necessarily included the capping region sequence of a given gene. The same "decapitation" method was used by Furuichi's group to locate the promoter and accurately identify the transcription start site on the Werner syndrome gene. A very similar procedure but employing a chemical reaction to remove m7G from capped mRNA was applied by Hayashizaki's group for high-throughput analysis of transcriptional starting points and frequencies of promoter usage in tissue-specific libraries of mouse genes.

Conclusion

As reflected in this short review, it is interesting to note that so many amazing findings followed the epochal discovery of caps on eukaryotic mRNAs. This flow of new, cap-related information continues even after 30 years and promises more surprises in the future. Clearly, molecular biology research is alive and well across the globe.

Clinical Relevance

Capping and caps are essential for eukaryotic life, and their absence is lethal. Thus no human diseases have yet been ascribed to capping defects. However, capping represents a novel target for drug development against yeast, fungi and some parasites because in these organisms the RNA 5' triphosphatase, unlike the mammalian enzyme, requires a cation cofactor. Among pathogenic viruses, capping by alpha viruses differs from cellular capping in that GTP is methylated before, rather than after, linkage to the 5' end of the RNA. Another antiviral possibility is the development of small molecules that interfere with cap snatching by influenza polymerase.

References

1. Shatkin AJ (1976) Capping of Eucaryotic mRNAs. *Cell* 9:645–653
2. Furuichi Y, Shatkin AJ (2000) Viral and Cellular mRNA Capping: Past and Prospects. *Adv Virus Res* 55:135–184
3. Gingras A-C, Raught B, Sonenberg S (1999) eIF4 Initiation Factors: Effectors of mRNA Recruitment to Ribosomes and Regulators of Translation. *Annu Rev Biochem* 68:913–963
4. Shuman S (2001) The mRNA Capping Apparatus as Drug Target and Guide to Eukaryotic Phylogeny. *Cold Spring Harbor Symp. Quant Biol* 66:301–312

mRNA Decay

Definition

mRNA decay describes the process by which mRNAs are degraded. In eukaryotic cells, the major pathways of mRNA decay start with the exonucleolytic shortening of the poly(A) tail.

► Polyadenylation

mRNA Processing

Definition

Cellular mRNA transcripts are made in the nucleus as precursor molecules that need to be processed extensively to be fully functional during translation in the cytoplasm. This processing includes the capping of the mRNA at its 5' end, excision of intron sequences (splicing) and addition of a polyA tail at the 3' end, and N6-methylation of specified adenosine residues.

► Nuclear Import and Export

► RNA Capping

mRNA Splicing

Definition

mRNA splicing comprises of the splicing out of introns and splicing together of exons in generation of mRNA to be used as a template for protein synthesis.

► Familial Hypercholesterolemia

mRNA Stability

Definition

mRNA stability characterizes a state maintaining mRNA structural integrity. Stable mRNAs are not readily degraded.

► Messenger RNA Stability

► RNA Capping

mRNA Translation

► RNA Capping

► Translation

► Translational Control in Eukaryotes

mRNP

Definition

The term mRNP refers to a messenger ribonucleoprotein complex that consists of an mRNA and specific RNA binding proteins, and is thought to protect mRNA from degradation and mediate transport of mRNA from nucleus to cytoplasm.

► Cap-Independent Translational Control

MS

► Mass Spectrometry

MS/MS

Definition

MS/MS refers to a twofold mass spectrometric analysis. A spectrum is generated after fragmenting a selected peptide (the parent ion) in a mass spectrometer. A mass spectrum of the fragments (daughter ions) is generated which may be used to define the sequence of the parent ion.

► Mass Spectrometry

► Proteomics in Cancer

MSA

► Multiple System Atrophy

MSH

Definition

MutS homologues (MSH) are genes that are all related to the bacterial mutS gene; their primary function is to identify and correct base mismatches ([► mismatch repair](#)) that have been erroneously introduced during DNA replication. Mutations in several human mismatch repair genes are linked to the pathogenesis of hereditary nonpolyposis colorectal cancer (HNPCC).

[► Hereditary Non-Polyposis Colorectal Cancer](#)

MSI

Definition

Microsatellite instability (MSI) means gain or loss of short repeat units within a perfect tandem, a repeat stretch that results in the length variation of the [► microsatellite](#) [e.g. (CA)₁₂ > (CA)₁₁].

[► Hereditary Nonpolyposis Colorectal Cancer](#)

mtDNA

[► Mitochondrial DNA](#)

Mucin

Definition

Mucin is a large glycoprotein with a high content of serine, threonine, and proline residues and numerous O-glycans, often occurring in clusters on the polypeptide.

[► Glycosylation of Proteins](#)

Mucocutaneous Candidiasis

Definition

Mucocutaneous candidiasis refers to an infection or disease of the mucous membrane and skin caused by the yeast like fungus *Candida albicans*.

[► Hyper-and Hypoparathyroidism](#)

Mucoviscidosis

[► Cystic Fibrosis](#)

[► Epistaxis in Cystic Fibrosis](#)

MudPIT

MudPIT (Multidimensional Protein Identification Technology) is a proteomics technology which couples two-dimensional chromatography of peptides in mass spectrometry-compatible solutions directly to tandem mass spectrometry, allowing for the identification of proteins from highly complex mixtures. Since the initial descriptions of MudPIT, this approach has been implemented in the analysis of whole proteomes, organelles and protein complexes.

[► Proteomics – from Proteins to Disease Mechanisms](#)

Multicloning Site

Definition

Multicloning site denotes a segment of DNA that contains a number of different unique restriction endonuclease sites that allow the insertion of a foreign DNA fragment by molecular cloning techniques.

[► Recombinant Protein Production in Mammalian Cell Culture](#)

Multidimensional Chromatography

Definition

Multidimensional chromatography designates the sequential use of different types of chromatography. The first chromatography might be based on ion exchange followed by a second chromatography based on a reverse phase separation. The multidimensional chromatography is a powerful tool for separating complex mixtures containing thousands of peptides.

► [Proteomics in Cancer](#)

Multidimensional NMR Spectroscopy

PETER SCHMIEDER

FMP, Institute for Molecular Pharmacology, Berlin, Germany

schmieder@fmp-berlin.de

Definition

Multidimensional NMR spectroscopy (1) forms the basis of the determination of the three-dimensional structure of biomolecules (proteins, DNA, RNA) (2) by providing the resolution necessary to analyze their complex spectra. One-dimensional NMR spectra result from a Fourier transform (FT) of a directly detected time domain signal, the ► [free induction decay](#) (FID) that is recorded at the end of ► [pulse sequence](#) during the ► [acquisition time](#). Higher-dimensional spectra are created by indirectly detecting further time domains. This is accomplished by systematic and independent variation of one or several delays in the pulse sequence while repeatedly acquiring an FID. In between the variable delays that result in the indirect detected dimensions are mixing sequences of varying complexity. They enable a transfer of magnetization *via* several mechanisms resulting in correlations between signals from different nuclei. Depending on the type of interaction used to accomplish this transfer the evaluation of the spectra allows for the extraction of either a resonance assignment or structurally relevant parameters.

Description

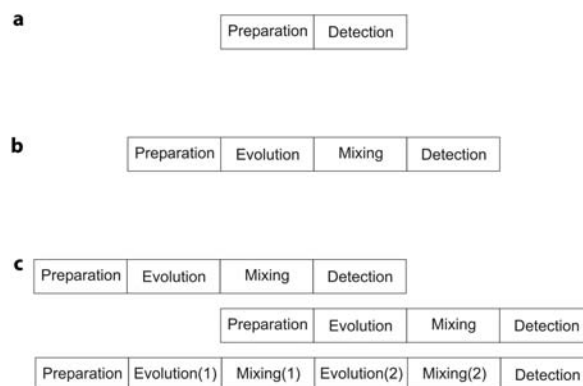
Basics

The basis of every modern NMR experiment is a pulse sequence, a series of radio frequency pulses and delays that ends with an acquisition time, during which the

FID is recorded. This is independent of whether solution state or solid state NMR is employed. For a one-dimensional experiment the pulse sequence can be divided in two major sections, the preparation period and the detection period (Fig. 1a). While the detection period always consists of the detection of the free induction decay (FID), the preparation period can be of varying complexity. In the simplest case it consists solely of the relaxation period followed by a pulse to create detectable magnetization. The relaxation period is necessary to allow the sample to return to its initial state before repeating the acquisition of the FID for improved signal-to-noise ratio. The preparation period can, however, also contain a large number of delays and pulses that together form a complex pulse sequence designed to prepare the desired state of magnetization prior to detection. This is, in particular, accomplished by varying the pulse phases from one acquisition of an FID to the next in form of a phase cycle.

Two-Dimensional NMR

The step from one-dimensional to two-dimensional NMR spectroscopy introduces two additional sections into the pulse sequence, namely the ► [evolution time](#) and the ► [mixing time](#) (Fig. 1b). The indirect detection of an additional time domain takes place during the evolution time. After completion of the acquisition of the FID including the repetitions necessary to perform



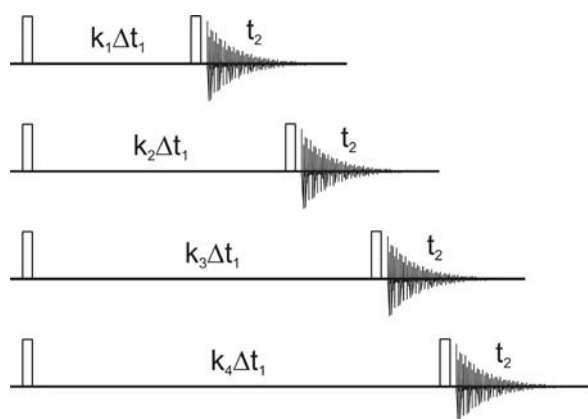
Multidimensional NMR Spectroscopy.

Figure 1 Schematic representation of multidimensional spectra. (a) A one-dimensional NMR experiment consists of a preparation and detection period. (b) A two-dimensional NMR experiment is created by inserting an evolution period and a mixing period. During the evolution period the frequencies of the indirect dimension are recorded, during the mixing time a transfer of magnetization takes place. (c) A three-dimensional NMR experiment is catenated from 2 two-dimensional ones and contains two evolution and two mixing periods.

the phase cycle of the experiment and to achieve sufficient signal-to-noise, the experiment is repeated numerous times (Fig. 2). A delay in the pulse sequence is incremented systematically from one experiment to the next and the intensities detected in the FID during the acquisition time are thus modulated according to the length of this delay. The result is a two-dimensional time domain data set, one dimension from direct detection, one dimension from systematic incrementation of a delay. It is processed using two subsequent Fourier transformations along both time domains. First the directly detected FIDs are converted into a series of one-dimensional spectra, the interferogram. Then the interferogram is converted into a two-dimensional spectrum that exhibits two independent frequency axes. Both axes will usually contain the chemical shift of nuclei in the sample, the type of chemical shift as well as the appearance of the two-dimensional spectrum results from the design of the mixing period. During the mixing time the magnetization that has been created during the preparation period and has been modulated by a frequency during the evolution time is converted into another type of magnetization that is subsequently detected during the acquisition. This is accomplished using some type of interaction (scalar or J-coupling, dipolar coupling) between one or several spins in the sample.

Homo- and Heteronuclear NMR

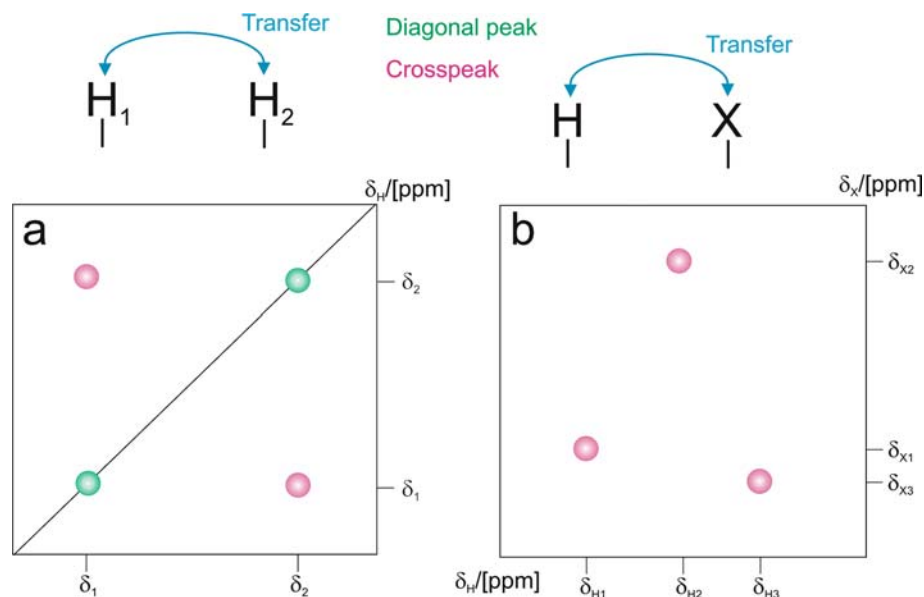
The simplest two-dimensional experiment possible in solution NMR is the COSY (correlation spectroscopy)



Multidimensional NMR Spectroscopy. Figure 2 The indirect dimension of a multidimensional experiment is created by systematic variation of a delay in the pulse sequence. Here, the simplest possible two-dimensional NMR experiment, the COSY, is shown. For each value $k_n \Delta t_1$ the full phase cycle is executed and a separate FID is recorded. k ranges from zero to a finite number (typically 128–1024 FIDs are recorded), Δt_1 is called the increment and determines the spectral window in the indirect dimension.

experiment; the pulse sequence is shown in Fig. 2. It is most commonly performed to create correlations between hydrogen nuclei (protons). The preparation period contains the relaxation period and a pulse to create detectable magnetization. During the evolution time a modulation by chemical shift takes place, in addition homonuclear, scalar coupling will be active and creates types of magnetization that enable a partial transfer from one spin to the next based on this interaction. This is accomplished by the subsequent pulse that represents the mixing period. Finally two types of signals are detected during acquisition. In one case, magnetization has not been transferred from one spin to another during the mixing time. The chemical shifts detected during the evolution time and the acquisition time are therefore the same and the frequencies resulting from the two Fourier transformations are identical. The signal will appear on the diagonal of the two-dimensional spectrum. The other type of signal results from magnetization that has been transferred from one spin to the next during the mixing time. The chemical shifts detected in the two dimensions will therefore differ and the signal will appear away from the diagonal. The first type of signal is termed “diagonal peak” while the second is termed “cross peak”. A schematic spectrum is shown in Fig. 3a. Since the frequencies on both axes result from the same type of nucleus, this kind of spectrum is called a “homonuclear spectrum”.

The occurrence of cross peaks is of major importance and represents the key feature of multidimensional NMR spectroscopy. In COSY-type spectra these peaks indicate a scalar coupling interaction between the nuclei that are represented by the two chemical shifts and thus allow the establishment of a connectivity between these nuclei, which forms the basis for resonance assignment. Other homonuclear spectra can be created by either modifying the manner of transfer *via* scalar coupling (e.g. in the so-called TOCSY experiment (total correlation spectroscopy)) or by utilizing a different kind of interaction. An important example for another type of interaction is the nuclear Overhauser enhancement effect (NOE). It results from the interactions of spins through space based on dipolar coupling and is dependent on the distance between the two nuclei involved in the interaction. Allowing for a NOE-effect during the mixing period results in a NOESY spectrum (nuclear Overhauser enhancement spectroscopy). In this kind of spectrum the appearance of a cross peak indicates spatial proximity of the nuclei represented by the chemical shifts. Moreover the intensity of the interactions is dependent on the distance and the analysis of NOESY spectra therefore yields structurally highly relevant information. Other types of interaction are utilized in solid state NMR, where dipolar coupling is of major importance to establish connectivities used for resonance assignment.



Multidimensional NMR Spectroscopy. Figure 3 Two types of two-dimensional NMR spectra exist. (a) In a homonuclear experiment, chemical shifts of the same type of nucleus (δ_H) are recorded in both dimensions. If no transfer takes place during the mixing period, the same chemical shift is recorded in both dimensions resulting in a diagonal peak. If there is transfer of magnetization, different chemical shifts are recorded and a cross peak appears, indicating a connection between the nuclei corresponding to the chemical shifts. (b) In a heteronuclear experiments two different chemical shifts (δ_H, δ_X) are recorded and in the case of a transfer of magnetisation during the mixing time, a cross peak appears, again indicating a connection between the nuclei involved.

Another type of spectrum is called a “heteronuclear spectrum” and is schematically depicted in Fig. 3b. Here the chemical shifts detected in the two time domains result from different types of nuclei. During the preparation period magnetization of the first type of nuclei is created. The chemical shifts of these nuclei are then recorded during the evolution time. Subsequently a transfer of magnetization between the two different types of nuclei is accomplished by proper choice of the mixing sequence. Mostly heteronuclear scalar coupling are used to create heteronuclear spectra in solution state NMR, while again dipolar interactions play a vital role in heteronuclear solid-state techniques. Unlike homonuclear spectra, heteronuclear spectra only contain cross peaks. If magnetization prepared for one nucleus during the preparation period is not transferred to the other type of nucleus during the mixing period, it will not be detected and thus only cross peaks with different frequencies on both axes can appear. Again the presence of a cross peak indicates an interaction between the two types of nuclei, which can be utilized to assign the chemical shift of the nuclei involved in the interaction or to extract structurally relevant information. In the case of proteins, two nuclei of major importance besides protons are carbon and nitrogen. Since the natural abundance of the two isotopes well suited for NMR spectroscopy, ^{13}C and ^{15}N , is quite

low, isotopic enrichment becomes mandatory. Two types of heteronuclear NMR spectra are of importance for the study of biomolecules using solution state NMR spectroscopy. One is the correlation between protons and a heteronucleus that is achieved by utilizing the direct scalar coupling between the two nuclei. The techniques to record these correlations in solution are called HMQC (Heteronuclear Multiple Quantum Correlation) or HSQC (Heteronuclear Single Quantum Correlation). The other type of spectra is the large family of triple resonance experiments (3, 4, 5) that are recorded utilizing the scalar interactions not only between protons and heteronuclei but also between heteronuclei. These spectra are standard spectra for the assignment of proteins in solution.

Three-Dimensional NMR

A major advantage of two-dimensional spectra compared to one-dimensional spectra is the increased resolution. Since the signals of interest are spread out over a two-dimensional plane rather than a one-dimensional line, the possibility of overlap is strongly reduced. Since overlap is a major obstacle in the analysis of NMR spectra of biopolymers, the introduction of two-dimensional techniques was instrumental in determining the structure of biomolecules using NMR. To further increase the resolution the techniques can be

extended to three-dimensional experiments (6). This is done by combining two two-dimensional experiments in the manner depicted in Fig. 1c. The resulting spectrum contains two evolution and two mixing periods. Both evolution periods are created, independently but otherwise identically, as in a two-dimensional experiment by variation of two separate delays in the pulse sequence. First a delay is systematically incremented and the acquisition of the FID is repeated for every increment to create the first indirectly detected dimension. Once the desired number of incrementations has been performed for this delay, it is set back to its original value. The other delay that is used to create the second indirectly detected dimension is now incremented once and the incrementation of the first delay is repeated, again with the acquisition of an FID for every increment. This is done until the variation of the second delay has been completed as well. The resulting three-dimensional data set is processed using a sequence of three Fourier transformations resulting in a three-dimensional spectrum with three independent chemical shift axes (Fig. 4). Again the spectra can be classified as homonuclear and heteronuclear spectra. Homonuclear spectra exhibit the chemical shift of the same nucleus on all three axes. These spectra are rarely recorded since the two transfer steps increase the number of possible correlations in the

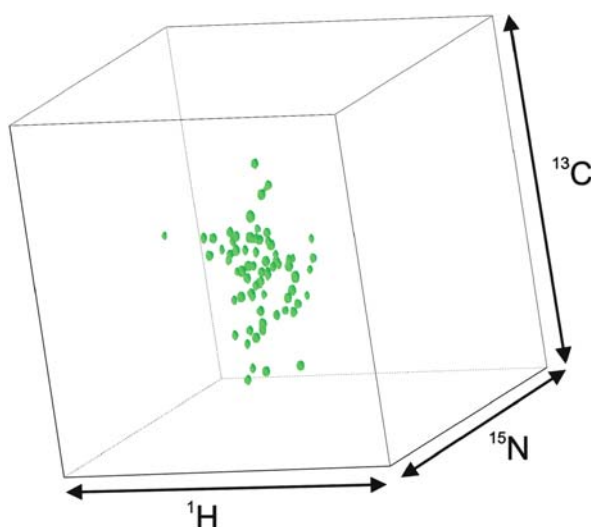
three-dimensional space considerably. The analysis is thus too complicated to be practical. Heteronuclear spectra exhibit two or three different nuclei on the three axes. In triple-resonance spectra of proteins in solution the three chemical shifts are those of protons, carbons and nitrogen. Heteronuclear edited NOESY spectra result from a combination of a two-dimensional heteronuclear correlation and a two-dimensional NOESY. The major advantage of these types of spectra is the increased resolution due to the third dimension that is not compromised by an increased number of correlations.

Four-Dimensional NMR

The extension of a three-dimensional experiment by another dimension is straightforward (7), again another delay in the pulse sequence is incremented independently of the other two delays. The pulse sequence now consists of three evolution times and three mixing times. Again only heteronuclear spectra are recorded with at least two dimensions exhibiting the chemical shift of nuclei other than protons. While the analysis of two-dimensional spectra is quite straightforward using either paper printouts or computer assistance, this is not as simple for three and four-dimensional spectra. They are therefore dissected into planes extracted from the higher-dimensional spectra. These planes correspond to two-dimensional spectra and are extracted by keeping the chemical shift in one or two dimensions constant. They can be analyzed as two-dimensional spectra, but since the information in the spectrum is spread out over a large number of planes, computer assistance for the analysis of the spectra becomes mandatory. This includes a database to contain all the information about chemical shifts and structurally relevant parameters.

Resolution in NMR Spectra

A major difference between a directly detected dimension and indirectly detected ones is the effort required to obtain sufficient resolution, i.e. to record a sufficient number of data points. There are only minor limitations on the size of the data array in the acquisition dimension since the acquisition of additional points does not necessarily extend the time of the experiment. In the indirect dimension a doubling of the number of data points results in a doubling of measurement time, since the experiments have to be repeated twice as often. The number of data points, however, is directly related to the resolution in the spectrum, therefore a high number of data points is desirable. A compromise between experiment time and resolution is required and the data array in the indirect dimensions is usually reduced. To compensate for the reduced number of data points, several solutions have been suggested. Either the size of the spectral window is reduced in a controlled manner thereby folding some



Multidimensional NMR Spectroscopy.

Figure 4 Three dimensional triple resonance experiment correlating the chemical shifts of amino protons, carbonyl carbons and nitrogens in a fully labeled protein. The name of the experiment shown here is HNCO; cross peaks appear as three-dimensional objects distributed in a three-dimensional space. The analysis of this type of spectrum is performed by dissecting the three-dimensional spectrum into two-dimensional planes.

of the resonances back into the spectral window but increasing the resolution of the spectrum or the resolution of the spectrum is enhanced using special processing techniques other than the Fourier transformation. Another approach is the recording of two or more indirect dimensions simultaneously. Two or more frequencies are then encoded in the FID of an indirect dimension and some means to disentangle the information is necessary. This approach, termed “reduced dimensionality”, works best with spectra that exhibit simple signal patterns and a limited number of peaks and has thus been successfully used in combination with triple resonance spectra.

Multidimensional NMR in Structural Genomics

In the context of structural genomics projects the acquisition and analysis of NMR spectra of proteins in solution in a high throughput manner is required (8). Several approaches for an automated acquisition, data analysis and structure calculation have therefore been presented. The assignment of all proton, carbon and nitrogen resonances is almost exclusively performed using triple resonance experiments. Because of their relatively high sensitivity and their good intrinsic resolution, these experiments are well suited for an automated resonance assignment. The analysis of heteronuclear three- and four-dimensional NOESY spectra provides the distance information essential for structure determination. These spectra are difficult to analyze automatically because of the high amount of signal overlap and sophisticated algorithms have been developed to cope with this problem. It has been shown to be very efficient to supplement the distance information from NOESY spectra with orientational restraints from residual dipolar couplings. Using strategies for automated analysis of the spectra and automated structure calculation, the time for a protein structure determination is in the range of several weeks.

References

1. Ernst RR (1992) Nuclear Magnetic Resonance Fourier Transform spectroscopy (Nobel Lecture). *Angew Chem Int Ed Engl* 31:805–823
2. Wüthrich W (2003) NMR Studies of Structure and Function of Biological Macromolecules (Nobel Lecture). *Angew Chem Int Ed Engl* 42:3340–3363
3. Montelione GT, Wagner G (1990) Conformation independent sequential NMR connections in isotope-enriched polypeptides by ^{13}C J Magn Reson 87:183–188
4. Kay LE, Ikura M, Tschudin R et al (1990) Three-dimensional triple-resonance NMR-spectroscopy of isotopically enriched proteins. *J Magn Reson* 89:496–514
5. Sattler M, Schleucher J, Griesinger C (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Progr NMR Spectr* 34:93–158
6. Oschkinat H, Griesinger C, Kraulis PJ et al (1988) Three-dimensional NMR spectroscopy of a protein in solution. *Nature* 332:374–376
7. Kay LE, Clore GM, Bax A et al (1990) Four-dimensional heteronuclear triple-resonance NMR spectroscopy of interleukin-1 beta in solution. *Science* 249:411–414
8. Szyperski T, Yeh DC, Sukumaran DK et al (2002) Reduced-dimensionality NMR-spectroscopy for high-throughput protein resonance assignment. *Proc Natl Acad Sci USA* 99:8009–8014

Multidimensional Protein Identification Technology

► MudPIT

Multidrug Resistance

ULRIKE STEIN

Max-Delbrück-Center for Molecular Medicine, Berlin, Germany
ustein@mdc-berlin.de

Definition

Multidrug Resistance and Cancer

To date, treatment of most locally confined malignant tumors is based on surgical and radiotherapeutic approaches. For advanced and metastatic tumors, ►chemotherapy is the most effective treatment. However, clinical success has been observed to differ from patient to patient. Some patients are cured, others respond transiently and a third group show incomplete responses. Furthermore, clinical oncologists noted that cancer patients treated with multiple anticancer drugs developed cross-resistance to many other chemotherapeutics to which they had never been exposed. As a consequence, the possibility of curing these patients with chemotherapy is dramatically reduced. This phenomenon of the ability of a tumor cell to resist typically lethal or sublethal doses of multiple, usually cytotoxic drugs is called multidrug resistance (MDR) (1, 2). MDR is defined as simultaneous resistance to structurally and functionally unrelated natural product anticancer drugs. MDR may occur intrinsically without any treatment or may be acquired during therapy. To

date, MDR of human tumors to a variety of chemotherapeutic agents represents the major cause of failure of cancer chemotherapy. Therefore, the ability to understand MDR and thus predict, circumvent and overcome drug resistance is desirable for the improvement of chemotherapy.

Characteristics

Cellular Mechanisms of Drug Resistance

Multiple mechanisms of drug resistance may occur in parallel or sequentially, leading to intrinsic (without any treatment) or acquired (therapy-induced) resistance towards cytotoxic drugs. Anticancer drugs have different possibilities for entering the cell. Uptake of hydrophilic drugs into the cell depends on transporters, carriers or channels, since these drugs are not able to cross the cell membrane by themselves. Defects in drug uptake proteins lead to reduced influx, resulting in lowered intracellular drug concentration; the cell becomes resistant to a single drug. Natural anticancer products such as ►**anthracyclines** (doxorubicin, daunorubicin, epirubicin), ►**Vinca alkaloids** (vinblastine, vincristine), ►**podophyllotoxins** (etoposide, teniposide) and ►**taxanes** (paclitaxel, docetaxel) are hydrophobic and enter the cell by diffusion across the cell membrane. Resistance towards these drugs develops by increased drug efflux depending on the activity of efflux pumps, such as transmembrane energy (ATP)-dependent transporter proteins. Thus, intracellular drug concentration is reduced and the cell becomes resistant to a panel of hydrophobic drugs. Furthermore, activation of proteins that are involved in metabolism or detoxification of drugs may cause drug resistance without affecting drug accumulation. Intracellular drug redistribution, e.g. the nucleocytoplasmic drug transport by nuclear pores is reported to contribute to drug resistance. Moreover, cytostatics-induced activation of nuclear proteins such as mismatch repair genes involved in enhanced repair of drug-induced DNA damage may lead to drug resistance. Cells may also become resistant to drug-induced cell death by activation of anti-apoptotic proteins and by affecting the cell-cycle and checkpoints.

Multidrug Resistance and ABC Transporters

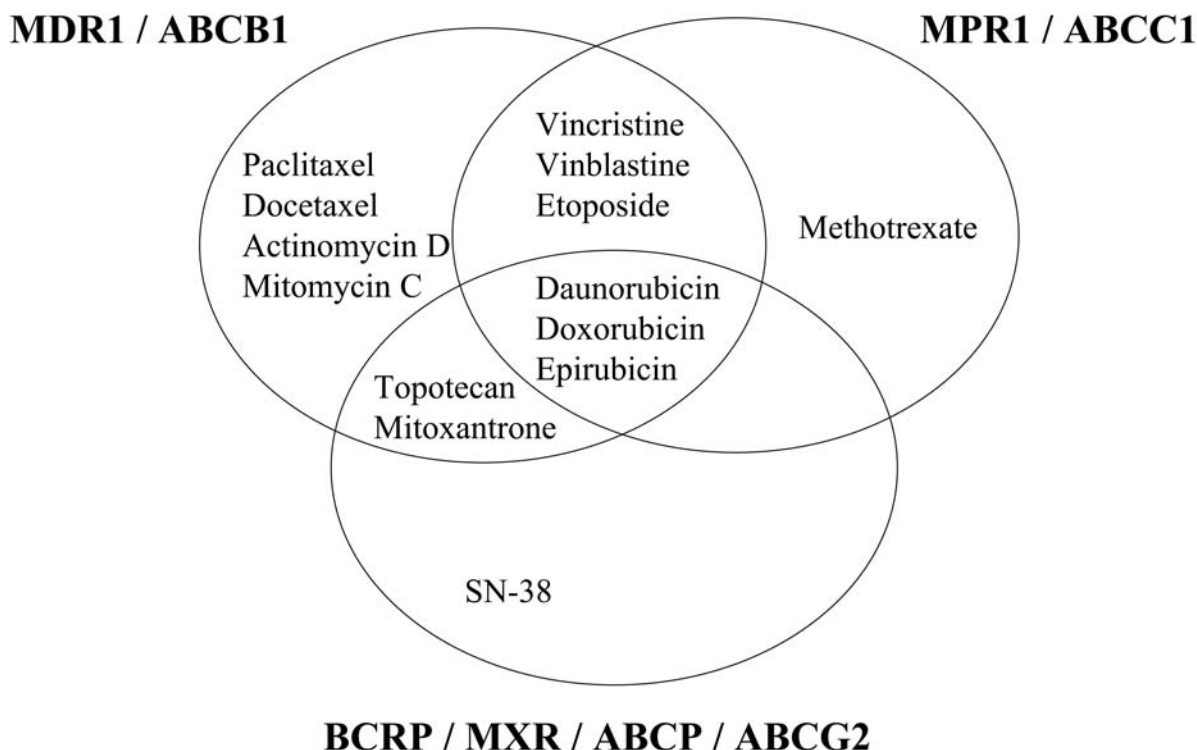
Thus, several mechanisms might be able to contribute to the drug resistance phenotype. However, within the last two decades one resistance mechanism was identified as very likely representing the most frequent cause of the development of MDR in cancer cells, the over-expression of ATP-dependent transporter proteins in the cell membrane acting as drug efflux pumps, thereby lowering the intracellular concentration of the cytotoxic drugs due to increased drug efflux. These drug efflux pumps belong to the intensively studied protein family of the transmembrane ►**ATP-binding**

cassette (ABC) transporters (3, 4). They transport multiple anticancer drugs such as anthracyclines, *Vinca* alkaloids, podophyllotoxins, taxanes, etc. out of the cell making it multidrug resistant. In the human genome, 48 ABC transporters have been identified so far. Based on sequence similarities – they share common features such as nucleotide binding sites and transmembrane domains – they have been classified (ABCA-ABCG). ABC transporters associated with the MDR phenotype are encoded by the following genes. The *multidrug resistance gene 1 (MDR1) alias ABCB1* encodes P-glycoprotein, causing the so-called classical, P-glycoprotein-mediated MDR, the genes for the *multidrug resistance-associated proteins 1-9 (MRP1-9) alias ABCC1-6 and ABCC10-12*, as well as the gene for the ►**breast cancer resistance protein (BCRP) alias ►ABC transporter** in placenta (ABCP) *alias* mitoxantrone-resistance protein (MXR) *alias* ABCG2, all lead to the so-called atypical, non-P-glycoprotein-mediated MDR. Their causal role in generation of the MDR phenotype was demonstrated for some of these MDR-associated ABC transporters, e.g. MDR1, MRP1 and BCRP; for others, their ability to transport multiple cytotoxic compounds was reported. Interestingly, the drug spectra of individual MDR-associated ABC transporters are, although overlapping, not identical (Fig. 1). Moreover, it has been shown for several transporters that their substrate specificities may vary due to defined point mutations within the MDR-associated genes.

Molecular Interactions

MDR1/P-Glycoprotein: Drug Transport, Gene Expression and Therapy-caused Induction

The ABCB subfamily of ABC transporters contains 11 members including the most prominent *MDR1* gene, which was first described more than two decades ago. *MDR1*/P-glycoprotein was the first cloned human ABC transporter for which the generation of the MDR phenotype was shown directly by gene transduction. The *MDR1* gene encodes P-glycoprotein, a drug efflux pump with two nucleotide binding sites and two transmembrane domains each consisting of 12 membrane spanning α -helical regions and believed to determine the substrate specificity of the drug transported. P-glycoprotein is able to bind a wide spectrum of hydrophobic, neutral or positively charged substrates, such as anthracyclines, *Vinca* alkaloids, podophyllotoxins and taxanes. After binding the drug, ATP hydrolysis leads to conformational changes in the protein as an essential prerequisite for the release of the substrate from the cell. A second ATP hydrolysis is necessary for resetting of the transporter molecule. *MDR1* gene expression has been analyzed in an enormous number of studies. High expression levels



Multidrug Resistance. Figure 1 Selected chemotherapeutic drugs which are substrates for a single or multiple multidrug resistance-associated ABC transporters MDR1 / ABCB1, MRP1 / ABCC1 and BCRP / MXR / ABCP / ABCG2 (from references 1–5).

have been repeatedly detected in normal tissues with excretory or secretory functions. In tumors of these tissues, such as colon, kidney, adrenocortical and hepatocellular cancers, over-expression of this protein has also been demonstrated. This is a major reason for the limited selection of chemotherapeutics for tumors of these organs. Moreover, for a variety of tumor entities, such as breast cancer, bone and soft tissue sarcomas and non-solid cancers, e.g. acute myelogenous leukemia, *MDR1* expression levels and treatment-caused inductions have been correlated to clinical outcome.

Furthermore, besides the basal expression levels of the P-glycoprotein, external factors such as components of multimodal cancer therapy are able to induce the expression of the *MDR1* gene. Thus, cytostatics, but also heat or radiation, may lead to elevation of *MDR1* gene expression. Several studies have shown, that in approximately 50% of all treated tumors inductions of *MDR1* expression have been observed. By identification of the *MDR1* gene promoter and generation of promoter deletion mutants thereof, drug responsive elements and heat responsive elements have been identified, mediating therapy-caused stress signals into transcriptional activation of the *MDR1* gene. This is in agreement with chemotherapy- or

hyperthermia-induced elevations of *MDR1* gene expression that have been detected in patients.

The MRP Family: Drug-Transport and Clinical Implications

The ABCC subfamily of ABC transporters consists of 12 members with 9 *MRP*-related genes (MRP1–9). The most studied member MRP1 was identified more than a decade ago in a non-P-glycoprotein expressing, but multidrug resistant human tumor cell line. MRP1 is structurally similar to P-glycoprotein, however, with an amino terminal extension of 5 membrane spanning α -helical regions. Like MRP1, MRP2, MRP3 and MRP6 also harbor these extra amino terminal transmembrane regions as compared with P-glycoprotein. The spectrum of hydrophobic natural anticancer drugs transported overlaps with those transported by P-glycoprotein and also includes compounds of anthracyclines, *Vinca* alkaloids and podophyllotoxins. Like P-glycoprotein, the substrate specificity of MRP1 and MRP2 is altered by single amino acid substitutions within the transporter proteins. However, the mechanism of drug transport is different. These MRPs transport anionic and neutral drugs conjugated to acidic ligands, such as glutathione, glucuronate or

sulfate. Alternatively, they can cause resistance to neutral drugs without conjugation but by co-transporting these drugs with free glutathione. MRP4, MRP5 as well as MRP7 lack the additional helical regions, but show higher sequence similarities to MRP1 than to P-glycoprotein or other ABC transporters. Both MRP4 and MRP5 are also organic anion transporters like MRP1, 2 and 3, which have been shown to pump nucleotide analogues.

The causal role for *MRP1* in conferring the MDR phenotype to previously chemosensitive cells has been shown by gene transduction. Basal expressions of *MRP1* have been found to be ubiquitous in human tissues and basal *MRP1* expression levels have also been found in almost all malignant tissues analyzed. However, the clinical importance of *MRP1* is still a matter of discussion. Studies have both confirmed and rejected a correlation of outcome and *MRP1* expression. However, for several tumor entities, correlations of high *MRP1* expression with clinical outcome parameters, such as response to chemotherapy and disease-free survival, have been reported, e.g. in mammary, lung and ovarian carcinomas, as well as for leukemias. In refractory hematological malignancies, increases in *MRP1* have been observed.

BCRP/ABCP/MXR/ABCG2: a Half-Transporter

The breast cancer resistance protein (BCRP) or ABC transporter in placenta (ABCP) or mitoxantrone-resistance protein (MXR) was originally identified in a mitoxantrone-resistant, but P-glycoprotein and MRP1-negative human carcinoma cell line and was named differently due to its almost simultaneous discovery by 3 independent groups. This gene, now named *ABCG2*, also encodes an ABC transporter and, together with 5 additional members, makes up the ABCG subfamily. This so-called “half-transporter” harbors only one nucleotide binding site and only one transmembrane domain with 6 membrane spanning α -helices regions, whereas “full-transporters” such as P-glycoprotein and the MRPs contain 2 nucleotide binding sites and 2 transmembrane domains, suggesting ABCG2 homodimerization may be required to gain full transport activity. Transduction of *ABCG2* cDNA caused the resistance phenotype. The drug spectra of ABCG2 considerably overlap with those of P-glycoprotein, including mitoxantrone, topotecan and doxorubicin, with however, differing efficiencies. ABCG2 showed high affinities for mitoxantrone, thus representing the long sought transporter for this cytotoxic compound. Defined point mutations leading to amino acid substitutions affect substrate specificity. However, other typical P-glycoprotein substrates, such as vincristine and taxanes, are not included in the ABCG2 drug spectrum.

ABCG2 has been mainly detected in human tissues such as placenta, colon, mamma and liver. So far, in human tumors, *ABCG2* expression has been most frequently found in carcinomas of the digestive tract, lung carcinomas and melanomas. Its clinical relevance remains to be elucidated.

Regulatory Mechanisms

Reversal of MDR

Failure of cancer chemotherapy is most of all linked to basal and/or induced expression of P-glycoprotein. Thus, the inhibition of P-glycoprotein in order to aim at reversal of the MDR phenotype has been extensively studied for more than two decades. Several approaches have been made to target the function and the expression of MDR-associated genes and proteins, such as the employment of specific antibodies, inhibiting the drug transport, the introduction of antisense oligonucleotides and ribozymes or, one of the newest developments, the transduction of small interference RNA molecules. However, the most promising attempts to reverse the MDR phenotype have been made with inhibitory compounds targeting P-glycoprotein function (5).

MDR Modulators of the First, Second and Third Generations

Numerous approaches have been made with so-called first-generation ►MDR modulators of P-glycoprotein. These are antagonists that have already been used for other purposes. First-generation modulators are themselves substrates for P-glycoprotein and compete with the cytotoxic drug for binding and efflux by the P-glycoprotein pump thereby limiting the efflux of the drug. As a consequence, intracellular drug concentrations increase, leading to elevated cytotoxicity; the cell becomes chemosensitive. These modulators worked with great success on cultured cells to overcome MDR. In the clinic, they produced disappointing results, due to the requirement for extremely high modulator concentrations to reverse MDR, which caused unacceptable toxicity. Furthermore, these modulators are not exclusively transported by P-glycoprotein, resulting in unpredictable pharmacokinetic interactions when coadministered with the drug.

►Second-generation MDR modulators were generated with the aim of specifically overcoming the P-glycoprotein-mediated resistance. Compared with first-generation modulators, they showed less toxicity and greater potency for inhibition of P-glycoprotein. Clinical trials have demonstrated that administration of second-generation modulators together with the anticancer drug may lead to reversal of the MDR phenotype when treating refractory cancers. However, as seen with the early modulators, some limitations such as

unacceptable toxicity and interaction with additional transporter molecules could not be avoided.

Thus, ►[third-generation MDR modulators](#) were developed, based on structure-activity relationships. These newly generated compounds inhibit P-glycoprotein more specifically and with greater potency and do not usually interact with other transporter molecules. They are not substrates for a defined ABC transporter themselves as are first- and second-generation modulators. Modulators of the third-generation bind noncompetitively to the pump thereby causing conformational changes in the transporter protein that hinder ATP hydrolysis. Consequently, drug transport out of the cell is prevented, leading to increased intracellular concentration of the cytotoxic drug and enhanced cytotoxicity. Moreover, it has been demonstrated for several of these newly created modulators that the pharmacokinetics of classical MDR-associated drugs, such as doxorubicin, vincristine, etoposide and paclitaxel, were not affected in patients. Based on these data obtained with the third-generation modulators, MDR of cancer cells might represent a potentially surmountable obstacle to improved chemotherapy. If the mechanisms of resistance can be overcome, the spectrum of conventional agents and of treatable tumor entities will certainly be extended.

References

1. Gottesman MM, Fojo T, Bates SE (2002) Multidrug resistance in cancer: Role of ATP-dependent transporters. *Nat Rev Cancer* 2:48–58
2. Leonard GD, Fojo T, Bates SE (2003) The role of ABC transporters in clinical practice. *Oncologist* 8:411–424
3. Dean M, Rzhetsky A, Allikmets R (2001) The human ATP-binding cassette (ABC) transporter superfamily. *Genome Res* 11:1156–1166
4. Borst P, Oude Elferink R (2002) Mammalian ABC transporter in health and disease. *Annu Rev Biochem* 71:537–592
5. Robert J, Jarry C (2003) Multidrug resistance reversal agents. *J Med Chem* 46:4805–4817

decreased drug accumulation in multi drug resistant cells.

- [ATP-Binding Cassette \(ABC\) Transporters](#)
- [Multi-Drug Resistance](#)

Multidrug Resistance - Associated Protein 1

Definition

Multidrug resistance-associated protein 1 (MRP1), alias ABCC1, belongs to the ABC transporter family, subfamily ABCC (16p13.1). It is an integral membrane protein of 1531 amino acids, acting as an energy-dependent drug efflux pump responsible for decreased drug accumulation in multidrug resistant cells. Additional MDR-related family members are MRP2–9, alias ABCC1–6 and ABCC10–12.

- [ATP-Binding Cassette \(ABC\) Transporters](#)
- [Multi-Drug Resistance](#)

Multifactorial or Common Diseases

Definition

Multifactorial or common diseases are caused by both oligo- and polygenic genetic factors, as well as unknown non-genetic (environmental) factors.

- [Mendelian Forms of Human Hypertension and Mechanisms of Disease](#)
- [Multifactorial or Common Diseases](#)
- [Predictive Testing and Genetic Counseling](#)
- [Repeat Expansion Diseases](#)

Multidrug Resistance Gene 1

Definition

Multidrug resistance gene 1 (*MDR1*), alias ABCB1, belongs to the ABC transporter family as the most prominent member, subfamily ABCB (7q21.1), encoding the gene product P-glycoprotein. It is an integral membrane protein of 1280 amino acids, acting as an energy-dependent drug efflux pump responsible for

Multifactorial Threshold

Definition

According to the multifactorial threshold model (or MF/T), the occurrence of a phenotype depends on a very large number of genes (each of equal minor and additive effect) and on environmental factors. An accumulation of these genes and environmental factors is tolerated, e.g. by the developing fetus to a point

termed the threshold beyond which there is the risk of malformation.

► [Cleft Lip Palate](#)

Multifunctional Enzymes

Definition

Multifunctional enzymes are proteins that contain several catalytic activities, usually at different sites on the enzyme surface.

► [Nucleotide Biosynthesis](#)

Multiple Alignments

Definition

Multiple alignments refer to a set of biosequences, arranged in a table, such that each row of the table consists of one sequence separated by gaps.

► [Protein Databases](#)

Multiple Endocrine Neoplasia Type 1

Definition

Multiple endocrine neoplasia type 1 (MEN1) is an autosomal predisposition to combinations of parathyroid tumors, pancreatic islet cell tumors, and pituitary hyperplasia or tumor formation. Other endocrine and non-endocrine neoplasias can be involved but with a lower frequency. It is an autosomal dominant disorder, but it may also occur sporadically as a result of new mutations. Synonym: Wermer's syndrome.

► [Hyper- and Hypoparathyroidism](#)

Multiple Sclerosis

Definition

Multiple sclerosis is a demyelinating disease of the human central nervous system.

► [Genetic Predisposition to Multiple Sclerosis](#)

Multiple Sequence Alignment

Definition

Multiple sequence alignments are computational alignments of three or more peptide sequences, such that conserved regions or patterns like the dominance of hydrophobic residues etc. are highlighted. Alignment is thus helpful for protein structure prediction. Sequence alignment also enables the inference of the evolutionary history of the sequences.

► [Protein Domains](#)

Multiple System Atrophy

Definition

Multiple System Atrophy (MS) designates a cause of parkinsonism that is not typically dopamine responsive, and which is associated with dysautonomia among other features.

► [Parkinson's Disease: Insights from Genetic Causes](#)

Multiplex Quantitative PCR

Definition

Multiplex quantitative polymerase chain reaction (PCR) refers to a method that is designed for the detection of exonic deletions or duplications. Exons are addressed by short fluorescent fragments generated by simultaneous PCR amplification to facilitate the detection of changes in relative dosage.

► [Hereditary Nonpolyposis Colorectal Cancer](#)

Multiplexed Expression Fluorescence in Situ Hybridization

► [Single-Cell Gene Expression Profiling: Cell-level Biology by Multiplexed Expression Fluorescence in Situ Hybridization](#)

Multiplexing

Definition

Multiplexing is a technical term that describes a collection of multiple signals of information on a carrier at the same time in the form of a single, complex signal and then recovers the separate signals at the receiving end.

► [Protein Microarrays as Tool for Protein Profiling and Identification of Novel Protein-Protein Interactions](#)

Multipotent

► [Pluri-/Multipotent](#)

Multipotent Stem Cells

Definition

Multipotent stem cells are cells that are present in developing and adult tissues, and maintain the ability to undergo asymmetrical division to produce one new self-maintaining stem cell and one daughter cell, which cycles and differentiates into several different cell types characteristic of the tissue.

► [Gut Epithelium](#)
 ► [Stem Cells - Overview](#)

Multipotent/Multipotency

Definition

Multipotency is a property of ► [precursor cells](#). It implies that the precursor cells can develop into multiple cell lineages within one germ layer or one tissue.

► [Gut Epithelium](#)
 ► [Neural Stem Cells](#)

Multi-Wavelength Anomalous Diffraction

Definition

Multi-wavelength anomalous diffraction denotes a method to determine the diffraction phases in protein crystallography, from several data-sets that are measured at two or more wavelengths near the absorption edge of natural or artificially incorporated anomalous scatterers. A tunable radiation source (synchrotron) is required for recording high anomalous and dispersive signals from the scatterer. The method allows determination of macromolecule structures of up to 200,000 M_r .

► [MAD Phasing](#)

Munc13

Definition

Munc13 is a presynaptic protein with an essential role in making synaptic vesicles competent for fusion with the membrane (synaptic vesicle priming). The priming activity appears to be activated by diacylglycerol, and is thought to be mediated by binding of Munc13 to syntaxin I.

► [Protein and Membrane Transport in Eukaryotic Cells](#)

Munc18

Definition

Munc18 is an absolute requirement for synaptic vesicle exocytosis. It interacts with syntaxin 1 and – possibly in concert with a yet unidentified Rab or Rab effector – is believed to be involved in ► [SNARE](#) complex formation.

► [Protein and Membrane Transport in Eukaryotic Cells](#)

Muscle Atrophy

Definition

Muscle atrophy (amyotrophy) is characterized by wasting of the muscles.

► [Duchenne Muscular Dystrophy](#)

Muscle Contraction

INGO MORANO

Max Delbrück Center for Molecular Medicine,
Department of Molecular Muscle Physiology, Robert,
Berlin and, University Medicine Berlin (Charité),
Johannes-Müller Institute for Physiology, Berlin,
Germany
imorano@mdc-berlin.de

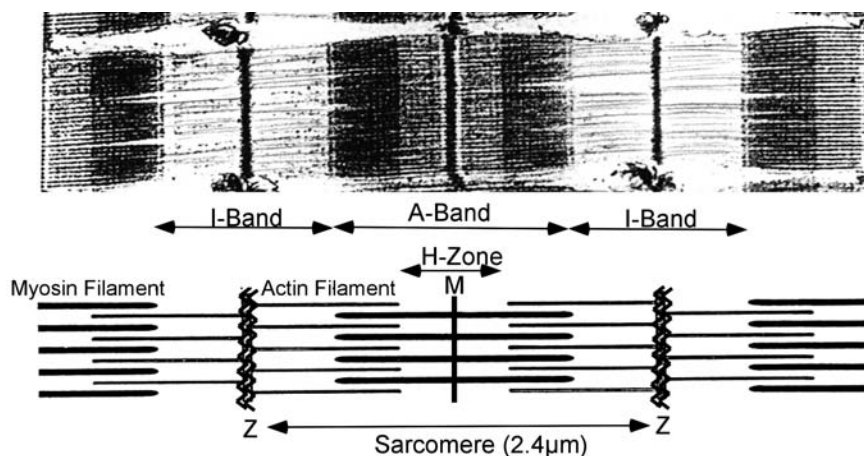
Definition

The main function of muscle is to contract, i.e. to generate force and to shorten, thus producing work and heat. Contractions can be isometric (force generation without shortening) or isotonic (shortening at constant load/force generation). According to their appearance in the light microscope, there are cross-striated muscle (skeletal muscle and cardiac) and smooth muscle types.

Characteristics

Roughly 40% of the body mass is skeletal muscle, another 10% smooth and cardiac muscle. Skeletal muscle is made up of muscle fiber bundles, which consist of muscle fibers (around 50 μm in diameter) and up to 12cm length. A muscle fiber consists of a plasma membrane (sarcolemma, with regular invaginations called the T-tubuli), an endoplasmic reticulum (sarcoplasmic reticulum consisting of longitudinal and terminal cisternae storing large amounts of Ca^{2+} , which

communicate with the T-tubuli, thus forming “triades”) and **myofibrils** (tubular cross-striated structures around 1 μm in diameter extending along the whole length of the muscle fiber). The sarcolemma is stabilized by the dystrophin-glycoprotein complex (DGC). The DGC consists of a cytoplasmic syntrophin complex, a transmembrane sarcoglycan complex, a transmembrane/extracellular dystroglycan complex (which binds to the extracellular matrix proteins laminin and agrin at the neuromuscular junction) and the huge cytoplasmic rod-shaped protein dystrophin (utrophin at the neuromuscular junction), which is linked to cytoskeletal actin and the myofibrils. Myofibrils mainly consist of longitudinally arranged contractile thin filaments (polymerized g-actin molecules, around 1 μm in length and 50 nm thick), thick filaments (polymerized Type II myosin molecules, around 1.6 μm in length and 100 nm thick), filaments made of titin and nebulin and vertically arranged Z- and M-lines (Fig. 1). The distance between the Z-lines is called a sarcomere and has a length of 2–2.4 μm in the resting muscle. The actin filaments anchor in the Z-lines, the myosin filaments locate within the sarcomeres vertically interconnected by a central M-line. Titin filaments connect Z- and M-lines (therefore called connectin), while nebulin filaments are tightly bound to the thin filament. Hence, the myofibrils show regions with high birefringence where thick and thin filaments overlap (“A-bands”) and low birefringence with no overlap (“I-bands” generated by actin filaments and “H-bands” generated by thick filaments) (Fig. 1). Besides, 7 vertically arranged N-lines in the I-band, and



Muscle Contraction. Figure 1 Structure of the myofibril. Top: Electron microscopy; Bottom: Schematic representation. Please note the fibrillar structure of the contractile apparatus, which is organized in sarcomeres in series. Actin (thin) filaments are anchored into the Z-line, while myosin (thick) filaments are fixed by a M-line in the middle of the sarcomere. The A-band corresponds to the overlapping region of thick and thin filaments. The I band corresponds to the non-overlapping region of actin filaments. The H-zone corresponds to the non-overlapping regions of the myosin filaments.

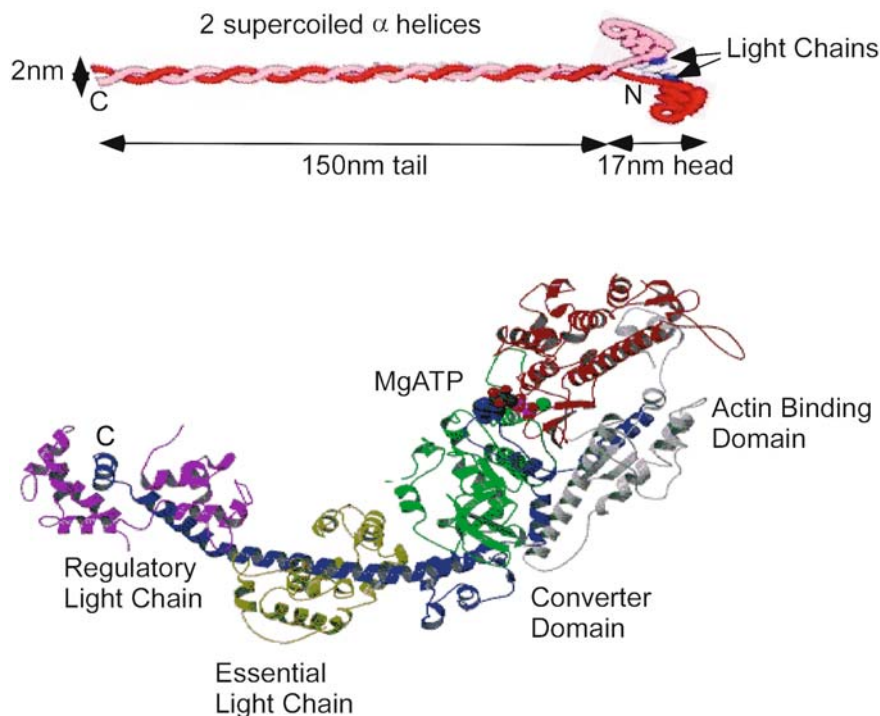
9 C-zones stripes made of protein C, -H, and -X with 43 nm spacing in the A-band exist in the sarcomer. The thin filament is associated with filamentous tropomyosin (35 nm length) and troponin complexes each composed of troponins I, C and T with an axial repeat of 38.5 nm. The Z-line with its characteristic zigzag structure is mainly composed of globular alpha-actinin, but contains many more structural and regulatory proteins. The M-band appears in longitudinal sections as a set of 3–5 prominent stripes spaced at about 22 nm axial intervals. The main non-myosin proteins of the M-band are a muscle-specific creatine kinase, M-protein and myomesin. ▶**Type II myosin** (myosin) is the motor protein of all muscle types. It is a hexamer composed of two heavy chains (MHC) and 4 light chains MLC (regulatory and essential MLC) (Fig. 2). The heavy chains (200 kD each) consist of a 150 nm alpha-helical C-terminal rod, which associates to form the thick filaments and a pear-shaped 20 nm N-terminal head domain with the ATP and actin binding sites and a converter domain (motor domain). The MLCs (16–28 kD) bind at the alpha-helical rod/head junction, the neck domain, which functions as a lever arm to amplify rotational movements of the converter domain during ATP hydrolysis (1). The thick filaments are of opposite polarity with the tails of the myosin molecules pointing

towards the center of the filament. The myosin heads project from the thick filament except for a 0.15 μm bare zone, where there are only overlapping myosin tails. Each thick filament contains about 300 myosin molecules. The heads project from the thick filament forming a helical array with a periodicity of 42.9 nm and an axial interval of 14.3 nm. Actin is a 5.5 nm globular protein (g-actin, 42 kD) with ATP bound. It polymerizes under physiological conditions to a filamentous structure (f-actin) with 13 actin molecules (with ADP bound) arranged on left-handed turns with a 36 nm crossover repeat.

Regulatory Mechanisms

Mechanochemical Energy Transformation

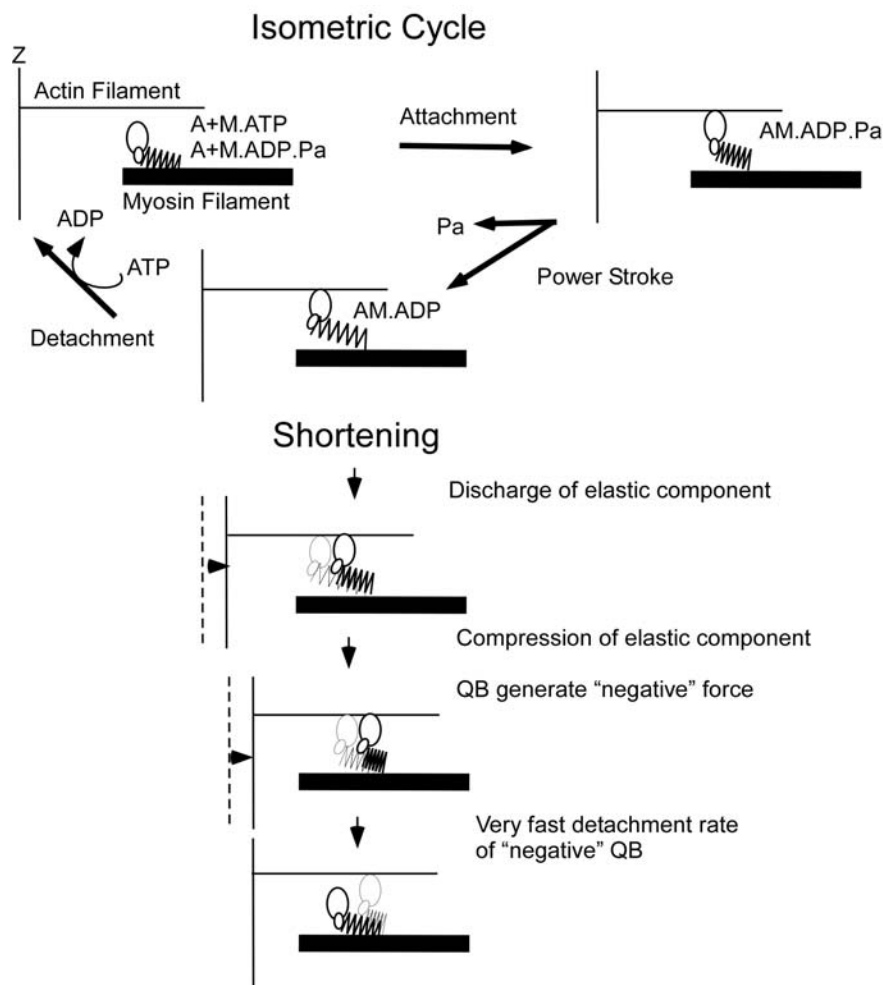
Muscle contracts by the relative sliding of thick and thin filaments (2). Thus the distance between Z-lines, the sarcomere, decreases during shortening. The sliding force is generated by the pear-shaped motor domain of the myosin molecules, which bind as the “▶cross-bridges” to the thin filament. Mechanochemical energy transformation is performed by a cyclic interaction of the cross-bridges with actin, in which cross-bridges bind, generate force upon a conformational change (the power stroke) and detach from the thick filaments (3). Each cycle is driven by the



Muscle Contraction. Figure 2 Structure of myosin. Top: Scheme of the native myosin molecule. The carboxyterminus (C) of the myosin molecule is an alpha-helical coiled coil 150 nm long, while the N-terminus (N) is the pear-shaped 17 nm motor domain. Bottom: 3D structure of the head portion of myosin (1). See text for details.

hydrolysis of 1 ATP molecule at the catalytic site of the myosin cross-bridge (Fig. 3). The ATP bound to myosin (M.ATP) is hydrolyzed very fast to ADP and Pi (M.ADP.Pi), which stick to the catalytic domain. M.ADP.Pi may bind to actin, initially in a pre-force generating state (AM.ADP.Pi). Actin accelerates the releases of Pi from the cross-bridge, which is associated with an increase in actin affinity and the power stroke of the cross-bridge. The power stroke is executed by a conformational change in the elastic lever arm of the myosin molecule of roughly 10 nm (1). Detachment of the force-generating cross-bridge from the actin filament is achieved by the release of ADP and the binding of ATP, which reduces the actin affinity of the cross-bridge. In the absence of ATP, the cross-bridges remain bound to the actin filament in a specific

force-generating conformation, the **rigor** (AM) state. In muscle, ADP release from the AM.ADP complex is the slowest (rate-limiting) step, which enables the cross-bridges to accumulate in force-generating states with high actin affinity. The isometric force of muscle contraction (F) during steady state therefore depends on the number of cross-bridges in the force-generating state and the force a cross-bridge can exert (F' , 1–4 pN). During activation, there is always a fraction of the total cycling cross-bridges (N) in force generating, and another in non-force generating states (Fig. 3). The fraction of force-generating cross-bridges “n” depends on the rate of cross-bridge detachment “g” and the rate of attachment “f”, namely $n = f/(f + g)$ (3). Thus F equals $F = F' \times N \times f/(f + g)$. The force of muscle contraction therefore also depends on the overlap



Muscle Contraction. Figure 3 Scheme of mechanochemical energy transformation during isometric force production and shortening due to a quick release. For simplicity, part of a sarcomere is shown with one Z-line, actin filament, myosin filament and a single cross-bridge with an elastic component. See text for details. A, actin; M, myosin; ATP, adenosine triphosphate; ADP, adenosine diphosphate; Pi, inorganic phosphate.

between thick and thin filaments. Furthermore, the kinetics of cross-bridge cycling and hence the fraction of cross-bridges in the force-generating state can be varied by the expression of genes coding for different myosin isoenzymes.

If an isometric contracting muscle is allowed to shorten quickly (in less than 1 ms, “quick release”), there is an instantaneous fall in tension, due to the discharge of the elastic element of the force-generating cross-bridges (Fig. 3). A quick release of roughly 1% of muscle length at optimal overlap, which equals a step size of around 10 nm per half-sarcomere, causes a complete fall of isometric contraction to zero force. Larger release amplitudes cause the compression of the elastic spring component of the cross-bridges (Fig. 3), thus generating force in the reverse direction. The fraction of these “negative” cross-bridges increases, while that of positive cross-bridges decreases, with increasing shortening velocities. The decline in positive cross-bridges with increasing shortening velocities, which can be directly demonstrated by stretching a muscle with very small amplitudes (stiffness measurements) during isotonic shortening, could partially explain the hyperbolic force-velocity relationship of muscle contraction, i.e. the experience that a small load can be lifted faster than a big load. Hence, to allow fast shortening, the detachment rate of “negative” cross-bridges (“ g_2 ”) should be very much higher than the detachment rate “ g ” of positive cross-bridges during **isometric contraction**. As a consequence, the **maximal shortening velocity** V_{max} of a muscle depends on “ g_2 ”. Different V_{max} values of fast and slow muscles

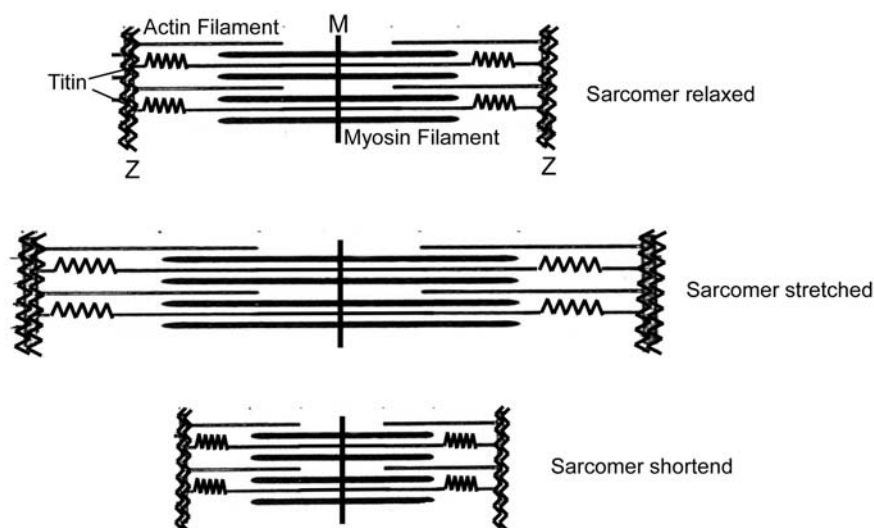
types (see below) are partly due to distinct gene expression of myosin isoenzymes with higher and lower g_2 values.

The work that can be performed by the cross-bridges is around 22 kJ/mol cross-bridges. Since the free energy change of ATP splitting is about 50 kJ/mol, cross-bridges convert roughly one half of the free energy of ATP as work; one half is liberated as heat.

The passive elasticity of a muscle fiber, i.e. the tension generated upon stretch during rest, depends mainly on titin filaments, which have spring-like elements, in particular in the I-band region (Fig. 4). Furthermore, compression of titin spring elements during sarcomere shortening could provide the restoring force that sets the sarcomere length to resting levels if activation ceases.

Excitation-Contraction Coupling

Skeletal muscle fibers are innervated by large myelinated nerve fibers, the motor neurons that originate from the anterior horn of the spinal cord. Each nerve fiber branches many times and stimulates several muscle fibers by specific contacts, the motor end plates. Acetylcholine (ACh) secreted from the motor neuron binds to nicotinic ACh-receptors (ACh-gated non-specific cation channels) in the sarcoplasm, thus increasing the open-probability of its channel domain. Depolarization of the subsynaptic sarcolemma elicits an action potential, which has a duration of around 1–5 ms and a conduction velocity of around 3–5 m/s and spreads all over the sarcolemma. Specific L-type calcium channels, located mainly in the T-tubular



Muscle Contraction. Figure 4 Scheme of a sarcomere with three filaments in resting, stretched and shortened states. Please note the elastic spring in the I-band region of titin, which becomes distended or compressed during stretching or shortening respectively.

membrane of the sarcolemma, physically link calcium release channels (ryanodine receptors) inserted into the terminal cisternae of the SR. These L-type calcium channels mainly function as voltage sensors (but cf “The Heart” in this volume), which change their conformation in response to an action potential and, as a consequence of its direct interaction, increase the open-probability of the RyR. Large amounts of Ca^{2+} are then released from the SR into the myoplasm, increasing the free Ca^{2+} from around 10^{-7} M in the resting state to above 10^{-5} M at maximal contraction. This triggers muscle contraction by binding to troponin C. The subsequent conformational changes in the whole troponin-tropomyosin complex, associated with a change in the location of tropomyosin on the thin filament, turns the actin filament from an “off” into an “on” state that permits the force-generating interaction with the myosin cross-bridges and contraction (4).

A single action potential elicits a transient contraction (twitch) with low force, while increasing the stimulation frequency causes prolonged contraction, first unfused, later fused, tetani, which generate much higher contractile forces than twitch contractions. This phenomenon is based (i) on the different durations of action potentials and twitch-time and (ii) on the much higher myoplasmic free Ca^{2+} reached during tetanic, as compared to twitch, contractions.

EC-coupling in smooth muscle differs from that in striated muscle types. Although a rise of free Ca^{2+} also triggers smooth muscle contraction, the Ca^{2+} receptor is calmodulin rather than troponin C. The Ca^{2+} -calmodulin complex activates myosin light chain kinase (MLCK), which specifically phosphorylates the 20 kD myosin light chain of smooth muscle myosin. Phosphorylation of the 20 kD myosin light chain allows actin activation of the myosin cross-bridge, leading to cyclic actin interaction and contraction (5).

Skeletal muscle fiber types mature according to their innervation with fast (high firing frequency) or slow (low firing frequency) alpha-motor neurons during maturation and develop distinct contractile and energetic features. Thus slow-twitch fibers (long twitch time) have high concentrations of enzymes for oxidative energy metabolism, while fast-twitch fibers (short twitch time) have high glycolytic activities. They express different myosin isoenzymes with low and high ATPase activities respectively. Furthermore, fast fibers have higher fiber volumes of the T-system SR as well as a higher Ca^{2+} uptake rate by the SR than the slow fibers. Fast and slow muscle types have similar mechanical efficiencies of around 35%. The distinct firing pattern of fast and slow motor neurons causes different spatial and temporal changes in intracellular free Ca^{2+} , which may be transduced into distinct changes in gene transcription and fiber maturation.

Thus the level of calcineurin activation as well as the activation of the different multifunctional Ca^{2+} -calmodulin-dependent kinases depend on the temporal mode of muscle activation (6).

Clinical Relevance

Use and disuse of muscle cause hypertrophy and atrophy, which cause increase and decrease in muscle strength respectively. Stimulation of the expression of insulin-growth factor-1 (IGF-1) and its paracrine/autocrine release upon increased muscle work, as well as the calcineurin-NFAT pathway, turned out to be important in skeletal muscle (7).

Hereditary muscle diseases comprise myotonic and dystrophic diseases (8). In particular mutations of the Na^+ and Cl^- channels cause myotonic syndromes. Mutations of the genes coding for β -myosin heavy chain, tropomyosin and nebulin are associated with central core disease and nemaline myopathy respectively. Myotonic dystrophy is associated with a trinucleotide (CTG) repeat on chromosome 19 coding for a protein kinase. Mutations of genes coding for proteins of the dystrophin-glycoprotein complex are associated with several forms of muscle dystrophy (MD), e.g. Duchenne and Becker MD (dystrophin mutations), limb-girdle MD (sarcoglycan mutations), congenital MD and Emery-Dreifuss MD (laminin mutations). Mutations of the Ca^{2+} channel and the ryanodine receptor are associated with malignant hyperthermia. Deterioration of the nicotinic acetylcholine receptor – by mutations or autoantibodies – causes myasthenic syndromes. Disorders of glycogenolysis, e.g. McArdle’s and Tarui’s diseases are caused by mutation of glycogen phosphorylase and phosphofructokinase, respectively.

References

1. Geeves MA, Holmes KC (1999) Structural mechanism of muscle contraction. *Ann Rev Biochem* 68:687–728
2. Huxley H, Hanson J (1954) Changes in the cross-striations of muscle during contraction and stretch and their structural interpretation. *Nature* 173:973–976
3. Huxley AF (1957) Muscle structure and theories of contraction. *Prog Biophys Biophys Chem* 7:255–318
4. Squire JM, Morris EP (1998) A new look at thin filament regulation in vertebrate skeletal muscle. *FASEB J* 12: 761–771
5. Somlyo AP, Somlyo AV (2000) Signal transduction by G-proteins, rho-kinase and protein phosphatase to smooth muscle and non-muscle myosin II. *J Physiol* 522:177–185
6. Buonanno A, Fields RD (1999) Gene regulation by patterned electrical activity during neural and skeletal muscle development. *Curr Opin Neurobiol* 9:110–120
7. Olson EN, Williams RS (2000) Remodeling muscles with calcineurin. *BioEssays* 22:510–519
8. Emery AE (2002) Muscular dystrophy into the new millennium. *Neuromuscul Disord* 12:343–349

Muscle Development

HANS-HENNING ARNOLD

Cell and Molecular Biology, Technical University of Braunschweig, Braunschweig, Germany
h.arnold@tu-bs.de

Definition

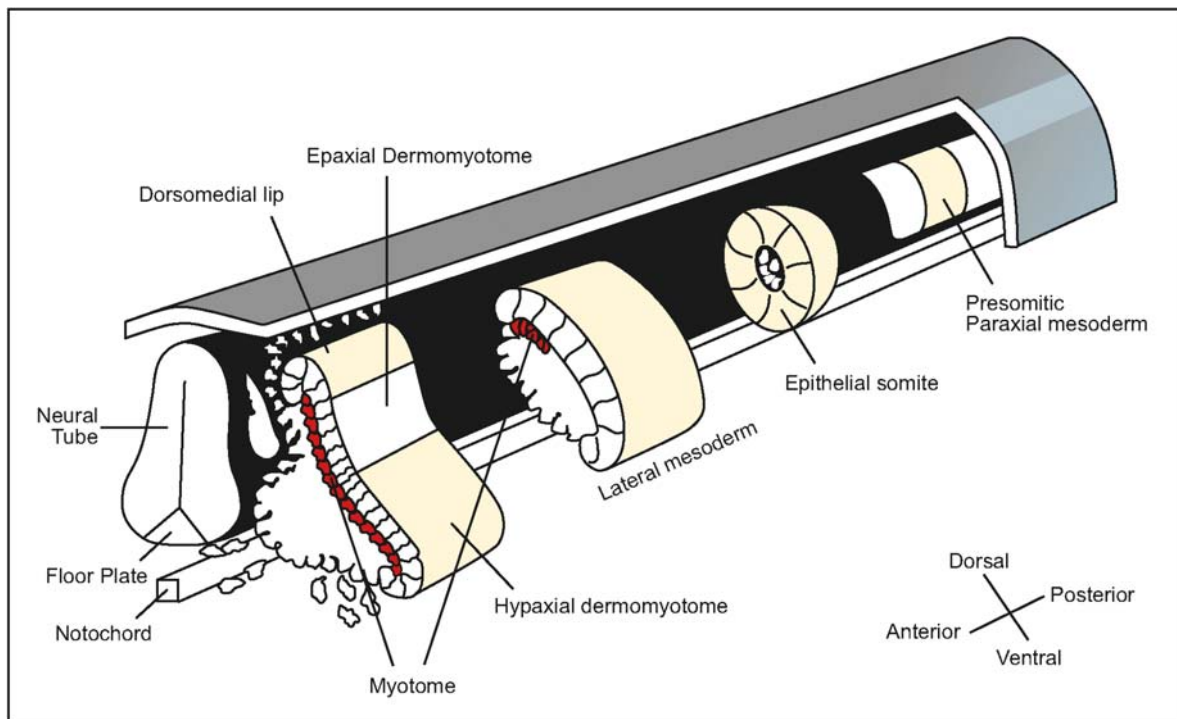
This essay focuses on development of the skeletal musculature in vertebrates and disregards the formation of other muscles, such as the ►myocardium and smooth muscles, as they follow entirely different developmental strategies. Specification of distinct cell types during embryogenesis is usually triggered by signals from adjacent tissues leading to a cascade of gene activations that ultimately generate the differentiated cellular phenotype. Development of the skeletal musculature conforms to this scheme and constitutes a paradigm for molecular mechanisms underlying ►lineage determination of cells as a first step and their entrance into the specific ►differentiation programme as a second step. Formation of skeletal muscle involves myogenic ►commitment of distinct mesodermal cell populations to generate proliferating myoblasts which eventually withdraw from the ►cell cycle, switch on the expression of muscle-specific genes, as for instance those of the contractile apparatus, and fuse into large multinucleated myotubes or fibers, the functional units of muscle. Cell culture models have been established which allow the study of both the commitment and the differentiation of muscle cells *in vitro*. A gene family encoding four myogenic regulatory factors (MRFs) is essential to control the recruitment of progenitor cells to the myogenic fate as well as their subsequent differentiation into contractile muscle fibers. The MRFs, referred to as Myf5, myogenin, Mrf4 and MyoD, are ►transcription factors of the ►basic helix-loop-helix type, which are expressed specifically in the muscle-forming regions of the embryo with overlapping but distinct spatiotemporal expression profiles. Genetically altered mouse models have helped to determine the developmental roles of individual MRF genes and other important factors in myogenesis.

Characteristics

All Skeletal Muscles are Derived from Defined Mesodermal Progenitor Cell Populations in the Vertebrate Embryo

The skeletal musculature, although subdivided into multiple individual muscles, constitutes by far the largest organ in the body. In vertebrates, most skeletal muscles including those of the trunk, limbs, tongue,

and ►diaphragm are derived from somites which are transient embryonic structures generated from unsegmented ►paraxial mesoderm on both sides of the ►neural tube in ►rostrocaudal sequence by progressive segmentation into epithelialized tissue blocks (Fig. 1). Head muscles, such as the jaw muscles, arise from paraxial head mesoderm and prechordal mesoderm. The myogenic precursors of facial muscles are initially present in the ►branchial arches. Shortly after somite formation, the ventral half of somites transforms to the mesenchymal sclerotome which gives rise to the axial skeleton of the vertebral column and ribs. The dorsal part of somites, adjacent to the surface ectoderm, develops the dermomyotome harboring progenitor cells for skeletal muscle and dermis. Beginning in cranial somites (at embryonic day 8.0 in the mouse embryo) and subsequently proceeding along the body axis, cells of the dermomyotome segregate first from the dorsomedial lip (adjacent to the dorsal neural tube) and somewhat later also from the ventrolateral lip to form the myotome underneath the still epithelial dermatome (Fig. 1). Although the myotome, the first differentiated skeletal muscle in the embryo, appears to be a homogeneous cell layer with no obvious histological distinction, it is clearly patterned along its dorso-ventral axis. While the dorsal myotome, frequently referred to as the epaxial myotome, later develops into the intrinsic back muscles, the ventral, hypaxial myotome grows into the somitic bud, which contributes the progenitors of intercostal and ventral body wall muscles. In somites at limb levels muscle progenitor cells delaminate from the ventral (hypaxial) dermomyotome and migrate to the limb buds providing the muscle masses for arms and legs. Migratory muscle precursors also delaminate from occipital/cervical somites to form the ►hypoglossal cord, which contributes to tongue and pharyngeal muscles and possibly to the diaphragm. Most of this information has been obtained from avian embryos but a growing body of evidence supports the notion that the same principles apply to myogenesis during mammalian development (for review, see 1). In summary, primary myogenesis in the vertebrate embryo is initiated in three distinct territories, the branchial arches, somites and limb buds in a defined temporal sequence (Fig. 2). Following the formation of skeletal muscle in the embryo, further growth and functional maturation of muscle fibers ensues during secondary myogenesis in the fetus by recruitment and differentiation of committed myoblasts. Innervation of muscles by motor nerves also occurs during this developmental period and contributes to the specification of ►slow and fast muscle fiber types. Postnatally, enlargement and regeneration of the skeletal musculature involves ►satellite cells, a population of myogenic stem cells that resides under the basal lamina of muscle fibers in a quiescent state,



Muscle Development. Figure 1 Schematic representation of somite development. Removal of surface ectoderm reveals the consecutive steps of somitogenesis. Note that the most differentiated somite consisting of mesenchymal sclerotome, myotome (in red), and epithelial dermomyotome is the oldest at the anterior end, while the paraxial mesoderm at the posterior end is still unsegmented. First myotomal cells appear at the dorsomedial lip.

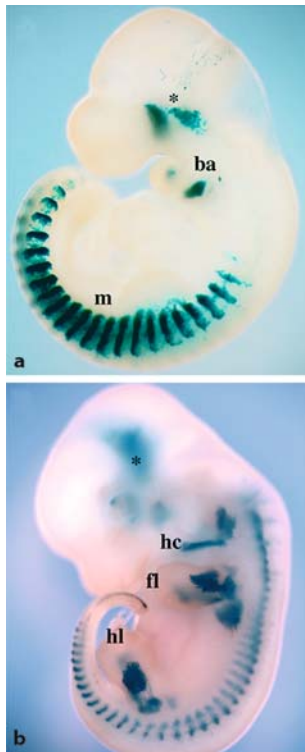
until they are activated through signals evoked by muscle activity or tissue damage.

Molecular Interactions

The MRF Family of Four Basic Helix-Loop-Helix Transcription Factors Dictates Skeletal Muscle Development

Skeletal muscle development requires the coordinated expression of numerous transcription factors but the myogenic regulatory factors (MRFs) are particularly important and play key roles in directing mesodermal progenitor cells to the myogenic lineage and mediating their subsequent differentiation into contractile muscle (for review, see 2). The potent myogenic activity of all four MRFs was first recognized when they were introduced individually into a variety of non-muscle cells thereby converting these cells to the muscle phenotype in an astonishingly dominant fashion. *In vivo*, however, the roles of individual MRFs in myogenesis appear to be more refined. While Myf5 and MyoD are absolutely essential to specify myoblasts from mesodermal progenitors, the two other members of this transcription factor family, myogenin and Mrf4, are necessary for differentiation of myoblasts. This has been concluded from [targeted gene disruptions](#) of individual MRF genes in mice and from pair wise

combinations in double mutant mice. During embryogenesis, Myf5 is the first of the MRF proteins to be expressed in branchial arches and somites, followed 0.5 and 1 day later by myogenin and Mrf4, respectively. MyoD expression in somites starts approximately 2.5 days after Myf5. Expression of all MRFs is progressively down-regulated during fetal muscle development and ceases postnatally with the exception of Mrf4, which increases in the fetus and persists in adult skeletal muscles. Consistent with these temporal patterns, Myf5-deficient mice fail to form myotomes, lack muscle precursor cells, and do not activate the downstream MRF genes, *myogenin* and *Mrf4*. Interestingly, MyoD expression starts at the correct developmental time in head and limbs and somewhat late in somites of *Myf5* mutants. Embryonic myogenesis proceeds almost normally from this point onwards. This suggests that MyoD can also recruit mesodermal cells to the muscle fate in a Myf5-independent pathway. In line with this apparent [genetic redundancy](#), double mutants lacking both Myf5 and MyoD are completely unable to develop any skeletal muscle precursor cells. In these double mutant embryos, cells that switch on the *Myf5* gene (without making the protein) and would normally be destined to become muscle are misplaced into adjacent territories and take on alternative fates,



Muscle Development. Figure 2 The skeletal muscle-forming regions in the mouse embryo marked by β -galactosidase activity (blue) from a *Myf5-nlacZ* transgene. (a) Myogenic cells are seen in branchial arches (ba) and myotomes (m) of a mouse embryo at embryonic day 9.5. (b) At day 12.5 myogenic cells are also present in fore (fl)- and hindlimbs (hl), and in hypoglossal cord (hc). The asterisks denote regions of ectopic transgene activity in head mesenchyme.

such as cartilage or dermis. These observations clearly indicate the unique roles of *Myf5* and *MyoD* genes in muscle cell determination, which is further supported by the demonstrated capability of both transcription factors to remodel [chromatin](#) and open up gene loci that will be transcribed during differentiation. In contrast to *Myf5* and *MyoD*, the *myogenin* gene is not required for the determination of muscle precursor cells but it is necessary for the differentiation of myoblasts. Mouse mutants lacking myogenin accumulate myoblasts in all muscle forming regions of the embryo but these cells fail to undergo fusion and do not express muscle-specific genes, such as those of the [sarcomeres](#). Although some evidence suggests that *Mrf4* is also involved in the terminal differentiation of myoblasts, its precise function in muscle development is not clear due to complicating cis-effects on the adjacent *Myf5* gene in all three available *Mrf4* mutant alleles. In particular, *Mrf4* gene function in postnatal muscle cannot be assessed in these mouse mutants as they usually die at birth. The fact that *Myf5* and *MyoD*

are expressed in undifferentiated but committed myoblasts, whereas myogenin and *Mrf4* expression only starts with the onset of differentiation or later (*Mrf4*) is entirely consistent with the genetic evidence for their respective roles in muscle development.

Members of the Myocyte Enhancer Factor 2 (MEF2) Family of Transcription Factors Collaborate with MRFs in Myoblast Differentiation

In order to activate the transcription of downstream target genes, all MRFs form heterodimers by combining with a related class of basic-helix-loop-helix transcription factors that are present in almost every cell type including myoblasts. These heterodimeric complexes then bind to a distinct DNA sequence motif, the so-called E-box, present in control regions of many muscle-specific genes, and activate their transcription. Interestingly, E-boxes are not only present in muscle genes but also in others which are not transcribed in muscle, raising the question as to how selective transactivation might be achieved. At least part of the specificity for muscle genes appears to be associated with a conserved peptide motif within the helix-loop-helix domain of MRFs that is required for the combinatorial function of MRFs with members of the MEF2 family in activating transcription and myogenesis (3). Indeed, MEF2 factors are essential for muscle differentiation in *Drosophila* as well as in mouse and the majority of skeletal muscle genes depend on both MRF and MEF2 family members for activation. Consistent with this notion, muscle-specific enhancers frequently contain E-boxes and MEF2-binding sites in close proximity allowing for cooperative interactions of the bound protein complexes. It is reasonable to assume that MRFs can also recruit MEF2 factors to the enhancer in the absence of a MEF2 DNA binding site and *vice versa*, since both proteins have been shown to form complexes *via* direct protein-protein interactions. Undoubtedly, formation of myoblasts and the correct execution of the complete differentiation programme of muscle cells involves multiple transcription factors in addition to MEF2 and MRF family members, but the available evidence strongly suggests that these proteins are absolutely crucial in a cascade of events that ultimately results in the fully functional skeletal muscle cell. In particular *Myf5* and *MyoD*, possibly in a parallel pathway, appear to act at the nodal point where the prospective myogenic fate of mesodermal progenitor cells is to be decided. It is interesting to note that the same players responsible for myogenesis in the embryo are reactivated in satellite cells, when these stem cells are challenged to make new muscle in the adult. It should also be mentioned that the transcription factors involved are subject to molecular modifications by several signaling pathways, which regulate their biological activity in various physiological and

pathological conditions, such as muscle aging, regeneration and hypertrophy (4).

Regulatory Mechanisms

The Transcriptional Circuitry Underlying Skeletal Muscle Development

The highly conserved structure of the four MRF proteins not only argues for a common ancestral gene but also for their similar, albeit not identical, biochemical properties and biological activities as transcription factors. Indeed, replacing the coding exons of the *myogenin* gene by the *Mrf4* coding sequence restores myoblast fusion indicating that Mrf4 protein can compensate for the loss of myogenin, if expressed at the correct developmental time and location. This then implies that the decision where and when muscle develops in the embryo is at least in part determined by the distinct spatial and temporal expression patterns of individual MRFs rather than by their intrinsic biochemical differences. Although precise molecular mechanisms of MRF gene regulations are not clear, cis-acting control elements are beginning to emerge. Since Myf5 functions at the top of the myogenic cascade marking the initiation of skeletal muscle development in the embryo and during muscle regeneration in the adult, its regulation represents the decisive step for myogenesis. *Myf5* and *Mrf4* genes are closely linked in one genetic locus and key regulatory elements are widely dispersed in a modular fashion over 140 kb of sequence. The complex set of control elements orchestrates Myf5 expression in various phases at each site of myogenesis under the influence of signals produced by adjacent tissues and reveals an unexpected heterogeneity of myogenic precursor cell populations. This heterogeneity in origin may be related to physiological differences of muscle groups in the adult organism and their differential susceptibility to myopathies. Distinct enhancers for Myf5 expression in branchial arches, limb buds and different domains in somites have been identified, but most cognate transcription factors are still unknown (5 and references therein). The early epaxial enhancer driving the earliest, transient wave of Myf5 expression in the epaxial dermomyotome contains a binding-site for the transcription factor Gli that mediates regulation by the sonic hedgehog (Shh) signal emanating from the notochord and floorplate. This is the first example of direct signal integration in *Myf5* gene expression and muscle development and supports the notion that axial structures (notochord and neural tube) are necessary for epaxial myogenesis, as shown by surgical removal in the chicken embryo. *Wnt* signals produced in the neural tube and dorsal surface ectoderm also promote myogenesis and have been implicated as positive regulators of Myf5 and MyoD

expression. BMP signals produced in lateral plate mesoderm inhibit MRF expression and may help to delimit the lateral border of skeletal muscle formation.

The *myogenin* gene appears to be a direct target of Myf5 consistent with a cascade model of sequential MRF gene activation. In addition to Myf5, MEF2 family members and the homeobox transcription factors *sine oculis* (*Six1* and *4*) are also required for correct spatiotemporal myogenin expression. Essential binding sites for these transcription factors are present in the promoter proximal region of the *myogenin* gene. Following the initial activation, myogenin probably stabilizes the differentiated muscle phenotype by maintaining its own expression and further stimulating expression of MEF2. Circumstantial evidence suggests that the *Mrf4* gene may also be controlled directly or indirectly by Myf5 and/or MyoD.

The *MyoD* gene is regulated by two collaborating enhancers located approximately 5 and 20 kb upstream of the promoter. Several essential E-box consensus binding sites have been identified within the core enhancer sequence, one of which seems to specifically mediate Myf5-dependent expression of MyoD in somites. Interestingly, in mouse double mutants lacking both Myf5 and the transcription factor Pax3 that is expressed in the dermomyotome, *MyoD* is not activated in somites and muscles are missing in the trunk. This is in contrast to single mutations in either *Myf5* or *Pax3*, which do not prevent MyoD expression. These observations provide evidence that *Pax3* and *Myf5* define two distinct myogenic pathways in somites and *MyoD* acts genetically downstream of these genes. However, no Pax3 binding sites have yet been detected in the *MyoD* gene enhancers suggesting that Pax3 action may be indirect. Alternatively, deficiency of Pax3 might affect cell proliferation, survival, or short-range migration, which may lead to loss of the myogenic progenitor cell population.

Myogenic Determination Occurs Independently in Somites and Limb Buds

Skeletal muscle development in the limbs differs from that in trunk considerably (for review, see 6). While most muscles of the body develop from the myotome, the muscle groups in limbs derive from the lateral dermomyotome and do not go through the myotomal intermediate. These myogenic progenitors delaminate from the epithelial dermomyotome and migrate as single cells into the limb buds without expressing Myf5 or MyoD until they have reached their destination in the limb mesenchyme. This independent regulation of MRF expression reflects distinct enhancer activities, which have been identified approximately 60 and 20 kb upstream of the *Myf5* and *MyoD* promoters,

respectively. While BMPs and Shh signals present in limb buds have been shown to affect positioning, proliferation and differentiation of muscle precursor cells in limb buds, transcription factors, which may integrate these signals in MRF gene control, are still elusive. The homeobox transcription factors Lbx (related to ladybird in *Drosophila*) and Mox2, which have no role in the myotome, are expressed in the migratory progenitor cells and mutations of these genes strongly reduce or eliminate Myf5 expression and muscle development in limbs. MyoD expression in the appendages, however, is controlled neither by Mox2 nor by Myf5. In contrast to myotomal myogenesis, Pax3 is absolutely essential for muscle development in limbs. Its role here appears to be at least partly mediated by the tyrosine-kinase-receptor c-met, which together with its ligand scatter factor (SF) is essential for the release and migration of muscle progenitors from the dermomyotome. Later, Pax3 also seems to be involved in controlling the homeostasis between proliferation and differentiation of myoblasts, possibly in response to graded BMP signals. The requirement of FGF signaling for correct limb muscle differentiation has been appreciated recently. The regulatory influence of other signaling molecules that are expressed plentifully in limb buds must await further investigations. In conclusion, the elucidation of molecular details underlying muscle development constitutes a prime example of complex developmental strategies of cell type specification and organ development. Moreover, this knowledge may help to establish rational approaches to therapeutic intervention in muscle-wasting diseases.

References

1. Buckingham M (2001) Skeletal muscle formation in vertebrates. *Curr Opin Genet Dev* 11:440–448
2. Arnold HH, Braun T (2000) Genetics of muscle determination and development. *Curr Top Dev Biol* 48:129–164
3. Black BL, Olson EN (1998) Transcriptional control of muscle development by myocyte enhancer factor-2 (MEF2) proteins. *Annu Rev Cell Dev Biol* 14:167–196
4. Winn N, Musar A, Rosenthal N (2002) Insulin-like growth factor isoforms in skeletal muscle aging, regeneration, and disease. *Cold Spring Harb Symp Quant Biol* 67:507–518
5. Buchberger A, Nomokonova N, Arnold HH (2003) Myf5 expression in somites and limb buds of mouse embryos is controlled by two distinct distal enhancer activities. *Development* 130:3297–307
6. Buckingham M, Bajard L, Chang T et al (2003) The formation of skeletal muscle: from somite to limb. *J Anat* 202:59–68

Muscular Dystrophy

Definition

Muscular dystrophy comprises of a group of inherited degenerative muscular myopathies involving progressive wasting of muscle tissue.

► Repeat Expansion Diseases

Muscular Hypotonia

Definition

Muscular hypotonia designates weakness of muscles.

► Prader Willi and Angelman Syndromes

Mutagen

Definition

Mutagen stands for an agent that causes mutations.

► Large-Scale ENU Mutagenesis in Mice

Mutagenesis

Definition

Mutagenesis comprises of the manipulation of genes by inducing random or specific mutations. ENU mutagenesis – inducing point mutations by alkylation of nucleic bases using N-ethyl-N-nitrosourea. Irradiation mutagenesis – inducing point mutations and deletions using gamma-irradiation. Insertional mutagenesis – inducing insertions by random or targeted integration of DNA sequences into the genome.

► Large-Scale ENU Mutagenesis in Mice

► Mutagenesis Approaches in the Zebrafish

Mutagenesis Approaches in Medaka

HIROYUKI TAKEDA

Department of Biological Sciences, Graduate School of Science, University of Tokyo, Tokyo, Japan
htakeda@biol.s.u-tokyo.ac.jp

Definition

Medaka, *Oryzias latipes* (order Beloniformes), is a small, egg-laying freshwater teleost fish (Fig. 1) found in brooks and rice paddies in eastern Asia, primarily in Japan. Breeding medaka in small basins has been one of the traditional hobbies in Japan, often seen in old Japanese prints known as Ukiyo-e of the Edo period (17th to 18th centuries). There has been a long history of medaka for culture and science in Japan (most information on medaka is found at the Medaka Fish Homepage, ►<http://biol1.bio.nagoya-u.ac.jp:8000/>).

Characteristics

The adult fish are about 3 cm long and the female lays a cluster of eggs (10–30 eggs) every day. The embryos develop externally, and the embryos and chorions are transparent. Thus, the phenotype of the mutants can be easily evaluated. The embryos hatch 7 days after fertilization at 25°C and they grow to sexual maturity in 2–2.5 months. These biological features, shared by zebrafish, have made the medaka a suitable model animal, particularly for developmental biology and genetics. There are several advantages of medaka over the commonly used zebrafish. First, the medaka genome is relatively smaller in size (800–1000 Mb), only one-third of the human genome and less than half



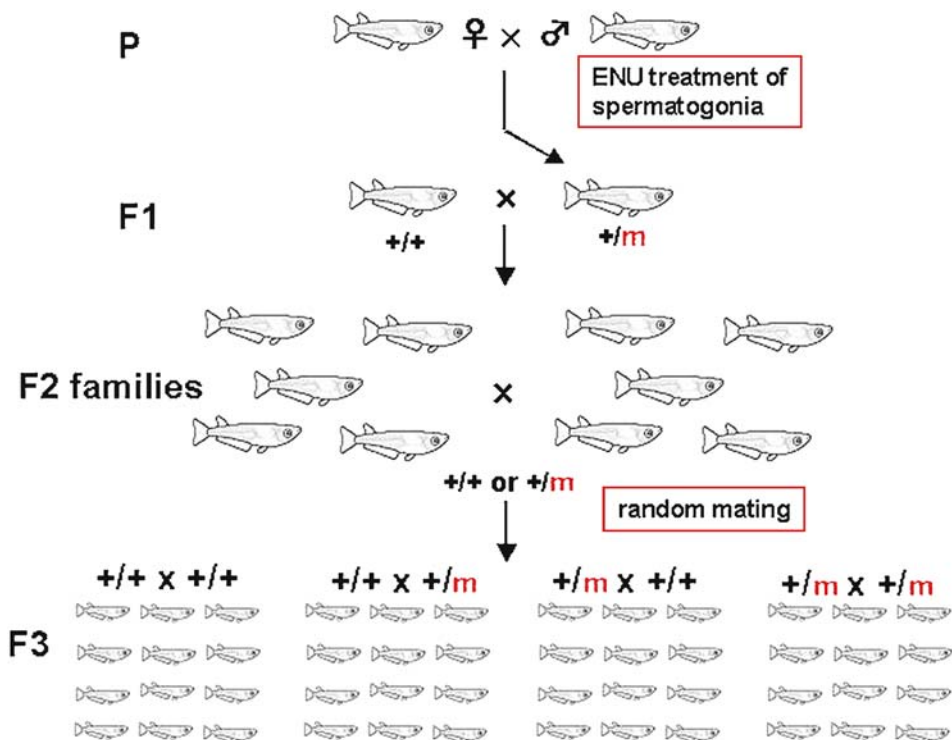
Mutagenesis Approaches in Medaka. Figure 1 Adult medaka. Female of orange-red variety with fertilized eggs. A cluster of eggs remains attached to the belly of the female for some hours after spawning at dawn. A photograph is a courtesy of Drs. Kiyoshi Naruse and Kenichi Ijiri (University of Tokyo).

the size of the zebrafish genome. Second, there are highly polymorphic ►[inbred strains](#) available in medaka, which can be used for mutagenesis screening and genetic mapping. Finally, unlike zebrafish (tropical fish), medaka (temperate fish) embryos can develop at wider range of temperature, 6–40°C, which facilitates isolation of ►[temperature-sensitive mutants](#).

Early comparative analysis revealed a genome-wide gene duplication in the ancestry of teleosts during evolution. Indeed, zebrafish and medaka have seven ►[hox clusters](#) while mouse has four. It has been thought that gene duplication could play a vital role in providing new genetic material for natural selection to act upon. Duplicated genes have been subjected to non-functionalization, neo-functionalization and sub-functionalization, providing a great genetic diversity in fish species. From the evolutionary point of view, zebrafish, closely related to carp, is a relatively old fish, 110–160 million years apart from the modern fish, medaka. Considering this long evolutionary distance between the two fish, one would expect that medaka and zebrafish would have a different repertoire of genes, and thus a different spectrum of mutant phenotypes would be obtained. Together with recent development of genomic resources (see below), the medaka becomes an attractive model system complementary to zebrafish. For a recent review, see (1).

Mutagenesis Approaches

The genetic screen is a means to completely analyze the genome, but based primarily on phenotype rather than sequence. The ability to perform genetic screens is one of the main reasons for using teleost fishes as vertebrate models. Early work on medaka has identified more than 90 spontaneous mutants, most of which were collected by Dr. Hideo Tomita and kept at the Laboratory of Freshwater Fish Stocks, Bioscience Center, Nagoya University (►<http://biol1.bio.nagoya-u.ac.jp:8000/mutant.html>). They are mostly pigmentation mutants. However, some organ-specific mutants are included, such as *pectoral fin-less* (*pl*), *eyeless* (*el*) and *double anal fin* (*Da*, a homozygous adult with ventralized caudal structures in the dorsal region) (2). In addition to spontaneous mutants, mutagenesis screens using ►[ENU](#) (N-nitroso-N-ethyl urea) or X-rays followed by three-generation crosses were carried out recently, in the expectation of novel mutants not found in zebrafish (3, 4). As with zebrafish, in mutagenesis screens, male fish mutagenized with ENU or X-rays were outcrossed to wild type females and the F1 offspring were used to establish F2 families. F2 siblings were intercrossed and the F3 progeny were scored at certain stages after fertilization for morphological alterations (Fig. 2). For the induction of



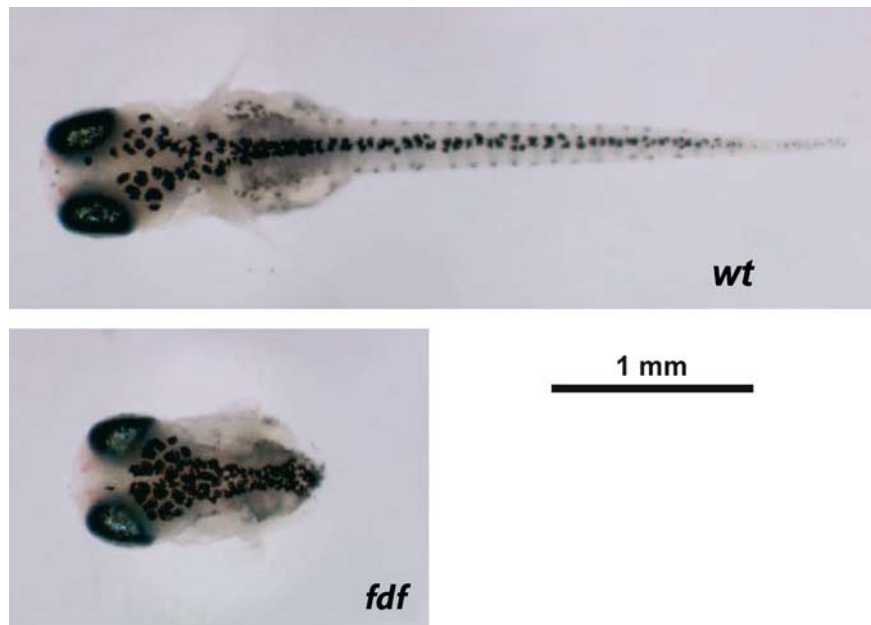
Mutagenesis Approaches in Medaka. Figure 2 Family inbreeding scheme for mutagenesis screening. Males mutagenized with ENU are used to produce F1 carriers. These are intercrossed to produce F2 families. Crosses between F2 siblings produce F3 clutches that can be screened for a new phenotype. Homozygosity for newly induced mutations occurs in 1/4 of the embryos in 1/4 of the F3 crosses.

mutations affecting single genes, chemical mutagens are generally preferred over X-rays. X-rays result in large deletions or chromosomal rearrangements and induce recoverable mutations at a much lower rate than ENU. Furthermore, ENU, which induces point mutations, offers the possibility of hypomorphic mutations in addition to complete null mutations, which might be useful for functional analysis of genes of interest. The mutagenesis rate using ENU is very high, around 1×10^{-3} . Thus, for most of the screens, ENU mutagenesis is the method of choice. Loosli et al (4) first reported the systemic mutagenesis approach to isolate embryonic-lethal developmental mutants in medaka, focusing on eye and brain phenotypes. In any screens reported so far, ENU induced mutations were recovered efficiently in medaka, suggesting that medaka, as well as zebrafish, is a suitable vertebrate for mutagenesis approaches. As expected, some of the medaka mutants exhibit phenotypes similar to those recovered from zebrafish screens. However, some, such as *Da*, *kepke* (the trunk and tail fail to develop) (1) and *headfish* (similar to *kepke*, Fig. 3), are remarkably different,

indicating a non-overlapping spectrum of embryonic mutant phenotypes between medaka and zebrafish.

Inbred Medaka Strains

The first step towards identifying genes responsible for mutant phenotypes involves the establishment of linkages between the mutated locus and a polymorphic marker. The wild populations of the medaka in Japan are divided into two major genetic groups, the northern and southern populations. The two populations are different in many morphological, behavioral and genetic characteristics. In spite of these, they normally cross and produce hybrid offspring. Sequencing comparisons of orthologous loci revealed ▶single nucleotide polymorphisms (SNPs) between the two populations at a frequency of 1% in ▶exons and 3% in ▶introns. Considering the 1–2% difference between human and ape, the SNPs between the two medaka populations are quite high. So far, twelve inbred medaka strains from various origins including wild populations have been established, e.g. HNI and Kaga from the northern populations and Hd-rR, Cab and



Mutagenesis Approaches in Medaka. Figure 3 Medaka tail-less mutant, *headfish* (*fdf*). Dorsal views of 9-day *wild type* and *fdf* embryos are shown. In *fdf*, the trunk and tail fail to development, which has not been reported in any zebrafish mutants.

AA2 from the southern populations. Unlike zebrafish, most of the medaka inbred lines are highly fertile and can be used for mutagenesis screening. Furthermore, since the inbred lines are highly polymorphic to those derived from other populations, they are extremely useful for genetic mapping of genes and mutants. Most of the inbred strains, produced by Dr. Hyodo Taguchi, have been maintained at the National Institute of Radiological Sciences (3) ([▶http://biol1.bio.nagoya-u.ac.jp:8000/RadMut.html](http://biol1.bio.nagoya-u.ac.jp:8000/RadMut.html)).

Large-Scale Analyses of mRNAs

The [▶expression sequence tag](#) (EST) approach is powerful in massive cloning of cDNAs as well as in large-scale characterization of cDNA sequences for functional genomics. Several groups have attempted large-scale isolation of ESTs from medaka embryos, adult, liver, ovary and so on. At March 2003, about 100,000 entries of medaka ESTs can be found in public databases. Some of the ESTs are being mapped to the [▶linkage map](#) and *in situ* hybridization screening with isolated ESTs is also underway ([▶http://medaka.lab.nig.ac.jp](http://medaka.lab.nig.ac.jp); [▶http://www.embl-heidelberg.de/mepd](http://www.embl-heidelberg.de/mepd)).

Linkage Map and Positional Cloning

Genetic linkage maps are very effective tools for identification of mutant responsible genes and for

comparative and evolutionary genomics. Usually, markers used for linkage maps are phenotypic traits, ESTs, [▶random amplification of polymorphic DNA markers](#) (RAPDs), [▶amplified fragment length polymorphic markers](#) (AFLPs) and [▶microsatellite markers](#). A medaka linkage map first described by Aida (1921) (5) demonstrated that the male-determining factor (Y locus) was linked with the gene controlling carotinoid deposition in [▶xanthophores](#) (R locus). In 2000, a detailed and genome-wide linkage map of medaka was constructed by the use of 633 markers (488 AFLPs, 28 RAPD markers and so on) (6). The linkage map utilized the high degree of polymorphism between the two inbred strains, AA2 and HNI, derived from southern and northern populations, respectively, and was constructed with a reference typing DNA panel from 39 cell lines derived from back cross progeny. The number of linkage groups was 24, which is the same as medaka's haploid chromosome number. Since markers in this linkage map are mostly derived from non-coding regions or anonymous, a large-scale mapping of randomly selected ESTs is now in progress. By the end of 2004, about 2,000 ESTs will be mapped onto the linkage map (K. Naruse and HT, a personal communication), and if the total genome size of medaka is 800 Mb, the estimated density of genetic markers would be around

470 kb per marker, which would provide reliable anchor points for ►positional cloning. Part of these mapping data is available at ►http://mbase.bioweb.ne.jp/~dclust/medaka_top.html. The data also demonstrate that there are large regions of ►conserved synteny between chromosomes of zebrafish, medaka and human (K. Naruse et al, a personal communication).

The detailed linkage map and high quality of BAC libraries that are prerequisite for identification of mutated genes are now available in the medaka system. Using these genetic resources, successful positional cloning for several spontaneous medaka mutants was reported in 2001. The first one was *b* locus; a mutation in *b* reduces melanin content in medaka and *b* turned out to encode a novel transporter protein that was also found in mouse (7). Second and third were *el* (eyeless) and *rs* (reduced scale) that encode the ►homeodomain protein Rx3 and the ectodysplasin-A receptor (EDAR), respectively (8, 9). More recently, the sex-determining region of the medaka Y chromosome has been positionally cloned (10). The region harbors a transcription factor, *DMY* that contains the highly conserved ►DM domain and plays a critical role in testis development. So far, *DMY* is the only gene identified as a non-mammalian counterpart of ►*Sry*, which is required for mammalian male development. Positional cloning of ENU-induced medaka mutants is now being carried out in several laboratories.

Genome Sequencing Project

To further promote mutant-based research and comparative genomics of medaka, a genome-sequencing project for medaka has started at the Academia Sequencing Center of the National Institute of Genetics (Mishima, Japan) in collaboration with the medaka research community. The project is supported by a Grant-in-Aid for Scientific Research on Priority Areas "Genome Science" from the Ministry of Education, Culture, Sports, Science and Technology of Japan. The strain for sequencing is a southern inbred strain, Hd-rR, and sequencing is being conducted by the whole-genome shotgun strategy. By May 2005, there was a 10× coverage of sequences completed (►<http://medaka.utgenome.org/>). In parallel, end sequencing of BAC clones and construction of BAC contig and mapping of SNPs and ESTs onto the linkage map are now underway in several laboratories in Japan and Germany (►<http://medaka.dsp.jst.go.jp/MGI/>). Hopefully, in the near future, all these data will be integrated to produce a high quality draft sequence of the medaka genome.

Clinical Relevance

It is already clear that fish orthologs exist for most human genomes and that there are large regions of conserved synteny between fish and human. Furthermore, unlike *Drosophila* and *C. elegans*, fish has all of the tissues (except for lung) afflicted by common human diseases and fish organs are functionally and morphologically similar to those of the human. As compared with targeting of genes for disruption in mice, ENU-induced mutations in fish are more relevant to human disease, because ENU often causes partial loss-of-function or hypomorphic mutations. Many common human diseases are not due to null mutations. Therefore, it is hoped that mutagenesis approaches in fish might unveil mechanisms and pathways directly relevant to human diseases and therapy. In fact, in mutagenesis screens of medaka, many organ-specific mutants (e.g. heart, liver, thyroid and so on) are being recovered (3, 4; HT, a personal communication; see also (11)). With the growing pool of medaka genomic resources that should facilitate rapid identification of the candidate genes, these medaka mutants could serve as good models for human diseases.

References

1. Wittbrodt J, Shima A, Scharl M (2002) Medaka – a model organism from the Far East. *Nat Rev Genet* 3:53–64
2. Ozato K, Wakamatsu Y (1994) Developmental Genetics of Medaka. *Dev Growth Differ* 36:437–443
3. Ishikawa Y (2000) Medakafish as a model system for vertebrate developmental genetics. *Bioessays* 22:487–495
4. Loosli F, Koster RW, Carl M et al (2000) A genetic screen for mutations affecting embryonic development in medaka fish (*Oryzias latipes*). *Mech Dev* 97:133–139
5. Aida T (1921) On the inheritance of color in a freshwater fish *Aplocheilichthys latipes* Temmick and Schlegel, with special reference to sex-linked inheritance. *Genetics* 6:554–573
6. Naruse K, Fukamachi S, Mitani H et al (2000) A detailed linkage map of medaka, *Oryzias latipes*: comparative genomics and genome evolution. *Genetics* 154:1773–1784
7. Fukamachi S, Shimada A, Shima A (2001) Mutations in the gene encoding *B*, a novel transporter protein, reduce melanin content in medaka. *Nat Genet* 28:381–385
8. Loosli F, Winkler S, Burgdorf C et al (2001) Medaka eyeless is the key factor linking retinal determination and eye growth. *Development* 128:4035–4044
9. Kondo S, Kuwahara Y, Kondo M et al (2001) The medaka *rs-3* locus required for scale development encodes ectodysplasin-A receptor. *Curr Biol* 11:1202–1206
10. Matsuda M, Nagahama Y, Shinomiya A et al (2002) *DMY* is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature* 417:559–563
11. Medaka Special Issue (2004) *Mech Dev*, vol 121

Mutagenesis Approaches in the Zebrafish

ALEXANDER GRUNDT, WOLFGANG ROTTBAUER
University Clinic Heidelberg, Inner Medicine III,
Heidelberg, Germany
wolfgang.rottbauer@med.uni-heidelberg.de

Definitions

► **Mutagenesis** is defined as a directed or random alteration in DNA at a specific site in the genome of an organism, to study the mutation's effect on organ morphology or function. Different mutagenesis approaches such as chemical, insertional and irradiation induced mutagenesis, have been used successfully for functional genome analysis in zebrafish.

Significant progress has been made over the last decade in sequencing the genomes of several vertebrates, including human, mouse, and zebrafish. In parallel, new approaches have been developed for elucidating the function of all genes in the genomes. The ► **genetic screen** is a tool to completely analyze the genome, but based principally on ► **phenotype** rather than sequence. By identifying unitary defects associated with ► **mutations**, and in a manner complete enough to reflect a large proportion of all genes involved in the target process, the screen can provide the logical elements of a complex biological process.

There are two opposing concepts for characterizing gene functions namely “reverse” and “forward” genetics. The idea of reverse genetics is to identify gene functions (phenotypes) by genetically manipulating known genes using transgenesis, gene “knock-out” and gene “knock-in” targeting techniques. In contrast, in forward genetic screens phenotypes are induced by random mutagenesis first, and then the gene mutation causing the phenotype is identified by ► **positional cloning** or ► **gene trap** approaches.

Characteristics

Large-scale forward genetic screens have been one of the most efficient methods for identifying novel genes and pathways controlling the development of embryonic form and function and inherited diseases. The first genetic saturation screens based on phenotypes have been performed in invertebrate model systems such as *Caenorhabditis elegans* and *Drosophila melanogaster*, resulting in an impressively large number of newly identified genes and pathways essential for fundamental processes in biology. Initially, in vertebrates such as *Mus musculus*, only small-scale screens were performed, since the number of animals that must

be raised, maintained and screened to achieve saturation in a screen is some millions. This problem has been solved by the successful use of zebrafish (*Danio rerio*) as a vertebrate model system. Because of its small size (from about 3 cm) and modest character the zebrafish can be easily managed in large numbers in laboratory environment. Several advantages, such as a short generation time of only three months and the fact that mature females lay hundreds of eggs at weekly intervals, mean that zebrafish are particularly suitable for genetic studies. The externally fertilized zebrafish embryos allow unobstructed access to the researcher and develop extraordinarily rapidly. Furthermore, simple visual screening of the synchronously developing and optically clear zebrafish embryos in the first 5 days after fertilization can reveal essential mutations in the development of most of the major organ systems. In addition, mutations can be induced highly efficiently throughout the zebrafish genome. So far, many small-scale forward screens and two large-scale screens have been carried out in zebrafish leading to the isolation of hundreds of unique and organ-specific phenotypes. Together, the two large-scale Boston and Tübingen screens (1, 2), starting from about 300 ► **ENU** (N-ethyl-N-nitrosourea) treated founder males, involved raising more than 5,000 F₂ families and analyzing more than 6,000 mutagenized genomes and selected more than 2,000 ► **recessively** inherited new development mutants for characterization. Today, basically three different approaches are used for forward genetic screens in zebrafish, ► **ENU mutagenesis**, ► **insertional mutagenesis** and irradiation mutagenesis. Reverse genetic screens in zebrafish were started recently.

ENU Mutagenesis Screens

Most of the zebrafish large-scale forward genetic screens were performed with the chemical mutagen N-ethyl-N-nitrosourea (ENU), which induces genome-wide ► **point mutations** in pre-meiotic ► **germ cells** (spermatogonia) in a random pattern. ENU transfers ethyl groups to O⁶-guanine, O⁴-thymine, N³-adenine and the PO₄-backbone of DNA. The repair of mismatches during replication and transcription resulting from lost H-bonds and steric interferences can lead to the exchange of the non-alkylated base, resulting usually in hundreds of point mutations throughout the targeted genome. To achieve an optimal ratio of mutagenicity to toxicity in zebrafish, several pre-screens have been performed, varying ENU concentrations and incubation times and using four different pigmentation loci as markers for the mutagenicity. In the most efficient protocol, adult male fish are mutagenized by exposing them three times for one hour periods within one week to 3 mM ENU in water. The estimated frequency of independent mutations per locus per gamete is $0.9\text{--}1.3 \times 10^{-3}$ and 2.1×10^{-6} per

base pair (bp) per gamete. In other words, about 1300 mutagenized genomes must be analyzed for one locus (gene) to find one mutation induced by ENU. Three weeks after ENU treatment, males are crossed to wild type females at weekly intervals and the progeny (F_1 founder) originating from mutagenized pre-meiotic germ cells are raised. The F_1 founder males are crossed again to wild type females. A random sibling cross of the F_2 progeny leads to 25% ▶**homozygous** wild type embryos, 50% ▶**heterozygous** embryos and 25% homozygous mutant embryos for each point mutation induced by ENU (Fig. 1A) (1, 2).

Haploid Screens

Phenotype detection of a recessive allele within a F_3 generation requires many crosses in F_2 families. To avoid the cumbersome step of screening thousands of progeny for recessive mutations in conventional F_2 screens (Fig. 1A), methods have been devised to uncover recessive alleles in a single generation by exploiting the ability to create ▶**haploid** or homozygous ▶**diploid** zebrafish embryos suitable for genetic screening. Zebrafish can live up to 3 days post-fertilization as haploid organisms. Recessive alleles in F_1 fish are therefore exposed within a generation, streamlining the screens for both size and time. In a haploid screen, female F_1 fish (derived from the cross between a wild type female and an ENU-mutagenized male) are squeezed gently to release their eggs, which are fertilized with ultraviolet (UV)-treated sperm to generate haploid embryos. UV treatment breaks down the parental DNA, without affecting the ability of sperm to activate eggs. A haploid clutch derived from a heterozygous female will contain 50% mutant and 50% wild-type embryos (Fig. 1B). The disadvantage of this approach is the induction of abnormal phenotypes in embryonic development usually on the second day post-fertilization, due to the single set of chromosomes rather than an ENU induced point mutation. Although haploids have the same body plan as diploids, they are shorter, contain more and smaller cells than diploids and can have other morphological defects that may complicate screening.

Homozygous Diploid Screens Using Early Pressure or Heat Shock Treatment

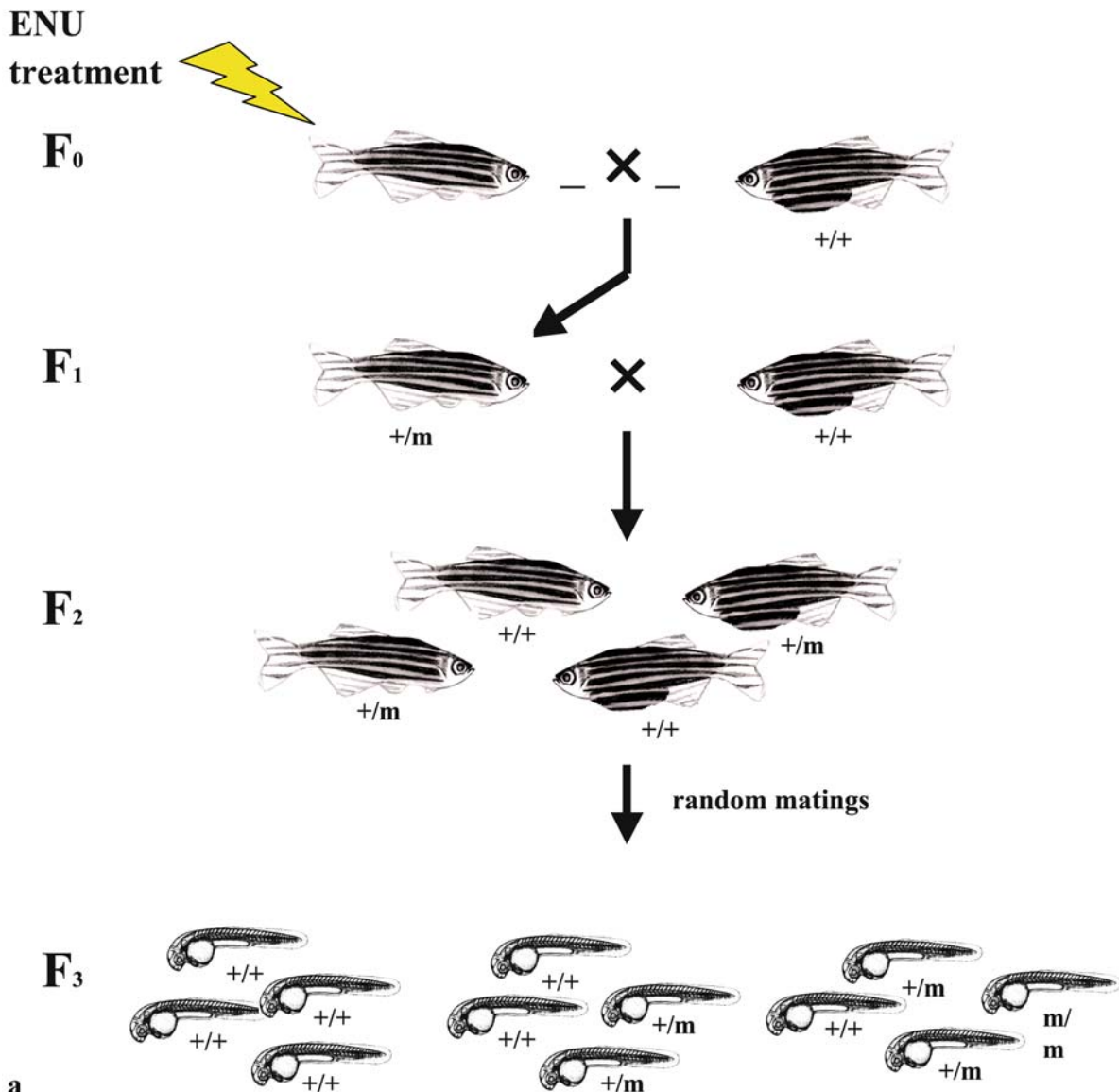
Eggs extracted from a female have completed meiosis I (the separation of homologous chromosomes) during ovulation, and initiate meiosis II (the separation of sister chromatids) on fertilization. ▶**Early pressure** (EP) applied to embryos a few minutes post-fertilization breaks down the meiotic spindle and the egg maintains both sister chromatids. Subsequently, eggs undergo their first mitosis as diploids, with two sets of maternal chromosomes. By contrast, ▶**heat-shock** (HS) treatment of eggs during the first cell-cycle inhibits the

first mitotic division. Eggs activated with UV-treated sperm enter the first mitotic division as haploids, abort mitosis by HS and enter the second mitotic division directly as diploids. In meiosis, recombination and chiasma interference occur between homologous chromosomes during alignment as tetrads. Because of this, there is on average one single crossover event per chromosome arm. Therefore, embryos that are derived from EP treatment will be homozygous for loci that are proximal to the crossover event that occurred at meiosis I and heterozygous for loci that are distal to it. Similar to haploid clutches, a gynogenetic diploid clutch that is derived from a heterozygous female and generated by HS will contain 50% mutant and 50% wild type embryos. However, in contrast to EP treatment, embryos generated by HS are homozygous at all loci. Therefore, they would be preferable to embryos generated by EP for use in genetic screens. However, embryos generated by HS usually have a poor viability of 10–20%.

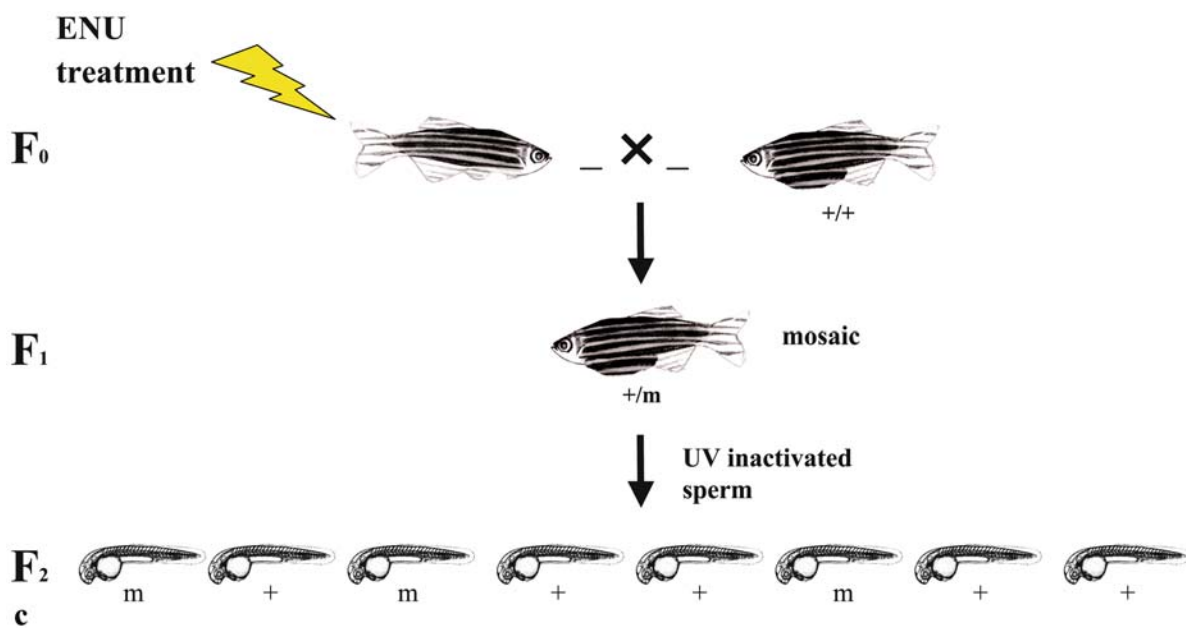
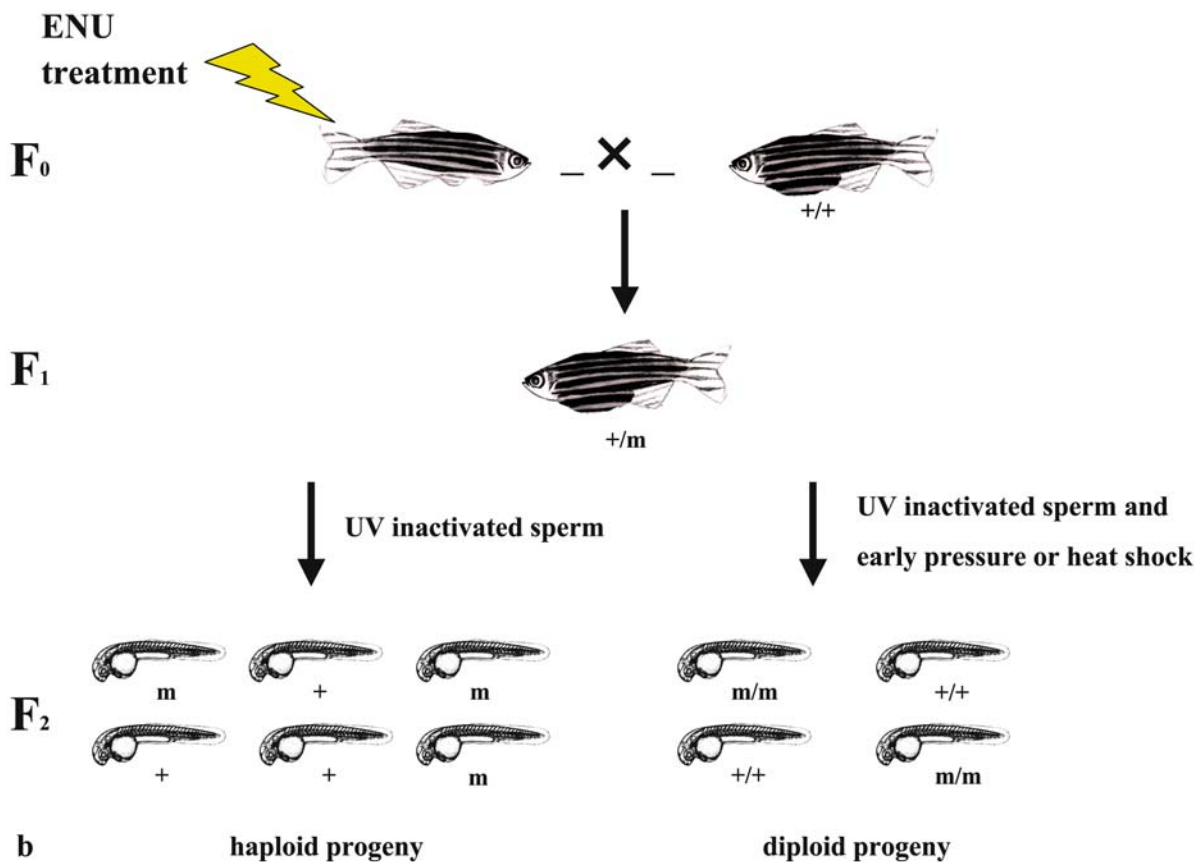
Mosaic Screens

Standard ENU mutagenesis involves mutagenizing pre-meiotic germ cells in adult males, then keeping these fish for several weeks to fix the mutation through several rounds of DNA replication before the development of mature sperm cells. As ENU alkylates the bases on only one DNA strand, mutations in the ENU-treated post-meiotic germ cells become fixed only during cell divisions after fertilization. Therefore, if the mutagenized males are bred immediately or early after exposure to ENU, the resulting F_1 generation will be genetically ▶**mosaic** for numerous ENU induced mutations, increasing the efficiency of identifying specific mutations due to an estimated tenfold greater mutational load in the mosaic F_1 generation. It is thought, that mosaic fish are able to carry a greater mutational load than non-mosaic heterozygous fish, because wild type cells within all tissues compensate for the heterozygous mutant cells. However, several mutations in haploid embryos and the unpredictable phenotypic representation within a clutch can complicate screening and allele recovery in mosaic screens. If the F_1 females are mosaic, a variable proportion of the F_2 haploid embryos are mutant (Fig. 1C). The exact number of mutant embryos depends upon the proportion of the F_1 germ-line cells that is heterozygous for the mutation. Nonetheless, mosaic screening has proven to be a robust and rapid method to identify several unique cardiac mutations (3).

Over the last decade, ENU saturation mutagenesis has proven to be one of the most efficient tools for forward genetics in zebrafish, revealing at least 2000 unique genes required for proper embryonic development. Several of these mutated genes have already been cloned, which assists in the dissection of the gene



Mutagenesis Approaches in the Zebrafish. Figure 1 ENU based crossing schemes to breed homozygous mutant zebrafish. (a) The classical three generation breeding scheme. ENU treated males are crossed to wild type females. The F₁ males are also crossed to wild type females and the F₂ progeny paired with each other. In the F₃ generation, recessive mutations are revealed. m, mutation, +, wild-type. (b) Generation of haploids and gynogenetic diploids. ENU treated males are crossed to wild type females. Eggs from F₁ females are released by gently squeezing and fertilized with UV inactivated sperm resulting in haploid embryos. Treating the fertilized eggs in the one-cell-stage with hydrostatic pressure or exposing them to heat results in diploid embryos. (c) Mosaic screen. Zebrafish males are crossed to wild type females immediately or early after exposure to ENU, resulting in a F₁ generation that will be genetically mosaic for numerous ENU induced mutations. Eggs from F₁ females are released by squeezing and fertilized with UV-inactivated sperm. Since the F₁ females are mosaic, a variable proportion of the F₂ haploid embryos will be mutant, depending upon the proportion of the F₁ germ-line cells that is heterozygous for the mutation.

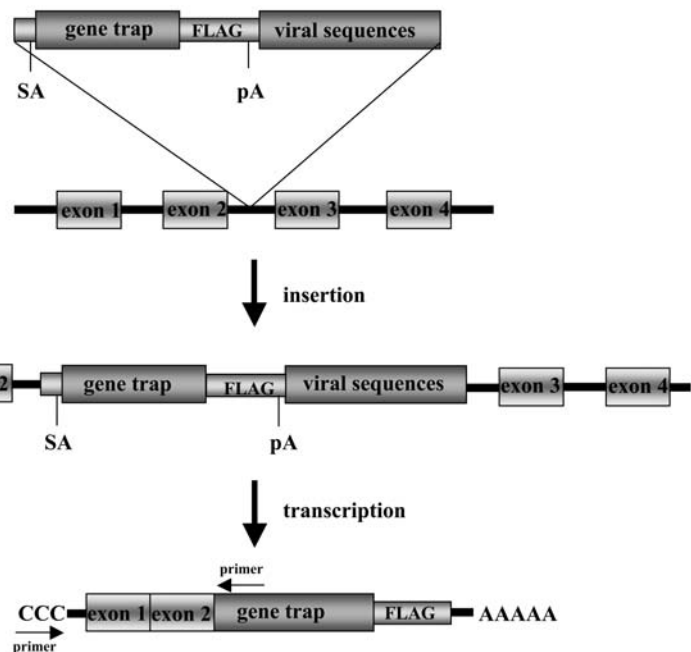


gene trap targeting construct

gene before the insertion

mutagenized gene

mRNA after insertion



Mutagenesis Approaches in the Zebrafish. Figure 2 Insertional mutagenesis using a gene trap approach. The viral insertion construct includes a gene trap (GT) at the beginning of the vector. The GT contains a splicing acceptor element (SA), the sequences for FLAG epitopes in all reading frames and a polyadenylation element (pA) to stop the transcription. After successful insertion into intronic sequences, as well as proper transcription and mRNA splicing, the gene trap is fused to upstream exon sequence of the targeted gene. This exonic sequence can be recovered by PCR using a CCC-anchor and GT specific primers (displayed as arrows) in the 5'-RACE. Using an anti-FLAG antibody the tissue specific expression pattern of the targeted gene can be evaluated by whole-mount immunostaining of zebrafish embryos or Western blot analysis of defined tissues.

networks that control early development and disease. Positional cloning of mutations generated with ENU has been highly successful, although it remains very laborious and expensive. However, the development of an excellent genomic infrastructure, including several meiotic and radiation hybrid (RH) maps of the zebrafish genome, where thousands of simple sequence length polymorphisms (SSLPs) and expressed sequence tags (ESTs) have been mapped to the 25 zebrafish chromosomes, as well as the ongoing sequencing of the zebrafish genome by the Sanger Centre, UK greatly facilitate positional cloning efforts.

Insertional Mutagenesis Screens

An alternative approach to chemical mutagenesis is insertional mutagenesis using gene trap (GT) approaches. Viral DNA containing a GT construct is injected into zebrafish embryos and then integrates randomly into the genome, thereby disrupting expressed sequence and serving as an anchor for primers to identify the disrupted gene by cloning and sequencing. The GT, a short open reading frame or reporter gene such as β -galactosidase, placed at the beginning of the viral construct provides all important splicing

elements, such as splicing acceptor and pA-element, to fuse to the last **exon** upstream of the targeted **intron** (Fig. 2). Therefore, the resulting mRNA, consisting of residues of the targeted gene and the GT, can readily be cloned, sequenced and identified by 5'-rapid amplification of cDNA ends (5'-RACE). Using a GT with FLAG epitopes in all three reading frames, the tissue specific expression pattern of the targeted gene can be estimated using *in situ* hybridization or Western blot analysis.

At first, two Moloney murine leukemia virus (MLV) based pseudotyped viruses with envelopes containing the vesicular stomatitis virus (VSV) G-protein, SFG and NK, which are able to infect zebrafish cells and integrate stable proviral DNA into the zebrafish genome, were used to infect mouse 3T3 fibroblasts to get high titer virus stocks. The virus stocks (about 10^9 cfu/ml) were microinjected into zebrafish embryos at the 512-2048 cell stage. After 24 h 50–80% of the embryos survived. The majority of these grew to adulthood. To detect germ-line transmission of proviral DNA, the injected fish were mated and the F₁ progeny tested for the presence of proviral sequences by PCR. The percentage of transgenic offspring from the

mosaic SFG founders was on average 24% and from NK founders 13%. The SFG founders transmit on average eleven proviral ►insertions to their F₁ progeny. However, only about one in 85 proviral insertions inactivates a gene essential for embryonic development. Therefore, to improve the efficiency of large-scale insertional screens, fish with multiple insertions were used to identify developmental mutation. For this purpose a new viral construct with a unidirectional VSV G-pseudotyped SFG-nLacZ virus as backbone was developed and transiently transfected into zebrafish cell lines A7 and F5 by lipofection (7.5 µg pCMV-G and 50 µl Lipofectamine). Using virus stocks from the most efficient cell clones, the average number of proviral insertions in the surviving F₀ founders was up to 25 and the F₁ gametes contained eight or more insertions with a 57% survival rate after 2 days. About 96% of all gametes contained multiple insertions (4). In one large-scale viral insertion screen, nine recessive and two dominant embryonic lethal mutations were obtained from an estimated 760 insertions in 92 F₂ families. In contrast to ENU mutagenesis, the targeting efficiency using viral insertion is about nine times lower. However, the targeted genes can be identified quicker and more easily by inverse PCR or 5'-Race. Estimating the size of the zebrafish genome to be approximately 1.6×10^9 bp, about 500,000 randomly placed integrations have to be analyzed in a saturation screen. A screening of about 100,000 randomly placed integrations resulting in approximately 1000 essential loci is in progress since 1999 in the Nancy Hopkins Laboratory at the MIT, Boston, USA. (5)

Ionizing Radiation Mutagenesis Screens

The mutagen selected to manipulate the founder fish affects the number and types of mutations that are passed on to subsequent generations. As outlined above, ENU induces mainly point mutations throughout the whole genome, although the mutation frequency can vary widely between loci. Ethyl methane sulphonate (EMS), although a potent mutagen in *Drosophila*, is much less potent in zebrafish. Ionizing radiation, such as γ -rays and X-rays, also induce genetic alterations, which can vary from point mutations through large genomic ►deletions to translocation events. Large genomic deletions and translocations, however, can alter more than one gene and generate several phenotypes in a clutch of embryos, complicating a robust and reliable identification of genes and mutations associated with phenotypes and gene functions.

Earlier forward genetic screens in zebrafish used predominantly γ -irradiation as mutagen. In these small-scale screens, haploid mutant embryos derived

from heterozygous F₁ females crossed to UV-treated sperm were used to identify specific phenotypes. Although elegant in principle, as described above, haploid screens have the disadvantage that specific phenotypes caused by a single mutation must be recognized against a high background of retarded and often abnormally developing embryos after the second day post-fertilization. Altogether, less than 30 loci essential for embryonic development have been identified so far using irradiation screens in zebrafish (6).

Reverse Genetic Screens – Target Selected Mutagenesis

In reverse genetics, gene functions (phenotypes) are usually identified by genetically manipulating known genes using transgenesis, gene “knock-out” and gene “knock-in” targeting techniques. Another reverse genetic approach, now used in zebrafish is target-selected mutagenesis. In this approach, male zebrafish are ENU mutagenized and paired with wild type females. Frozen sperm and genomic DNA from several thousand heterozygous F₁ founders is then collected and cryo-preserved, providing a mutant library from which mutations of almost all zebrafish genes should be identifiable. Now these samples can be evaluated for heterozygous mutations in any gene of interest. Using the ENU mutagenesis protocol described above, recessive embryonic lethal mutations per locus occur at a frequency of one in 1600 mutagenized genomes, so that the cDNA of 1600 founders has to be analyzed to find one gene knockout mutation in the gene of interest. The method described was used successfully to identify mutations in the zebrafish *rag1* gene (7). With the establishment of this technology, the zebrafish is now the second vertebrate model organism in which reverse genetics can be applied on a large scale. The advantage of multiple alleles, including partial loss-of-function mutants, may establish target-selected mutagenesis as the preferred approach.

Clinical Relevance

Using forward large- and small-scale mutagenesis approaches in zebrafish thousands of unique phenotypes have been identified, resulting in more than a 1000 different genetic loci mutagenized, of which an impressive number have already been cloned and characterized. Improving the efficiency of insertional mutagenesis using better integration systems and of positional cloning by finishing the zebrafish genome sequencing project, the functional aspect of all estimated 35,000 vertebrate genes might be revealed in the near future. Many of the identified zebrafish mutants already serve as animal models for human genetic diseases.

References

1. Driever W, Solnica-Krezel L, Schier AF et al (1996) A genetic screen for mutations affecting embryogenesis in zebrafish. *Development* 123:37–46
2. Haffter P, Nusslein-Volhard C (1996) Large scale genetics in a small vertebrate, the zebrafish. *Int J Dev Biol* 40:221–227
3. Solnica-Krezel L, Schier AF, Driever W (1994) Efficient recovery of ENU-induced mutations from the zebrafish germline. *Genetics* 136:1401–1420
4. Chen W, Burgess S, Golling G et al (2002) High-throughput selection of retrovirus producer cell lines leads to markedly improved efficiency of germ line-transmissible insertions in zebra fish. *J Virol* 76:2192–2198
5. Amsterdam A, Burgess S, Golling G et al (1999) A large-scale insertional mutagenesis screen in zebrafish. *Genes Dev* 13:2713–2724
6. Fritz A, Rozowski M, Walker C et al (1996) Identification of selected gamma-ray induced deficiencies in zebrafish using multiplex polymerase chain reaction. *Genetics* 144:1735–1745
7. Wienholds E, Schulte-Merker S, Walderich B et al (2002) Target-selected inactivation of the zebrafish *rag1* gene. *Science* 297:99–102

Mutagenesis Approaches in Yeast

MICHAEL BREITENBACH, PETER LAUN
Department of Cell Biology, University of Salzburg,
Austria
michael.breitenbach@sbg.ac.at

Definition

The single-celled, free-living eukaryotic microorganism, *S. cerevisiae*, has been studied as a convenient model system for molecular biology and cell biology since the 1970s. It is hardly necessary to repeat here the advantages of the yeast system over many other biological model systems, such as an extremely short generation time, growth under many well-controlled conditions, well-researched metabolism, a detailed genetic map based on ►meiotic mapping and a fully developed sexual cycle, ease of transformation with plasmids and with foreign DNA, and so on. Yeast is perhaps the organism, in which the construction of ►knockout and ►knock-in mutations is most highly developed and most easily done. Based on these techniques, a huge number of advanced techniques for functional analysis of genes has been developed. All this became even more important when the first eukaryotic chromosome sequence (chromosome III of *S. cerevisiae*, (1) and the first eukaryotic genome sequence [that of *S. cerevisiae*, in April, 1996 (2)]

became known. It was relatively easy to annotate the yeast genome, as evidenced by the fact that a fully annotated yeast genome sequence was made publicly available shortly after sequencing was finished (2, 3). Problems remain, but they are minute compared to the problems of annotating the human genome. At the time of assembling and annotating the yeast genome sequence, more than half of the genes that were clearly recognizable were of unknown function, constituting a major challenge to the yeast community to analyze these unknown functions. Most surprisingly, many human genes that had been identified by the study of human hereditary disease, shared close homologs with yeast. Even several years later, with many more human “disease genes” discovered, still at least 40% of single-gene determinants of human heritable diseases find homologues in yeast (4). Given these facts, it is obvious why one would like to study these genes in yeast. It is possible to reveal functional data that are not easily obtained by the study of human cells and which could help to understand the ►pathomechanism of the disease, which is a prerequisite for any medical intervention. Examples will be given (see below and Table 1). However, this is not the only scenario in which the yeast system is useful. It is also possible to introduce authentic human genes (that have no yeast homologs) into yeast cells in order to study their behavior. A prominent example is huntingtin, another the apoptosis inducers of the bax family. And finally, we want to mention that the yeast system is also successfully being used to define virulence factors of fungal pathogens and drug targets that are needed to develop new antifungal antibiotics.

Characteristics

The roughly 6000 protein-encoding genes on the *S. cerevisiae* genome consist to a large degree of gene pairs, which are functionally redundant or have similar but slightly different functions, thus reflecting a gene duplication event that occurred relatively recently in evolutionary history. For this reason, the basic set of genes needed for a free-living single-celled organism is somewhat smaller than 6000. Roughly, the yeast genes can be divided into a large group of eukaryotic ►housekeeping genes defined as genes for which homologs are clearly recognizable in all eukaryotes for which genome sequences have been determined and a smaller group of fungi-specific genes, which are needed for the specific lifestyle of fungal organisms. Homologs of such genes are found in close relatives of *S. cerevisiae* that have been sequenced (*S. mikatae*, *S. bayanus*, *S. paradoxus*, *S. kudriavzevii*) or even in distantly related fungi (*S. pombe*), but not in plants or animals. Studying the non-coding regions of homologous genes in closely related species is a powerful tool for annotation that could also be applied to the

Mutagenesis Approaches in Yeast. Table 1 Examples of functional gene analysis in yeast

Expression profiling	Spellman, Sherlock et al., <i>Mol Biol Cell</i> 9:3273–97 (1998) Hughes, Marton et al., <i>Cell</i> 102:109–26 (2000) Werner-Washburne, Wylie et al., <i>Genome Res</i> 12:1564–73 (2002)
Bioinformatics	Barry, Fichant et al., <i>Yeast</i> 12:1163–78 (1996) Wood, Rutherford et al., <i>Comp Funct Genom</i> 2:143–154 (2001) Oshiro, Wodicka et al., <i>Genome Res</i> 12:1210–20 (2002) Cliften, Sudarsanam et al., <i>Science</i> 301:71–6 (2003) Kellis, Patterson et al., <i>Nature</i> 423:241–54 (2003) Salzberg, <i>Nature</i> 423:233–4 (2003)
Virulence factors of pathogenic fungi	Breitenbach, Crameri et al., <i>Chem Immunol</i> 81:5–9 (2002)
Large protein complexes	Gavin, Bosche et al., <i>Nature</i> 415:141–7 (2002) Ho, Gruhler et al., <i>Nature</i> 415:180–3 (2002)
Two hybrid analysis	Legrain and Selig, <i>FEBS Lett</i> 480:32–6 (2000) Schwikowski, Uetz et al., <i>Nat Biotechnol</i> 18:1257–61 (2000) Ito, Chiba et al., <i>Proc Natl Acad Sci USA</i> 98:4569–74 (2001) Walhout and Vidal, <i>Methods</i> 24:297–306 (2001)
Conditional alleles	Gari, Piedrafita et al., <i>Yeast</i> 13:837–48 (1997) Bogengruber, Briza et al., <i>FEM Yeast Res</i> 3:35–43 (2003)
GFP fusions	Cormack, <i>Curr Opin Microbiol</i> 1:406–10 (1998) Belmont, <i>Trends Cell Biol</i> 11:250–7 (2001) Guedener and Hegemann, http://mips.gsf.de/proj/yeast/info/tools/hegemann/gfp.html (2003)
Huntingtin	Krobitsch and Lindquist, <i>Proc Natl Acad Sci USA</i> 97:1589–94 (2000)
Bax-family involved in apoptosis	Frohlich and Madeo, <i>FEBS Lett</i> 473:6–9 (2000)
Mismatch repair pathways and colon carcinoma; nucleotide excision repair and Xeroderma pigmentosum	Friedberg, Walker and Siede, <i>DNA repair and mutagenesis</i> , ASM Press, Washington, DC (1995)
Mitochondrial disease	Wallace, <i>Science</i> 283:1482–8 (1999)
Frataxin	Babcock, de Silva et al., <i>Science</i> 276:1709–12 (1997) Foury and Cazzalini, <i>FEBS Lett</i> 411:373–7 (1997) Koutnikova, Campuzano et al., <i>Nat Genet</i> 16:345–51 (1997) Rotig, de Lonlay et al., <i>Nat Genet</i> 17:215–7 (1997) Pandolfo, <i>Neuromuscul Disord</i> 8:409–15 (1998) Miranda, Santos et al., <i>FEBS Lett</i> 512:291–7 (2002) Rotig, Sidi et al., <i>Trends Mol Med</i> 8:221–4 (2002)

human genome once our closest biological relative, the chimpanzee, has been sequenced. This method is also a powerful tool to define functional regulatory motifs in the non-coding upstream regions of genes.

The disease gene homologs in the yeast genome obviously belong to the eukaryotic housekeeping genes. In many cases the disease-causing alleles are not nulls, but ►**hypomorphs**, as could be expected because the housekeeping genes often have vital functions.

The DNA repair pathways of eukaryotic cells are remarkably well conserved (despite some additional functional diversification in higher eukaryotes as compared to *E. coli* and *S. cerevisiae*) and their defects give rise to some interesting disease patterns. They will be mentioned only briefly because they are now general textbook knowledge. About ten years ago the eukaryotic ►**mismatch repair** pathway was elucidated by researchers studying repair in yeast and by human molecular geneticists studying the hereditary cancer predisposition HNPCC (hereditary non-polyposis colon cancer). These lines of research converged in a spectacular way. The human genes, *MSH2* and *MLH1*, were isolated by *positional cloning* based on family studies. They showed homology to the newly cloned yeast mismatch repair genes and the actual biochemical activity in mismatch repair was subsequently shown through yeast studies. Subsequent work yielded a coherent picture of the disease. Cancer predisposition is caused by the heterozygote state of the individual (therefore the pattern of inheritance is autosomal dominant, although the mutation is a loss of function mutation). The primary event causing cancer is the loss of heterozygosity at the locus involved which then leads to mutation accumulation in those cells and eventually to the cancer-causing mutations in human ►**proto-oncogenes** etc. Human mismatch repair recognizes frameshifts and short repeat elongation caused by slippage of DNA polymerase very well. This model nicely explains why the tumors, but not the patients present with an increase in ►**microsatellite** heterogeneities caused by un-repaired unstable repeats of tandem short nucleotide sequences. A very similar story could be told about nucleotide excision repair and the hereditary cancer predisposition and neurological diseases, Xeroderma pigmentosum (8 complementation groups) and Cockayne syndrome (2 complementation groups).

A large number of methods has been developed during the past 10 years, in particular in connection with the European and international joint effort to perform a genome wide functional analysis of yeast.

Part of this effort was to supply the yeast community with a collection of *bona fide* deletion mutants corresponding to all known yeast genes. This monumental task was accomplished by developing high-

throughput PCR methods for making precise deletions of the ORFs creating about 6000 heterozygous diploid deletion strains marked with ►*KanMX4* (G418 resistance) and with a “molecular barcode” (5). From the primary heterozygous deletion strains, haploids of both mating types (“sexes”) and the homozygous diploids were derived by classical meiotic analysis. In about 17% of the cases, the genes were haplo-lethal (“essential”) and the only strains available are the heterozygous diploids. On the other hand, haploinsufficiency on rich media (meaning that a heterozygous diploid would be unviable) is practically unknown in yeast, with the possible exception of the genes coding for the core histones. The huge collection of deletion strains is curated and is commercially available from EUROSCARF (►<http://www.uni-frankfurt.de/fb15/mikro/euroscarf/>). In addition, a set of cognate genes on plasmids and of disruption cassettes is available.

The bar codes are unique 20-mer sequences inserted together with *KanMX4*. One can apply selection of any kind to a mixture of all viable haploid deletion strains and analyze bulk DNA from the survivors on special microarrays containing the complementary barcodes. In this way, a relatively simple experiment can determine which genes on the yeast genome are needed for survival under certain conditions (e.g. heat shock), which is important functional information (5). About 600 strains that carry genes with affinity tags are available for the community, enabling researchers to purify and analyze large protein complexes by micro-bore HPLC/ion spray mass spectrometry. The global organization of metabolism (and its regulation) in many cases works through such large protein complexes (some of them larger than ribosomes!) and there can be little doubt that this is a general eukaryotic principle. The protein-protein interactions that build up the large protein complexes have also been mapped in a systematic way by applying various kinds of two-hybrid analyses.

Generally applicable methods have been developed to create useful conditional mutants for each essential gene of yeast. This is based on the bacterial tTA transactivator, which is expressed under the CMV promoter in yeast and by PCR-based insertional mutagenesis replacing part of the cognate promoter of the gene in question by the tetO₇-CYC1 hybrid promoter. Binding of the transactivator is abrogated by doxycycline so that gene expression can be titrated by adding doxycycline, which doesn't influence cellular metabolism at the concentration used, but switches off the target gene. A large number of phenotypic tests (limited only by the imagination of the experimenter) can be performed with the deletion mutants or with the conditional mutants just described. An indispensable part of functional analysis, which is generally applicable and easy to do, is the

determination of the intracellular location of the gene product. This is generally done by N-terminal or C-terminal GFP (green fluorescent protein) fusion and viewing the live cells under a laser-confocal microscope. Problems arising before the method optimally worked for yeast cells were overcome by creating several new forms of GFP by *in vitro* mutagenesis and by devising the appropriate yeast vectors that could express the fusion proteins from a plasmid or preferentially from a chromosomally integrated version. Quantum yield had to be improved by site-directed mutagenesis of this co-factor-independent naturally fluorescent protein, because it has been shown that many essential yeast proteins exist in the cell in only 1–10 copies whilst over-expression of the GFP fusion protein often leads to incorrect localization. Therefore it is preferable to express the protein in question under its own promoter on the chromosome. Differently colored versions of GFP were developed to enable double GFP-labeling and co-localization studies *in vivo*. Codon usage had to be optimized to ensure normal expression of the fusion protein in yeast. Finally, experience was gained pertaining to the biological function of GFP-fusion proteins. In many (but not all) cases the fusion proteins still complement the corresponding deletion mutation, which is of great advantage, because experiments can be performed in a cell expressing only the GFP fusion protein without competition by the wild type version.

With the advent of genome-wide transcriptional expression analysis on DNA microarrays and the corresponding new bioinformatics tools, functional analysis gained another dimension. Knowing not only the location of the protein but also the time course of expression under various physiological and developmental conditions can lead to a reliable prediction of function. An excellent example is the large group of ribosome-associated proteins involved in ribosome biogenesis, many of which are also needed for translational regulation and many of which were predicted by pattern matching techniques using microarray data. We have mentioned a few methods of major importance for functional analysis, but it must be stressed that many more exist. Many of these methods were pioneered in yeast, and in many cases the yeast cell serves as a testing ground for foreign (human) genes, because of the ease of experimentation in yeast.

Clinical Relevance

As an example we describe the successful functional analysis of frataxin, the yeast orthologue of a human disease gene *FRDA*. Friedreich's ataxia is a severe autosomal recessive disease that presents with neurological symptoms and frequently with dilated heart disease. Typically, the onset of the disease is in young adults and there is a very large phenotypic variability.

The disease is the most frequent of the ataxias with an incidence of about 2×10^{-5} in all studied populations. When the frataxin gene (*FRDA*) was identified through positional cloning in afflicted families in about 1996, the complete genome sequence of yeast had just been published and it became immediately clear that a very close homolog existed in yeast. Two competing groups undertook a functional analysis of the gene and, in fact, the yeast data made it very clear that frataxin is located in mitochondria in both systems. The defect leads to respiratory deficiency through loss of the mitochondrial genome. The mitochondria in the mutant cells over-accumulate iron, although the cellular iron is not increased. It is still unclear, whether frataxin is the mitochondrial iron transporter or maybe is in some other way involved in iron homeostasis. This question will also be answered easily with the help of the yeast system in the near future. Thus, although *FRDA* is a nuclear gene showing Mendelian inheritance, Friedreich's ataxia is a typical mitochondrial disease. It is striking in this example that all of the phenotypic peculiarities mentioned above can be explained by a pathomechanism based on known mitochondrial function. Iron over-accumulation in mitochondria leads to an excess of free radical or ROS (reactive oxygen species) production, which in turn leads to mitochondrial mutation and respiratory deficiency. The affected organs are those rich in mitochondria (heart) and those where other mitochondrial diseases cause typical symptoms (specific parts of the nervous system). The primary mutation identified in the diseased human allele is not a complete loss of function, but a hypomorph caused by trinucleotide GAA expansion in the first intron of the gene. The earlier the age of onset of the disease the larger is the number of the GAA repeats. The ▶intrinsic mutation leads to a diminished synthesis of an otherwise WT gene product. Typical patients are a somatic mosaic of different size trinucleotide repeats. The phenotypic variability is caused by this mosaicism and subsequently by the slow segregation of defective and still functional mitochondrial genomes during mitotic cell cycles, as in other mitochondrial diseases. As there are several thousand mitochondrial genomes per cell, this takes many cell generations. The gene deletion was studied in the robust yeast cell that can easily survive as a respiratory deficient "petite" cell. While the K.O. mouse devoid of frataxin showed embryo-lethality, the disease allele (230 GAA repeats) when introduced into the mouse, was stable there and caused no apparent phenotype.

We have begun to understand the pathomechanism of Friedreich's ataxia and "for the patients and their families the possibility of developing a rational treatment of the disease may turn out to be a legitimate hope". This treatment would be an intervention prohibiting the severe oxidative stress in the diseased cells.

References

1. Oliver SG, van der Aart QJ, Agostoni- Carbone ML et al (1992) The complete DNA sequence of yeast chromosome III. *Nature* 357:38–46
2. Goffeau A, Barrell BG, Bussey H et al (1996) Life with 6000 genes. *Science* 274:563–567
3. Mewes HW, Albermann K, Bahr M et al (1997) Overview of the yeast genome. *Nature* 387:7–65
4. Oliver SG (2002) Functional genomics: Lessons from yeast. *Philos Trans R Soc Lond B Biol Sci* 357:17–23
5. Giaever G, Chu AM, Ni L et al (2002) Functional profiling of the *Saccharomyces cerevisiae* genome. *Nature* 418:387–391

Mutant (Phenotype)

Definition

Mutant is a term that is applied to a gene or phenotype altered by mutation; and/or an individual carrying a mutation.

- Cell Polarity
- Large-Scale ENU Mutagenesis in Mice

Mutation

Definition

Mutation denotes a process through which genes undergo a structural change that potentially results in a change in the function of one or more genes. Examples include chromosome rearrangement, deletion, insertion, inversion, nucleotide repeat expansion, nucleotide substitution, point mutation (transition or transversion), translocation, and transposition.

Mutations arise by the action on the DNA of endogenous factors (e.g. DNA Polymerase Incorporation errors, DNA transaction mistakes, and oxidation) and exogenous factors (e.g. X-ray irradiation, UV irradiation, and chemical mutagens). Deletions or insertions of bases in protein-encoding DNA sequences lead to a totally different amino acid sequence, as downstream of the mutated position the reading frame is shifted (frameshift mutation). Base exchanges in the first and second codon positions frequently result in an exchange of the amino acid incorporated at the mutated codon. Base exchanges that lead to the incorporation of

chemically related amino acids are referred to as neutral mutations. There is no change in the amino acid sequence when a base exchange, mostly in the third codon position, creates a codon that specifies the same amino acid (silent mutation).

- Base Excision Repair
- DNA Helicases
- Fragile X Syndrome
- Huntington's Disease
- Large-Scale ENU Mutagenesis in Mice
- Mendelian Forms of Human Hypertension and Mechanisms of Disease
- Mutagenesis Approaches in the Zebrafish
- Protein Interaction Analysis: Suppressor Hunting
- Repeat Expansion Diseases
- Reverse Transcriptase
- Ribosomes
- Two-Hybrid System

Mutation Rate

Definition

Mutation rate refers to the rate at which new inheritable changes in DNA arise in the population in a given gene.

- Duchenne Muscular Dystrophy

Mutator Phenotype

Definition

Mutator phenotype defines the condition of having an elevated spontaneous mutation rate.

- Base Excision Repair

Myalgia

Definition

Myalgia means pain in one or several muscles.

- Mitochondrial Myopathies

Myelin

Definition

Myelin is a three-layered protein-lipid-protein complex that is produced by Schwann cells (PNS) and oligodendrocytes (CNS) and ensheaths many axons with multiple layers. Myelin insulates the axonal membrane, and speeds nerve conduction by restricting action potentials to the so-called nodes of Ranvier (interruptions at regular intervals in the myelin sheath that contain Na^+ and K^+ channels in high density).

- Glial Cells and Myelination
- Neurons

coronary artery and often characterized by severe chest pain.

- Cardiac Signaling: Cellular, Molecular and Clinical Aspects

Myocardium

Definition

Myocardium refers to the striated muscle components of the heart, between endo- and epicardium, which perform the pumping action.

- Heart
- Muscle Development

Myelination

Definition

Myelination refers to a process by which nerve fibres are enwrapped by myelin sheaths enhancing the conduction of nerve impulses.

- Glial Cells and Myelination
- Multiple Sclerosis
- Schizophrenia Genetics

Myocytes

Definition

Myocytes are single muscle cells, which form the muscle fibers.

- Ion Channels/Excitable Membranes

Myeloperoxidase

Definition

Myeloperoxidase is a myeloid cell enzyme that is responsible for the hydrolysis of hydrogen peroxide in presence of halide ions.

- Morbus Wegener

Myofibrillogenesis

Definition

Myofibrillogenesis describes the formation of myofibrils and cylindrical elements that are formed by a series of sarcomeres.

- Limb Girdle Muscular Dystrophies

Myocardial Infarction

Definition

Myocardial infarction describes a sudden interruption or insufficiency of the supply of blood to the heart, typically resulting from occlusion or obstruction of a

Myofibrils

Definition

Myofibrils are filamentous contractile protein bundles (1.2 μm diameter) running along the long axis of a muscle cell, and are composed of numerous myofilaments.

- Heart
- Muscle Contraction

Myofilament

Definition

Myofilament defines a longitudinally arranged filamentous protein of the myofibril (actin, myosin, nebulin, and titin filaments).

- Heart
- Muscle Contraction

Myoglobinuria

Definition

Myoglobinuria refers to the urinary excretion of the oxygen binding myoglobin. Myoglobinuria is caused by rhabdomyolysis.

- Mitochondrial Myopathies
- Rhabdomyolysis

Myosin

Definition

The myosins comprise of a large family of motor proteins. Myosin is a heteropolymer composed of two heavy chains (approximately 200,000 MW) and two pairs of light chains. The light chain pairs are called light chain 1 (25,000 MW) and light chain 2 (19,000 MW). Myosin combines with actin. The complex formed is called actomyosin.

- Focal Complexes/Focal Contacts
- Molecular Motors
- Muscle Contraction
- Type II Myosin

Myotome

Definition

Myotome refers to the central ►somite compartment in a vertebrate embryo, which is populated by cells from the dermomyotome and differentiates into skeletal muscle. The term is also used to anatomically define a muscle, or group of muscles, derived from one somite and innervated by a single segment of a spinal nerve.

- Somitogenesis

Myotonic Dystrophy Protein Kinase

Definition

Myotonic dystrophy protein kinase (DMPK) is the product of a gene encoding a serine/threonine protein kinase, which contains the instable (CTG)_n repeat causing myotonic dystrophy type 1 in its 3' UTR exon.

- Myotonic Dystrophy

Myotonic Dystrophy Type 1

DERICK G. WANSINK, BÉ WIERINGA

Department of Cell Biology, Nijmegen Center for Molecular Life Sciences, Radboud University Nijmegen Medical Center, Nijmegen, The Netherlands
r.wansink@ncmls.ru.nl
b.wieringa@ncmls.ru.nl

Synonyms

Dystrophia myotonica; Steinert's disease

Definition

►Myotonic dystrophy (DM), first described about one hundred years ago, is the most common form of muscular dystrophy among adults, with an estimated incidence of 1:8000 (2). Manifestation of the disease is highly variable and typically consists of a combination of muscular, neural and endocrine features, each of which may vary in severity with much variation in the age of onset. Myotonia, muscle degeneration, cataracts, testicular atrophy, endocrine abnormalities, cardiac conduction defects, hypersomnia, neuropsychiatric disturbances, smooth muscle dysfunction and insulin resistance are among the most prominent features (1, 2). Although generally considered a muscle disorder, it should be mentioned here that brain-based problems in DM, including behavioral and personality changes, and excessive daytime sleepiness form perhaps the major burden of disease manifestation for patients and their environment.

Characteristics

DM patients have been categorized into different subgroups on the basis of age of onset and presenting symptoms (Fig. 1). Mild or senile DM has its onset in middle to old age (50 years or over) and is characterized by the occurrence of cataract and minimal or no muscle abnormality. Classical or adult DM has its

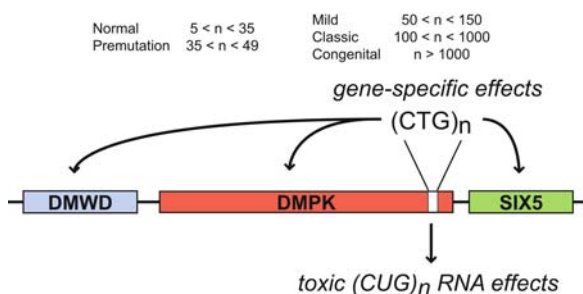
onset in adolescence or early adult life and is characterized by muscle weakness and myotonia, often accompanied by cataracts, gonadal atrophy and minor intellectual impairment. The most severe form of the disease is the congenital form, characterized by a high neonatal mortality and symptoms like facial weakness, delayed motor development, hypotonia, mental retardation and respiratory distress. Novel histopathological findings point to a situation where in congenital cases consequences of developmental arrest of myofiber or CNS differentiation are intermixed with cell-degenerative manifestations. In late onset cases the cell-degenerative effects of disease predominate, leading to muscle wasting, brain atrophy and white matter lesions. About one decade ago, the genetic mutation underlying myotonic dystrophy (now called ►DM1, MIM 160900) was identified as an expanding (CTG)_n repeat on chromosome 19q13.3. Above a critical threshold, this (CTG)_n repeat has a greater than 95% chance to expand from one generation to the next, causing 'anticipation', i.e., progressive earlier onset of disease and worsening of symptoms in subsequent generations (2, 6). Importantly, shortly after the development of (CTG)_n repeat diagnostics, a category of patients with a myotonic dystrophy phenotype similar to DM1, but

without an expanded (CTG)_n repeat, were identified (3). The mutation in patients with this form of DM, designated ►DM2 or PROMM (MIM 602668), was recently characterized as an expanding (CCTG)_n repeat and mapped to chromosome 3q21.3. How (CTG)_n or (CCTG)_n repeat expansions cause the overtly similar, but clinically distinct constellations of features in DM1 and DM2 is still an intriguing open question.

Cellular and Molecular Regulation

The DNA-Chromatin Level

The chromosomal segment in which the unstable (CTG)_n repeat in DM1 patients is located represents a gene-dense area on human 19q13.3. This entire area is evolutionarily well conserved in other mammalian species like mouse (situated on proximal chromosome 7) (1). The (CTG)_n repeat itself is contained within the last 3'-untranslated exon of the human ►myotonic dystrophy protein kinase (*DMPK*) gene, but also located just 5' of the homeodomain-protein-encoding gene ►*SIX5* (Fig. 1). Furthermore, immediately upstream of the *DMPK* gene in both species lies the *DMWD* gene. It has been reported that, possibly due to the presence of weak polyadenylation sites, transcription of the *DMWD* gene can extend into the *DMPK* gene, thus creating another possible transcription unit in which the repeat area is included. The (CTG)_n repeat proper is located within a relatively short segment of DNA with distinct features, characterized by the presence of a CpG island and a Dnase I hypersensitive site. Presumably, this segment, flanked by two binding sites of the zinc-finger protein CTCF, serves as an insulator element by blocking enhancer-promoter communication necessary for independent transcription regulation of *SIX5* and *DMPK*. Other studies suggested a role for the expanded (CTG)_n repeat in abnormal chromatin folding and assembly. In congenital DM1, in contrast to classic DM1 with an adult onset, CpGs in this insulator are hypermethylated and the insulator function is abolished. Although there are species differences at the nucleotide sequence level, it has been shown that expression of each of the three genes *Dmwd*, *Dmpk*, and *Six5* is independently and similarly regulated in mice and humans. Exactly how (CTG)_n repeat expansion affects transcriptional efficiency for each of these genes in patients is not known. How then is (CTG)_n repeat length at the DNA level, or (CUG)_n repeat dosage at the RNA expression level, determined and correlated to the severity and variability in manifestation of clinical features? All we currently know is that unaffected individuals have fewer than 37 repeats, while repeat sizes in affected individuals range from ~100 into the thousands (Fig. 1). Global disease severity and age of onset correlate reasonably well with repeat length in blood: alleles in the 35–49 repeats range are considered



Myotonic Dystrophy Type 1. Figure 1 Schematic representation of the DM1 locus and effects of an unstable (CTG)_n-repeat.

The DM1 locus is strongly conserved between mouse and man and contains three genes located closely together in a gene-dense region: *DMWD*, *DMPK* and *SIX5*. *DMWD* encodes a WD-repeat protein, *DMPK* a serine/threonine protein kinase and *SIX5* a homeodomain transcription factor. An unstable (CTG)_n repeat is situated in exon 15 of the human *DMPK* gene, whereas the mouse gene contains a (CTG)₂(CAG)₂(CTG) sequence at the cognate position. The length of the repeat in blood ranges from 5 to several thousands of CTG triplets and is correlated with the age of onset and severity of DM1 disease manifestation (see box). Effects of the (CTG)_n repeat may be exerted *in cis* via chromatin effects on the levels of *DMWD*, *DMPK* or *SIX5* gene products, or - when expressed as *DMPK* mRNA with a long expanded (CUG)_n repeat - *in trans* via toxic influence on splicing or transport of any other cellular transcript.

premutation alleles, mild or very late onset of disease is seen in individuals with (CTG)_{50–150} repeats, “classic” disease manifestation is associated with (CTG)_{100–1000} and in congenital DM1 patients the repeat contains several thousands of CTG triplets (1, 2). Patients with moderate or long expansions demonstrate age-dependent, expansion-biased mosaicism in repeat length, presumably mediated *via* multiple small length change events. Thus, besides being highly unstable in germ line tissue, repeat length also varies considerably within and between somatic tissues. For example, repeats in muscle cells (skeletal muscle and heart) are generally greater than in blood lymphocytes, but currently no clear picture exists for the relative stability in other tissues, simply because this has not been studied in a large enough cohort of patients.

Mouse models of repeat expansion are more easily accessible and presumably mimic the molecular mechanisms underlying expansion in DM1 patients fairly accurately (6). From in depth study of these models it has become clear that progressive repeat expansion occurs upon aging, and that there is tissue-type preference and cell-type dependence. Early in embryogenesis there may be a tendency to repeat contraction. Together with genetic background effects that have been observed, this is indirect evidence to support involvement of trans-acting modifiers in repeat instability. Direct evidence for the existence of such factors was obtained from study of expanding repeat models showing that deficiency or proficiency for DNA repair proteins Msh2, Msh3 and Msh6 may have differential effects on somatic repeat instability, particularly in post-mitotic events. Involvement of the DNA repair, replication and recombination machinery is now a salient feature of many reports. Better understanding of the relationship between repeat instability, and the role of the DNA repair, replication and recombination machinery therein, may therefore ultimately reveal new entry sites for somatic therapy.

The RNA Level

An important stimulus for understanding DM1 molecular pathogenesis was the finding that *DMPK* transcripts bearing long (CUG)_n repeats in their 3'-UTR were retained in nuclear foci of cells from myotonic dystrophy patients. As a consequence, cytoplasmic *DMPK* transcripts with long (CUG)_n repeats drop to almost undetectable levels, which has been shown to be essentially a post-transcriptional process. In addition, next to the cis-effect of the (CUG)_n repeat on *DMPK* mRNA production, dramatic trans-effects on production and use of other mRNAs with profound influence on alternative splicing and translation of mRNAs have been reported (4). This RNA gain-of-function mechanism, also termed ‘►RNA dominance’, is currently the favored model

to explain the major features of DM1 disease manifestation. This RNA dominance theory was further strengthened by the recent identification of a (CCTG)_n repeat as the genetic defect in DM2, the notion of DM symptoms in the different mouse models with long repeats (the HSA^{LR}-mouse, which has only the repeat area in common with human patients and the DM300 mouse, which has complete transgenic copies of the human repeat area (6)) and by the finding that long (CUG)_n repeats in differentiating myoblast cultures may have severe consequences for late myotube differentiation.

RNA dominance effects are best explained by a mechanism whereby, at the level of RNA, the expanded (CUG)_n repeat interferes with processing or translation of many (primary) non-*DMPK* transcripts. In this mechanism, interaction with RNA-binding proteins, like CUG-BP, ETR-3 or muscleblind proteins, may play a determining role leading to overall perturbation of nuclear RNA metabolism (4). Circumstantial evidence suggests that misregulation of CUG-BP by the (CUG)_n repeat may have the most severe consequences, with effects on late skeletal muscle differentiation, chloride-channel evoked myotonia, and insulin resistance *via* transcription and splice factor regulation. In another scenario, a (CUG)_n length-dependent activation of dsRNA-activated protein kinase (PKR), with toxic effects on myocyte differentiation or survival has been proposed. Since also the neurodegenerative disease spinocerebellar ataxia type 8 (SCA8) is caused by an untranslated (CTG)_n repeat, it may be that SCA8, DM1 and DM2 (where the (CCUG)_n repeat is expressed in the intron of primary *ZNF9* transcripts) share a common disease pathway (4).

The Protein Level *DMPK*

Central in the DM1 area is the *DMPK* gene (MIM 605377; Fig. 1), specifying a serine/threonine protein kinase, which is expressed mainly in skeletal muscle, heart, stomach and in particular regions of the brain (5). *DMPK* transcripts are subject to extensive cell-type dependent alternative splicing. This process is conserved between mouse and man and results in six major splice isoforms (1). All isoforms contain a (CUG)_n repeat in the 3'-UTR of the mRNA. A minor *DMPK* splice isoform, lacking the (CUG)_n repeat was identified in humans. Alternative splicing defines presence or absence of the five-amino-acid motif val-ser-gly-gly-gly and the nature of the C-terminus, for which three different variants exist among the major isoforms. The VSGGG-motif and different C-termini confer distinct cellular properties to the six *DMPK* isoforms related to function(s) at the endoplasmic reticulum and the mitochondrial outer membrane. According to sequence homology in the kinase domain, *DMPK* is a member of the AGC group of kinases and is

most homologous to myotonic dystrophy kinase-related Cdc42-binding kinase and Rho kinase. Various *in vitro* substrates were reported for DMPK, including the dihydropyridine receptor, DMAP, MKBP, phospholemman, CUG-BP and myosin phosphatase targeting subunit, but whether members of this rather diverse set of proteins are *in vivo* substrates needs to be investigated. Because of its presumed role in cell motility, contraction or cell-organelle dynamics, and the observed loss of DMPK in DM1 patients, the DMPK protein is still a likely candidate for being involved in various aspects of disease manifestation, perhaps most probably the developmentally-determined disease features in muscles and brain of congenital DM1 (6).

SIX5

The (CTG)_n repeat is located virtually in the promoter of the *SIX5* gene (formerly known as *DMAHP*), a vertebrate homologue of *sine oculis*, a gene encoding a homeodomain transcription factor essential for development of the visual system of *Drosophila*. Transcription of *SIX5* initiates downstream of the DM1 repeat, excluding inclusion of the (CUG)_n repeat in *SIX5* mRNA. Experimental studies using patient material have indicated that, as for *DMPK*, expansion of the (CTG)_n repeat results in down-regulation of *SIX5* levels. *Six5* is mainly expressed in skeletal muscle, heart, eye, kidney, liver, lung and small intestine. Expression patterns of *Six5* and *Dmpk* are essentially overlapping, albeit that *Six5* levels are generally 5–50 fold lower than those of *Dmpk*. A screen for downstream targets of *Six5* revealed several candidate genes expressed in somites, skeletal muscle, brain and meninges. In *Drosophila*, the *Six5* homologue *D-Six4* was found to play a role in gonad and muscle development. In conclusion, the presumed function of *Six5* in tissues involved in DM1 pathology (e.g., skeletal muscle, eyes, testis), and the finding that expansion of the (CTG)_n repeat has a negative effect on *SIX5* expression levels has opened the possibility that DM1 should be considered a multigenic (-locus) disease. Evidence to support this hypothesis was found in two independently generated *Six5*^{-/-} transgenic mouse models, which display ocular cataracts (6).

Clinical Relevance

DM1 is a highly complex and variable multisystemic disorder. We know now that a rather unpredictable dynamic behavior of mutations in (CTG)_n repeat length in somatic tissues of gene carriers is underlying disease manifestation. Individual differences in genetic background, cell type and developmental or adult age may play an important role in the dynamics of the repeat. This makes it currently virtually impossible to establish reliable genotype-phenotype correlations for individual

members of DM1 families. Prediction of disease development based on repeat length typing in blood is therefore impossible, and considered unethical. Based on current knowledge, our future hope for treatment is based on (i) a reduction of disease incidence in DM families *via* accurate and early carrier-diagnosis and (ii) the development of therapeutic strategies. Given the widespread occurrence of problems in many tissues and organs, our efforts must be primarily directed towards the use of pharmacological rather than gene-therapeutic approaches.

References

1. Groenen P, Wieringa B (1998) Expanding complexity in myotonic dystrophy. *BioEssays* 20:901–912
2. Harper PS (2001) Myotonic dystrophy. 3WB Saunders, London, UKrd edn.
3. Ranum LP, Day JW (2002) Myotonic dystrophy: clinical and molecular parallels between myotonic dystrophy type 1 and type 2. *Curr Neurol Neurosci Rep* 2:465–70
4. Ranum LPW, Day JW (2002) Dominantly inherited, non-coding microsatellite expansion disorders. *Curr Opin Genet Dev* 12:266–271
5. Ueda H, Ohno S, Kobayashi T (2000) Myotonic dystrophy and myotonic dystrophy protein kinase. *Progr Histochem Cytochem* 35:187–251
6. Wansink DG, Wieringa B (2003) Transgenic mouse models for myotonic dystrophy type 1 (DM1). *Cytogenet Genome Res* 100:230–242

Myotonic Dystrophy Type 2

Definition

Myotonic dystrophy type 2 (DM2), also called proximal myotonic myopathy (PROMM), is a tetranucleotide repeat disorder very similar to myotonic dystrophy type 1, which is caused by an instable (CCTG)_n repeat in the zinc finger protein 9 (ZNF9) gene.

► [Myotonic Dystrophy](#)

Myristoylation Signal

Definition

Myristoylation signal is a sequence motif in a protein that causes an enzyme to attach a myristoyl group (a fatty acid chain) to that protein.

► [Two-Hybrid System](#)

N-Acetyl-D-Mannosamine

Definition

ManNAc is a specific substrate for the synthesis of N-acetylneuraminic acid, the essential precursor of bacterial capsular polysialic acid. N-Acetyl-D-mannosamine is used for the synthesis of sialic acid. It is also a synthetic intermediate for a number of carbohydrate-derived families of biologically active compounds and pharmaceutical candidates.

- ▶ Biochemical Engineering of Glycoproteins
- ▶ Glycosylation of Proteins

they are membrane-bound and require activation by cytosolic factors that are recruited by specific receptor-mediated signaling events. The catalytic units, the Nox proteins, are multidomain proteins that contain heme and flavine. They differ between tissues and depend on different activators. The best investigated system is the one responsible for the respiratory burst of macrophages. It comprises gp91^{phox} (=Nox2) as catalytic unit p22^{phox}, the GTP binding protein Rac and p47^{phox}, p40^{phox} and p67^{phox} as phosphorylation-dependent activators.

- ▶ Free Radicals

N-Acetylneuraminic Acid

- ▶ Biochemical Engineering of Glycoproteins
- ▶ Sialic Acid

NADPH + H⁺/NAD⁺

Definition

Abbreviations for Nicotinamide adenine dinucleotide phosphate (reduced, oxidized, respectively): this is a chemical co-factor required for reactions where a molecule is reduced, for example, the reaction catalyzed by dihydrofolate reductase. In this process, the reduced form NAPH + H⁺ is converted to the oxidized form NAD⁺.

- ▶ DNA Ligases
- ▶ Nucleotide Biosynthesis

N-Acylation

- ▶ Amide-Linkage

NADPH Oxidase

Definition

NADPH oxidases are members of a group of complex proteins that produce superoxide radicals. Typically,

Naked Cuticle

Definition

Naked (NKD)1 and NKD2 are mammalian homologs of Drosophila Naked Cuticle, which negatively regulates Wnt signaling by binding ▶Dishevelled. Nkd contains an EF-like calcium binding domain.

- ▶ Wnt/Beta-Catenin Signaling Pathway

'Naked' DNA

Definition

"Naked" DNA defines, in general, pure DNA without any protein(s) attached. DNA is "naked" when purified.

►DNA-based Vaccination

NALD

►Infantile Refsum Disease

Narcolepsy

THOMAS POLLMÄCHER

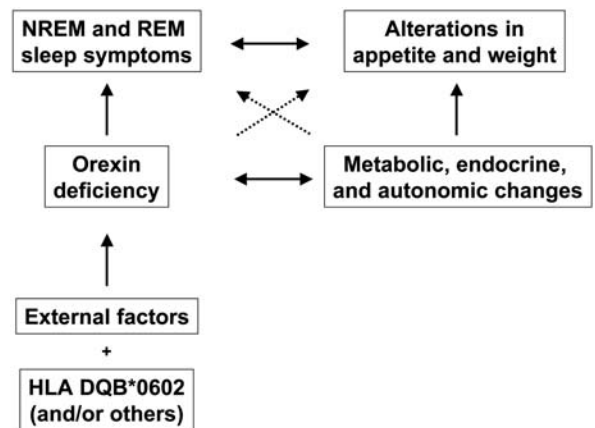
Center of Mental Health, Clinical Center Ingolstadt,
Ingolstadt, Germany
thomas.pollmaecher@klinikum-ingolstadt.de

Synonyms

Narcolepsy-cataplexy; Maladie de Gélineau

Definition

Narcolepsy is a rare human sleep disorder (life time prevalence about 1:2000) characterized by 2 major symptoms. These are excessive daytime sleepiness of episodic, often irresistible character, and ►cataplexy (4), a sudden and short-lasting loss of muscle tone triggered by emotions. The pathogenetic background is complex and multifactorial (Fig. 1). Narcolepsy has a genetic background closely associated to the ►human leukocyte antigen (HLA) DQB*0602, which can be found in more than 95% of Caucasian patients. Due to this close HLA association, autoimmunity has been suggested to play a causative role. Although positive evidence for immune processes directed against self-antigens is still lacking, recent studies have identified a putative target, which is a small group of hypothalamic cells producing ►orexins (also called hypocretins). These are recently discovered neuropeptides involved in the regulation of feeding, metabolism and sleep. The vast majority of patients with narcolepsy display a complete and selective acquired orexin deficiency.



Narcolepsy. Figure 1 Current pathogenetic model of narcolepsy. Based on a genetic predisposition external factors induce an orexin deficiency (e.g. through an autoimmune process). Orexin deficiency leads to sleep related symptoms and disturbances of endocrine and metabolic networks which interact in a yet poorly defined manner (for details see text).

However, the definitive pathophysiological role of these peptides remains to be unraveled. Although there is no cure for narcolepsy, there are a number of effective drug treatments that considerably improve sleepiness and cataplexy.

Characteristics

Narcolepsy was first described by Gélineau in the late 19th century and since then has remained a stable clinically defined disease entity. However, as discussed in detail below, the cause and many aspects of the pathophysiology still remain an enigma, although recently there has been tremendous progress in this field.

Symptoms

The hallmarks of narcolepsy are excessive daytime sleepiness and cataplexy. Sleepiness in narcoleptic patients is severe and might lead to irresistible sleep attacks. However, the patients are not continuously sleepy; therefore, their vigilance is not impaired between periods of increased sleep pressure. These might occur at multiple times during the day, often in a more or less regular temporal pattern. Typical daytime sleep episodes are short (10–20 min) and refreshing. The second clinical hallmark is cataplexy defined as a sudden and short-lasting (seconds to minutes) loss of muscle tone, which is triggered by strong emotions. The most typical trigger is laughter, but anger and surprise may also provoke a cataplectic attack. It is important to note that patients are fully conscious during cataplexy and aware of their surrounding. Due

to the strong dependence on triggers, the frequency of cataplectic attacks is highly variable within and between patients. The same holds true for severity, which varies between very subtle weakness in the knees or blurring of speech to the complete inability to remain in the upright position.

Other symptoms of narcolepsy include automatic behavior (the continuation of behavioral acts while almost sleeping), ►**hypnagogic hallucinations** (frightening, realistic hallucinations while falling asleep), ►**sleep paralysis** (a short-lasting inability to move upon awaking) and disturbed night sleep. Only very recently has it been recognized that narcoleptic patients do not only suffer from symptoms related to sleep-wake regulation, but in addition display endocrine and metabolic abnormalities (2). Most prominent are a high incidence of obesity and an increased rate of type II diabetes. The underlying causes are not fully understood, but they are probably related to the deficiency in hypothalamic orexin production described in detail below.

Epidemiology and Course of Narcolepsy

There are no robust estimates about the frequency of the disorder, but lifetime prevalence is likely to be around 0.05% worldwide. Typically, disease onset is in late adolescence and early adulthood, but children may also be affected, and a second, smaller peak of incidence occurs in the 3rd to 4th decade of life. In the majority of patients sleepiness precedes cataplexy by several years. Narcolepsy is a life-long chronic disorder, although the severity of symptoms might show considerable fluctuations. In the vast majority of patients narcolepsy induces considerable disability.

Since the very early descriptions it has been known that narcolepsy sometimes runs in families. The risk of first-degree relatives developing narcolepsy is 10–40× greater than in the general population. Due to the low overall lifetime prevalence, however, the absolute risk is still below 1%. Although these figures indicate a causative genetic contribution, no clear-cut inheritance pattern could be established. The fact that the majority of monozygotic twins investigated are discordant for narcolepsy clearly suggests that non-genetic factors are pivotal.

Differential Diagnosis

Recognition of typical narcolepsy is an easy task if the physician is aware of daytime sleepiness and cataplexy as the hallmarks of the disorder. However, cataplexy might occur for the first time years after daytime sleepiness, or cataplexy might be so subtle that it escapes the patients' or the physicians' attention. In these cases polysomnographic sleep recordings are very helpful to distinguish narcolepsy from other disorders of excessive hypersomnolence. The typical

and quite specific finding in narcoleptic patients is an early occurrence of rapid-eye-movement (REM) sleep during sleep episodes. These so-called sleep-onset ►**REM** episodes can be detected during night sleep or during daytime naps. HLA typing is only helpful to exclude narcolepsy in clinically doubtful cases because the typical genotype (HLA-DQB*0602, for details see below) is present in more than one third of normal people. The absence of orexin-A in CSF (for details see below) is probably highly specific for narcolepsy, although normal orexin levels do not definitively exclude the diagnosis.

Cellular and Molecular Regulation Immunogenetics of Narcolepsy

Based on the knowledge of familial cases of narcolepsy Japanese and European researchers conducted immunogenetic studies in the 1980s and reported a more than 90% association of narcolepsy with the human leukocyte antigen (HLA) DR2 carried by about one third of healthy people. The association is even stronger to a related class II antigen (DQB1*0602) and represents the tightest link of a disease to an immunogenetic marker that has ever been discovered. Additional HLA subtypes also increase (DQB1*0301) or reduce (DQB1*0601 and 0501) disease susceptibility (3). Familial clustering of narcolepsy cannot be explained exclusively by sharing of HLA haplotypes, because some families are negative for the DQB0602, suggesting that non-HLA genes modulate disease susceptibility. Although very little is known about these genes so far, there are interesting candidates such as genes related to the tumor necrosis factor- α cytokine system.

The strong association of narcolepsy with an HLA marker has suggested that autoimmune processes might play a pathophysiological role. Although no clear-cut evidence supporting this assumption could be established so far, the recent discovery of a complete deficiency in orexin production has renewed the interest in the autoimmune hypothesis of narcolepsy.

Orexins and Narcolepsy

In the 1970s an animal model of narcolepsy was discovered in dogs, where cataplexy and, to a lesser extent also sleepiness, are transmitted as an autosomal recessive trait with full penetrance. It took until 1999 for the underlying genetic defect to be identified as a missense mutation in a neuropeptide receptor. This ORX-2 receptor was one of 2 receptors for a pair of novel hypothalamic neuropeptides (1). Orexin-A and B, also called hypocretin-1 and -2, are derived from one common precursor peptide called preproorexin. At the same time as the ORX-2 receptor defect was discovered in narcoleptic dogs, it was found that mice with a targeted disruption of the preprohypocretin

gene display periods of behavioral arrest resembling cataplexy.

Very shortly after these findings in animals, it was reported in humans that narcolepsy is associated with a complete loss of orexin production in the vast majority of patients affected, suggesting that in the human disorder orexins play a pivotal pathophysiological role. In contrast to animals, however, defective orexin signaling is due to a secondary loss of peptide production, rather than due to a genetic defect. No mutations have been found in the human orexin receptors, nor is there an association between narcolepsy and hypocretin polymorphisms in humans. Only one single case of a child with very early disease onset at 6 months of age was found to carry a mutation in the ►[preproorexin](#) gene.

Hence, human narcolepsy in the vast majority of cases is associated with a secondary, acquired orexin deficiency. But, it is unclear what causes this defect. On the level of the hypothalamus, neuroanatomical studies, which so far have been based on very few brains, suggest that, indeed, orexin production is almost or completely absent. However, it is unclear, whether this is due to a lack of orexin-producing neurons or due to an arrest of orexin production within the hypothalamus. One group described gliosis in this area, which is compatible with an earlier inflammatory process, but no signs of active inflammation have been found.

Although it is now clear that orexin deficiency plays an important role in the pathophysiology of narcolepsy, details of this role have been poorly defined so far. This is not surprising, because these peptides were discovered only five years ago and their physiological functions are just starting to be understood. Acute administration of orexins stimulates food intake, and the orexin system interacts with multiple endocrine networks including the hypothalamo-pituitary-adrenal and the hypothalamo-pituitary-somatotropic axes (5). However, a pivotal role in the physiological regulation of food intake is less likely, because mice with defective orexin-signaling are either of normal weight or even obese. Obesity is also a typical feature of narcolepsy in humans, who in addition display endocrine and metabolic abnormalities such as type II diabetes and hypo►[leptinemia](#).

Orexins have potent arousing properties, which might explain why orexin deficient narcoleptic patients are sleepy. However, the symptomatology of narcolepsy is much more complex and the common denominator of the divergent symptoms is a disturbance of the transitions between the different states of vigilance (wakefulness, ►[nonREM](#) sleep and REM sleep). These transitions might be too fast (e.g. from wakefulness to nonREM sleep during a sleep attack) or they might be qualitatively disturbed (e.g. the occurrence of cataplexy

which represents an incomplete and inappropriate transition from wakefulness to REM sleep). On the level of classical neurotransmission, sleepiness might be due to a reduced stimulatory effect of orexins in the dopaminergic system. Cataplexy, in contrast, might be caused by a reduced input of orexin-producing neurons to cholinergic and noradrenergic brain stem centers involved in REM sleep regulation. However, how orexins really control these transitions between vigilance states remains to be elucidated.

Clinical Relevance

Although narcolepsy is rare, it is of considerable clinical relevance due to its chronic and disabling course. Moreover, it is a prototypic disorder of sleep-wake regulation important for our general understanding of sleep in health and disease. Finally, narcolepsy qualifies as a very attractive disease model to study the complex interactions between genes and environment, because the disorder involves a highly specific peptide deficiency occurring on a clearly defined immunogenetic background.

References

1. Beuckmann CT, Yanagisawa M (2002) Orexins: from neuropeptides to energy homeostasis and sleep/wake regulation. *J Mol Med* 80:329–342
2. Dalal M, Schuld A, Beiting P et al (2002) Neuroendocrine and metabolic aspects of narcolepsy. *Somnologie* 6:95–100
3. Lin L, Hungs M, Mignot E (2001) Narcolepsy and the HLA region. *J Neuroimmunol* 117:9–20
4. Scammell TE (2003) The neurobiology, diagnosis, and treatment of narcolepsy. *Ann Neurol* 53:154–166
5. Taylor MM, Samson WK (2003) The other side of the orexins: endocrine and metabolic actions. *Am J Physiol Endocrinol Metab* 284:E13–17

Narcolepsy-Cataplexy

► Narcolepsy

Native State

Definition

The native state is the fully folded biologically active form of a protein in its unperturbed conformation.

► Protein Folding

Natural Antisense Transcripts

ANDREAS WERNER

Institute of Cell and Molecular Biosciences, University of Newcastle, Newcastle upon Tyne, UK
andreas.werner@ncl.ac.uk

Definition

Natural antisense transcripts overlap their related sense transcripts in reverse complement orientation. Both coding and non-coding antisense RNAs are found. Natural antisense transcripts are very heterogeneous in length and may be spliced. If sense and antisense transcripts are expressed in the same cell, RNA-RNA hybrids are believed to be formed. Nucleic acid base pairing guarantees the exquisite specificity of this interaction. The RNA duplexes trigger various mechanisms that lead to the down-regulation of both RNAs or the silencing of the gene at a transcriptional level.

Genome sequencing efforts and computational screens have identified about 2000 overlapping [transcriptional units](#) that give rise to sense/antisense pairs (1). The physiological relevance of [bi-directional](#) transcription at the observed scale is unclear. Only a few examples are experimentally characterised, however increasing evidence suggests that natural antisense transcripts have a pivotal role in gene regulation. An altered ratio of sense/antisense RNA may result in an aberrant expression pattern of the related protein. The pathological phenotype depends upon the nature of the protein.

Natural antisense transcripts as described above relate to cis-encoded RNAs. Trans-acting molecules such as [microRNAs](#) that do not exhibit full complementarity are not discussed here. Nevertheless, natural antisense RNAs and microRNAs may feed into similar cellular gene silencing pathways.

Characteristics

Cellular Consequences of RNA-RNA Hybrids

The transcription of a natural antisense RNA causes down-regulation of sense RNA encoded protein. The RNA-RNA duplexes trigger two cellular mechanisms, [RNA interference](#) and RNA editing, that lead to the degradation of both RNA strands and eventually to gene silencing. RNA interference involves an RNase-III-type recognising double stranded RNA (Dicer), which cleaves the target into small oligonucleotides of 21–23 base pairs. The oligonucleotides trigger various mechanisms that eventually suppress the expression of the cognate gene. In human (and many other eukaryotic species) an RNA-induced silencing complex (RISC) is

formed that degrades both strands of the targeted RNA. Other mechanisms involve the silencing of gene expression at transcriptional and translational level. The mechanism of RNA interference is well conserved throughout eukaryotes. However, considerable variations may occur between organisms.

Double stranded RNA is also recognised by an enzyme that converts adenosine (A) into inosine (I), called ADAR (adenosine deaminase acting on RNA). Such modifications may lead to [point mutations](#) or altered splice sites. Alternatively, edited mRNA is prevented from nuclear export and eventually degraded. Recent evidence suggests that I-rich (edited) RNA binds to components of the [DNA methylation](#) machinery and leads to [epigenetic](#) gene silencing. The link between RNA interference/RNA editing and gene silencing at the transcriptional level by chromatin modification is hypothesised, however, direct experimental evidence is missing. A model that amalgamates RNA interference and RNA editing to explain how double stranded RNA leads to gene silencing is presented in Fig. 1. The number of RNA-RNA hybrids is a critical determinant of which of the indicated pathways is activated.

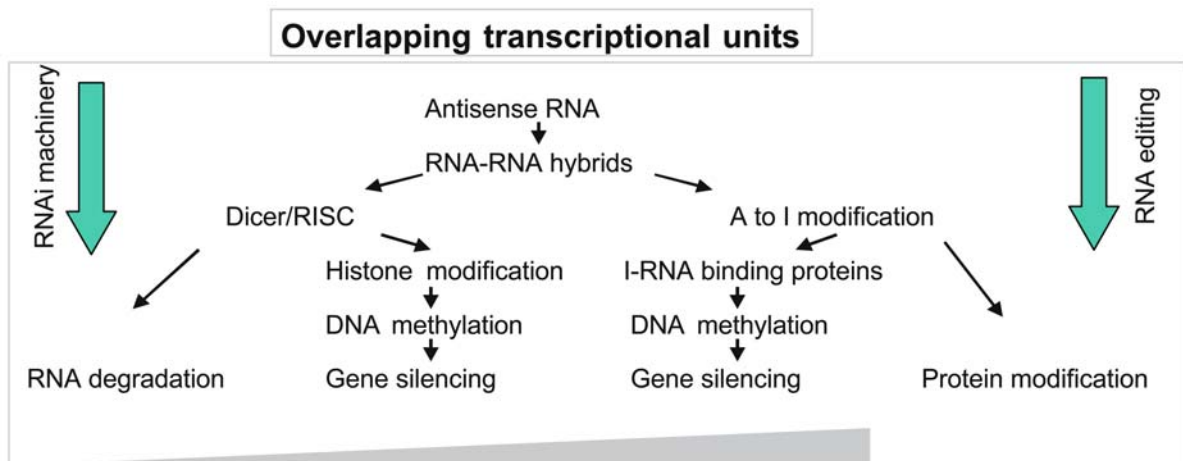
Natural antisense transcripts also interfere with the processing of the related sense RNA. [Splicing](#), [polyadenylation](#) and the export of the mRNA from the nucleus to the cytoplasm are affected. In the cytoplasm, translation efficiency is reduced by RNA duplex formation. However, natural antisense transcripts are thought to exert their regulatory function predominantly in the nucleus.

Physiological Consequences of Antisense RNA

The consequences of antisense RNA expression are relatively well established in the context of epigenetic gene silencing phenomena including DNA methylation (2), [genomic imprinting](#) (3) and [X chromosome inactivation](#) (4).

DNA methylation is closely linked to the transcriptional activity of specific areas of the genome. [Promoter](#) regions of genes that are actively transcribed are poorly methylated. The expression of antisense transcripts can induce the methylation of the sense promoter and silence the gene. An individual suffering from α -[thalassemia](#) was recently found to display a normal α -globin (*HBA2*) gene that was silenced by a juxtaposed antisense transcript. The antisense RNA derived from the widely expressed *LUC7L* gene that was incorrectly localised a few hundred base pairs downstream of *HBA2*. The expression of the antisense transcript leads to hypermethylation of the *HBA2* promoter and concomitantly to the silencing of the gene (2).

Imprinted genes express either the maternal or the paternal [allele](#). More than 60 human imprinted genes

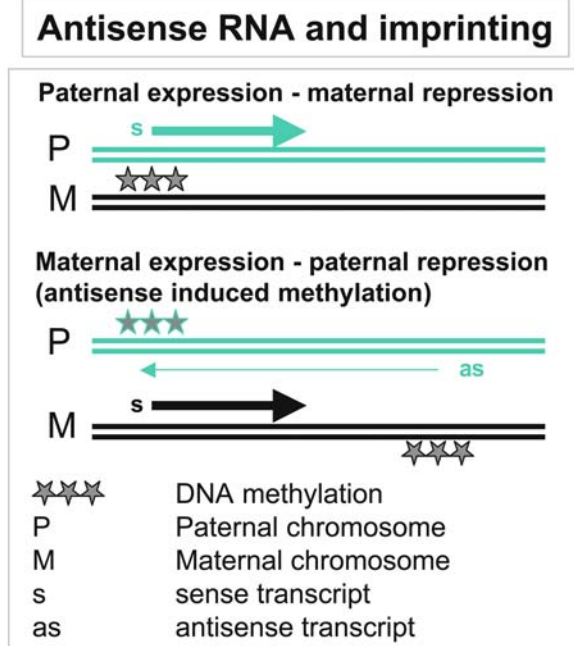


Natural Antisense Transcripts. Figure 1

have been reported so far. They are grouped in clusters and often display both DNA methylation and antisense transcripts. The phenomenon of genomic imprinting is restricted to marsupials and placental mammals. An elegant hypothesis integrates DNA methylation, antisense transcription and maternal/paternal expression. In contrast to the methylated maternal DNA, the paternal chromosome becomes actively demethylated after fertilisation; transcriptional inhibition of paternal genes due to methylated promoters is abolished. In order to reinforce mono-allelic maternal gene expression, an antisense transcript is expressed. Accordingly, most of the antisense transcripts within imprinted gene clusters are paternally expressed (Fig. 2). A mouse gene knock out study focusing on the paternally expressed *Air* antisense transcript has corroborated the importance of antisense expression in imprinting. *Air* is expressed in antisense direction of the maternally expressed genes *Slc22a2*, *Slc22a3* and *Igf2r* and overlaps with the first exon of *Igf2r*. Premature termination of the *Air* RNA resulted in both maternal and paternal expression of *Slc22a2*, *Slc22a3* and *Igf2r*. The transgenic animals showed 15% reduced birth weight; a phenotypic alteration shared by mice with bi-allelic *Igf2r* expression (3).

Dosage compensation is necessary to balance the expression of X chromosome encoded genes in XX (female) and XY (male) individuals. A process called X chromosome inactivation leads to the silencing of one X allele in females. A large non-coding RNA (*Xist*) recruits proteins that modify histones and DNA to one of the X chromosomes thus leading to the suppression of transcriptional activity. An antisense transcript (*Tsix*) plays a crucial role in dictating which X chromosome becomes inactivated.

The antisense transcripts that have been characterised in the context of ►DNA methylation, imprinting or X chromosome inactivation constitute less than 5% of the



Natural Antisense Transcripts. Figure 2

estimated 2000 overlapping transcriptional units (1). This leaves a large number of antisense transcripts that carry the potential to interfere very specifically with the expression of the sense-encoded protein. A general hypothesis predicts that an antisense RNA is an instrument for tight spatial or temporal regulation of the sense transcript. Several reports have quantified sense/antisense transcription of selected genes, under varying physiological conditions and confirm the regulatory impact of the antisense transcripts. Generally, the impact of the antisense RNA is restricted to the

regulation of the cognate sense transcript. The physiological consequences of a given antisense RNA are therefore intimately related to the biological importance of the sense-encoded protein. Natural antisense transcripts are thought to trigger the degradation of the related sense transcript. Consequently, aberrant regulation of antisense RNA expression and mutations affecting the sense transcript may result in similar phenotypes. A few examples of a rapidly increasing list of experimentally characterised bi-directionally transcribed genes are described below.

The cardiac muscle contains two forms of myosin heavy chains, denoted α MHC and β MHC. The proportion of the two proteins is tightly regulated during development and may be altered in pathological situations. The genes encoding the two proteins are located closely together (~5 kb). The promoter driving α MHC expression shows bi-directional activity giving rise to the α MHC encoding transcript and a β MHC related antisense RNA. This arrangement couples the expression of α MHC with the silencing of β MHC. The proportion of the three transcripts, α MHC encoding mRNA, antisense β MHC RNA and β MHC encoding mRNA, are altered in response to physiological and pathophysiological changes (development, diabetes) (5).

The gene encoding the basic fibroblast growth factor 2 (bFGF-2) is transcribed in both directions. Both transcripts give rise to a protein, the sense transcript to bFGF-2 and the antisense transcript to a DNA repair enzyme of the MutT family called GFG. Antagonising effects of the two gene products have been observed on both the RNA and protein levels. Patients with endometriosis show altered bFGF sense/antisense RNA ratios in endometrial stromal cells. On the protein level, the normal pituitary expresses the antisense encoded protein (GFG) but not bFGF-2 whereas pituitary adenomas express bFGF-2 and have reduced levels of GFG (6).

The expression of thyroid hormone receptors (TRs) including the two splice isoforms TR α 1 and TR α 2 is highly regulated. TR α 1 binds thyroid hormone whereas TR α 2 antagonises thyroid hormone action. A natural antisense transcript influences the proportion of the two splice isoforms and consequently the responsiveness of the cell to thyroid hormone (7).

Natural antisense transcripts are a widespread phenomenon in both prokaryotes and eukaryotes. However, prokaryotic antisense RNAs, despite the fact that bacteria show extensive bi-directional transcription, hardly compare to those of eukaryotes. Whereas prokaryotic antisense transcripts often exert their activity without the aid of proteins, the handling of RNA-RNA duplexes in eukaryotic cells requires specific protein machineries. The RNA interference pathway in particular is limited to eukaryotes. Since this mechanism and other aspects of bi-directional transcription have only very recently been discovered and fully

recognised, a general picture of gene regulation by natural antisense transcripts is still emerging.

References

1. Yelin R, Dahary D, Sorek R et al (2003) Widespread occurrence of antisense transcription in the human genome. *Nat Biotechnol* 21:379–386
2. Tufarelli C, Stanley JA, Garrick D et al (2003) Transcription of antisense RNA leading to gene silencing and methylation as a novel cause of human genetic disease. *Nat Genet* 34:157–165
3. Sleutels F, Zwart R, Barlow DP (2002) The non-coding Air RNA is required for silencing autosomal imprinted genes. *Nature* 415:810–813
4. Lee JT (2003) Molecular Links between X-Inactivation and Autosomal Imprinting: X-Inactivation as a Driving Force for the Evolution of Imprinting? *Curr Biol* 13: R242–254
5. Haddad F, Bodell PW, Qin AX et al (2003) Role of antisense RNA in coordinating cardiac myosin heavy chain gene switching. *J Biol Chem* 100:37132–37138
6. Asa SL, Ramyar L, Murphy PR et al (2001) The endogenous fibroblast growth factor-2 antisense gene product regulates pituitary cell growth and hormone production. *Mol Endocrinol* 15:589–599
7. Hastings ML, Ingle HA, Lazar MA et al (2000) Post-transcriptional regulation of thyroid hormone receptor expression by cis-acting sequences and a naturally occurring antisense RNA. *J Biol Chem* 275:11507–11513

NC

NC (nucleocapsid) is the retrovirus Gag protein that forms the viral core. NC interacts with the viral genome RNA.

► [Retroviruses](#)

NCAM

► [Adhesion Molecules](#)

NCBI

Definition

The National Center for Biotechnology Information (NCBI) was established in 1988 as a national resource (USA) for molecular biology information. NCBI

creates public databases (e.g. BLAST, Entrez), conducts research in computational biology, develops software tools for analyzing genome data, and disseminates biomedical information. The resources are accessible online (<http://www.ncbi.nlm.nih.gov/>).

- [Cleft Lip Palate](#)
- [Protein Databases](#)

NCGR

Definition

NCGR stands for the National Center for Genome Resources (<http://www.ncgr.org>).

- [Protein Databases](#)

NCV

Definition

Nerve conduction velocity (NCV) is a test of the speed of conduction of impulses through a nerve. Abnormal results are symptomatic for inherited neuropathies or demyelinating diseases.

- [Hereditary Neuropathies, Motor and/or Sensor](#)

Nearest Neighbour

Definition

The nearest neighbour algorithm is a simple nonparametric learning algorithm in machine learning, and belongs to the most popular methods used in statistical pattern recognition. The method can make discrete predictions by reporting the relative frequency of the classes in the neighbourhood of the prediction point. The procedure uses a “training dataset” to make predictions on new unlabelled data.

- [EST Mining for Expression Analysis](#)

Nearest Neighbour Model

Definition

Nearest neighbour model designates a model in which DNA sequence dependent properties are assumed to be

well represented by sums of parameters characteristic of neighboring base pairs.

- [Thermodynamic Properties of DNA](#)

Necrosis

Definition

Necrosis is the unprogrammed death of living cells, tissues, or an organ in the body. Cells that may not yet have reached their full life span are hit by an external noxa that interrupts some of their vital functions or disrupts their physical integrity. The intracellular contents are also spilled into the microenvironment. There are many causes for necrosis including mechanical injury, infection, cancer, infarction, inflammation, oxidative stress or radiation.

- [Apoptosis](#)
- [Apoptosis, Regulation and Clinical Implications](#)
- [Inflammatory Response](#)

Neighbour-Joining

Definition

Neighbour-Joining is a frequently used algorithm for construction of phylogenetic trees. Based on a multiple sequence alignment, the algorithm calculates distances between proteins, and groups them in a way that the total distance in the final tree is minimal.

- [Sequence Annotation in Evolution](#)

Nemo Like Kinase

Definition

Nemo like kinase (NLK) designates a mouse kinase that is related to the *Drosophila* Nemo polarity gene. NLK is an essential co-activator of Wnt signaling during development.

- [Wnt/Beta-Catenin Signaling Pathway](#)

N-End Rule

Definition

The N-end rule gives an estimate of the stability of a certain protein, based on the amino acid exposed at the

N-terminus. An E3 (Ubr1) ubiquitin ligase recognizes the N-terminus of proteins with different affinity based on the exposed amino acid. During the next step, the Ubr1-interacting E2 (Ubc2) ubiquitinates the protein.

► [Ubiquitination](#)

Neocentromeres

Definition

The term neocentromeres refers to centromeres that arise at new and previously unspecified locations on chromosomes.

► [Centromeres](#)

Neocortex/Cortical Column

Definition

Evolutionarily, the neocortex is the latest generated part of the brain consisting of the two hemispheres that are subdivided into sensory, motor and associational areas. Sensory cortices are characterised by functional units running perpendicular from the surface to the white matter, the so-called cortical columns. Neurons within a cortical column are thought to possess similar functional properties.

► [Brain](#)

► [Neurons](#)

Neonatal

Definition

Neonatal describes a life span shortly after birth; in humans it refers to the first 4 weeks after birth.

► [Cardiac Signaling: Cellular, Molecular and Clinical Aspects](#)

Neonatal Adrenoleukodystrophy

Definition

Neonatal adrenoleukodystrophy (NALD) is a peroxisome biogenesis disorder, with clinical signs and

symptoms resembling those observed in ► [Zellweger syndrome](#) but less severe.

► [Cerebro-Hepato-Renal Syndrome](#)

► [Peroxisomal Disorders](#)

Neoplasia

Definition

Neoplasia describes an abnormal or uncontrolled condition of cell proliferation, which may lead to either benign or malignant tumors.

► [Splicing](#)

Nephrocalcinosis

Definition

Nephrocalcinosis refers to renal lithiasis characterized by diffusely scattered foci of calcification in the renal parenchyma.

► [Hyper- and Hypoparathyroidism](#)

Nephrogenic Mesenchyme

Definition

Nephrogenic mesenchyme consists of mesenchymal cells adjacent to the tips of the branching ureteric bud that will form the nephrons during kidney organogenesis.

► [Kidney](#)

Nephrolithiasis

Definition

Nephrolithiasis refers to the presence of renal calculi (kidney stones).

► [Hyper- and Hypoparathyroidism](#)

Nephron

Definition

Nephron defines the basic, structural and functional unit of the mature kidney, responsible for the purification and filtration of the blood. A nephron consists of three major parts: the glomerulus, the Bowman's capsule, and the tubule (which is further divided into the proximal and distal tubule and the Loop of Henle). A mature kidney consists of about one million nephrons.

► [Kidney](#)

Nephropathy

Definition

Nephropathy means disease of the kidneys (renal disease). In diabetes, this may ultimately lead to kidney failure.

► [Diabetes Mellitus](#), [Genetics](#)

NER

► [Nucleotide Excision Repair](#)

Nerve Conduction Velocity

► [NCV](#)

NES

Definition

NES (nuclear export signal) refers to a leucine-rich sequence, which can be recognized by a group of exportin proteins, called ► [exportin 1](#) or Crm1.

► [NFκB Pathway](#)

► [Nuclear Import and Export](#)

Nestin

Definition

Nestin is an intermediate filament protein that is essential for proper cytoskeletal formation during neurogenesis and myogenesis. Nestin is expressed in neural ► [precursor cell](#) populations. It shows an interesting but problematic overlap with glial fibrillary acidic protein (► [GFAP](#)).

► [Neuronal Stem Cells](#)

N-Ethylmaleimide-Sensitive-Factor

► [NSF](#)

NeuN

Definition

NeuN stands for "Neuronal nuclei", and is an antigen expressed in the nuclei of many, but not all, neuronal cell populations in the brain. It is a widely used marker to demonstrate the neuronal identity of a cell.

► [Neuronal Stem Cells](#)

Neural Crest Cells and their Derivatives

CHAYA KALCHEIM

Department of Anatomy and Cell Biology, Hebrew University-Hadassah Medical School, Jerusalem, Israel
kalcheim@nn-shum.cc.huji.ac.il

Definition

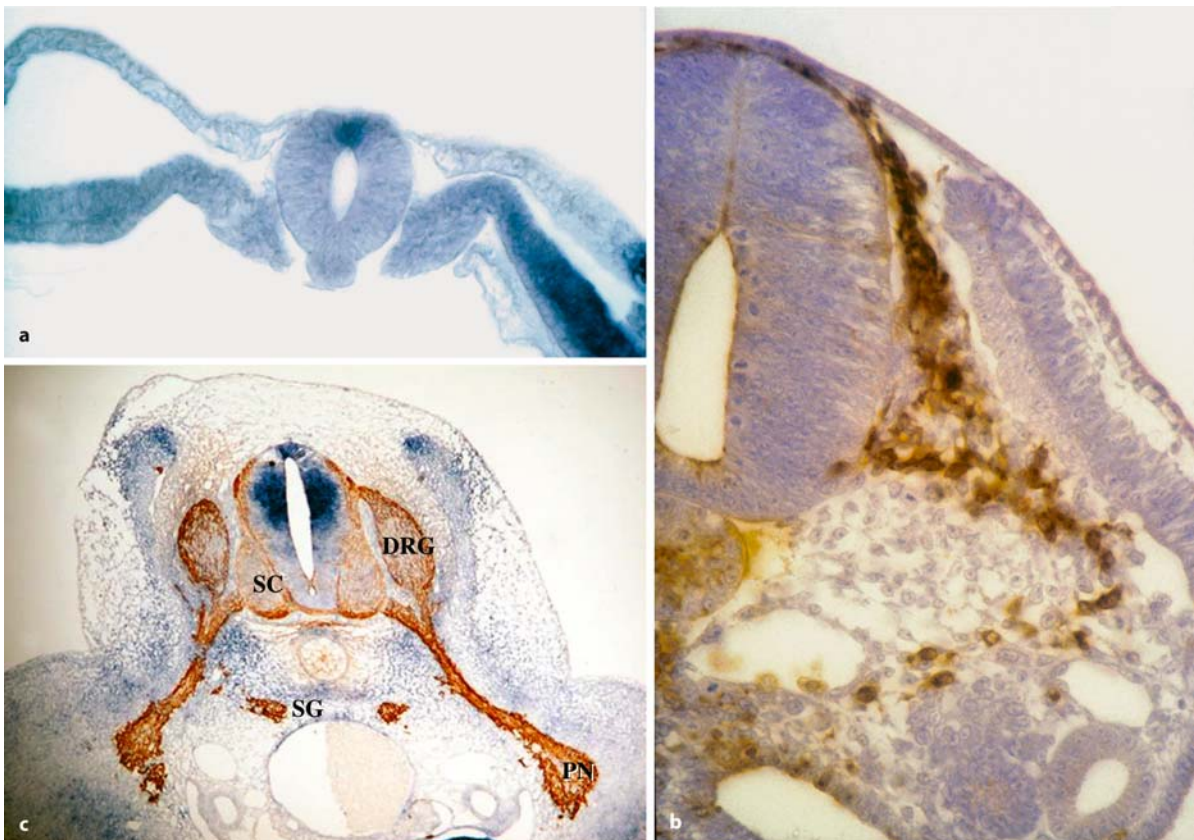
The neural crest (NC) is a major cell type arising in the lateral tips of the neural folds during the process of neurulation. Upon fusion of the folds and formation of the ► [neural tube](#), the NC temporarily resides in its dorsal midline, being an integral part of the pseudostriated neuroepithelium (1) (Fig. 1a). The significance

of the NC stems from the immense variety of derivatives that this discrete group of cells yields during ontogeny. Most of the peripheral nervous system arises from the NC including all the glia of the ►sensory ganglia, all neurons of the dorsal root ganglia and the majority of neurons of the cranial sensory ganglia. The ►autonomic nervous system derives entirely from the NC, comprising all sympathetic and parasympathetic branches and also the specialized ►enteric nervous system. Lying along peripheral nerves are the ►Schwann cells, a subset of peripheral crest-derived glia (Fig. 1c). In addition to specific neural derivatives, NC cells also develop into distinct endocrine and paraendocrine cells, among which the adrenomedullary chromaffin cells are the most studied ones. Notably, except for the retinal pigment, all the melanocytes of the body, which provide the vertebrate organism with a rich diversity of colors, also arise from NC. Most strikingly, at cranial levels of the axis, the NC gives rise to most of the skeleton and connective tissue of the head, face and neck. So, it appears that during evolution, the NC, rather

than cells of mesodermal origin that normally differentiate into skeletal derivatives, was able to provide a skull to cover and protect the increasing volume of the vertebrate brain (2). Furthermore, recent evidence supports the notion that the NC is also the source of molecular information that patterns certain aspects of craniofacial development through interactions with adjacent tissues (3).

Characteristics

The development of the NC embodies a multistage process initiated by the specification of neuroepithelial cells to a NC fate and followed by the separation of these progenitors from the neural tube and the onset of cell migration. Migratory progenitors follow stereotyped pathways, which differ among axial levels and species. After reaching their homing sites, NC cells finally differentiate into distinct derivatives. A central question concerning the mechanisms that control the migration of NC cells is whether there is a relationship between the directionality of cell migration and the



Neural Crest Cells and their Derivatives. Figure 1 Successive stages in NC development. (a) Slug-positive NC cells (blue product) are located in the dorsal midline of the neural tube following specification but preceding emigration. (b) In a slightly older embryo, NC cells expressing the HNK-1 epitope (brown color) are migrating through the somite. (c) Two days later, NC cells have formed the sensory dorsal root ganglia (DRG), the sympathetic ganglia (SG) and Schwann cells lining along the peripheral nerves (PN). Abbreviation: SC, spinal cord.

degree of cell specification. More precisely, do NC cells migrate in a stochastic manner along the different paths to become specified by local cues present at their homing sites, or alternatively, are there predetermined progenitors responding differentially to signals expressed along the pathways leading them to migrate in a directional fashion. Although generally considered as multipotent stem cells at the population level, it has become increasingly evident that already at the beginning of emigration from the neural tube, the NC is composed of heterogeneous populations of cells, some multipotent and others already restricted to different degrees in their developmental potential, including precursors committed to one particular fate. These results suggest that environmental signals encountered during migration and homing are likely to operate both by instructive and also by permissive mechanisms on target cells with varying degrees of developmental restriction. Thus, unraveling the relative influence of the environment *vis-à-vis* the state of commitment of the migrating progenitors is an essential issue for understanding how an initially discrete population of embryonic cells segregates into such an extensive variety of derivatives.

Cellular and Molecular Interactions

NC Specification

Formation of the NC is the result of interactions between cells of the neural plate and the non-neural ectoderm. Using expression of the *Slug* transcription factor to follow NC generation, experiments were performed in which pieces of epidermal ectoderm and neural plate were brought into contact. This juxtaposition resulted in the generation of *Slug*-positive NC cells, which were not observed in explants of each tissue separately. The possibility that the mesoderm also plays a role in NC generation has also been put forward. What molecules mediate induction of the NC phenotype? In avian embryos, treatment of naïve neural plate explants with BMP4 and BMP7, two members of the TGF β superfamily, were shown to stimulate the formation of NC cells. A \blacktriangleright BMP gradient was also suggested to act in amphibians; yet other studies highlighted the need for additional factors of the FGF and Wnt families whose exact roles remain to be resolved (4).

Emigration of NC Cells from the Neural Tube

Although being initially an integral part of the neuroepithelium (Fig. 1a), specified NC cells segregate from the dorsal midline of the neural tube and adopt a mesenchymal morphology, which is required for generating cell movement away from their source. A key issue for understanding early NC ontogeny is, therefore, elucidating the mechanisms that regulate the separation of these cells from the CNS epithelium.

Being basically a process of \blacktriangleright epithelial-to-mesenchymal conversion, the onset of migration of NC cells is spatiotemporally coordinated with the development of the \blacktriangleright somites that subsequently serve as substrates for the migrating progenitors at trunk levels of the axis.

It was found that an interplay between BMP4 and its inhibitor noggin in the dorsal neural tube generates graded concentrations of the former that in turn trigger the delamination of NC progenitors. Consistent with this suggestion, disruption of the gradient by over-expression of noggin inhibited emigration of HNK-1-positive NC cells without affecting their state of specification. BMP-4 activity is required for maintenance of a variety of dorsal neural tube genes among them *cadherin 6B*, *rhoB*, *Pax3*, *Msx 1* and *2*, etc. suggesting that at least some of these molecules are likely candidates to mediate a BMP-dependent cascade leading to NC delamination. Recently, a pivotal role for the cell-cycle in NC delamination was suggested, based on the observation that NC cells synchronously emigrate in the S-phase of the cell-cycle and that successful G1-to-S transition is a prerequisite for cells to delaminate (5).

Migration of NC Cells

NC cells advance through pathways of diverse natures. They move on extracellular matrices, along basement membranes that delimit various epithelia and blood vessel endothelia and are also able to advance by intermixing with mesenchymal cells of mesodermal origin (Fig. 1b). Most notably, several embryonic anlagen were found to inhibit the migration of NC cells and recent evidence supports the notion that these inhibitory mechanisms are pivotal to the patterning of NC migration, which in turn determines accurate morphogenesis of its derivatives.

The identity and mode of action of molecules that govern NC migration are beginning to be unraveled. In the trunk, migration of NC cells is characterized by a polarized entrance of progenitors into the rostral, but not into the caudal, domain of each somite. Embryonic manipulations of the neural tube and somitic environments have shown that differences inherent in the somitic mesoderm are responsible for the metamereric pattern of cell movement and for the consequent formation of segmented ganglia. It has been well documented that the Delta-Notch pathway is required for the formation of intra- and inter-somitic boundaries, thereby specifying initial rostrocaudal identity of somites and indirectly, the segmental migration of crest progenitors. Additional molecules were found to be differentially expressed in the somites. The majority of them are located in somitic regions that are avoided by the migrating cells, suggesting that migration of NC cells in the trunk is largely patterned by inhibitory signals. These areas include the caudal halves of the sclerotome, which convey segmental patterning, the

paranotochordal mesenchyme and the dermomyotomal epithelium. Signals with the potential to inhibit the migration of NC cells include ►**extracellular matrix**-associated molecules such as chondroitin sulfate proteoglycans bearing chondroitin 6-sulfate chains, versican, collagen IX and F-spondin. Other proteins may act *via* a cell contact-mediated mechanism. These include specific ephrins (Eph ligands), whose localization is restricted to the caudal halves of the sclerotome in several species, and Eph receptors. The interaction between them was shown to have a repulsive function towards NC cells and growing neurites. Notably, these proteins exert an avoidance behavior only when presented in a gradient-like fashion or in stepwise discontinuous concentrations but not uniformly. These results suggest that restriction to migration of NC cells on Eph ligand members is not an absolute behavioral pattern but reflects a sensitivity to the presence of a boundary between relatively repulsive and permissive substrates.

NC Cell Differentiation

Sensory and Sympathetic Lineages

Ongoing research on the molecular requirements for differentiation of peripheral neurons in sensory and sympathetic ganglia has led to the identification of neurogenic factors whose activities are specific either to sensory or to autonomic neuroblasts. Neuregulins are EGF-like growth factors that signal through tyrosine kinase receptors of the ErbB family. Mice homozygous for either the neuregulin-1, *erbB2* or *erbB3* genes reveal severe hypoplasia of the sympathetic ganglia, suggested to result from a migratory defect that leads to a failure to colonize the sympathetic primordia. In addition, BMP-2 and BMP-4, two factors present at sites of sympathetic ganglion development, were shown to act instructively on promotion of autonomic neurogenesis. The MASH-1 transcription factor is a nuclear regulatory protein of the ►**basic helix-loop-helix** class, expressed in autonomic but not sensory progenitors. Consistently, analysis of homozygous embryos with a deletion in this gene revealed that the MASH-1 product is involved in promoting differentiation of a sympathetic neuronal progenitor following its arrival in the sympathetic primordia, but it does not affect sensory neurogenesis.

The generation of sensory neurons from the NC requires the function of another class of bHLH transcription factors, the neurogenins. Neurogenin-1 was shown to be necessary for the development of proximal cranial sensory neurons while neurogenin-2 was required for distal neurons. In dorsal root ganglia, neurogenin-1 accounts for the generation of most neurons expressing the neurotrophin receptor *trkA* characteristic of small nociceptive neurons, whereas the function of neurogenin-2 is necessary for generating

trkB- and *trkC*-expressing cells that transmit proprioceptive, temperature and pressure sensations. Moreover, neurogenin-1 is required for the activation of a cascade of downstream bHLH factors including *NeuroD*. While over-expression of both factors in *Xenopus* embryos can induce ectopic neurogenesis, the timing of expression of *NeuroD in vivo* suggests that it is more likely to play a role in neuronal differentiation, while neurogenins are more likely to behave as neuronal determination genes for the sensory lineage (6).

Enteric Progenitors

The intrinsic innervation of the gastrointestinal tract derives entirely from two distinct levels of the NC, the vagal level from somitic regions 1–7 and the sacral level caudal to somite 28. While the vagal cells migrate in a rostrocaudal direction to colonize the entire length of the gut, the sacral contribution is restricted rostrally to the level of the umbilicus. It is now well established that both vagal and sacral cells have the capacity to generate neurons as well as glial cells. In recent years, significant progress has been achieved in understanding several factors that affect development of enteric progenitors. Availability of mouse mutants lacking innervation of the distal hindgut was also valuable in this respect because aganglionic megacolon is a congenital disease of humans (Hirschsprung's disease), which can be either sporadic or familial. The *lethal spotted* and *piebald lethal* mouse loci were found to encode for the endothelin-3 peptide and the endothelin-B receptor respectively. Another essential factor is glial-derived neurotrophic factor (GDNF), which signals through the c-ret receptor. Deletion of either gene caused a dramatic lack of enteric innervation. Although an initial colonization of the esophagus and stomach occurred in these mutants, these cells disappeared by the time of birth, consistent with a role of GDNF in cell survival. In contrast, no innervation of more distal regions of the gut could be detected at any stage of embryogenesis, suggestive of a primary defect in gut colonization. Based on the relative severity of the GDNF/c-ret mutations when compared to the endothelin mutations, one can postulate that GDNF acts earlier in development on a less restricted population of NC cells. *In vitro* studies revealed both mitogenic and differentiative effects of GDNF and, although endothelin-3 was also found to be mitogenic for enteric NC cells, its primary effect was to modulate the action of GDNF by inhibiting neuronal differentiation of NC cells, thus maintaining the precursor pool and ensuring that sufficient cells are present to colonize the entire gut. Furthermore, neurotrophin-3 (NT-3), a factor with multiple activities on proliferation of early NC cells and differentiation of sensory and sympathetic progenitors was also found to influence the development of NC-derived cells in the gut. NT-3 was shown to stimulate

an increase in the proportion of both neurons and glial cells by affecting their survival and/or differentiation but not as a mitogen.

Pigment Cells

Melanocytes, the pigment forming derivatives of the NC, develop following progenitor migration along a dorsolateral pathway underneath the ectoderm. This spatially distinct distribution of NC cells could result from two alternative mechanisms. First, multipotent NC cells could undergo progressive restrictions as they advance along this path or after they colonize the skin and differentiate into melanocytes in response to local cues. Alternatively, NC cells may be fate-restricted prior to dispersion, which, for melanocyte progenitors, in fact begins relatively late, after migration of the main neurogenic stream of cells has taken place. ► **Clonal analysis** of NC cells taken from the lateral path revealed the coexistence of melanoblasts along with cells expressing neuronal traits. The latter were, however, found to be relatively few in number and were removed by an episode of ► **apoptosis**. On the other hand, melanogenic precursors back-grafted into the migratory paths of early embryos precociously and preferentially dispersed along the lateral pathway, suggesting an early specification towards the melanogenic fate. Consistent with this notion, the microphthalmia (*mitf*) gene product, which encodes a bHLH transcription factor that regulates melanogenesis, is expressed in avian embryos both prior to and during NC dispersal along the lateral path.

Two well-known mouse coat color mutants with defects in melanogenesis are *white dominant-spotting* (W) and *Steel* (Sl), which encode for the c-kit receptor and for steel factor, respectively. While steel factor is expressed along the pathway of melanoblast migration, c-kit is expressed by the migrating NC cells. There is evidence that different alleles of steel factor are required to support survival and/or migration of NC cells along the subectodermal space. Another ligand-receptor system involved in melanogenesis is endothelin-3 and the endothelin-B and B2 receptors. Mutations in the first two genes produce both aneural megacolon and defects in pigmentation. *In vitro* studies revealed that endothelin signaling elicits generalized NC cell proliferation and also melanogenic differentiation.

Future Perspectives

During the past few years, significant progress has been achieved in uncovering the molecular basis of NC specification, delamination, migration and differentiation. In spite of having attained only an initial knowledge of these intricate processes, new tools are rapidly emerging which will greatly assist in this endeavor. Screens at the genomic level could potentially lead to the identification of an array of

NC-specific genes characteristic of particular developmental stages and processes of interest (7). These genes could then be assembled into functional networks. To understand their function, genomic approaches should be complemented with state of the art techniques in experimental embryology. The latter field has also developed significantly. For example, the avian embryo, which is particularly suited for cell biological studies, can now also be exploited for spatiotemporally controlled gene misexpression by ► **electroporation** of DNA constructs, dsRNAs or antisense morpholinos. The evolutionary conservation of genes and genetic networks should enable us to exploit the most successful techniques in each species in order to obtain an integrated picture of the mechanisms underlying NC ontogeny.

► Bone and Cartilage

References

1. Le Douarin NM, Kalcheim C (1999) *The Neural Crest*, 2nd edn. Cambridge University Press, New York
2. Gans C, Northcutt RG (1983) Neural crest and the origin of vertebrates. A new head. *Science* 220:268–274
3. Couly G, Creuzet S, Bennaceur S et al (2002) Interactions between Hox-negative cephalic neural crest cells and the foregut endoderm in patterning the facial skeleton in the vertebrate head. *Development* 129:1061–1073
4. Knecht A, Bronner Fraser M (2003) Induction of the neural crest: a multigene process. *Nat Rev Gene* 3:453–461
5. Burstyn-Cohen T, Kalcheim C (2002) Association between the cell cycle and neural crest delamination through regulation of G1-S transition. *Dev Cell* 3:383–395
6. Ma Q, Fode C, Guillemot F et al (1999) Neurogenin 1 and Neurogenin 2 control two distinct waves of neurogenesis in developing dorsal root ganglia. *Genes Dev* 13:1717–1728
7. Gammill LS, Bronner-Fraser M (2002) Genomic analysis of neural crest induction. *Development* 129:5731–41

Neural Development

THOMAS MÜLLER

Department of Medical Genetics, Max Delbrück Center for Molecular Medicine, Berlin, Germany
thomu@mdc-berlin.de

Definition

The nervous system consists of several structural and functional subdivisions. The central nervous system, comprised of the brain and spinal cord, fulfils complex

tasks like cognition, acquisition of memory and behavior. The peripheral nervous system consists of somatic and autonomic divisions. The somatic division gathers and transmits sensory information from the periphery to the central nervous system. The autonomic division mediates visceral sensation and motor control of the viscera, smooth muscles and exocrine glands. Cells of the nervous system derive from the neural plate, with the exception of a few components of the peripheral nervous system that are generated from ectodermal placodes.

The tasks performed by the mature nervous system depend on precise connections between neurons. The neurons and their connections are generated during embryonic and postnatal development. Neural development proceeds in several steps. During gastrulation, the primitive ectoderm gives rise to the neural plate that contains neural progenitor cells. During neurulation, the edges of the neural plate rise and fuse to form the neural tube. Neuroepithelial cells acquire a particular molecular identity, which reflects their position along the antero-posterior and dorso-ventral axes. They then produce postmitotic neurons and glial cells in a spatially and temporally controlled manner. Neurons migrate to their final positions, mature, extend axons and form synapses with their target cells. Synaptic connectivity is subsequently refined by activity dependent mechanisms.

Characteristics

For a developmental biologist, it is always amazing to see how the mammalian nervous system, composed of thousands of neuron types, billions of neurons and trillions of specific connections, develops with only little variability between individuals of a particular species. How can this be achieved, given that only a limited number of genes are available that can be used to specify all neuronal characteristics? Many mechanisms involved in neural development are widely conserved and are used in flies and worms, in frogs and fishes, in birds and humans. The observed increase in the complexity of the nervous system during evolution is not paralleled by an increase in genome complexity. Thus, strategies must have evolved which allowed the controlled development of more and more complex brains using only a limited number of regulatory genes.

Development of the nervous system depends on the expression of particular genes at particular positions and times during development. Both exogenous signals and the cell-intrinsic programmes regulate the spatial and temporal pattern of gene expression in the developing nervous system. Neural development can be viewed as a series of inductive interactions between groups of cells. Due to their restricted diffusion capacity, inducing signals act over limited distances

only and the signals received by a particular cell therefore depend on its position. The competence of a cell to respond to an inducer depends on its repertoire of receptors and signal transducing molecules, which in turn reflects its developmental history.

The complexity of the nervous system develops gradually from the simple ►neural plate. By progressive compartmentalization, the developing nervous system is subdivided into increasingly smaller units. Compartmentalization is a basic developmental strategy that is used repeatedly to generate diversity. Furthermore, molecules act in combination, thus generating a complex code. For instance, combinations of transcription factors define progenitor cells in the spinal cord. Finally, signals can be used repeatedly in consecutive steps of development, for example, FGF8 (fibroblast growth factor 8), which patterns the neural tube at different developmental stages.

Initially, progenitor cells of the developing nervous system possess a high level of plasticity that allows them to adapt their fate to extracellular cues. Subsequently, negative and positive feedback mechanisms consolidate fate decisions and thereby guarantee the stability of developmental processes. A second mechanism that ensures stability is redundancy. For instance, several signals can be employed to provide directional cues during axonal pathfinding. In this way, organisms can compensate for genetic and epigenetic variability and ontogenesis can proceed reproducibly. We will now consider in more detail the following successive steps of neural development: (1) induction and patterning of the nervous system, (2) proliferation and differentiation of neural cells, (3) guidance of neuronal migration and axon growth, (4) synapse formation.

Regulatory Mechanisms & Molecular Interactions Induction and Patterning of the Nervous System

Early in development, the neural tube is established and the neural progenitors assume positional identities. This process can be described in three major developmental steps: induction of neural progenitor cells in the neural plate, formation of the neural tube and establishment of the antero-posterior and dorso-ventral axes.

Induction of the Nervous System

The primitive embryonic ectoderm (epiblast) gives rise to neural tissue and epidermis. Neural induction, i.e. the specification of neural tissue from naïve ectodermal cells, starts at the late blastula stage and proceeds during ►gastrulation. It depends on two signals, one provided by the organizer, which resides in the epiblast in the anterior region of the primitive streak, and the other provided by the extraembryonic hypoblast, called the anterior visceral endoderm in the mouse.

As yet, there is no consensus about the molecular mechanisms that control neural induction. Diverse signaling pathways might cooperate. In frogs and fish, abrogation of ►BMP (bone morphogenetic protein) signaling promotes neural specification and high levels of BMP suppress neural development. The “neural default model” proposes that ectodermal cells become neural unless exposed to BMPs. It is suggested that BMP antagonists suppress BMP activity and allow neural induction. However, studies in other species suggest that suppression of BMP activity is not sufficient to induce neural identity and that earlier signals, most likely FGFs, establish a “preneural” state prior to gastrulation. Furthermore, suppression of Wnt signaling seems to be required for neural induction (1).

Anterior and Posterior Neural Plate

In parallel with neural induction, an early antero-posterior pattern is established in the neural plate. Nodals, FGFs, Wnts, BMPs and retinoic acid have been proposed as factors that impose a posterior character on the neural plate. The anterior visceral endoderm and the anterior axial mesoderm produce inhibitors of such signals, for instance Noggin, Chordin, Cerberus, Frzb1 and Dickkopf-1. These molecules protect the anterior neural plate from posteriorization. The presence of the transcription factor Otx2 in the anterior visceral endoderm and the anterior mesendoderm is essential to induce and maintain anterior character in the neural plate. A combination of several signals establishes the early antero-posterior division of the neural plate (2).

Neurulation

During the process of neurulation, the neural plate bends at the ventral midline. The lateral edges of the neural plate thicken, fold upwards and form the neural folds, which eventually fuse to create the neural tube. The original latero-medial axis of the neural plate becomes the dorso-ventral axis of the neural tube. At the time of neural tube closure, the neural crest becomes specified at the border of the neural plate and the surface ectoderm. Subsequently, the neural crest cells emigrate to form the peripheral nervous system. The current knowledge about molecules involved in neurulation was reviewed recently (3).

Patterning of the Neural Tube

Two general sets of patterning mechanisms along the antero-posterior and dorso-ventral axes establish a grid-like organization in the neural tube. Along the antero-posterior axis, six subdivisions of the central nervous system can be distinguished morphologically in the mouse at embryonic day E9.5 (five vesicle stage): (i) telencephalon, (ii) diencephalon (telencephalon and

diencephalon together constitute the prosencephalon or forebrain), (iii) mesencephalon (midbrain), (iv) metencephalon (which will give rise to the pons and cerebellum), (v) myelencephalon (which will give rise to the medulla) and (vi) the spinal cord. Metencephalon and myelencephalon together form the rhombencephalon (hindbrain).

Different paradigms are used for the patterning along the antero-posterior axis. The myelencephalon is segmented into rhombomeres. The mes- and metencephalon are patterned by gradients of ►morphogens, which are secreted by the isthmus organizer at the junction of mid- and hindbrain. ►Morphogens are paracrine signals that act over distance in a concentration-dependent manner. In the forebrain, the two mechanisms have been combined; the forebrain is segmented into prosomeres, which are patterned by morphogens that are released by the anterior neural ridge and the zona limitans intrathalamica.

Non-neural tissues correspond to early signaling centers that pattern neural tissues along the antero-posterior axis. Signals from non-neural ectoderm, the organizer, axial and paraxial mesoderm are involved. In response to these, secondary signaling centers arise within the neural tube (2). The isthmus organizer is a secondary signaling center that is formed at the mid-/hind-brain border and secretes the patterning signals FGF8 and Wnt1. Another signaling center, the anterior neural ridge, arises at the border of the anterior neural plate and the ectoderm. FGF8 is an important morphogen secreted by this structure. Reduction in the FGF8 level leads to forebrain truncations. The zona limitans intrathalamica, a further secondary signaling center, forms in the diencephalons. One important morphogenic molecule produced in the zona limitans intrathalamica is Sonic hedgehog (Shh).

Along the dorso-ventral axis, four subdivisions, the floor, basal, alar and roof plates, can be distinguished in the entire neural tube. Initially, signals that establish the dorso-ventral polarity are provided ventrally by the prechordal plate and notochord and dorsally by the surface ectoderm. The ventral signaling center is subsequently transferred from the prechordal plate and notochord to the floor plate. Similarly, the dorsal signaling center is transferred from the surface ectoderm to the roof plate.

Shh is the principal ventralizing signal (4). It is initially expressed by the axial mesoderm and later by the floor plate. Shh acts as a morphogen, i.e. it induces ventral neuronal and glial cell types in a concentration dependent manner. Members of the TGFβ (transforming growth factor-β) family provide dorsalizing signals (5). BMPs, dorsalin-1 and growth differentiation factors (GDFs) are secreted in the roof plate. BMPs induce dorsal neuron types in naive spinal cord explants.

Homeobox and bHLH Transcription Factors Define Regional Identity

A general readout of the patterning signals in the neural tube is the expression of patterning genes encoding mainly ►homeobox and ►basic-helix-loop-helix transcription factors in restricted progenitor domains. In the ventral spinal cord, for instance, graded Shh signaling specifies the identity of progenitors by regulating the expression of transcription factors that fall into two major groups, a set of class I proteins that are repressed by Shh signaling and a set of class II proteins that are induced by Shh. These proteins function primarily as transcriptional repressors and their selective cross-regulatory interactions help to establish specific neural progenitor domains and to sharpen the boundaries between these domains. The established subdivisions manifest themselves by the presence of a specific combinatorial code of transcription factors. This code subsequently instructs the fate of the cells generated by the progenitors, i.e. different neuronal and glial cell types (6).

A particular subfamily of homeobox factors encoded by the genes of the *Hox* clusters patterns the hindbrain and the spinal cord. The hindbrain is segmented into rhombomeres and the anterior border of the expression of many *Hox* genes coincides with rhombomere boundaries. A combinatorial code of Hox factors thus defines the rhombomere identity and the types of neurons generated therein. The most posterior rhombomeres and the spinal cord are not obviously segmented. Nevertheless, expression of *Hox* genes shows antero-posterior specificity and *Hox* gene expression instructs axial level specificity in the generation of motor neuron subtypes (7). A hallmark of *Hox* gene expression in the hindbrain is the regulation by retinoic acid. Furthermore, ephrins and their Eph-receptors are expressed in specific rhombomeres and are targets of patterning genes in the hindbrain. Eph signaling is suggested to mediate repulsion between cells of neighboring rhombomeres and to compartmentalize the hindbrain.

Generation of Neuronal and Glial cells

In parallel to the patterning mechanisms that control regional identity, mechanisms exist that control the timing of proliferation and differentiation of progenitor cells in the neural tube.

Proliferating progenitor cells of the central nervous system reside within the ►ventricular zone of the neural tube. They have stem cell features, i.e. they can self-renew, are multipotent and give rise to multiple types of postmitotic neurons and glial cells. Postmitotic neurons migrate out of the ventricular zone and into the mantle layer of the neural tube. Different neuron types, oligodendrocytes and astrocytes are generated in a

defined temporal order and at particular positions in the neural tube. Neuronal cells are generally produced in an overabundance and the superfluous cells are removed by apoptosis.

Proliferation and differentiation of neural progenitor cells is regionally and temporally controlled during development. Cell numbers in the nervous system are determined by the decision of a dividing progenitor cell to generate two proliferative progenitors or to generate one or two postmitotic daughter cells. ►Proneural transcription factors and Notch signaling influence the decision as to whether a neuroepithelial cell continues to proliferate or begins to differentiate.

Proneural bHLH Factors

Members of the achaete-scute and atonal families, such as Mash1, Math1 and Neurogenins 1 and 2, are bHLH transcription factors that have proneural activity (8). They function as heterodimers with ubiquitous bHLH factors known as E proteins. Once proneural factors have accumulated at a sufficient level, they induce cell-cycle exit and differentiation. The expression of proneural genes in individual neural progenitors is only transient. Proneural genes are down-regulated before progenitor cells exit the proliferative zone and begin to differentiate. The ability of proneural genes to promote neuronal differentiation thus relies on the induction of target genes that implement the differentiation programmes. Genes of the bHLH family, such as *NeuroD*, or homeobox genes, such as *Hb9*, can consolidate the neuronal differentiation programme.

Proneural genes are often expressed in restricted progenitor domains in the neural tube. This observation and the structural diversity of these factors suggested that they possess not only proneural activity but also specify neuronal identity. The dorsal embryonic spinal cord provides an example of the functional differences between the different bHLH factors Math1, Neurogenin-1 and -2 and Mash1. The factors are expressed in discrete, neighboring and non-overlapping progenitor domains that produce distinct types of dorsal interneurons. The analysis of Math1 and compound Neurogenin-1 and -2 mouse mutants has shown that these proneural factors are indeed essential for the correct specification of specific interneurons. Thus, proneural bHLH genes appear to determine also the acquisition of particular neural fates.

Proneural factors do not only trigger the differentiation of neural progenitors, but are also instrumental in coupling the onset of differentiation with cell-cycle exit. Induction of the cell-cycle exit by proneural factors coincides with the expression of genes that encode cyclin-dependent kinase inhibitors, such as p16, p21 and p27, which inhibit the entry into S-phase.

Notch Signaling Mediates Lateral Inhibition

The activity of proneural factors has to be tightly controlled to preserve the progenitor pool over the extended period of neurogenesis. This is achieved by an inhibitory feedback mechanism called ►**lateral inhibition**, i.e. differentiating cells provide signals to their neighbors that instruct them to remain progenitors. Proneural factors not only induce differentiation but also the expression of genes that encode the Notch ligands Delta and Serrate/Jagged. The presence of these ligands in one cell activates Notch in the neighboring cells. Notch activation induces the expression of target genes, for instance those that encode the Hairy-Enhancer-of-split/Hes bHLH factors, which inhibit the proneural factors. Thus, a cell that receives Notch signals maintains its progenitor status. Since cells that express proneural factors and Notch ligands differentiate and leave the ventricular zone, Notch is only transiently activated, which allows the selection of new cells for the next round of differentiation.

Lateral inhibition has been well studied in *Drosophila* and illustrates the mechanism by which single neural precursor cells are selected in the ectoderm from a cluster of equivalent cells. Subsequently, Notch regulates the acquisition of different fates by the two daughter cells that are generated by a single precursor cell. In this context, Notch signaling is regulated by other proteins, among them the cytoplasmic protein Numb that binds to Notch and inhibits Notch signaling. During division of a precursor cell, Numb becomes asymmetrically distributed to one of the two daughter cells where it antagonizes Notch signaling. The difference in Notch activity determines the distinct cellular identities of the daughter cells.

Two Numb homologs, Numb and Numb-like, are found in vertebrates. Compound Numb/Numb-like mutant mice show hyperproliferation of cortical progenitor cells and impaired neuronal differentiation (9). The proportion of progenitor cells that gave rise to two progenitor daughter cells, as opposed to progenitors that generated one or two differentiating daughter cells, was greatly increased. Thus, Numb and Numb-like control asymmetric cell divisions in the nervous system of vertebrates and have similar functions to *Drosophila* Numb.

Patterning and Differentiation

The expression of most patterning genes ceases when neural precursor cells exit the cell-cycle and begin to differentiate. Patterning information is passed from the progenitors to their progeny by induction of a new set of transcription factors, which instruct the subsequent differentiation of postmitotic neurons. This mechanism is understood best in the spinal cord for the induction of motor neurons. There, the expression of the *Pax6*, *Nkx6* and *Olig2* genes defines the motor neuron

progenitor domain. *Olig2* acts as a bHLH factor that has patterning function and also induces the proneural bHLH factor Neurogenin-2. Together these factors induce in the emerging motor neurons the synthesis of transcription factors, such as *Lhx3*, *Isl1/2* and *Hb9* that specify and consolidate the motor neuron identity. Depending on the axial level, motor neurons subsequently acquire different identities. Axial level specific expression of genes of the *Hox* clusters specifies this identity. Visceral motor neurons that innervate sympathetic ganglia develop only on thoracic spinal cord levels. Their differentiation is induced by *Hoxc9*. The lateral motor column arises only at limb levels and contains motor neurons that innervate limb muscles. Differentiation of brachial lateral motor column neurons is specified by *Hoxc6* (7).

Generation of Glial Cells

Glia comprises 10–20% of the cells in the *Drosophila* nervous system and more than 90% of the cells of the human brain. The major glial types in the central nervous system are ►**oligodendrocytes** and ►**astrocytes**, which derive from the neuroepithelium. Oligodendrocytes form myelin sheaths, which allow faster conductance. Astrocytes are diverse in phenotype and function and, for instance, maintain the blood-brain barrier, produce neuropeptides and modulate synaptic transmission. Glia of the peripheral nervous system comprises ►**Schwann cells**, the myelinating cells in peripheral nerves, and non-myelinating cells, for instance the satellite cells in the ganglia. Like peripheral neurons, ►**glial cells** of the peripheral nervous system derive from the neural crest.

Oligodendrocyte precursor cells are specified in the neural tube by a mechanism that is similar to the specification of neurons. In the spinal cord and hindbrain, oligodendrocytes derive from the same set of progenitors that generated motor neurons at earlier stages. The functional analysis of *Olig-1* and *-2* genes provided an insight into the molecular link between motor neuron and oligodendrocyte development (10). In the mouse spinal cord, *Olig-2* is essential for specification of motor neurons between E9 and E10.5 and for the subsequent specification of oligodendrocyte precursors at E12.5. *Olig-1/2* positive progenitors in the brain and spinal cord appear to be restricted in their fate and generate neurons and oligodendrocytes but not astrocytes. Astrocytes are thought to arise in the central nervous system after neurogenesis has ceased. On a molecular level, the specification of astrocytes has not been characterized.

Peripheral glial cells differentiate from pluripotent ►**neural crest cells**. *Sox10*, a member of the ►**high-mobility-group transcription factor** family, is a key regulator in the differentiation of peripheral glial cells (11). In *Sox10* mutant mice, neuronal cells form in

dorsal root ganglia, but Schwann cells and satellite cells are not generated. Sox10 controls the expression of *Notch1* in neural crest cells. The expression of activated Notch in undifferentiated neural crest cells inhibits neuronal and promotes Schwann cell differentiation. Thus, Sox10 could control the acquisition of glial fates in the peripheral nervous system *via* induction of Notch, which subsequently promotes glial differentiation. Sox10 also controls the expression of *ErbB3*, a Neuregulin-1 receptor in Schwann cells, that controls growth and proliferation during their development. Motor and sensory neurons produce Neuregulin-1 and this neuron-derived signal adjusts the number of Schwann cell progenitor cells during development. Neuron-derived Neuregulin-1 also influences the thickness of the myelin sheath and thus regulates Schwann cell function in adulthood.

Neurotrophic Factors and Neuronal Survival

Neurons are produced in an overabundance during development and their number is adjusted by apoptosis in both vertebrates and invertebrates. What might be an evolutionary advantage of this? The peak of neuronal cell death is observed during the establishment of neuronal connections. The complexity of the nervous system is high and the correct connectivity might not be ensured solely by neuronal identity and controlled axonal pathfinding. Instead, correct connectivity appears to be achieved by an elimination of those neurons that made inappropriate contacts. In addition, the final numbers of neurons in a neuronal circuit can be adjusted by this mechanism.

The ►**neurotrophic factor hypothesis** suggested that the target of innervation of a particular neuron produces a limited amount of a neurotrophin, i.e. a molecule that is essential for the survival of the neuron. Originally it was assumed that neurotrophic factors are essential nutrients for neurons, but their subsequent characterization demonstrated that they are signaling molecules that act anti-apoptotically. The classic neurotrophins are members of the nerve growth factor family and signal *via* the Trk tyrosine kinase receptors and *via* the low affinity NGF receptor p75; p75 acts as a Trk co-receptor but can also signal on its own. Other factors, many of which signal *via* ►**tyrosine kinase receptors**, for instance glial-derived neurotrophic factor (GDNF) and hepatocyte growth factor (HGF), were also shown to possess neurotrophic activity. Genetic analysis in mice has demonstrated that neurotrophic factors act in a neuron type specific manner. For instance, muscle spindle derived neurotrophin-3 (NT3) is required for the differentiation and survival of proprioceptive sensory neurons.

Mutational analysis in mice has not provided clear evidence for a function of neurotrophic factors in the regulation of neuronal survival in the central nervous

system. This may be due to the redundant actions of various neurotrophins. Ectopic application of neurotrophic factors can, however, reduce *in vivo* and *in vitro* cell death of neurons. The importance of neurotrophic factors goes beyond a regulation of apoptosis and includes functions in synaptogenesis and synaptic plasticity (12).

Guidance of Neuronal Migration and Axon Growth

Neurons and glial cells can migrate from their place of origin to their final position over large distances. Within the central nervous system, two principal modes of neuronal migration can be observed, radial migration, in which cells migrate from the ventricular zone towards the surface of the brain, and tangential or non-radial migration. Most structures of the brain integrate neurons that arrive *via* radial and tangential migrations. Cues used to direct tangentially migrating neurons appear to be similar to those that direct axon pathfinding. Distinct molecular mechanisms are used during radial migration.

Radial Migration

Processes of radial glial cells guide migrating neurons during the formation of the cortical plate. Radial glial cells arise throughout the neural tube during early development. Their cell bodies are located in the ventricular zone and they elaborate a process that spans the wall of the neural tube from the ventricular (apical) to the pial (basal) surface. The term “glial” can be misleading, since these cells represent progenitor cells that can give rise to neurons and glial cells.

Layers in the cortical plate are established according to an inside-outside pattern, where deeper layers are formed by neurons of early origin and superficial layers by later born neurons. Cohorts of migrating neurons pass over previously developed neurons until they reach the marginal zone, where they detach from the radial glia. Cajal-Retzius cells are a particular neuronal cell type that originates before the cortical plate forms and resides in the marginal zone. These cells secrete Reelin, a high affinity ligand for receptors of the LRP family, VLDLR and LRP8 (also known as ApoER2). The Reelin signal is transduced by the intracellular adaptor protein Disabled-1. Mice lacking Reelin, the VLDLR and LRP8 receptors or Disabled-1 have similar migration defects in the cortex and in other brain regions. Reelin and its receptors are therefore important components in the control of radial migration of neurons.

►Axon Guidance

Many neurons extend axons over great distances before they form synapses on the appropriate target cells. Some general strategies are employed to guide growing axons to their target area. One strategy is the use of

pioneer axons, which correspond to axons that grow early in the embryo and serve as a scaffold for later growing axons. Another strategy is to break the long journey of an axon into segments and to use guidance cues for each of the intervals. Furthermore, axons grow in bundles before successively smaller groups of axons diverge at specific sites along the path. This strategy allows the iterative use of a limited number of signaling molecules for axon guidance. Furthermore, axons are able to modulate their responsiveness along their route and guidance cues are often multifunctional depending on cellular context.

The growth cone is the essential cellular compartment for the steering of axons. It constitutes a specialized terminal apparatus and fulfils both sensory and motor functions. Growth cones contain numerous receptors for environmental cues and transmit them to cytoskeletal proteins and actin-based motors that generate either forward movement or growth cone collapse (13). Molecules that Direct Axon Pathfinding

Growth of axons is guided by their substrate preferences and by directional cues. Substrates for growing axons can be extracellular matrix, cell surface or other axons in axon fascicles. Molecules of the extracellular matrix like laminins, collagens and fibronectin are recognized by receptors of the integrin family. Proteoglycans can modulate axon growth through interactions with neural adhesion molecules. Cadherins and immunoglobulin-like adhesion molecules mediate differential cell-cell adhesion.

Directional cues may act over long or short distances, may be bound to the extracellular matrix or to the cell surface or may be soluble and produced by a spatially defined source. They can act as chemoattractant or chemorepellent. Here, I will briefly summarize the major families of guidance cues and their receptors (13).

Netrins were discovered as soluble chemoattractants for vertebrate commissural axons. Across species, their function of attracting axons toward the midline has been conserved. Netrins can also act as repulsive factors. One netrin receptor, deleted-in-colorectal-cancer (DCC), mediates attraction to netrins but can also participate in repulsion. The second netrin receptor, *unc-5*, transmits repulsive signals alone or in combination with DCC.

Slits are secreted proteins that signal through Roundabout (Robo) family receptors. Slits and Robo receptors were identified as mediators of short-range signals that repel axons at the midline. Slit was also purified as a factor that stimulates sensory axon branching and elongation.

Semaphorins are a family of ligands encoded by more than 20 genes in mammals and were initially purified as inducers of growth cone collapse in sensory

neurons. Semaphorins seem to act predominantly as short-range or contact-mediated repulsive cues. Semaphorins can be secreted, attached to the cell surface through a GPI-anchor or be transmembrane proteins. Transmembrane semaphorins can back-signal, i.e. they can also function as receptors. Semaphorins signal through multimeric receptor complexes. Transmembrane semaphorins bind to receptors of the plexin family. Secreted semaphorins require neuropilins as co-receptors.

Ephrins were identified in the search for graded signals that instruct the formation of the topographic map of retinotectal projections in the optic tectum. They signal through a large family of tyrosine kinase receptors, the Eph-receptors. Ephrins are membrane anchored either by a GPI-anchor (ephrinA) or through a transmembrane domain (ephrinB), and they can function as back-signaling receptors. Ephrins and Eph-receptors mediate repulsive and attractive cues.

Recently, morphogens such as Shh, BMPs and Wnts and neurotrophins like NGF were also shown to possess axon guidance activity. The canonical signal transduction pathways by which these factors control cell fate are well studied and involve transcriptional regulation. Non-canonical pathways coupled directly to the cytoskeleton are expected to mediate the axon guidance activity of these morphogens.

Synapse Formation

How are precise synaptic connections between specific pre- and postsynaptic neurons established? In 1963, Roger Sperry proposed that “highly specific cytochemical affinities” between an axon and its target cell determine the establishment of axonal connections. One key feature of this model is that the precise connectivity develops as a predetermined process and is independent of neural activity. Instead, the establishment of proper connections between a neuron and its target is thought to rely on a recognition process, i.e. the matching of adhesive surface molecules. These molecules might include cadherins, immunoglobulin superfamily cell adhesion molecules and ephrins and Eph receptors (14, 15). Further synapse formation relies on signals exchanged between axon and target cell. Activity-dependent mechanisms refine the pattern of connectivity after synaptic connections are established. The [▶neuromuscular junction](#) is the best-understood synapse and provides a paradigm for understanding the molecular mechanisms of formation and activity-dependent refinement of synapses.

Initiation of Synapse Formation

The neuromuscular junction constitutes a specialized cholinergic synapse between a motor neuron and a skeletal muscle cell (16). It comprises a motor axon

terminal, a muscle fiber and glial cells. All three acquire specialized features in the neuromuscular junction. The axon terminal is rich in synaptic vesicles, which are clustered at dense patches on the presynaptic membrane called active zones, where they can fuse with the plasma membrane to release neurotransmitter into the synaptic cleft. The surface of the muscle fiber forms a series of indentations opposite to the active zones, called junctional folds. These postsynaptic sites are rich in acetylcholine receptors and specialized cell adhesion and signaling molecules and contain an intricate cytoskeletal network. A muscle fiber (also called a myotube) arises through fusion of myoblasts and is therefore a multinucleated cell. mRNAs coding for components of the postsynaptic apparatus, such as the different subunits of the acetylcholine receptor, are only transcribed in synaptic nuclei underlying the neuromuscular junction after the synapse has formed. The formation of the neuromuscular junction is initiated when a motor axon approaches a muscle fiber. The axon terminal senses some as yet unidentified cue on the muscle cell, the contact is made and pre- and post-synaptic differentiation begins. Most synaptic components of the motor neuron and the myotube are already synthesized before the synapse forms. For example, motor neurons form synaptic vesicles and produce neurotransmitters in the absence of muscle. Uninnervated myotubes synthesize functional acetylcholine receptors, which can form clusters and associate with components of the cytoskeleton. Exchange of signals between motor axon and myotube is however required to coordinate the assembly of the pre- and post-synaptic apparatus in a spatially and temporally controlled manner.

Agrin is a large proteoglycan that is secreted by motor neurons and incorporated into the synaptic basal lamina. Agrin induces postsynaptic differentiation and the number, size and density of acetylcholine receptor clusters is severely reduced in *Agrin* mutant mice. The observed phenotype indicates that Agrin provides the initial signal for the assembly of the pre- and post-synaptic apparatus.

The muscle-specific tyrosine kinase (MuSK) appears to be required for the recognition of Agrin and for the formation of neuromuscular junctions. Agrin induces the kinase activity of MuSK, can be chemically crosslinked to MuSK, but does not bind MuSK directly. The phenotype of *MuSK* mutant mice is very similar to that of *Agrin* mutants, i.e. postsynaptic specialization is severely impaired. Moreover, motor axons behave abnormally in *Agrin* and *MuSK* mutant mice, i.e. they continue to sprout after reaching the muscle and do not initiate presynaptic differentiation.

Rapsyn is required for postsynaptic differentiation and acts downstream of the MuSK signal in the muscle. It is

a peripheral membrane protein that colocalizes with acetylcholine receptors. MuSK concentrates at postsynaptic sites and becomes activated in the absence of Rapsyn, but clustering of acetylcholine receptor and other postsynaptic components fails. MuSK is thus a critical component of a primary postsynaptic scaffold, which recruits Rapsyn, as well as acetylcholine receptors and other proteins.

Postsynaptic specialization of neuromuscular junctions includes synapse specific gene expression, which appears to be independent of Rapsyn and the clustering of acetylcholine receptor. Neuregulin-1 (also known as acetylcholine receptor inducing activity, ARIA) can stimulate acetylcholine receptor expression in muscle cells *in vitro* and was proposed to regulate synapse-specific gene expression. While transcription of genes encoding acetylcholine receptor subunits is increased in synaptic nuclei, their expression is suppressed in non-synaptic nuclei. Activity induced changes in intracellular Ca^{2+} concentrations are thought to trigger a signaling cascade that reduces acetylcholine receptor gene expression in non-synaptic nuclei.

The muscle fiber organizes the differentiation of the presynaptic axon terminal. The change in transmission efficacy, accumulation of synaptic vesicles and formation of active zones require retrograde signals from the muscle fiber. Little is known about the nature of these signals. Candidate molecules include cell adhesion molecules like N-CAM and N-cadherin, soluble factors like brain-derived neurotrophic factor (BDNF) and components of the synaptic basal lamina like laminin-11.

Synapse Elimination at Neuromuscular Junctions

In the adult skeletal muscle, one motor neuron innervates many muscle fibers, but the majority of muscle fibers are typically innervated by a single axon. During development, however, several axons synapse with one muscle fiber, and all but one axon contact are eliminated postnatally. This is not a consequence of neuronal death.

Polyneuronal innervation of target cells and subsequent elimination of synapses are often observed when groups of neurons with similar or identical function innervate homogenous populations of target cells. This is, for instance, the case when motor neurons innervate individual muscle fibers. Apparently, mechanisms of axon guidance and target specificity do not explicitly instruct the connectivity pattern within a particular muscle. Synapse elimination occurs gradually over time and requires muscle activity. The process of synapse elimination seems to begin in the postsynaptic membrane, with a decrease in density of acetylcholine receptors and other postsynaptic proteins. The muscle fiber becomes gradually unresponsive to stimulation by

this axon terminal. Loss of the postmitotic apparatus causes the retraction of the motor axon, but the nature of the retrograde signal is currently unclear. Synapse loss is initiated by an activity dependent competition between different motor axons that innervate a single muscle fiber (18). This form of heterosynaptic depression satisfies the Hebbian rule; synaptic efficacy is stable or strengthened if pre- and post-synaptic activities are coincident and weakened if they are non-coincident. The molecular mechanisms leading to synapse destabilization are as yet unclear, but it is known that repetitive elevation of Ca^{2+} levels in the muscle fiber is necessary and sufficient for heterosynaptic depression.

Concluding Remarks

Developmental neuroscience is an exciting and diverse field of research. Recent advances in technologies such as mouse genetics, microarray expression profiling and imaging have promoted our molecular understanding tremendously. In this essay, I have tried to give a general overview of developmental strategies and molecular mechanisms that guide neural development. More detailed information is provided in the review articles cited. For more comprehensive descriptions of neural development, I would refer the reader to two textbooks (18, 19).

References

1. Wilson SI, Edlund T (2001) Neural induction: toward a unifying mechanism. *Nat Neurosci* 4:1161–1168
2. Wilson SW, Houart C (2004) Early steps in the development of the forebrain. *Dev Cell* 6:167–181
3. Copp AJ, Greene ND, Murdoch JN (2003) The genetic basis of mammalian neurulation. *Nat Rev Genet* 4:784–793
4. Jacob J, Briscoe J (2003) Gli proteins and the control of spinal-cord patterning. *EMBO Rep* 4:761–765
5. Lee KJ, Jessell TM (1999) The specification of dorsal cell fates in the vertebrate central nervous system. *Annu Rev Neurosci* 22:261–294
6. Shirasaki R, Pfaff SL (2002) Transcriptional codes and the control of neuronal identity. *Annu Rev Neurosci* 25:251–258
7. Dasen JS, Liu JP, Jessell TM (2003) Motor neuron columnar fate imposed by sequential phases of Hox-c activity. *Nature* 425:926–933
8. Bertrand N, Castro DS, Guillemot F (2002) Proneural genes and the specification of neural cell types. *Nat Rev Neurosci* 3:517–30
9. Li HS, Wang D, Shen Q et al (2003) Inactivation of Numb and Numblike in embryonic dorsal forebrain impairs neurogenesis and disrupts cortical morphogenesis. *Neuron* 40:1105–1118
10. Rowitch DH (2004) Glial specification in the vertebrate neural tube. *Nat Rev Neurosci* 5:409–19
11. Britsch S, Goerich DE, Riethmacher D et al (2001) The transcription factor Sox10 is a key regulator of peripheral glial development. *Genes Dev* 15:66–78
12. Huang EJ, Reichardt LF (2003) Trk receptors: roles in neuronal signal transduction. *Annu Rev Biochem* 72:609–642
13. Huber AB, Kolodkin AL, Ginty DD et al (2003) Signaling at the growth cone: ligand-receptor complexes and the control of axon growth and guidance. *Annu Rev Neurosci* 26:509–563
14. Shen K (2004) Molecular mechanisms of target specificity during synapse formation. *Curr Opin Neurobiol* 14:83–88
15. Benson DL, Colman DR, Huntley GW (2001) Molecules, maps and synapse specificity. *Nat Rev Neurosci* 2:899–909
16. Sanes JR, Lichtman JW (1999) Development of the vertebrate neuromuscular junction. *Annu Rev Neurosci* 22:389–442
17. Fitzsimonds RM, Poo MM (1998) Retrograde signaling in the development and modification of synapses. *Physiological Reviews* 199878:143–170
18. Kandel ER, Schwartz JH, Jessell TM (2000) *Principles of Neural Science*. McGraw-Hill, New York
19. Zigmond MJ, Bloom FE, Landis SC et al (2002) *Fundamental Neuroscience*. Academic Press, San Diego

Neural Networks

Definition

Neural networks are paradigms for computing based on the processing/memory abstraction of human information processing. The approach is represented by a multi-layer structure composed of widely connected nodes with the ability to learn. Neural networking derives meaning from complicated or imprecise data, and can be used to extract patterns too complex to be detected by other techniques.

► [EST Mining for Expression Analysis](#)

Neural Plate

Definition

Neural plate describes a portion of the dorsal ectoderm that is specified to become neuroectoderm during gastrulation. Its cells become columnar by appearance. This region of the embryo is called neural plate.

► [Neural Development](#)

Neural Stem Cells

GERD KEMPERMANN

Gerd Kempermann, Max-Delbrück-Center for Molecular Medicine (MDC) Berlin-Buch, Berlin, Germany
gerd.kempermann@mdc-berlin.de

Definition

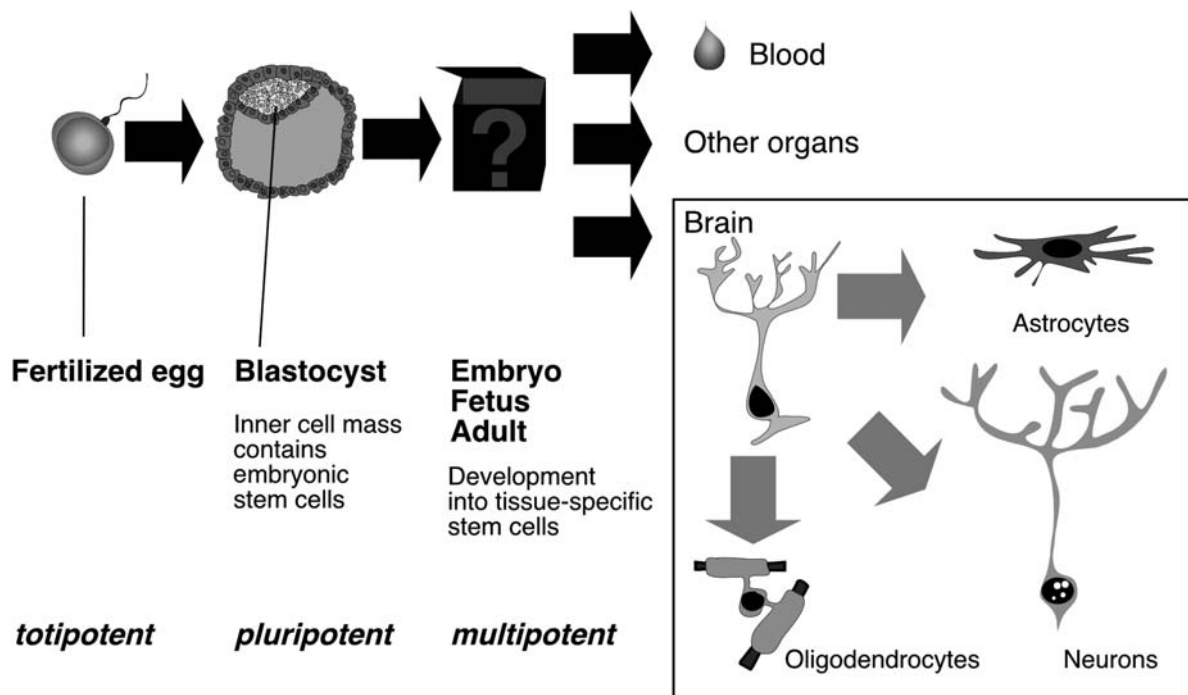
Neural **stem cells** are self-renewing and undifferentiated cells, from which the nervous system or parts of it can develop. Neural stem cells are **multipotent**, which means that their potential for differentiation is limited to cell types of their own germ layer, the neuroectoderm, but not of mesoderm or endoderm. Embryonic stem cells, in contrast, are **pluripotent**, because they can generate cells for all three germ layers. The main lineages of neural tissues into which neural stem cells can develop are neurons, astrocytes and oligodendrocytes. Brain cells that are not of neuroectodermal origin, such as microglia, macrophages, lymphocytes, meningeal cells and the vasculature are not derived from neural stem cells. Neural

stem cells can be found in the developing brain, but they persist during adulthood. In two regions of the adult brain, hippocampus and olfactory bulb, neural stem cells continue to produce neurons throughout life (**adult neurogenesis**) (Fig. 1).

Characteristics

General Principles

A number of defining principles apply to neural stem cells as to all other types of stem cells. In general, stem cells have two characteristics, the ability to self-renew and the potency to generate two or more different tissue or cell types (multipotency). Self-renewal means that cell division generates at least one identical copy of the mother cell. Symmetric stem cell divisions produce two identical copies; asymmetric divisions yield one stem cell identical to the mother cell and one cell which is relatively more determined towards a specific lineage of differentiation than the stem cell. These daughter cells show reduced self-renewal but increased lineage restriction and are referred to as "**progenitor cells**". Irrespective of their limited **self-renewal**, progenitor cells are more proliferative than stem cells and can vastly expand the number of new cells. They are thus often referred to as "transiently amplifying progenitor cells". Cells with different "degrees of stemness" can be



Neural Stem Cells. Figure 1 Neural stem cells are stem cells within the neuroectoderm. They generate neurons, oligodendrocytes and astrocytes. Like all other somatic or tissue-specific precursor cells they are multipotent. Embryonic stem cells, in contrast, are found in the inner cell mass of the blastocyst, before the germ layers have formed. Embryonic stem cells are pluripotent. Neural stem cells can be found in the developing and adult brain.

identified in the brain (and elsewhere), but in a concrete situation the categorization of a cell as a stem cell or a progenitor cell is impossible. The term “▶precursor cell” is used as an umbrella term encompassing both stem and progenitor cells. In most contexts of brain research the use of the term “stem cell” can be misleading, because tests of “stemness” have not actually been carried out.

Stem cells are cells in which more of the genomic information is accessible than in differentiated cells. Increasing tissue and lineage restriction of precursor cells is equal to an increasing limitation of this accessibility. It is not known how the stem cell potential is regulated on the chromatin level. But precursor cells are heterogeneous and their developmental potential is context-sensitive with regard to both space and time. Precursor cells from the ventricular wall produced hippocampal granule cells upon implantation into the dentate gyrus. Hippocampal precursor cells, when implanted into the olfactory system, generated new olfactory neurons. The earlier in development the transplantation occurs the greater the developmental potential. Early neural precursor cells implanted in later stage embryos integrated site and time specifically. In contrast, precursor cells from later stages implanted into earlier developmental stages failed to differentiate to the same extent as the early precursors. This argues in favor of an increasing limitation of potentials with progressing development. In principle, temporal identity is less flexible than spatial identity.

Neural Stem Cells *In Vitro* and *In Vivo*

In the developing and adult ▶brain, stem cells are found in the germinative matrices that originate from the ventricular wall.

Two protocols exist to bring neural stem cells into culture, adherent cultures and cell suspensions in which aggregates form (so-called “neurospheres”). The ability to form these aggregates alone does not prove the “stemness” of the cells. Rather it has to be demonstrated that single cells isolated from such spheres can clonally generate new secondary spheres. Cells are grown in serum-free conditions with fibroblast-growth factor 2 (FGF2) or epidermal growth factor (EGF) as mitogens. Both isolation strategies differ fundamentally from the isolation of hematopoietic stem cells for which a number of surface antigens that can be used for a prospective isolation based on the binding of specific antibodies are known. Stem cells are then defined as cells of a particular antigen profile (1). A similar strategy has been applied to identifying stem cells in the brain but had only a low yield. The relationship of the isolated cells to stem cells identified by the other methods is not known.

In vitro, multipotency is routinely assessed after inducing differentiation in stem cell cultures by withdrawing growth factors and adding serum to the culture

medium. The detection of cells from the three main neural lineages is taken as proof of multipotency. In most studies the identification of neurons, astrocytes and oligodendrocytes is based on the immunocytochemical detection of key antigens, such as beta-III-tubulin, Map-2 or ▶NeuN for neurons, O4 for oligodendrocytes and glial fibrillary acidic protein (▶GFAP) for astrocytes. Electrophysiology has also been used to confirm the function of stem cells and their progeny.

One important concern besides the marker sensitivity and specificity is that the culture conditions themselves might induce multipotency in the precursor cells. There is so far no direct evidence that neural stem cells are multipotent *in vivo*, because this conclusion is based on an analogy drawn from *in vitro* studies. Transplantation studies have generally indicated that local cues direct the realization of the developmental potential. Similarly, it has been argued that culture conditions determine the properties of the precursor cells, with different sets of culture conditions yielding incommensurable results. This does not mean that precursor cell identities are “culture artifacts”, but that precursor cell properties can be determined by external cues. This is in contrast to the view that “stemness” is entirely an intrinsic property. The extreme version of this idea is that “stem cell” is not a cellular property at all, but a functional state (2). In this sense, a population of cells, susceptible to these local stemness-inducing cues circulates through the body and context-dependently performs different types of stem cell function.

Neural Stem Cells in the Adult Brain

Although neural stem cells can be isolated from many brain regions including white matter tracts, in only two neurogenic regions of the adult brain, the ▶hippocampus and the olfactory system, are new neurons generated throughout adulthood. This “adult neurogenesis” has been confirmed for most scientifically relevant mammalian species including primates and humans. The functional relevance of adult neurogenesis is under intensive investigation. The hypothesis for the hippocampus is that the new neurons contribute to the adaptation of the hippocampal network to meet the levels of complexity and novelty experienced by the individual, thus optimizing information processing during learning. Consequently, a failure of adult hippocampal neurogenesis has been connected with hippocampus-related disorders such as dementias, major depression and temporal lobe epilepsy.

Except for the dentate gyrus and the olfactory system the adult brain appears to be non-neurogenic under normal conditions. Reports about physiological neurogenesis in other brain regions have so far remained controversial.

The lack of prospective stem cell markers causes particular problems when studying stem cells *in vivo*.

Intermediate filament ►**nestin** can be used to identify putative stem or progenitor cells, but neither sensitivity nor specificity are known. Nestin shows a surprising overlap with the expression of the astrocytic marker GFAP.

Neural stem cells of the adult subventricular zone have astrocytic properties, express GFAP and have a cilia-bearing process contacting the ventricular surface. They are called B cells and generate a population of transiently amplifying progenitor cells, called C cells, which in turn produce migrating neuroblasts, called A cells (3). The latter continue to divide while they migrate along the rostral migratory stream to the olfactory bulb, where they differentiate into two types of interneurons.

The putative neural stem cells of the adult dentate gyrus show astrocytic properties and GFAP-expression, and have a radial glia-like morphology. They are named type-1 cells (4). Like B cells of the subventricular zone they rarely divide, but generate populations of transiently amplifying progenitor cells that, depending on their expression of immature neuronal markers, can be distinguished as type-2a, type-2b and type-3 cells. Regulatory stimuli affect these stages differently. After exit from the cell-cycle the new neurons mature into functional granule cells over a period of several weeks. The populations of the putative transiently amplifying progenitor cells of both neurogenic regions are negative for GFAP, but positive for doublecortin and the polysialylated form of the neural cell adhesion molecule (PSA-NCAM). When adult hippocampal neurogenesis is stimulated, the expansion occurs in these intermediate populations. Similarly, an intermediate progenitor cell exists in the glial lineage.

Many nestin-expressing cells can be found throughout the adult brain, but it is not known which of these have precursor cell properties.

Molecular Interactions

Neural stem cells during development are characterized by their “positional identity”, that is the role they play within a coordinate system defined by rostro-caudal, left-right and ventral-dorsal axes. Depending on environmental cues, precursor cells take on different identities. In the developing brain, ►**bone morphogenic proteins** (BMPs) are expressed at the dorsal midline, whereas sonic ►**hedgehog** (Shh) is expressed at the ventral midline. Their antagonistic effects define positional identity in the ventral-dorsal orientation. As development proceeds, cells seem to remain sensitive to some of these cues; ventralized stem cells could be re-specified to a dorsal phenotype after delayed exposure to BMPs, indicating that regional specification is not terminally fixed in progenitor cells. Wnt signaling, as well as different forms of fibroblast growth factor (FGF) and retinoic acid, which are secreted by the

organizer regions of the neural tube, define the rostro-caudal axis.

Within their given position, precursor cells are instructed to divide and expand. Many of the morphogens mentioned such as the BMPs and Shh double in function as mitogens. For Wnt signaling a gradient exists that induces precursor cell proliferation near the site of Wnt expression in the roof plate and exit from the cell-cycle at greater distances. Consequently, cell differentiation during development shows a gradient inverse to Wnt signaling (5). In the dorsal brain Wnt acts as a mitogen downstream of BMPs. The receptors for FGF2, EGF and transforming growth factor alpha (TGF α) are expressed throughout the developing nervous system and they strongly induce proliferation. But even classical mitogens such as FGF2 are involved in defining differentiation patterns.

►**Basic helix-loop-helix** proteins (bHLH) are involved in balancing the exit from the cell-cycle and the initiation of differentiation. Notch expression maintains stem cell properties and inhibits neuronal differentiation by tilting the balance of *Hes* gene expression against the neurogenic isoforms of *Hes*. Similarly, activation of the LIF/STAT pathway maintains precursor cells in a multipotent state. Besides its function in determining positional identity, Shh is expressed in brain regions that undergo late and massive cell expansions, most notably the cortex. Here BMP signaling acts pro-differentiation.

Most newly generated cells during brain development die by apoptosis, especially during the early gestational expansion phases. This early phase is also characterized by symmetric divisions of the precursor cells. Pax6 and Emx2 are two examples of transcription factors that are involved in regulating the transition between the expansion phase and the neurogenic phase during development. Both maintain symmetric divisions. After the transition, mostly neurons are generated in midgestation. Glia is predominantly produced in late gestation and postnatally. The latter two stages are characterized by asymmetric cell divisions. *In vitro* data suggest that during brain development, neural precursor cells dramatically change their responsiveness to extracellular factors. Whereas exogenously applied BMPs induce apoptosis in rat E13 precursor cells, they promote neurogenesis as well as gliogenesis around E16 and only gliogenesis after birth (6).

BHLH transcription factors like Mash1, Ngn 1/2 and NeuroD are required for neuronal differentiation and the inhibition of gliogenesis. But bHLH functions are temporally specific, too. Olig2, for example, is co-factor to Ngn2 for motor neuron differentiation, but later in development together with Nkx2.2 involved in inducing the production of oligodendrocytes. Thus, during development oligodendrocytes and astrocytes might not generally be derived from a common glial

precursor cell but rather from neural precursor cells with different positional identities (7).

Both Notch and LIF signaling that maintain stem cell properties early in development take on new roles in determining cell fate in late development, when both act pro-gliogenically. The current hypothesis is that progenitor cells in the neuronal lineage as well as differentiated neurons secrete factors that direct neural precursors towards gliogenesis. The neurogenic regions in contrast might be defined by a "stem cell niche" comprising astrocytes secreting pro-neurogenic factors (8). In addition, the neurogenic niche of the adult hippocampus is characterized by a close proximity of the putative stem cells to the vasculature, much like precursor cells in the subventricular zone have contact with the ventricular surface. Circulating factors might be involved in determining precursor cell identity and activity in these regions throughout adulthood.

Generally, however, only limited information is available today about the transcriptional regulation of stem cell activity in the adult brain. The reason is that study of normal adult neurogenesis requires normal embryonic brain development. Due to the redundant use of many transcription factors at different developmental stages most conventional mutant models cannot be used; their developmental phenotype precludes the clear establishment of an adult phenotype.

Regulatory Mechanisms

Neural stem cell activity is regulated on several conceptual levels from transcriptional regulation as described above to behavior of the entire individual. Neural stem cells in the adult brain take part in neural plasticity, i.e. the structural adaptation of the brain according to functional demands. They respond to many physiological and pathological stimuli, and a vast literature describes these effects. On a behavioral level physical activity and exposure to environmental enrichment stimulate precursor cell activity and adult neurogenesis *in vivo*. Neurotransmitters from different transmitter systems, hormones, circulating and locally secreted growth factors, etc. can all influence the different stages of neuronal development. The underlying pathways are still largely unknown as are the genetic and molecular determinants of stem cell activity in the adult. What has become clear, however, is that development continues throughout life. Neural stem cell activity and neuronal development in the adult can be seen to be in line with embryonic brain development. Regulation of adult neurogenesis appears to be more strongly influenced by external stimuli than embryonic development. But many transcription factors known from embryogenesis appear to play a role in adult neurogenesis and its regulation as well. Pax6 for example is a likely candidate to maintain the neurogenic potential of radial glia-like precursor cells

both in development and adulthood. In the adult subventricular zone the BMP antagonist noggin inhibits gliogenesis and promotes neurogenesis. Many more details of these molecular pathways have to be elucidated in the adult in order to relate them to the observed macroscopic net regulation of adult neurogenesis. To date, it is difficult to link external regulatory cues to specific effects on the precursor cell level. However, the finding, that stem cells are regulated in the adult brain makes it promising to apply the knowledge about embryonic brain development to the situation in the adult.

► [Neural Development](#)

► [Stem Cells: an Overview](#)

References

1. Weissman IL, Anderson DJ, Gage F (2001) Stem and progenitor cells: origins, phenotypes, lineage commitments, and transdifferentiations. *Annu Rev Cell Dev Biol* 17:387–403
2. Blau HM, Brazelton TR, Weimann JM (2001) The evolving concept of a stem cell: entity or function? *Cell* 105:829–841
3. Alvarez-Buylla A, Garcia-Verdugo JM, Tramontin AD (2001) A unified hypothesis on the lineage of neural stem cells. *Nat Rev Neurosci* 2:287–293
4. Filippov V, Kronenberg G, Pivneva T et al (2003) Subpopulation of nestin-expressing progenitor cells in the adult murine hippocampus shows electrophysiological and morphological characteristics of astrocytes. *Mol Cell Neurosci* 23:373–382
5. Megason SG, McMahon AP (2002) A mitogen gradient of dorsal midline Wnts organizes growth in the CNS. *Development* 129:2087–2098
6. Mehler MF, Mabie PC, Zhu G et al (2000) Developmental changes in progenitor cell responsiveness to bone morphogenetic proteins differentially modulate progressive CNS lineage fate. *Dev Neurosci* 22:74–85
7. Lu QR, Sun T, Zhu Z et al (2002) Common developmental requirement for Olig function indicates a motor neuron/oligodendrocyte connection. *Cell* 109:75–86
8. Song H, Stevens CF, Gage FH (2002) Astroglia induce neurogenesis from adult neural stem cells. *Nature* 417:3944

Neural Tube

Definition

Neural tube describes an epithelial embryonic tube with a central lumen that gives rise to the development of the entire central nervous system (brain and spinal cord) and to the neural crest.

► [Muscle Development](#)

► [Neural Crest Cells and their Derivatives](#)

► [Neural Development](#)

Neuraminic Acid

- Sialic Acid

Neuraminidase

- Sialidase

Neurodegeneration

Definition

Neurodegeneration means the change of neural tissue to a lower or less functionally active form.

- Alzheimer's Disease
- Repeat Expansion Diseases

Neuroectoderm

Definition

Neuroectoderm designates the primordial tissue in the early developing embryo that produces the central nervous system.

- Axis Formation – Formation and Function of the Dorsal Organizer
- Neural Development

Neurofibrillary Tangles

Definition

Neurofibrillary tangles (NFT) are one of the histopathologic hallmarks of Alzheimer disease (AD). NFTs are formed intracellularly in degenerating neurons by helical filaments of abnormally phosphorylated microtubulus-associated tau proteins.

- Alzheimer's Disease

Neurofibromas

Definition

Neurofibromas are benign tumours that are associated with the peripheral nerve sheath as a hallmark feature of Neurofibromatosis. Neurofibromas consist of Schwann cells (60%–80%), fibroblasts, perineural cells, axons and mast cells.

- Neurofibromatosis Type 1 (NF1), Genetics

Neurofibromatosis Type 1 (NF1), Genetics

MEENA UPADHYAYA, DAVID N. COOPER
Institute of Medical Genetics, University of Wales
College of Medicine, Cardiff, UK
upadhyaya@cardiff.ac.uk
cooperdn@cardiff.ac.uk

Definition

► Neurofibromatosis Type 1 (NF1), Genetics is one of the most common dominantly inherited neurogenetic disorders, affecting about 1 in every 3500 individuals worldwide. The condition is characterised by multiple café-au-lait spots, benign ►neurofibromas and Lisch nodules (1). The café-au-lait spots tend to arise during infancy with ~80% of individuals with NF1 manifesting café-au-lait spots by one year of age. Dermal neurofibromas tend to arise during adolescence and are always benign. Lisch nodules are benign nodules on the surface of the iris and do not affect vision. They are best detected using an ophthalmic slit lamp examination. Other clinical manifestations include abnormalities of the cardiovascular, gastrointestinal, renal and endocrine systems, major orthopaedic problems, facial and body disfigurement, cognitive deficit and malignancy in the form of tumours of the peripheral nerve sheath and central nervous system. The less common features of NF1 include optic pathway gliomas (15%), ►plexiform neurofibromas (25–40%), malignant peripheral nerve sheath tumours (10%), learning problems (30–50%), tibial pseudoarthrosis (5%), dysplastic scoliosis (5%) and renal artery stenosis (1–2%). Individuals with NF1 have a 4- to 5-fold increased risk of developing malignancy (2). Clinical expression is highly variable even within the affected members of a family with identical mutations in the ►NF1 gene. The appearance of many manifestations of NF1 appears

to be age-dependent, although NF1 is fully penetrant by the age of 5.

Characteristics

NF1 Tumours

Neurofibromas are a hallmark feature of NF1 and are associated with the peripheral nerve sheath. Plexiform neurofibromas are associated with major nerve trunks and usually appear in young children; they differ from neurofibromas in having an expanded extracellular matrix. About 10% of these become transformed to malignant peripheral nerve sheath tumours (►MPNSTs), which are soft tissue tumours (sarcomas). Benign and plexiform neurofibromas comprise ►Schwann cells (60–80%), fibroblasts, perineural areas, axons and mast cells. About 15% of NF1 patients have optic gliomas, which rarely become symptomatic. Other rare forms of tumour associated with NF1 include pheochromocytomas, rhabdomyosarcomas, juvenile myeloid leukaemia and myelodysplastic syndrome. Several studies have demonstrated that it is only Schwann cells in the neurofibromas that exhibit ►somatic mutations (3).

Cellular and Molecular Regulation

The *NF1* Gene

The *NF1* gene, located at 17q11.2, spans 350 kb of genomic DNA, contains 60 exons and encodes a 12 kb mRNA transcript (Fig. 1). Introns 1 and 27b are both large, each being approximately 60 kb in size. Three genes (*EVI2A*, *EVI2B* and *OMGP*) are embedded in intron 27b, each comprises two exons. *EVI2A* is a small gene, encoding a 232 amino acid polypeptide, which is expressed in the brain and bone marrow. It encodes a putative transmembrane protein according to sequence analysis. The *EVI2B* gene is also a small gene that encodes a 448 amino acid protein; it is

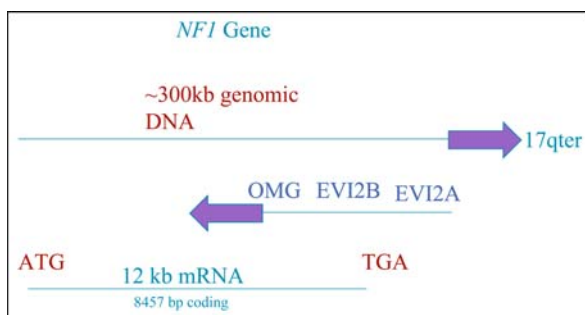
expressed exclusively in the bone marrow. The *OMGP* gene encodes a 416 amino acid cell adhesion molecule that is expressed primarily in oligodendrocytes. These genes are transcribed in the opposite orientation to the *NF1* gene. The role of these genes (if any) in regulating the *NF1* gene is, however, not known.

The *NF1* gene promoter is located in a CpG island and exhibits a high degree of sequence conservation with other species. The 3' untranslated region (UTR) of the human *NF1* gene is 3.5 kb in length and has been highly conserved during evolution, suggesting that it may be important for mRNA stability or translational efficiency.

The *NF1* Gene Product (►Neurofibromin)

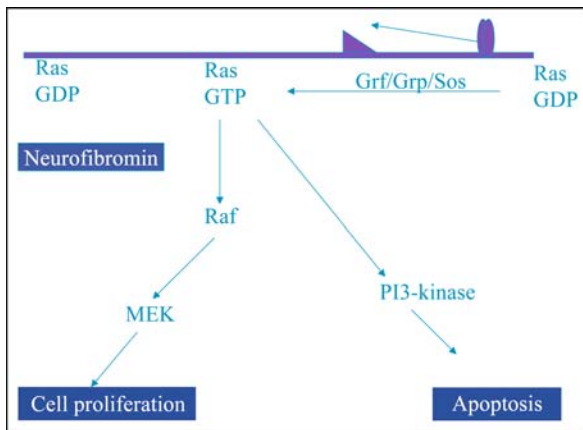
Neurofibromin, the *NF1* gene product, is ubiquitously expressed albeit at low levels. There are three alternatively spliced *NF1* exons (9a, 23a and 48a); 9a is specifically expressed in the brain whereas 48a is predominantly expressed in muscle. In most tissues, there appear to be equal amounts of mRNA with and without exon 23a. The functional significance of these three splice forms is however unclear. The expression of neurofibromin is highest in the central nervous system and it has been shown to be associated with tubulin.

Neurofibromin exhibits structural and sequence similarity to an evolutionarily conserved family of proteins, the mammalian GTPase activating protein (GAP)-related proteins. The most highly conserved region of the protein is the NF1 GAP-related domain (GRD), encoded by exons 20–27a. This is the only domain of neurofibromin to which a function has been ascribed. Activated p21-ras (ras with bound GTP) can transform fibroblasts and increases cell proliferation *in vitro*. Ras is anchored to the cell membrane through isoprenylation of its carboxyl CAAX box, which is mediated by farnesyl transferase. Neurofibromin has been shown to down-regulate cell growth by inactivating p21-ras (ras-GDP). Hence, inactivation of neurofibromin has been postulated to lead to up-regulation of ras activity and increased cellular proliferation. Co-localisation of neurofibromin with ras has however not been demonstrated. The role of activated ras in cell signalling is complex (Fig. 2). Activated ras sends signals along the Raf-mitogen-activated protein (MAP) kinase pathway thereby promoting cellular proliferation. Activated ras also sends signals through the PI3 (phosphatidylinositol 3) kinase and PKB (protein kinase B, also known as Akt) complexes and nuclear factor-kappa B (NF-kappa B), thereby exerting antiapoptotic influence. Activation of the ras pathway has been detected in a variety of human tumours. The inappropriate activation of these downstream effectors has also been reported in some NF1-deficient cells and tumours. It is still unclear whether Ras regulation is the primary and/or only function of neurofibromin.



Neurofibromatosis Type 1 (NF1), Genetics.

Figure 1 *NF1* gene spans approximately 300 kb of genomic DNA and comprises 60 exons. Three embedded genes *EVI2A*, *EVI2B* and *OMGP* located in intron 27b are transcribed in the opposite orientation to the *NF1* gene. The *NF1* transcript is 12 kb in size, of which 8.4 kb represents the coding region.



Neurofibromatosis Type 1 (NF1), Genetics.

Figure 2 Activated Ras sends signals along the Raf-mitogen-activated protein (MAP) kinase pathway thereby promoting cellular proliferation. Activated Ras also sends signals through the PI3 (phosphatidylinositol 3) kinase and PKB (protein kinase B, also known as Akt) complexes and nuclear factor-kappa B (NF-kappa B), thereby exerting antiapoptotic influence. Reduced levels of neurofibromin lead to increased signalling through *Raf* and PI3-kinase.

NF1 Gene Germline Mutations

The mutation rate for the *NF1* gene (~1 in 10,000 gametes/cell/generation) is approximately 10-fold higher than that found for most other inherited disease genes. The germline mutational spectrum of the *NF1* gene associated with disease expression in NF1 patients has been well characterised, with new mutations occurring in about one half of all NF1 individuals (4). Over 500 germline mutations have been identified so far and some 80% are truncating lesions (4). No clustering of mutations is apparent within the *NF1* gene. Recently, three separate mutations in the *NF1* gene have been identified in a Portuguese family. About 30% of mutations are associated with altered mRNA splicing.

The characterisation of mutations in the *NF1* gene has been a challenging task owing to the size of the gene, the absence of any obvious mutational clustering and the wide diversity of mutations (4).

The *NF1* gene promoter is highly conserved evolutionarily. No pathological mutation has so far been detected in this region. Similarly, only one mutation has been identified in the 3'UTR. Approximately 95% of all sporadic cases of NF1 result from mutation of the paternal chromosome. The paternally inherited mutations appear to be subtle lesions (<20 nucleotides) whereas the vast majority of *NF1* gene deletions are maternally derived.

NF1 ►Pseudogenes and ►Germline Mosaicism in the NF1 Gene

NF1-related sequences have been identified on a total of 7 human chromosomes viz. 2q12-q13, 12q11, 14p11-q11, 15p11.2, 18p11.2, 21p11-q11 and 22p11-q11. These *NF1* homologues represent non-processed pseudogenes, the result of a series of partial duplications of the functional 17q11.2-located *NF1* gene followed by translocation. The potential for such highly homologous sequences to interfere in the mutational analysis of the functional *NF1* gene should not be underestimated.

Two affected NF1 sibs with unaffected parents were shown to have an identical intragenic deletion in the paternal chromosome. 10% of the father's spermatozoa possessed the same constitutional mutation. ►**Mosaicism** for large deletions has been demonstrated in a number of typical NF1 patients using lymphocyte FISH analysis. Patients with segmental NF1 manifest NF1 features that are confined to a particular part of the body. It is believed that the molecular basis of this condition may be the somatic mutation of the *NF1* gene in an early stage of foetal development.

NF1 Gene Somatic Mutations

To date, specific somatic microlesions have only been reported in 42 NF1-associated tumours. Somatic *NF1* gene lesions have also been identified in several tumour types not usually associated with NF1, but it is not yet known just how *NF1* gene mutations contribute to the formation of either NF1-associated or unrelated sporadic tumours. Information on the somatic mutational spectrum of the *NF1* gene may help to identify other functional domains of neurofibromin, may improve our currently inadequate understanding of genotype/►**phenotype** relationships in NF1 and could also indicate whether the nature and location of germline mutations determines the type of "second hit", as has been reported in ►**familial adenomatous polyposis**.

NF1 – a ►Tumour Suppressor Gene

Genetic and biochemical evidence indicates that *NF1* is a tumour suppressor gene (TSG). Multiple neurofibromas from an individual NF1 patient harbour different somatic mutations indicating that the 'second hit' is an independent event in each neurofibroma. The somatic mutational spectrum of the *NF1* gene may thus be an important factor in the variable clinical expression of NF1. It follows that the factors that influence somatic mutation rates can in some sense be regarded as potential modifiers of NF1.

A number of different genetic mechanisms may give rise to the inactivation of the normal wild type *NF1* allele in an NF1-associated tumour, namely large or small DNA rearrangements, various subtler sequence

alterations, genomic instability at either nucleotide or chromosomal level or hyper-methylation of the gene promoter region. Whilst the acquisition of microsatellite instability may be a relatively late event in NF1-related malignancy, it may be comparatively infrequent in NF1 tumorigenesis.

It is still unclear if cells heterozygous for mutations at the *NF1* locus contribute to tumour formation or whether tumour development can only be initiated from cells with both *NF1* alleles inactivated. Mutations of the wild type allele are most likely to occur during cell division, but adult nerve sheath cells are known not to divide at a high rate. Whether the *NF1* second hit is acquired at an earlier developmental stage, with another subsequent event required to stimulate the growth of the tumour is unknown. Furthermore, the number of non-allelic gene lesions (subsequent to the initial *NF1* mutations) that are required for development of neurofibromas or for progression to MPNSTs is also unclear. Although dermal neurofibromas never become malignant, second hit mutations at the *NF1* locus have been detected. The nature of other non-genetic factors that might be involved in NF1 tumorigenesis is also unknown. Multiple sequential mutations of different genes involved in the progression towards a malignant phenotype have been demonstrated for colon cancer. Several studies have provided evidence that *TP53* and *NF1* mutations may exert a synergistic effect in some tumours both in humans and mice. In a panel of 11 MPNSTs, deletions of *TP53* and *CDKN2A* have been identified. Microsatellite instability (MSI) has been reported in NF1-associated neurofibromas and also intriguingly, the occurrence of NF1 in mismatch repair-deficient individuals with homozygous *MLH1* mutations.

Elucidation of the tumorigenic pathway is obviously a much more complex task than simply identifying a predisposing lesion in a given gene. Complete understanding will only come with the elucidation of the cellular consequences of those lesions in terms of the altered levels of expression of many other genes. Information on the gene expression patterns of cells in different developmental contexts and in response to different environmental conditions or disease states is now being obtained using microarray-based expression analysis.

Molecular Diagnosis

Owing to the extreme variability of clinical expression, NF1 disease progression cannot really be predicted. There has been little demand for prenatal diagnosis possibly because the majority of couples would want to know the clinical severity of the foetus's condition and this is something that cannot be predicted. There is nevertheless an urgent need for an inexpensive, accurate, DNA-based test for NF1. The development of a preimplantation test for NF1 has been initiated.

Animal Models of NF1

Neurofibromin is highly conserved at the amino acid sequence level in mouse, *Drosophila* and yeast. *Nf1* +/- heterozygous mice do not have an obvious phenotype; however, they are predisposed to develop leukaemia and pheochromocytomas. Although they manifest decreased learning ability, including poor performance in tests involving spatial learning and memory, these deficiencies can be overcome by extended training. Mice homozygous for *Nf1* mutations are embryonic lethals, having cardiac abnormalities and dying at approximately 12–14 days of gestation. Neurons harvested from the peripheral nervous systems of these animals show an ability to survive in culture without growth factor stimulation. Haematopoietic stem cells exhibit an increased proliferative ability and are more sensitive to granulocyte-macrophage colony-stimulating factor (GM-CSF) than either heterozygous or wild type cells. *Nf1* chimeric mice develop neurofibromas that are derived from *Nf1* -/- cells. Recent work has demonstrated that it is the interaction of Schwann cells carrying mutations on both the alleles of the *Nf1* gene with *Nf1* +/- mast cells that results in tumour formation, whereas the interaction with *Nf1* ++ mast cells does not result in the formation of tumours (5). Mouse models have further shown that NF1 malignancy is associated with homozygosity for mutations in the *TP53* gene.

The *Drosophila NF1* gene product is approximately 60% homologous with human neurofibromin. Homozygous *Nf1* null mutants are 20–25% smaller than wild type flies.

Further detailed studies on the interaction of neurofibromin with p21Ras, tubulin, growth factors or other proteins will also be important for the elucidation of the normal function of neurofibromin. By understanding the pathophysiology of NF1, effective therapies can be devised and targeted for the various features of NF1.

Clinical Relevance

Genotype/Phenotype Correlations

No clear relationship between genotype and clinical phenotype has yet been discerned for NF1. This may be due either to the influence of modifier loci or to the variable nature, location and developmental timing of the somatic mutations that determine the rate of progression and severity of disease in different tissues.

► **Genetic counselling** for NF1 patients is problematic, owing to the marked inter- and intra-familial variation in NF1 expression. Identical *NF1* gene mutations can occur in unrelated patients with very different phenotypes (4). The first genotype/phenotype studies demonstrated that NF1 patients with large deletions tended to have dysmorphic features and learning disability. NF1 patients with whole gene deletion are at higher risk of developing MPNSTs. In a recent study, we have also

found that NF1 individuals with missense mutations had a relatively lower risk of developing Lisch nodules when compared to people with nonsense or frameshift mutations. This tentative finding awaits confirmation in another NF1 patient population.

The exploration of genotype-phenotype correlations in NF1 is still however in its infancy due to the extensive mutational heterogeneity of the gene and because large-scale mutation screening is still laborious owing to the size and complexity of the gene. *NF1* mutations do not appear to be a major determinant of the common NF1 disease features. The corollary to this is that mutations in the *NF1* gene have been identified in families classified as rare variants of NF1 and these include Watson syndrome, neurofibromatosis-Noonan syndrome, familial café-au-lait spots, spinal neurofibromatosis and Leopard syndrome.

References

1. Friedman JM, Gutmann DH, Mac Collin M et al (1999) Neurofibromatosis: Phenotype, Natural History, and Pathogenesis. Johns Hopkins, University Press
2. Cichowski K, Jacks T (2001) *NF1* tumour suppressor gene function: narrowing the GAP. Cell 104:593–604
3. Serra E, Rosenbaum T, Winner U et al (2000) Schwann cells harbor the somatic *NF1* mutation in neurofibromas: evidence of two different Schwann cell subpopulations. Hum Mol Genet 9:3055–3064
4. Upadhyaya M, Cooper DN (1998) Neurofibromatosis Type 1: From Genotype to Phenotype. BIOS Scientific Publishers, Oxford pp 65–88
5. Zhu Y, Ghosh P, Charnay P et al (2002) Neurofibromas in NF1: Schwann cell origin and role of tumour environment. Science 296:920–922

Neurofibromin

Definition

Neurofibromin is the Ras-GTPase-activating protein encoded by the *NF1* gene, and consists of 2818 amino acids.

► [Neurofibromatosis Type 1 \(NF1\), Genetics](#)

Neurogenic Genes

Definition

The loss of function of all genes that are involved in the signal transduction of the Notch pathway cause an excess of neural tissue at the expense of epidermis in

Drosophila. This phenotype is called neurogenic, and all genes that cause this phenotype after inactivation are called neurogenic genes.

► [Notch Pathway](#)

Neuromuscular Junction

Definition

The neuromuscular junction constitutes a specialized cholinergic synapse between a motor neuron and a skeletal muscle cell.

► [Neural Development](#)

Neurons

JOACHIM LÜBKE, DIRK FELDMEYER
Institute of Medicine, Research Centre Jülich,
Jülich, Germany
j.luebke@fz-juelich.de
d.feldmeyer@fz-juelich.de

Definition

Neurons are the computational units of the central nervous system (CNS). The human ► [brain](#) contains about 10^{13} neurons. They represent individual elements and do not form a syncytial network as was originally proposed by Camillo Golgi and others. In contrast, Ramón y Cajal strongly supported the theory of neurons as individual elements that are connected via specialised contact sites.

Neurons are very heterogeneous with respect to both structure and function. In general, neurons fall into two groups. The majority (80%) are excitatory neurons, while the remainder are inhibitory ► [interneurons](#), which themselves represent a very heterogeneous class. In principle, neurons can be further characterised as local circuit or as projection neurons. Local circuit neurons are those within a given brain region, e.g. thalamus or ► [neocortex](#). Projection neurons connect various brain areas or regions, e.g. cortico-cortical, thalamocortical, sensory periphery with a higher brain region. In general, despite their heterogeneity all neurons consist of three basic structural compartments, the ► [soma](#) (cell body) of varying shape and size, a more or less elaborate dendritic tree that represents the input structure of the neurone and finally the ► [axon](#) resembling the output structure of the neurone that

always terminates in a highly specialized structure, the so-called ►**synapse** (from the Greek words for ‘together’ (syn) and ‘to clasp’ (haptein), introduced by Sir Charles Scott Sherrington). These synapses are the key elements of communication between individual neurons within a given neuronal network.

Characteristics

Development

The CNS is a derivative of the neuroectoderm, which develops very early during foetal development (third week *in utero*). It is generally accepted that the neurons of the CNS derive from various germinal zones. Among these are the so-called ►**ventricular zone**, where most of the excitatory neurons are generated and the ►**medial/lateral ganglionic eminence**, from where most of the inhibitory interneurons originate. The progenitor cells for excitatory neurons follow a vertical migration pathway along radial glial processes, while inhibitory interneurons migrate either horizontally or tangentially towards their final target region. In general, the migration of interneurons precedes that of excitatory neurons. In most mammalian species, the neocortex for example shows an inside first – outside last organisation of neuronal layers, meaning that early-generated neurons are located in deeper layers (1). Recently, evidence accumulated that both neurons and glial cells, the two major structural elements of the CNS, derive from common progenitor cells. During late prenatal and early postnatal development, transient neuronal populations appear to exist that may play important roles during neuronal migration and layer formation, axonal path-finding and synaptogenesis. After the establishment of the mature neuronal networks, these neurons may undergo selective cell death or they will be transformed into different neuronal cell types.

Structural Compartments of Neurons

Soma

Every neurone has a cell body (somatic region) of a different size and shape (3 µm in van G’huchten cells in the olfactory bulb up to 60 µm in a giant Betz pyramidal cell in the motor cortex). The largest element within the cell body is the nucleus, the largest biochemical compartment housing the genome of the neurone. The nucleus is constantly active, providing all the genetic information necessary for the protein biosynthesis required for information processing and storage (immediate early genes etc.). The neuronal cytoplasm contains all the structural elements found in a somatic cell that are indispensable to maintain neuronal metabolism and survival. In addition, the neuronal cell body produces and contains structural elements such as neurofilaments that are necessary for trafficking of membrane structures (vesicles) and receptor proteins (ion channels, G-proteins etc.) to the

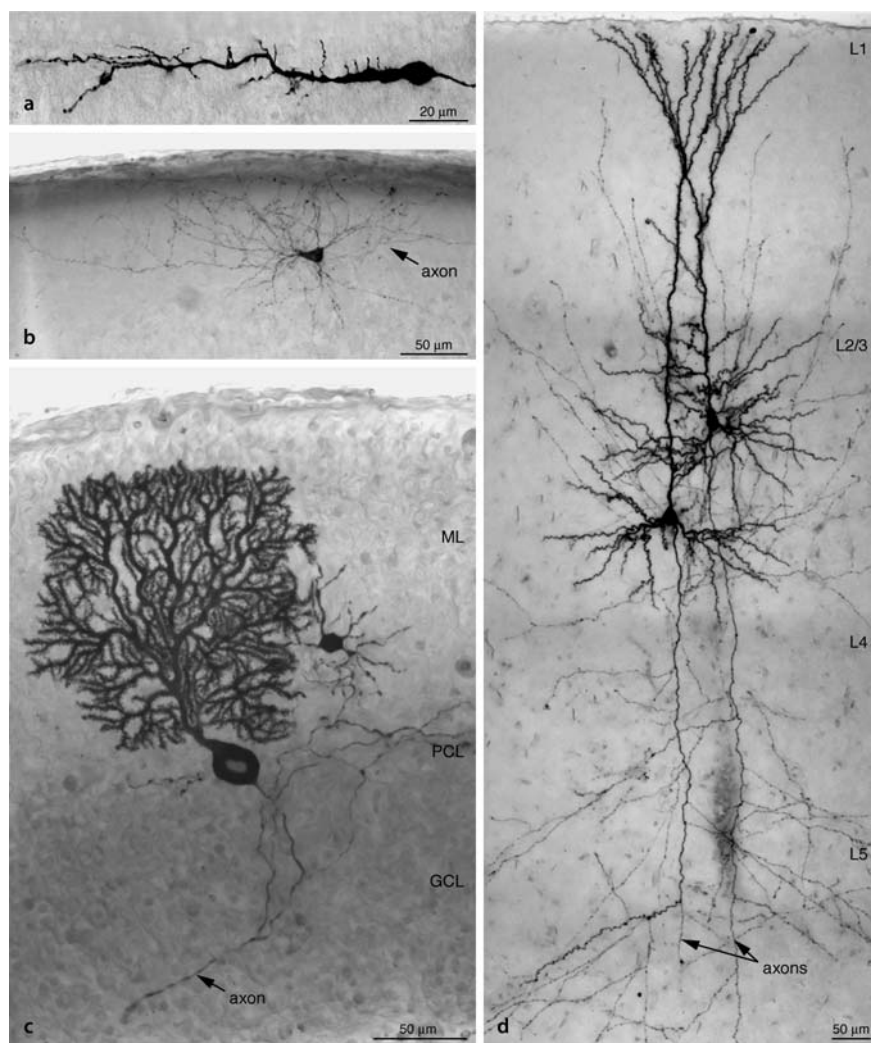
other neuronal compartments (axon and ►**dendrites**). Furthermore, the neuronal soma is an input station for synaptic signals, but exclusively for synaptic terminals originating from inhibitory interneurons. This specific termination of inhibitory inputs is one of the major sources of inhibition within a given neuronal network.

Dendrites

The vast majority of neurons possess a dendritic tree that is more or less elaborate. However, a very small fraction of neurons (e.g. dorsal root and sympathetic ganglion cells in the peripheral nervous system, PNS) completely lack a dendritic compartment. Dendrites always emerge from the soma and serve as an enlargement of the surface area (i.e. the input area) of a neurone. The geometric organisation of the dendritic tree is shaped by the neuronal function and is therefore highly variable. Neurons may have unipolar, bipolar, multipolar, radially symmetric or asymmetric dendritic configurations (Fig. 1). Excitatory neurons, in particular those of higher brain areas, possess a highly structured compartment that is almost exclusively located on dendrites, the so-called dendritic ►**spines**. The structures form only excitatory synapses. The spine morphology is also extremely variable ranging from very short ‘stubby spines’ to elongated ‘spine-like protrusions’. In contrast, inhibitory interneurons have either smooth or sparsely spiny dendrites. Like the soma, dendrites contain neurofilaments for membrane and protein trafficking and occasionally internal stores for Ca^{2+} , an important intracellular messenger.

Dendrites represent the main input region of a neurone, integrating thousands of excitatory and inhibitory inputs arising from either local (within a brain region or area) or distant sources (thalamus, sensory periphery etc.). Generally excitatory inputs are located more distally making an integration of inputs necessary, while inhibitory inputs are found in more proximal segments on the dendrites.

It has long been thought that dendrites act like passive cables but it is now well established that they have a large ion channel complement that plays important roles in synaptic transmission, signalling between neuronal compartments and integration of synaptic inputs. In particular, active propagation of the Na^+ ►**action potential** generated in the axon back into the dendrite has been demonstrated to play a role in regulating the ►**synaptic efficacy**, one step thought to be involved in memory formation. Furthermore, a second action potential initiation zone exists in apical dendrites of pyramidal cells where Ca^{2+} spikes are generated (in contrast to Na^+ action potentials at the axon). These Ca^{2+} action potentials may be important for the intracellular integration of synaptic inputs from different sources (2).



Neurons. Figure 1 Morphologies of different types of neurons in the rat neocortex and cerebellum. (a) A Cajal-Retzius cell in layer 1 of the developing neocortex, a neurone that plays an important role in the layer formation of various brain regions. The soma and dendrites of Cajal-Retzius cells are oriented parallel to the pial surface. Note the long spine-like protrusions emanating from the primary dendrite. (b) Inhibitory layer 1 interneurone with a dense axonal domain that is almost exclusively located in layer 1 of the neocortex. The synaptic connectivity of local interneurons is extremely high. (c) Purkinje cell of the cerebellum with its prominent fan-like dendritic tree projecting throughout the molecular layer (ML); the spine density of the Purkinje cell is extremely high. Next to the Purkinje cell is an inhibitory cerebellar stellate cell that is located in the middle of the molecular layer. The stellate cell was stained simultaneously with the Purkinje neurone and was presynaptic to it. PCL, Purkinje cell layer; GCL granule cell layer. (d) Two synaptically coupled excitatory pyramidal cells in layer 2/3 of the neocortex. Both pyramidal cells have prominent apical dendrites that terminate in an elaborate tuft in layer 1. The shorter basal dendrites emerge at the bases of the pyramidal cell somata. Both apical and basal dendrites are densely covered with spines. The axons show a layer-specific projection pattern; collaterals are found in layer 2/3 and layer 5 but rarely in layer 4. These collaterals may project over several mm of the cortical surface. All staining was by biocytin/neurobiotin injections via patch-clamp micropipettes.

Axons

Axons are the main output structures of neurons and are structurally their most complex compartment. Axons can be categorised as local circuit axons (most interneurons, see Fig. 1b; Chandelier cells project

locally within the range of the dendritic tree; field span ~ 0.5 mm) and projection axons (pyramidal cells of the neocortex, see Fig. 1d; the axons of giant Betz pyramidal cell form the cortico-spinal tract; length ~ 0.7 – 0.8 m). The axon emerges either directly from the soma or from a

primary dendrite with a thick unmyelinated ►axon initial segment that is thought to be the action potential initiation zone. From there numerous axonal collaterals emerge; however each neuronal cell type shows a highly specific projection pattern that is regulated by the function of the neurone within the network in which it is embedded (e.g. a columnar projection as shown for spiny stellate neurons).

Axons differ widely with respect to their diameter (<1–20 µm). Furthermore, they display differences in their degree of myelination from a thick ►myelin sheath (as in motoneurone axons in the PNS or axons of the pyramidal tract) to no myelination. The myelin sheath is formed by the invagination of the axon into the cytoplasm of a Schwann cell (in the PNS) or a process of an oligodendrocyte (in the CNS), where it is suspended by two layers of plasma membrane (the ‘mesaxon’). The mesaxon elongates and spirals around the axon, each turn forming a lamella of myelin. The cytoplasm and extracellular space between each turn of mesaxon are reduced to a minimum and very closely packed. Unmyelinated axons have a very small diameter of 1 µm or less. In the CNS, unmyelinated axons are rare; in the PNS so-called C fibres involved in pain sensation, post-ganglionic neurons of the autonomic nervous system and olfactory nerves have unmyelinated axons. The myelin sheath is periodically interrupted by the so-called nodes of Ranvier. At these sites the axonal membrane contains Na⁺ and K⁺ ion channels at an extremely high density. The channel density is particularly high in the axon initial segment, which is supposed to be the action potential initiation zone on the axon. This is a prerequisite for reliable and fast signal propagation, as it enables a saltatory transmission of excitation (3). A thick axon diameter goes with a high degree of myelination and thus with a fast action potential propagation, i.e. signal transduction. A demyelination will lead to a slowing of action potential propagation and eventually to a complete failure of signal transmission. Along axons in higher sensory brain areas, presynaptic membrane specialisations (swellings), the so-called axonal boutons are formed. These constitute the presynaptic elements of neuronal signalling. Boutons form *en passant* synapses with different compartments of target neurons (soma, dendrites, dendritic spines and axon initial segments). Axonal boutons are highly specific with respect to their target structures.

Synapses

The central element for signalling between individual neurons is the synapse. At the synapse, incoming electrical signals (action potentials) are translated into a chemical signal (►neurotransmitter release), which in turn is transformed into a postsynaptic potential change, the excitatory or inhibitory postsynaptic potential (►EPSP/IPSP).

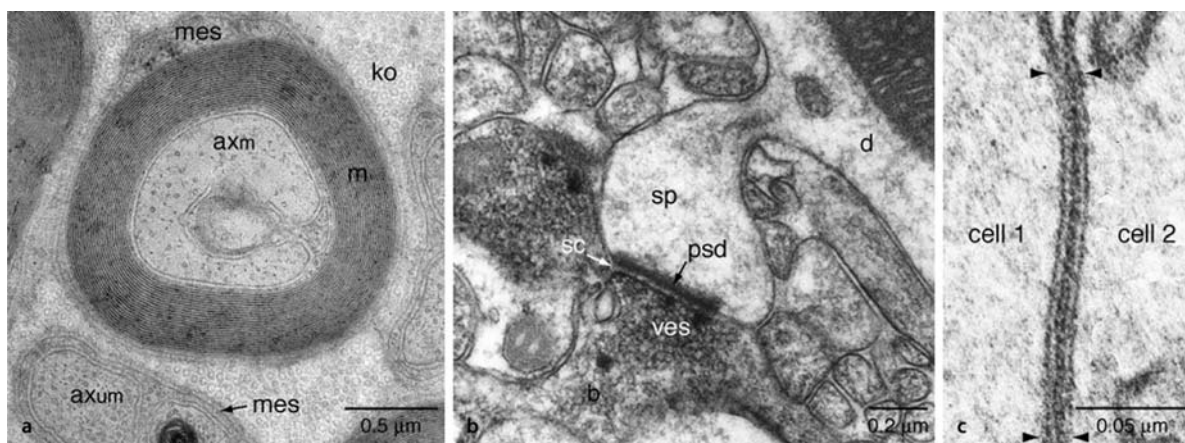
Structurally, a synapse always consists of a presynaptic terminal, a synaptic cleft and a postsynaptic target zone. Both the pre- and postsynaptic sites are highly specialized structures. The presynaptic site contains a pool of synaptic vesicles (containing the ►neurotransmitter) and a dense accumulation of various proteins involved in the neurotransmitter release process (e.g. the SNARE complex, Ca²⁺ channels etc.) and its metabolism (transporters etc.) (4). Once released the neurotransmitter molecules diffuse through the synaptic cleft (width 15–30 nm) to their target structures, the ligand-gated receptors (ion channels of G-proteins) that are embedded in a specialised area, the ►postsynaptic density (PSD), in the postsynaptic membrane. Apart from the signalling molecules, the PSD also contains various structural components that stabilise the synaptic complex and the arrangement and trafficking of the postsynaptic receptors. PSDs can be either continuous or perforated (periodic interruption of the PSD complex). Recently, it has been shown that synapses with perforated PSDs display a more efficient synaptic transmission (4).

These ‘chemical’ synapses fall into two major categories, inhibitory and excitatory synapses. Structurally, excitatory synapses are so-called ‘symmetric’ synapses because pre- and post-synaptic densities show a complete overlap while inhibitory synapses are ‘asymmetric’ (i.e. no overlap). In addition, synapses can be subdivided into those displaying ‘slow’ synaptic transmission (*via* G-proteins) and ‘fast’ transmission (*via* ligand-gated ion channels).

Neurotransmitter release at chemical synapses occurs *via* the Ca-dependent fusion of synaptic vesicles with the presynaptic membrane. However, it is currently not clear whether only one or several vesicles are released per action potential arriving at the presynaptic terminal. These two ►modes of release have been termed uni-vesicular and multi-vesicular, respectively. There is growing evidence that both mechanisms of neurotransmitter release exist at different synapses in the CNS (5).

Because of their structural and functional properties chemical synapses are subject to various regulatory mechanisms such as short- and long-term plasticity. The molecules responsible for the regulation of synaptic transmission are located both pre- and post-synaptically. Thus at ‘chemical’ synapses, both the probability of neurotransmitter release (which is Ca²⁺-dependent) and the postsynaptic receptor complement can be adapted to the state of activity in a given brain region.

Neurons may also communicate *via* a second type of synapse, the so-called ‘electrical’ synapse or ►gap junction. Gap junctions can be characterised as a specialised apposition zone of the somatic, dendritic or axonal membranes of two adjacent neurons that contain numerous hemi-channel proteins, the so-called



Neurons. Figure 2 (a) Electron microscopic image showing a cross-section through a myelinated (axm) and an unmyelinated axon (axum) in the peripheral nervous system. The myelin (m) sheath is formed by the invagination of the axon into the cytoplasm of a Schwann cell (glial cell), where it is suspended by two layers of plasma membrane (the 'mesaxon', mes). The mesaxon elongates and spirals around the axon, each turn forming a lamella of myelin. The cytoplasm and extracellular space between each turn of the mesaxon are reduced to a minimum, and very closely packed. Unmyelinated axons have a very small diameter of $1\mu\text{m}$ or less. They are only embedded in cytoplasm of a Schwann cell that forms a mesaxon but no myelin. (Photograph courtesy of A. Schardt/E. Niksch). (b) Electron micrograph of a synaptic contact between a presynaptic axonal bouton (b) and a postsynaptic dendritic spine (sp) that are separate by a narrow synaptic cleft (sc). A pool of neurotransmitter vesicles (ves) is clearly visible in the presynaptic bouton. The postsynaptic side is characterised by an electron dense structure, the postsynaptic density (psd), which contains structural proteins that anchor neurotransmitter receptors in the membrane opposite the neurotransmitter release site. Dendritic shaft, d. (c) Gap junction between two dendrites as seen in the transmission electron microscope. The opposing arrowheads indicate the gap junction coupling site between the two cells. The thin line in the middle represents the apposition of the gap junction hemi channels in the membranes of the two cells.

► **connexons.** Two connexons form a functional channel through which ions and small molecules can diffuse. Functionally, gap junctions serve to transmit electrical activity between connected neurons but also serve to interchange signalling molecules such as inositol triphosphate and Ca^{2+} . Signalling between gap-junction coupled neurons is always fast but undergoes heavy low-pass filtering, i.e. the signal waveform is slowed down and reduced in amplitude. Action potentials will appear as small 'spikelets' in the coupled neurone.

During early brain development, neurons communicate predominantly *via* gap junctions. This electrical communication is gradually replaced with communication *via* 'chemical' synapses during late postnatal development. However, it has been shown recently that even in the adult, distinct types of inhibitory interneurons in the neocortex and hippocampus form independent networks *via* gap junctions. These networks may play important roles in the oscillations of the electrical brain activity and rhythm generation that have been associated with arousal, sleep and attention (6).

Both types of synaptic transmission may occur in parallel at the same synaptic connection. Thus neuronal

communication within a network may occur *via* both signalling pathways indicating a high degree of complexity within the nervous system.

► **Biochemical Engineering of Glycoproteins**

► **Fragile X Syndrome**

References

1. Rakic P (1995) Radial versus tangential migration of neuronal clones in the developing cerebral cortex. *Proc Natl Acad Sci USA* 92:11323–11327
2. Stuart G, Spruston N, Häusser M (1999) *Dendrites*. Oxford University Press, Oxford
3. Debanne D (2004) Information processing in the axon. *Nat Rev Neurosci* 4:304–316
4. Cowan M, Südhof TC, Stevens C (2001) *Synapses*. Johns Hopkins University Press
5. Xu-Friedmann M, Regehr W (2004) Structural contributions to short-term synaptic plasticity. *Physiol Rev* 84:69–85
6. Hormuzdi SG, Filippov MA, Mitropoulou G et al (2004) Electrical synapses: a dynamic signaling system that shapes the activity of neuronal networks. *Biochim Biophys Acta* 1662:113–137
7. Synapse Web Home page ► <http://synapses.mcg.edu/>

Neuropathy

Definition

Neuropathy is a general term used to describe disorders of the peripheral nervous system (peripheral neuropathy). Peripheral neuropathy can be associated with poor nutrition, a number of diseases, pressure or trauma, or can be caused by exposure to toxins and certain drugs.

- Diabetes Mellitus, Genetics
- Hereditary Neuropathies, Motor and/or Sensor

Neurotransmission

Definition

Neurotransmission designates a process where chemical substances called neurotransmitters (such as dopamine, serotonin and glutamate) are released at a nerve terminal as a result of the nerve impulse. Transmitters cross the synapse (gap between nerve cells) and bind to receptors on the second cell, thereby exciting it.

- Neurons
- Schizophrenia Genetics

Neurotransmitter

Definition

Neurotransmitters are signalling molecules (glutamate, γ -amino butyric acid (GABA), serotonin, noradrenalin, dopamine, neuropeptide Y etc.) released from the presynaptic terminal in order to bind to postsynaptic receptor proteins that may be either ligand-gated ion channels or ligand-gated G-proteins.

- Neurons

Neurotransmitter Release

Definition

Neurotransmitter release is a Ca^{2+} -dependent process by which the neurotransmitter (glutamate, GABA,

acetylcholine etc.) is released from synaptic vesicles after having docked to and fused with the presynaptic membrane.

- Neurons

Neurotrophic Factor Hypothesis

Definition

The neurotrophic factor hypothesis suggests that the target of innervation of a particular neuron produces a limited amount of a neurotrophin, i.e. a molecule that is essential for the survival of the neuron. Neurotrophic factors include the classic neurotrophins BDNF (brain derived neurotrophic factor), NT3 (Neurotrophic factor 3), NT4/5 and NGF (nerve growth factor) as well as other signalling molecules. Many of them signal through tyrosine kinase receptors.

- Neural Development
- Neurotrophic Factors
- Receptor Serine/Threonine Kinase

Neurotrophic Factors

ANNA MARIA CALELLA, LILIANA MINICHIELLO
European Molecular Biology Laboratory,
Mouse Biology Unit, Monterotondo, Italy
minichiello@embl-monterotondo.it

Definition

The nerve growth factor (NGF) was the first neurotrophic factor to be discovered about 50 years ago. Its identification provided, among other things, the first support for the “neurotrophic factor hypothesis”, which holds that target cells secrete limited amounts of an essential trophic factor which is taken up by nerve endings *via* binding to specific cell surface receptors. Neurons that do not receive sufficient amounts of trophic factors will die by ►apoptosis. Today, NGF is only one of the numerous secreted factors discovered that have the ability to support the survival of neurons. These secreted trophic proteins have been grouped into different classes. The neurotrophins (NTs) are the best-studied class and thus far four neurotrophins have been identified in mammals, NGF, brain derived neurotrophic factor (BDNF), neurotrophin-3 (NT-3), and neurotrophin-4 (NT-4). Other classes of proteins able to

promote neuronal survival include members of the transforming growth factor β (TGF β) family, the interleukin 6-related cytokines, fibroblast growth factors (FGFs) and hepatocyte growth factor (HGF) (Table 1). Among these classes, the neurotrophin family is of significant importance because of its well-known physiological role in neuronal survival and the widespread expression of its family members both during development and in the adult nervous system. Here we will focus mainly on the characteristics, molecular interactions and regulatory mechanisms of this class of neurotrophic factors.

Characteristics

Structural Features

The neurotrophins together with transforming growth factor β belong to the “**cystine knot**” superfamily of polypeptide growth factors. The crystal structures of NGF and TGF β 2, have in fact revealed that although these proteins show very low sequence similarities, they share a common overall protomer topology which is based on three disulphide bonds (“cystine knot”) and four-stranded β sheets (1). All cystine knot growth factors acquire, upon dimerization, a well-defined hydrophobic core, which stabilizes the structure. In fact, their active forms exist only as **dimers**, homodimers or heterodimers. The interfaces used to form the dimers are in each case quite different, suggesting that dimerization is an important event and plays a fundamental role in regulating the binding of the different growth factors to their specific receptors and in modulating their biological activity.

Expression Patterns

In the peripheral nervous system (PNS), as the target-derived neurotrophic factor theory suggests, neurotrophins are expressed and secreted by the target tissues (which can be skin, muscle, glands or neurons) and regulate the survival and differentiation of the innervating neurons. The first NT expressed in the PNS during embryogenesis is NT-3, which supports the survival and differentiation of some sensory and sympathetic neurons, many of which become dependent on NGF during later stages of development. Likewise, BDNF promotes the survival of a subset of dorsal root ganglia (DRG) neurons, and other populations of sensory neurons like those of the vestibular organ (2). The expression of neurotrophins in the mammalian **brain** is regulated during development and by neuronal activity. During perinatal development, NT-3 shows the highest expression levels of all NTs, particularly in the **hippocampus**, neocortex and **cerebellum**, whereas BDNF expression levels in those same regions are low at birth and increase during the first weeks of postnatal development. NGF, on the other hand, behaves similarly to BDNF in the hippocampus, while in the other brain

structures NGF expression changes little during postnatal development. The NT-4 expression, on the other hand, is high in similar structures during postnatal development and continues thus into adulthood.

Functions: more than Survival Promoting Factors

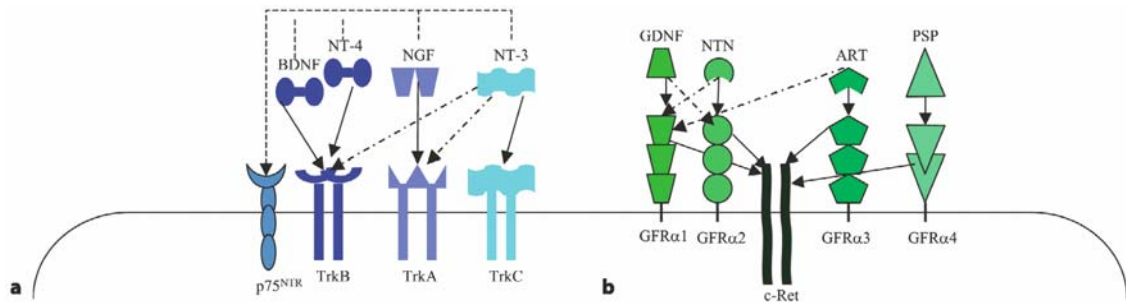
The neurotrophins were originally identified as neuronal survival factors, and the analysis of mouse strains carrying mutations in the neurotrophin genes has extensively proved that sensory and sympathetic neurons require the trophic support of these factors secreted by the innervated targets (3). For instance, deletion of NGF causes loss of sympathetic neurons and specific populations of sensory neurons, confirming the pioneering studies of Levi-Montalcini in the early 1960s, where, using antibodies against NGF, it was first demonstrated that neurotrophic factors were essential mediators of sympathetic and sensory neuron survival. A more complex picture appears in the central nervous system (CNS), where the survival of neurons depends on multiple neurotrophic factors. Although originally identified as survival promoting factors, the biological effects of neurotrophins also include regulation of axonal and dendritic outgrowth, **cell migration**, **synapse** formation and synaptic modulation (3). The trophic activity of other families of neurotrophic factors are less well characterized, e.g., the TGF β subfamilies were identified more than twenty years ago as factors controlling diverse cellular processes, including cell proliferation, differentiation, cell-specification and apoptosis during development and in mature tissues in various species from flies to worms to mammals. It is only recently that TGF β has been recognized to also play unique roles in the developing and adult nervous system. Furthermore, the effect of TGF β as a survival-promoting factor has been shown mainly in **synergism** with established neurotrophic factors such as neurotrophins, fibroblast growth factor and particularly GDNF (4).

Molecular Interactions

In order to mediate their diverse biological functions, neurotrophins interact with **receptor tyrosine kinases** belonging to the Trk receptor family of which there are three members termed TrkA, TrkB, and TrkC. The binding of neurotrophins to Trk receptors is highly specific; NGF interacts selectively with TrkA, BDNF and NT-4 with TrkB, NT-3 with TrkC and, to a lesser extent, with TrkA and TrkB (Fig. 1a). The activation of Trk receptors, as for other receptor tyrosine kinases, occurs upon binding of neurotrophin ligand, which triggers receptor dimerization (homodimers for the Trk family) and autophosphorylation of tyrosine residues in the activation loop of the kinase domain. This is followed by phosphorylation of specific cytoplasmic tyrosines that act as docking sites for **adaptor**

Neurotrophic Factors. Table 1 Partial list of neurotrophic factors and their functions in the nervous system

Neurotrophic factors	Functions in the nervous system
Neurotrophin class:	
NGF	Survival of sympathetic neurons (SCG), and sensory neurons (DRG and trigeminal). Survival of basal forebrain cholinergic neurons.
BDNF	Survival of most neuronal populations in the PNS including vestibular, nodose, petrosal, geniculate, DRG and trigeminal neurons. Differentiation and plasticity function in the CNS
NT-4	Survival of nodose, petrosal, and geniculate sensory neurons. Plasticity function in the CNS.
NT-3	Survival of embryonic sympathetic, DRG and trigeminal sensory neuron precursor cells. Survival of adult sympathetic neurons that innervate the external ear and pineal gland and of cochlear sensory neurons that innervate the organ of Corti.
Transforming growth factor β class:	
TGF- β 2, β 3	Neuron and glial cell survival and differentiation cooperating with other neurotrophic factors.
Bone morphogenetic proteins and growth/differentiation factors:	
BMP2	Survival and differentiation of striatal GABAergic neurons, induction of cholinergic properties in CNS neurons.
OP-1	Enhances dendritic growth from cerebral cortical neurons.
BMP4/7	Induction of catecholaminergic properties in PNS and CNS.
GDF-5, GDF-15	Neurotrophic effects on midbrain dopaminergic neurons.
GDNF family:	
GDNF	Survival of sub-populations of peripheral autonomic and sensory as well as spinal motoneurons, dopaminergic and noradrenergic neurons.
Neurturin, Artemin	Survival of peripheral sympathetic and sensory neurons as well as midbrain dopamine neurons.
Persephin	Supports CNS dopamine and motor neurons.
Interleukin 6 class:	
IL6	Survival of cholinergic neurons.
IL11	Neuronal differentiation of hippocampal progenitors.
LIF	Cholinergic nerve differentiation, survival of motoneurons and sympathetic neurons.
OM	Cholinergic nerve differentiation.
CNTF	Cholinergic nerve differentiation, survival of cholinergic neurons and motoneurons, astrocyte differentiation.
CT-1	Cholinergic nerve differentiation, survival of cholinergic and dopaminergic neurons.
Fibroblast growth factor class	Survival of motoneurons and some subpopulations of DRG.
Hepatocyte growth factor	Survival of motoneurons, synergic activity with other neurotrophic factors for the survival of sensory and sympathetic neurons.



Neurotrophic Factors. Figure 1 Neurotrophic factors and their receptors. (a) Neurotrophins interact with two classes of receptors the Trk receptor tyrosine kinases and the receptor p75^{NTR}. The figure illustrates the interactions of each member of the neurotrophin family with the Trk proteins and with receptor p75^{NTR}. Strong interactions are indicated with solid arrows, weaker interactions with broken arrows. Abbreviations: NGF (nerve growth factor), NT (neurotrophin), BDNF (brain derived neurotrophic factor). (b) The TGFβ proteins, GDNF subfamily, activate the transmembrane Ret tyrosine kinase via different glycosyl phosphatidylinositol-linked GFRα receptors. Strong interactions are indicated with solid arrows, weaker interactions with broken arrows. Abbreviations: GDNF (glial-derived neurotrophic factor), NTN (neurturin), ART (artemin), PSP (persephin), GFRα (GDNF family receptor α).

molecules, leading to the activation of several downstream cascades, including the Ras/mitogen-activated protein kinase (Ras/MAPK) cascade, the phosphatidylinositol 3-kinase (PI3K) pathway, and the phospholipase Cγ (PLCγ) cascade, which mediate the diverse biological functions of neurotrophins. The NTs are also able to bind to another class of receptor called p75^{NTR}, but unlike Trk receptors, p75^{NTR} exhibits predominantly low affinity binding for all NTs. The p75^{NTR} receptor belongs to the family of tumor necrosis factor (TNF) receptors. Paradoxically, the activation of this receptor by neurotrophins has been shown to promote neuronal cell death rather than survival (5). Other classes of neurotrophic factors like the GDNF subfamily, whose members include neurturin, artemin, persephin and GDNF, transduce signals in a more complex manner, *via* a common receptor subunit, the c-Ret receptor and a distinct lipid-anchored subunit of the GDNF receptor (GRFα) (Fig. 1b).

Regulatory Mechanisms

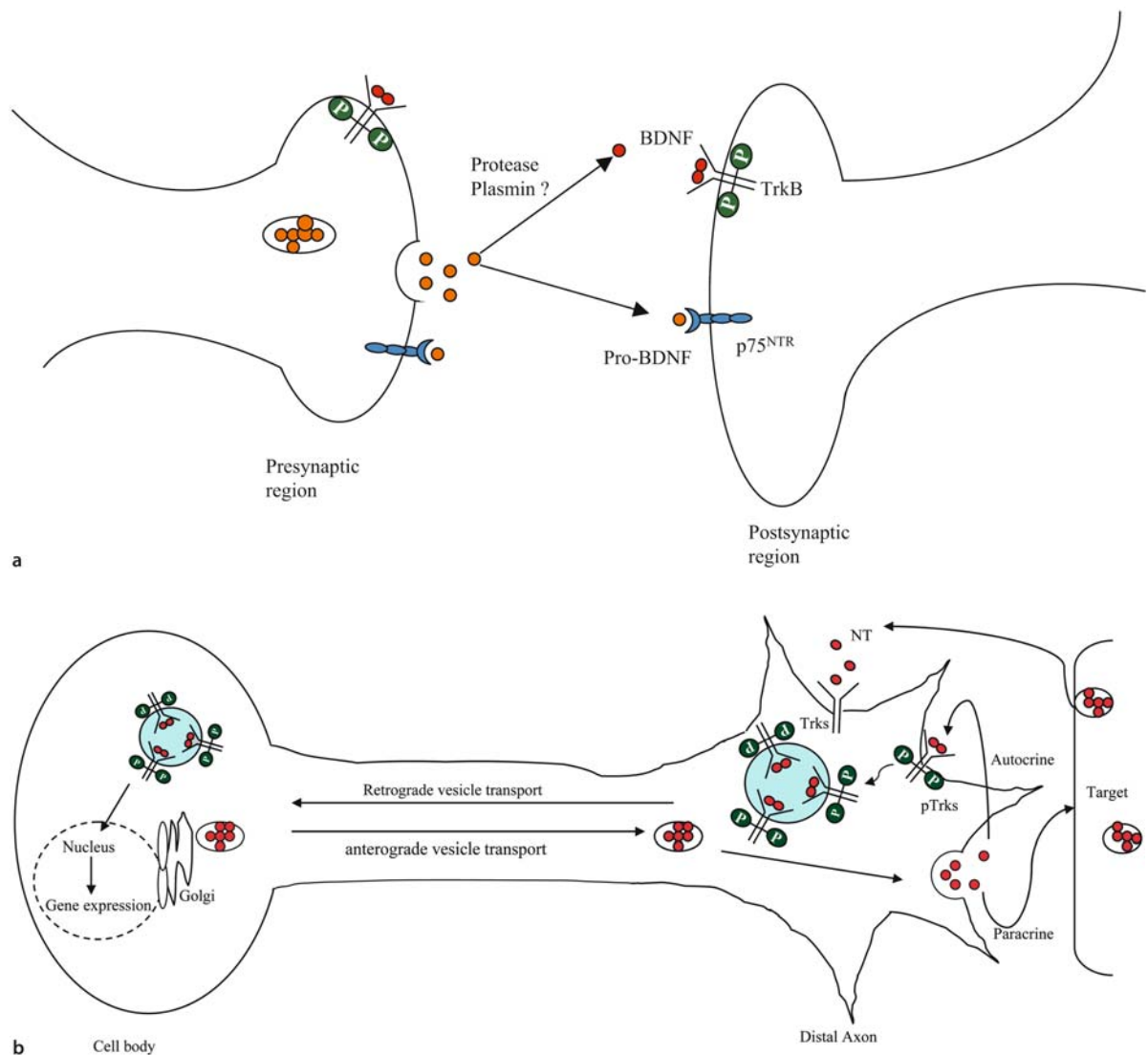
Cleavage of Neurotrophins

The NTs are synthesized as pro-neurotrophins or precursors of about 270 amino acids (aa). The pro-region (N-terminal fragment of about 120 aa) is proteolytically removed in either the trans-Golgi by furin or in secretory granules by pro-protein convertases to produce the mature proteins of about 12 kD in size. It has also been shown that proNGF and proBDNF can be cleaved *in vitro* by extracellular proteases such as MMP7 and plasmin. Recently, it has been shown that pro-neurotrophins may preferentially bind p75^{NTR}, whereas mature neurotrophins preferentially bind Trk receptors. Moreover, this differential interaction elicits opposite functional effects in that proNGF-binding p75^{NTR} induces apoptosis, whereas

mature NGF binding Trk receptor usually enhances cell survival. Secretion of pro-neurotrophins at synapses could account for regulation of synaptic transmission and plasticity by different mechanisms. This may be particularly relevant at hippocampal synapses for BDNF/TrkB signaling, which is known to regulate **LTP**. Pro-neurotrophins in the pre- or post-synaptic compartment may act *via* p75^{NTR} and elicit effects distinct from those mediated by TrkB and mature BDNF, or they may be simply cleaved at synapses resulting in active ligands for Trk receptors. Plasmin is a good candidate protease for the cleavage of proBDNF not only because it has been shown *in vitro* to cleave proBDNF at the appropriate site, but also because it is expressed at synapses (Fig. 2a).

Secretion

Compared to other regulatory mechanisms little is still known about the modes and sites of neurotrophin secretion. It is known that upon synthesis NTs are packed into secretory vesicles. Two types of vesicle secretion have been described, constitutive or regulated, depending on whether their secretion occurs in absence or presence of specific triggering factors respectively. The vesicles of the regulated pathway fuse to the plasma membrane only in a Ca²⁺-dependent manner. However, due to the important role of BDNF in modulating synapse efficacy there is great interest in the identification of the sites and mode of BDNF secretion. Mainly for this reason most progress has been made recently concerning neuronal BDNF secretion at axonal and dendritic sites. Recent data suggest that BDNF secretion depends critically on high intracellular Ca²⁺-levels that can be triggered by electrical activity. For more details see reference (6).



Neurotrophic Factors. Figure 2 Regulatory mechanisms of neurotrophins. (a) Extrasynaptic cleavage of proBDNF. Uncleaved pro-neurotrophins may act on p75^{NTR} on pre and post-synaptic membranes and elicit effects distinct from those mediated by Trk receptors. ProBDNF could be cleaved extracellularly from protease like plasmin, implicated in learning and memory. (b) Neurotrophin signaling. Retrograde neurotrophin signaling, endosome-mediated model. Trk receptors on distal axons are activated upon binding to neurotrophins. The ligand-receptor complex internalizes through clathrin-mediated endocytosis. Some of the vesicles become specialized endosomes and are transported retrogradely to the cell body using a dynein-dependent and microtubule-dependent transport mechanism. The vesicle-associated Trk receptor remains autophosphorylated and capable of promoting a signal at nucleus level. Through the anterograde transport NT vesicles can also be transported from the cell soma to the axon terminals. After pre-synaptic secretion, NT actions can be autocrine or paracrine, depending on the site of receptor expression.

Retrograde/Anterograde Transport

Following secretion, constitutive or regulated, by the innervated target tissue, NTs bind to axonal neurotrophin receptors and are endocytosed in vesicles, so called "signaling endosomes". They are then retrogradely transported, according to the classical view, in the axon to the cell soma where they can

activate signaling cascades promoting survival (Fig. 2B). The retrograde transport was described first for NGF in sympathetic and sensory neurons of the PNS and in cholinergic neurons of the CNS. Subsequently, this mechanism was shown also for the other members of the NGF family, BDNF, NT-3 and NT-4, in the PNS as well as in the CNS neurons. However, recent studies

have shown that endogenous BDNF and NT-3 can also be transported anterogradely in the axon towards synaptic terminals where they are stored in vesicles, released and elicit postsynaptic effects. This mechanism, especially relevant for BDNF, contrasts with the traditional role of NGF as a retrogradely transported target-derived survival factor. BDNF, being anterogradely transported by central or peripheral neurons, can exert multiple roles, a neuron-transmitter like role for rapid postsynaptic functions, a trophic factor for postsynaptic targets such as skin and muscle, a source of BDNF for astrocytes, especially in injured brain etc. This alternative mechanism of trafficking described for these molecules has opened up many questions regarding the function of neurotrophins as well as of other neurotrophic factors in the biology and pathology of the nervous system (6).

Neurotrophins and Disease

Neurotrophin genes have been implicated in neurological or psychiatric disorders. A recent study has linked a ►[polymorphism](#) in the pro-domain of BDNF with depression, bipolar disorders and schizophrenia. This polymorphism is caused by a single amino-acid change, from valine (Val) to methionine (Met), at position 66 in the pro-domain of BDNF. In patients with bipolar disorder or depression, the Val ►[allele](#) seems to confer greater risk for the disease, whereas in patients with schizophrenia, the Met allele seems to be associated with impaired memory function. In animal models several lines of evidence have implicated neurotrophins in depression, for example restraint stress leads to decreased expression of BDNF in the hippocampus and administration of BDNF to the midbrain or hippocampus results in antidepressant effects. In ►[Huntington's disease](#) decreased production of BDNF has been shown; this could lead to insufficient neurotrophic support for striatal neurons, which then die. NGF has been implicated in the pathogenesis of ►[Alzheimer's Disease](#). Mutations in TrkA, the high affinity receptor for NGF, have been identified as the cause of a human syndrome, congenital insensitivity to pain and anhidrosis (CIPA). This disease is an ►[autosomal-recessive disorder](#) in which individuals exhibit defects in thermoregulation, anhidrosis (absence of sweating), mental retardation, unexplained fever and a failure to react to noxious stimuli (7).

References

1. Murray-Rust J, McDonald NQ, Blundell TL et al (1993) Topological similarities in TGF-beta 2, PDGF-BB and NGF define a superfamily of polypeptide growth factors. *Structure* 1:153–159
2. Fritsch B, Silos-Santiago II, Bianchi LM et al (1997) Effects of neurotrophin and neurotrophin receptor disruption on the afferent inner ear innervation. *Semin Cell Dev Biol* 8:277–284
3. Bibel M, Barde YA (2000) Neurotrophins: key regulators of cell fate and cell shape in the vertebrate nervous system. *Genes Dev* 14:2919–2937
4. Unsicker K, Krieglstein K (2002) TGF-betas and their roles in the regulation of neuron survival. *Adv Exp Med Biol* 513:353–374
5. Huang EJ, Reichardt LF (2001) Neurotrophins: roles in neuronal development and function. *Annu Rev Neurosci* 24:677–736
6. Lessmann V, Gottmann K, Malsangio M (2003) Neurotrophin secretion: current facts and future prospects. *Prog Neurobiol* 69:341–374
7. Egan MF, Kojima M, Callicott JH et al (2003) The BDNF val66met polymorphism affects activity-dependent secretion of BDNF and human memory and hippocampal function. *Cell* 112:257–269

Neutrophils

Definition

Neutrophils (also referred to as neutrophil granulocytes) form the biggest subpopulation of granulocytes, a population of white blood cells. The name results from the relatively neutral colour that their granules have when routinely stained. Neutrophils arise from the bone marrow, are released into the circulation and have an important function as the first line of cellular defense. Neutrophils are highly motile. They are attracted by cytokines expressed by activated endothelium, mast cells and macrophages during an infection. Neutrophils quickly gather at the site of an infection. When activated, they release reactive oxygen species and proteinases to defend invading bacteria. Furthermore, they can act as phagocytes to debride the wound.

- [Apoptosis](#)
- [Morbus Wegener](#)
- [Wound Healing](#)

NEXT

Definition

NEXT stands for Notch Extra-Cellular Truncation. It is the name for the Notch intermediate created during activation of the receptor by proteolytic cleavage.

- [Notch Pathway](#)

NF1 Gene

Definition

The NF1 (neurofibromatosis 1) gene is located at 17q11.2; it spans 300 kb genomic DNA and contains 60 exons. The gene encodes for the Ras-GTPase-activating protein ►[neurofibromin](#).

The NF1 gene is classified as a tumor suppressor gene, since decrease or absence of NF1 protein is, for example, seen in malignant tumor cell lines from patients.

►[Neurofibromatosis Type 1 \(NF1\), Genetics](#)

NFT

►[Neurofibrillary Tangles](#)

NFκB

Definition

Nuclear factor kappa B (NFκB) proteins comprise of a family of structurally-related eukaryotic transcription factors, which are involved in a large number of normal cellular and organismal processes, such as immune and inflammatory responses, developmental processes, cellular growth, and apoptosis. They are also important during disease states, including cancer, arthritis, chronic inflammation, asthma, neurodegenerative diseases, and heart diseases.

►[Inflammatory Response](#)

►[NFκB Pathway](#)

NF-κB Pathway

CLAUS SCHEIDEREIT, DANIEL KRAPPMANN
Max Delbrück Center for Molecular Medicine,
Berlin, Germany
scheidereit@mdc-berlin.de

Definition

The nuclear factor-kappaB (NF-κB, also called NF-κB/Rel) family consists of five evolutionarily conserved

mammalian ►[transcription factors](#), p50, p52, p65/RelA, c-Rel and RelB, which are widely expressed in human tissues. NF-κB family members form various dimeric complexes and the most prominent NF-κB species is a p50-p65 heterodimer. In resting cells, NF-κB is sequestered in the cytoplasm by inhibitory proteins (IκBs). A large number of immunological, inflammatory, morphological and mitogenic stimuli (Table 1) induce destruction of IκBs, which is followed by nuclear entry of transcriptionally active NF-κB. All known stimulatory pathways converge at the IκB kinase (IKK) complex, which is composed of two catalytic subunits (IKKα and IKKβ) and a regulatory (IKKγ, NEMO) subunit. Upon activation, the IKK complex phosphorylates IκBs, which are then targeted for rapid ubiquitination and for subsequent degradation by the 26S proteasome. Released NF-κB dimers are imported into the nucleus, where they induce gene expression through high affinity binding to DNA response elements related to a decameric consensus sequence (GGGRNYYCC), located in ►[promoter](#) or ►[enhancer](#) regions of target genes (Table 1). NF-κB-dependent gene regulation is involved in diverse cellular processes, including differentiation, apoptosis, proliferation and immune and inflammatory responses. Consequently, deregulation of the pathways that promote NF-κB activation is often found associated with tumor development and inflammation and diseases of the vascular system, the immune system and the epidermis.

Characteristics

NF-κB is a prototypic transcription factor whose activation involves serine/threonine phosphorylation, ubiquitin conjugation and proteasomal proteolysis of inhibitors (1, 2, 3, 4). A characteristic feature of the NF-κB regulatory system is the multiple steps at which activation, as well as post-inductive deactivation, can be tightly controlled. NF-κB pathway regulation can occur at the following levels: (i) the relative expression levels and dimerization of the various NF-κB/Rel family members, (ii) the binding to the family of inhibitory IκB molecules, (iii) the induced release from IκBs through their phosphorylation by IκB kinases (IKKs), the rate limiting step in the activation pathway, which triggers IκB proteolysis by the ►[ubiquitin-proteasome pathway](#), (iv) the autoregulatory termination of pathway activity by NF-κB-stimulated IκB expression, (v) the fate of precursor proteins, which are modified by proteasomal proteolysis and (vi) subunit-specific regulation of NF-κB transcriptional activity by phosphorylation and association with nuclear IκB family members and by interaction with histone acetylases or deacetylases.

NF-κB Pathway. Table 1 A brief selection of NF-κB inducing agents and NF-κB target genes. For more extensive lists of inducers and target genes see <http://people.bu.edu/gilmore/nf-kb>

NF-κB inducers	NF-κB target genes
Inflammatory cytokines	Inflammatory cytokines
Interleukin -1 α/β	Interleukin-1 α/β
Interleukin-2	Interleukin-2
Lymphotoxin α/β	Interleukin-6
Tumor Necrosis Factor (TNF) α/β	Lymphotoxin α/β
	Tumor Necrosis Factor α/β
Immune receptor signaling	TRAIL
CD28	
CD40	Chemokines
B cell receptor (BCR)	Macrophage inflammatory protein 1 α/β
T cell receptor (TCR)	Macrophage inflammatory protein 2
	RANTES
Microbial pathogens	
Bacteria (e.g. <i>H. pylori</i> , <i>S. aureus</i>)	Cell surface receptors
Bacterial products (e.g. lipopolysaccharides, enterotoxin)	Tumor Necrosis Factor receptor-1
	B cell receptor (Immunoglobulin κ-light chain)
Viruses (e.g. HIV, HTLV, EBV)	T cell receptor (β-chain)
Viral products (e.g. HIV-TAT, HTLV-Tax, EBV-LMP)	Interleukin-2 receptor (α-chain)
	Chemokine receptor 7
	FAS/CD95
Stress conditions	
UV-light	Cell adhesion molecules
γ-irradiation	Intracellular adhesion molecule-1
Hydrogen peroxide	Vascular adhesion molecule-1
Hypoxia	P-selectin
Chemical agents	Apoptosis regulators
Phorbol ester	Bfl1/A1
Okadaic acid	Okadaic acid
Calyculin A	Inhibitors of apoptosis (IAPs)
Chemotherapeutic drugs	Cell proliferation
Daunorubicin	Cyclin D1/D3
Doxorubicin	p21-CIP1

NF-κB Pathway. Table 1 A brief selection of NF-κB inducing agents and NF-κB target genes. For more extensive lists of inducers and target genes see <http://people.bu.edu/gilmore/nf-kb> (Continued)

NF-κB inducers	NF-κB target genes
	Transcription factors/auto-regulation
	IκBα
	p105/NFκB1
	p100/NFκB2
	RelB
	MAIL
	JunB
	B-ATF

Molecular Interactions

NF-κB Dimerization

All NF-κB family members share a conserved amino-terminal Rel homology domain (Rel domain) of approximately 300 amino acids (1) (Fig. 1). The Rel domain contains the dimerization interface, the DNA binding domain to recognize the decameric consensus sequences and the ►nuclear localization signal (NLS) for interaction with the nuclear import machinery. In addition, p65, cRel and RelB contain a carboxy-terminal transcription activation domain (TAD), which communicates with general factors of the transcription initiation complex. In contrast, p50 and p52 lack a TAD and require heterotypic association to act as transcriptional activators. p50 and p52 are synthesized as cytoplasmic precursor proteins (p105 and p100, respectively) and the generation of the mature transcription factors depends on a proteolytic processing reaction, that removes sequences C-terminal to the Rel domain (3, 4). The processing end-point is close to a glycine rich region (Fig. 1).

IκB Association

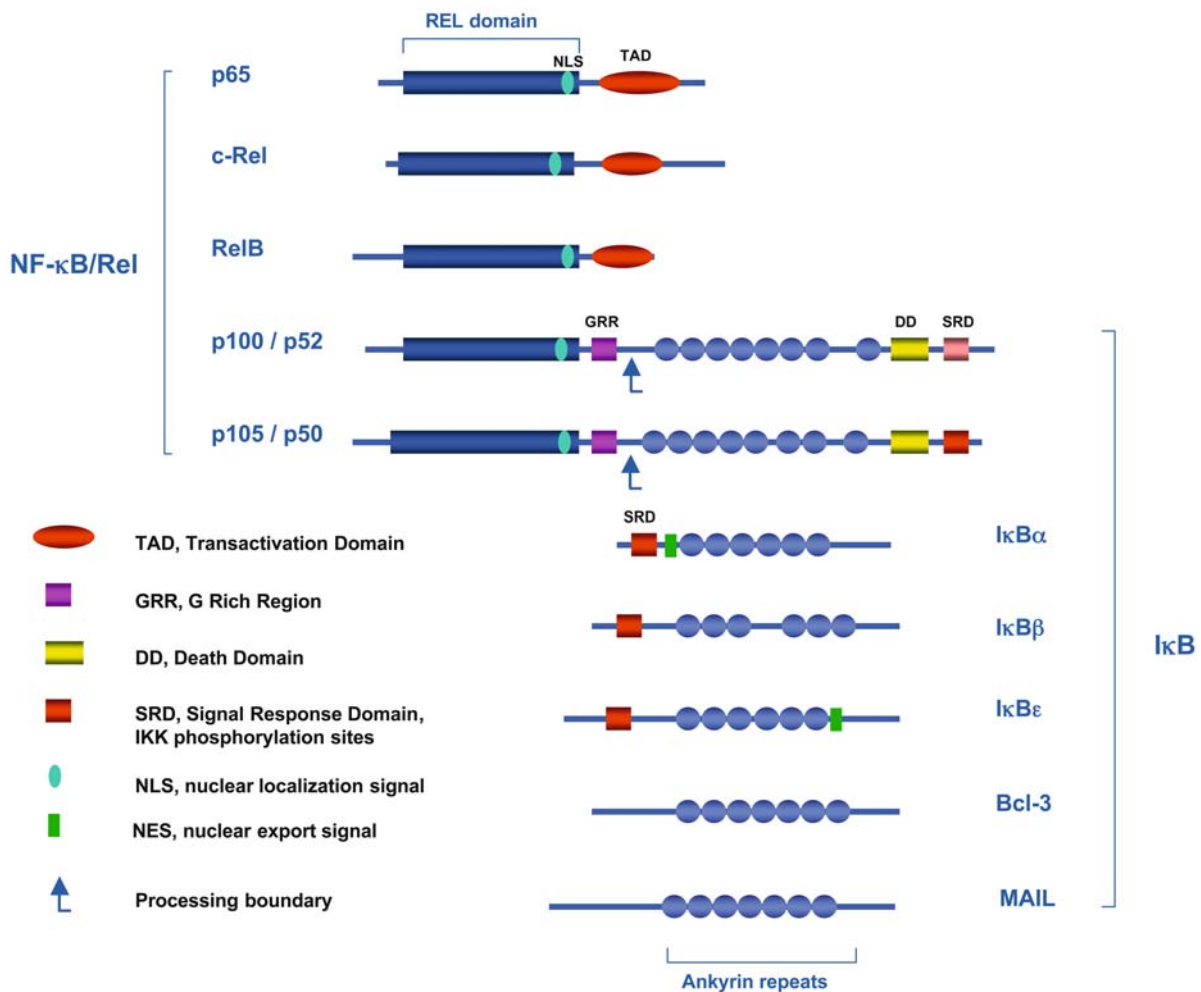
The activity of NF-κB is tightly controlled by its association with IκB proteins (Fig. 1). The IκB family is composed of the two cytoplasmic precursor molecules p105 and p100, the small cytoplasmic inhibitors IκBα, IκBβ and IκBε and the predominately nuclear IκB family members Bcl-3 and MAIL. The common structural motif of the IκBs is the ankyrin repeat domain (ARD), an interaction surface that mediates high affinity binding to the Rel domain of NF-κB dimers (3). While IκBα, IκBβ and IκBε preferentially interact with NF-κB heterodimers containing the p65 or c-Rel subunit, Bcl-3 and MAIL bind with highest affinity to p50 or p52 homodimers. p105 and p100 can bind to all NF-κB subunits, however p100 appears to be the only high affinity IκB for RelB.

In resting cells, cytoplasmic IκBs prevent nuclear entry of NF-κB by shielding the NLS at the carboxy-terminus of the Rel domain. Rapid IKK dependent proteasomal destruction of IκBs in response to activating agents produces accessibility and recognition of the NLS by importins, which drag NF-κB complexes to the nucleus. In contrast, Bcl-3 and MAIL associate with NF-κB-p50 or p52 dimers in the nucleus and function as co-regulators to affect expression of a subset of NF-κB regulated genes.

The genes coding for the inhibitory molecules *ikb*, *ikbe*, *p105/nfkb1* and *p100/nfkb2*, are primary NF-κB target genes (Table 1). Newly synthesized IκBα can enter the nucleus and dissociate NF-κB from DNA. A ►nuclear export signal (NES) on IκBα, which interacts with the nuclear export receptor Crm1, serves for export of IκB-NF-κB complexes. Thus, IκBα gene induction contributes to an auto-regulatory feedback mechanism and ensures that NF-κB activity is transient and ceases upon removal of the stimulus (3). Expression of the nuclear IκB homologue MAIL is strongly increased by a subset of NF-κB inducers (e.g. lipopolysaccharides and IL-1) and may positively modulate a secondary transcriptional response of NF-κB p50 or p52 homodimers, which escape the inhibitory action of *de novo* synthesized IκBα. Besides their inhibitory role in resting cells, the molecular interactions with IκBs establish a fine-tuning of the NF-κB transcriptional response in activated cells.

IκB Kinase Complex

Biochemical and genetic evidence identified the IKK complex as the central point of convergence of all known physiological NF-κB activating stimuli (2, 3, 4). The kinase complex is composed of three subunits: (i) the catalytic subunits IKKα and IKKβ, which hetero- or homodimerize through their ►leucine zippers and directly phosphorylate the SRD of the IκBs and (ii)



NF- κ B Pathway. Figure 1 The proteins encoded by the NF- κ B and I κ B multigene families of vertebrates. AR, ankyrin repeats; DB, DNA binding domain; DD, death domain; Dim, Dimerization domain; GRC, G rich region; NLS, nuclear localization sequence; RHD, Rel homology domain; SRD, signal response domain. The vertebrate NF- κ B proteins p105, p100, p50, p52, p65, c-Rel and RelB share a conserved DNA-binding and dimerization domain (Rel Domain) of about 350 amino acids which contains a nuclear translocation signal. Frequently used synonyms are RelA for p65, NFKB1 for p105, and NFKB2 or I κ B-10 for p100. p65, c-Rel and RelB contain C-terminal transcriptional activation domains (TAD). The p50 and p52 subunits are produced by processing of the precursor proteins p105 and p100. The C-terminal portions of the precursors contain repeated sequence motifs (*ankyrin repeats*), which are the characteristic conserved feature of the prevalently cytoplasmic group of I κ B proteins, I κ B α , I κ B β and I κ B ϵ and their predominantly nuclear homologues Bcl-3 and MAIL. Both precursors contain a conserved *death domain* (DD) a protein-protein interaction motif, which confers binding of I κ B kinases. Both precursors and the prevalently cytoplasmic I κ Bs carry a signal response domain (SRD), which upon phosphorylation by IKK attracts β ^{TRCP} ubiquitin ligases. Processing boundary, processive proteolysis limit close to a GRR. For conservation of NF- κ B and I κ B components in insects (1).

the functionally essential IKK γ /NEMO component, which has been proposed to recruit upstream activators into the complex (3). Dimeric IKK α / β molecules bind to IKK γ *via* a short peptide sequence in the very carboxy termini. IKK γ oligomerizes and a conserved **zinc-finger domain** at its carboxy terminal end is required for activation of the complex by a subset of

stimuli (3). Whereas IKK α and IKK β seem to be functionally redundant *in vitro*, studies in gene deficient mice reveal a prominent role of IKK β for rapid NF- κ B activation by pro-inflammatory stimuli and various other agents (2, 3). In contrast, IKK α contributes to an alternative NF- κ B pathway that induces processing of the precursor molecule p100

(see below) and whose function might be limited to prolonged processes, such as differentiation (3, 4). However, TNF α induces rapid translocation of IKK α to the nucleus, where it contributes to cytokine-induced chromatin-modification and gene activation (3, 4, 5). For IKK α , β and γ , the [▶chaperones](#), heat shock protein 90 (Hsp90) and Cdc37, are stoichiometric constituents of the IKK complex and are required for activation of the complex by TNF α , IL-1 or PMA. Hsp90 is bound *via* its co-chaperone Cdc37 to the kinase domain of IKK α and β (3).

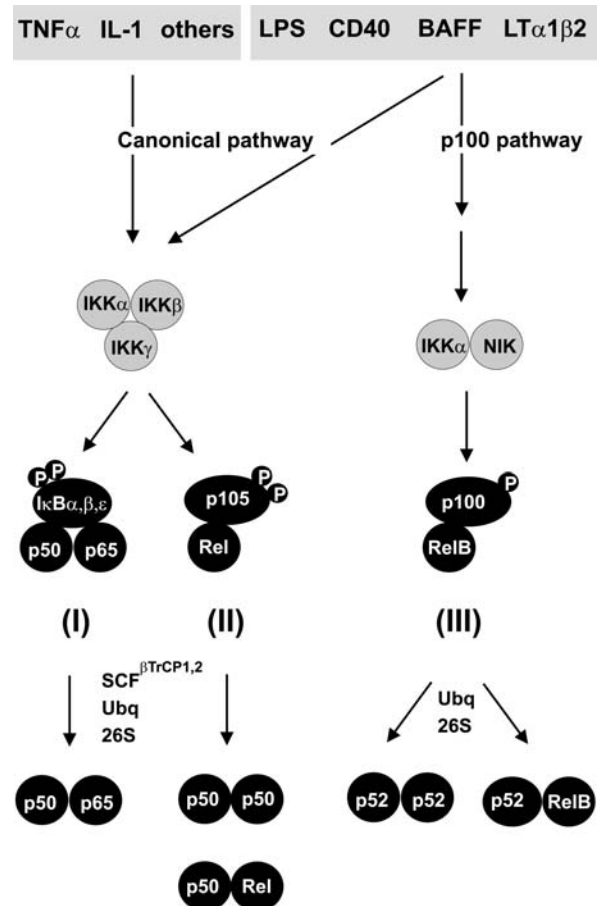
Regulatory Mechanisms

Molecular Mechanisms of NF- κ B Activation

The activation of NF- κ B activity by all known physiological stimuli depends on the activation of IKK. It is assumed that the activation of IKK kinase activity involves phosphorylation of IKK α and IKK β at serine residues in the activation loop in the conserved kinase domain. Several mitogen activated protein 3 (MAP3) kinase kinases, including NF- κ B-inducing kinase (NIK), mitogen-activated protein/ERK kinase kinases (MEKK1 and MEKK3), TGF- β -activated kinase 1 (TAK1) and mixed lineage kinase 3 (MLK3) have been suggested as upstream kinases and phosphorylate the IKK complex and activate NF- κ B when over-expressed or when analyzed *in vitro* (3, 4). Phosphorylation of the activation loop serines may also be achieved by trans-autophosphorylation of the kinases in an induced proximity mechanism. IKK α and IKK β are not biochemically equivalent and IKK α activation loop phosphorylation is not required for IKK activation by TNF α or IL-1 (3). Gene ablation experiments have established that IKK β and IKK γ are required for NF- κ B activation by proinflammatory stimuli, while IKK α is essential for morphogenic signals but also contributes to cytokine signaling (2, 3). For the TNF α pathway, recruitment of TNF receptor associated factors 2 or 5 (TRAF2, TRAF5) and receptor interacting protein 1 (RIP1) initiates downstream signaling. Furthermore, the IKK complex was shown to associate directly with the TNF receptor 1 complex and genetic evidence suggested that MEKK3 is needed for TNF induction (3).

The activated IKK complex phosphorylates I κ B α , β or ϵ at conserved serines in amino terminal sequences of the molecules. p105 contains a similar phosphorylation site in its carboxy terminus, which is phosphorylated by activated IKK as well. The phosphorylated sites in I κ B α , β , ϵ or p105 are bound by an SCF E3 ubiquitin ligase complex *via* F box containing substrate recognition molecules β TrCP1 and β TrCP2 (1, 2). The substrate I κ B molecules are then poly-ubiquitinated and completely degraded by the 26S proteasome (Fig. 2) canonical pathway (I). In contrast to the signal-induced destruction of the full-length p105

molecule, a basal processing reaction continuously converts a part of the cellular p105 pool to p50 by ubiquitin and proteasome dependent proteolysis of its C-terminal half. Depending on the relative expression levels of the NF- κ B subunits, p105 either binds its own processing product p50 or other Rel factors. Induced p105 degradation liberates these Rel factors (Fig. 2,



NF- κ B Pathway. Figure 2 General NF- κ B activation mechanisms. Most NF- κ B inducing agents rapidly activate the canonical pathway, which depends on the classical I κ B kinase complex; the IKK β and IKK γ subunits are essential. Small I κ B molecules (e.g. I κ B α) are phosphorylated, ubiquitinated after recognition by SCF $^{\beta$ TrCP E3 ubiquitin ligase complexes, degraded by the 26S proteasome (I) and release p50-p65 heterodimers. p105 complexes are phosphorylated as well, which results in SCF $^{\beta$ TrCP-dependent ubiquitination, proteasomal destruction of p105 and release of p105-bound Rel subunits (II). The p100-specific pathway ("novel pathway") is activated by a limited number of stimuli, which can also activate the canonical pathway. The p100 pathway specifically requires IKK α and NIK. Phosphorylation and ubiquitination of p100 results in its partial proteolysis by the proteasome, to generate the processing product p52, which associates with RelB (III).

(II)). The canonical pathway, which is induced by pro-inflammatory cytokines and various pathogenic agents is fast and results in degradation of I κ B α , β , ϵ and p105 within minutes to less than an hour and without requiring *de novo* protein synthesis. The vast number of rapidly activated NF- κ B target genes (Table 1) are activated by the canonical pathway and the predominant NF- κ B form, the p50-p65 heterodimer. A growing number of ligands, including lymphotoxins, B cell activating factor (BAFF) and CD40 ligand, which all activate the canonical cascade, also induce the p100 pathway. The latter, also known as the "novel pathway" results in induced proteolytic processing of p100 to p52 (3, 4). IKK α and NIK are required for phosphorylation of serines in a carboxy terminal signal response domain of p100, which in turn is essential for ubiquitination and degradation of the C-terminal half of p100 by the 26S proteasome. Induced p100 processing liberates p52-RelB heterodimers, since RelB is largely bound to p100 (Fig. 2, (III)). In contrast to the canonical pathway, induced p100 processing occurs only after hours of stimulation and requires the translation of unknown intermediates. It is assumed that the LT β receptor activated p100 pathway induces select target

genes, including the chemokines, B lymphocyte chemokine (BLC), secondary lymphoid tissue chemokine (SLC) and EB11-ligand chemokine (ELC) (3, 4).

Mechanisms of NF- κ B Dependent Transactivation

The release of NF- κ B proteins from cytoplasmic complexes with I κ B proteins and precursors allows nuclear import, but is not *per se* sufficient for full transcriptional activity of NF- κ B. A number of studies have highlighted the importance of p65 phosphorylation. p65 is the target of a variety of kinases and the sites of phosphorylation are both in the Rel domain and in the TAD (3, 5). One important site is serine 276, which is phosphorylated depending on the stimulus either by ►protein kinase A (PKA) or by mitogen- and stress-activated kinase-1 (MSK1). Phosphorylation at this site increases the association with transcriptional co-activators CBP and p300 and the transactivation potential (3, 5). Mice deficient in protein kinase C ζ (PKC ζ), glycogen-synthase kinase-3 β (GSK3 β) and TANK-binding kinase-1 (TBK1) all reveal impaired NF- κ B transcriptional activities, implying that phosphorylation of p65 or other subunits by these kinases may be functionally essential. Furthermore, p65 has

NF- κ B Pathway. Table 2 NF- κ B/Rel and IKK knockout phenotypes. An overview of the major physiological functions of NF- κ B and IKK gene products as revealed by homologous recombination in mice

Gene	Lethality	Affected Function
p50*		immune response (B cell function)
p52**		maintenance of peripheral B cell populations and normal spleen architecture
p65 (RelA)	embryonic lethality	anti-apoptotic function, liver
c-Rel		immune response (T and B cell function)
RelB	often premature death	normal function of hematopoietic system PP formation
p50*/p52**	premature death	osteoclast differentiation
p50*/ReB	as RelB, but stronger	immune response
IKK α	death at birth	normal limb formation, hyperproliferation of epidermal keratinocytes, embryonal PP formation
IKK β	death at E14.5	anti-apoptotic function, same as p65
IKK γ	male: death at E12.5-13.0 female: incontinentia pigmenti	same as p65 and IKK β
IKK α /IKK β	death at E12	neural tube closure, hepatocyte apoptosis, neuronal epithelium (hind brain), spinal cord and dorsal root ganglia formation

PP, Peyer's patches

* NF κ B1 (p105/p50)

**NF κ B2 (p100/p52)

been shown to be acetylated by histone acetyltransferase (HAT) activities, including the p300/CBP and p300/CBP-associated factor (PCAF) co-activators in a signal-dependent manner (3, 5). Acetylated p65 can be deacetylated by histone deacetylase 3 (HDAC3). Deacetylation may be important for nuclear-cytoplasmic shuttling of p65-I κ B α complexes, because acetylated p65 interacts poorly with I κ B α (3, 5). Another requirement for the transcriptional induction of NF- κ B target genes is inducible chromatin modification exerted by IKK α (4, 5). In response to TNF α , a fraction of the cellular IKK α translocates to the nucleus, where it is recruited to promoters of responsive genes, as was shown for I κ B α and IL-6. IKK α contains a nuclear localization sequence. It is unknown how the nuclear import and recruitment of IKK α to promoters is regulated. It has been proposed that IKK α is required for histone H3 phosphorylation at serine 10 and subsequent acetylation of histone H3, which in turn is required for efficient transcription activity (4, 5).

Physiological and Pathological Regulatory Networks Which Depend on IKK and NF- κ B Activity

The biological responses to IKK and NF- κ B signaling largely depend on transcriptional up-regulation of a network of direct target genes that contain NF- κ B consensus sites in their promoter or enhancer regions. The functional groups of these genes (Table 1) underscore the range of cellular and physiological functions of IKK and NF- κ B, which include innate and adaptive immunity, inflammatory cytokine production, differentiation, programmed cell death and cell-cycle progression (3). A number of knockout studies have substantiated that individual NF- κ B genes are generally required for proper immune cell function (6) (Table 2). Protection against embryonic liver cell apoptosis is a further important function shared by p65, IKK β and IKK γ . Other functions include bone morphogenesis and skin proliferation and differentiation (Table 2). In the skin, IKK α regulates the differentiation of keratinocytes. However, recent studies have shown that IKK α does not require its kinase activity and does not activate NF- κ B to maintain its differentiation function (3). Thus, in addition to NF- κ B, IKK α also appears to regulate other as yet unknown effectors.

Since NF- κ B activates the expression of inflammatory cytokines and chemokines, it plays a key role in chronic inflammation. NF- κ B-associated pathological conditions include a wide spectrum of autoimmune and inflammatory diseases, such as rheumatoid arthritis, psoriasis and other skin conditions, asthma, Crohn's disease and atherosclerosis. Due to its anti-apoptotic and cell division activating effects, NF- κ B has been recognized to be of importance in the resistance of tumors to chemo- or radio-therapy (7). Constitutive

activation of IKK and NF- κ B is considered to be a critical step in the pathogenesis of certain lymphomas and leukemias (7).

References

1. Silverman N, Maniatis T (2001) NF- κ B signaling pathways in mammalian and insect innate immunity. *Genes Dev* 15:2321–2342
2. Hatada EN, Krappmann D, Scheidereit C (2000) NF- κ B and the innate immune response. *Curr Opin Immunol* 12:52–58
3. Ghosh S, Karin M (2002) Missing pieces in the NF- κ B puzzle. *Cell* 109:S81–96
4. Yamamoto Y, Gaynor RB (2004) I κ B kinases: key regulators of the NF- κ B pathway. *Trends Biochem Sci* 29:72–79
5. Chen LF, Greene WC (2004) Shaping the nuclear action of NF- κ B. *Nat Rev Mol Cell Biol* 5:392–401
6. Gerondakis S, Grossmann M, Nakamura Y et al (1999) Genetic approaches in mice to understand Rel/NF- κ B and I κ B function: transgenics and knockouts. *Oncogene* 18:6888–6895
7. Rayet B, Gelinis C (1999) Aberrant rel/nfkb genes and activity in human cancer. *Oncogene* 18:6938–6947

NGF

Definition

NGF stands for Nerve Growth Factor. It regulates development and survival of nerve cells.

- Growth Factors
- Neurotrophic Factors

N-Glycan

- N-Linked Oligosaccharide

NHEJ

- Non-Homologous End Joining

Nick-Closing Enzymes

- DNA Topoisomerases

Nicotine

Definition

Nicotine is the primary component found in smoke from tobacco products, which acts on the brain and induces addiction. Nicotine shares discrete neurochemical and functional properties with other addictive drugs. It accelerates the release of dopamine in the nucleus accumbens and the amygdala as well as other limbic structures. Addiction to nicotine is an enormous health, social, and financial burden.

► [Addiction, Molecular Biology](#)

Nieuwkoop Center

Definition

Nieuwkoop center refers to a conceptual domain that generates intercellular signaling molecules to induce the dorsal organizer in early embryonic development.

► [Axis Formation – Formation and Function of the Dorsal Organizer](#)

Ni-NTA

Definition

Ni-NTA stands for Ni^{2+} -Ions bound to nitrilotriacetic acid. Ni-NTA resin is used to purify proteins containing a 6xHis tag (► [His-tag/6xHis](#)), since the tag binds to divalent cations; other cations such as Co^{2+} are also used. The bound protein is eluted from the resin using high concentrations of imidazole. Imidazole mimics the histidine side chain of the tag and thus competes for the binding sites on the resin.

► [Recombinant Protein Expression in Bacteria](#)

Nintra

Definition

Nintra is the activated form of Notch, consisting of the intracellular domain of Notch released through the sequence of proteolytic cleavages.

► [Notch Pathway](#)

Nitrosative Stress

Definition

By analogy to “oxidative stress”, nitrosative stress is characterized by overproduction of nitric oxide ($\cdot\text{NO}$). It is often complicated by simultaneous production of superoxide anions, which results in the formation of peroxynitrite and other reactive nitrogen species.

► [Free Radicals](#)

Nkd

► [Naked Cuticle](#)

N-Linked Oligosaccharide

Definition

N-linked oligosaccharide (N-glycan) designates a glycan in amide linkage to the side chain of asparagine residues on a glycoprotein.

► [Glycosylation of Proteins](#)

NLK

► [Nemo Like Kinase](#)

NLS

► [Nuclear Localization Signal](#)

NMD

► [Nonsense-Mediated RNA Decay](#)

NMDA Receptor

Definition

The ligand for the NMDA (N-methyl-D-aspartate) receptor is the neurotransmitter glutamate, which is the most important excitatory transmitter in the ►brain. NMDA is the synthetic agonist of the NMDA receptor. The NMDA receptor is a ligand- and voltage-gated ionic channel. It seems to be involved in synaptic plasticity, target recognition, and toxic effects of excessive glutamate, important e.g. during stroke.

►Addiction, Molecular Biology

- 3D Structure Determination by NMR
- Protein/DNA interaction
- Structure-Based Drug Design

NMR-based Screening

WOLFGANG JAHNKE

Novartis Institutes for BioMedical Research,
Novartis Pharma AG, Discovery Technologies,
Basel, Switzerland
wolfgang.jahnke@novartis.com

NMR

Definition

NMR stands for Nuclear Magnetic Resonance. This spectroscopic technique enables the identification and characterization of molecular interactions. Other applications of biomolecular NMR are elucidation of protein structure and dynamics. Atomic nuclei containing a non-vanishing nuclear spin quantum number display separation of energy states in a static magnetic field (Zeeman splitting). Transitions between these energy states (resonance) can be induced by absorption of electromagnetic radiation of the appropriate frequency ($\Delta E = h \cdot \nu$).

- 3D Structure Determination by NMR
- NMR-Based Screening
- Protein-Ligand-Interaction by NMR
- Protein Databases

Definition

►NMR-based screening refers to the identification of ligands for a pharmaceutically relevant target protein by the use of NMR spectroscopy (1, 2, 3, 4). NMR screening detects binding of ligands to a protein, rather than inhibition of protein function: NMR screening provides a binding assay, not a functional assay. NMR screening can be performed by protein observation, ligand observation or observation of a reporter ligand. A particular strength of NMR screening is its ability to robustly detect even weak protein-ligand interactions. This makes NMR screening particularly suitable for ►fragment-based ligand design (FBLD), a method for lead generation that is complementary to high-throughput screening. The lead compound can be further optimized with respect to potency, physico-chemical, pharmacological and pharmacokinetic properties and can, after successful optimization, be tested in clinical studies.

Description

Fragment-based Ligand Design

Like high-throughput screening (HTS), fragment-based ligand design is a method used to identify ligands for a target protein. Unlike HTS, which randomly screens available compounds for activity, FBLD attempts to design ligands based on structural knowledge and on knowledge of fragments with binding propensity to the target protein. FBLD can thus be considered a marriage of screening and ►structure-based drug design. FBLD comprises several steps: library design, fragment screening, structure elucidation and hit optimization (1, 5).

FBLD starts with the design and assembly of an appropriate compound library. This can be a general diverse library for use against a variety of target proteins, a targeted library, which is screened against a particular target class (e.g. kinases, proteases) and

NMR Spectroscopy

Definition

NMR spectroscopy is a method used to analyse the structure of molecules as large as proteins. When exposed to a magnetic field, the atomic nuclei in molecules can change their spin state in a resonance experiment. The resonance frequency depends on the local chemical environment. Furthermore, the information about the adopted spin states is transmitted to the neighboring nuclei via spin-spin coupling. This distance-dependent information can be used to learn about the spatial arrangement and thus the 3D structure of large molecules e.g. proteins.

which is biased to resemble known ligands for this target class, or can stem from a virtual library which is initially filtered by *in silico* screening. Whatever library is designed, care is taken to ensure that the compounds are “drug-like”, in the sense that they resemble fragments of marketed drugs or do at least not violate Lipinski’s “Rule of 5” and that they are readily amenable to chemical modification and linkage (6). Once the library is established, it is screened for compounds with binding affinity to the target protein. Due to their small size, which limits molecular interactions, the affinity of fragments to the target protein is generally weak, with dissociation constants often in the high micromolar or even millimolar range. Therefore, methods are needed which are sensitive enough to detect these weak interactions yet which are robust enough to avoid false positive detection. NMR spectroscopy is such a method, and NMR-based screening is particularly useful for this screening step. Other suitable screening methods include ►x-ray crystallography and ►mass spectroscopy. When binding fragments are identified, structural knowledge is sought to characterize the structure of the protein-fragment complex. This can be ideally done by x-ray crystallography if the complex crystallizes, or by NMR spectroscopy. Based on structural knowledge of the complex, an additional fragment that binds in the vicinity of the original fragment can be identified by second-site screening, so that both fragments can be subsequently linked to yield a high-affinity ligand, or the original fragment can be chemically modified and extended to yield a high-affinity ligand (1).

FBLD can be superior to HTS since the FBLD library is, as a virtual library, much larger than libraries used for HTS. If a protein binding site consists of three sub-sites, and each sub-site is probed by 1000 compounds, a virtual library of $1000 \cdot 1000 \cdot 1000 = 10^9$ (1 billion) compounds is screened, assuming that three independently identified fragments can eventually be linked to yield a single high-affinity ligand. However, actually only a thousand compounds need to be screened to gain access to the virtual library of a billion compounds. Libraries for HTS generally include only 10^5 – 10^6 compounds. NMR spectroscopy is an important component of fragment-based ligand design since it provides a sensitive and robust assay to identify weak binding affinity in fragments as well as the possibility of elucidating structural details on protein-fragment complexes.

NMR Screening Techniques

All NMR screening techniques described below perform binding assays rather than functional assays. This has the advantage that no functional assay has to be developed for screening and that in fact no knowledge about the biological function of a novel protein is

necessary. NMR provides a generic binding assay, which can be applied to all protein targets without the need for individual adjustment.

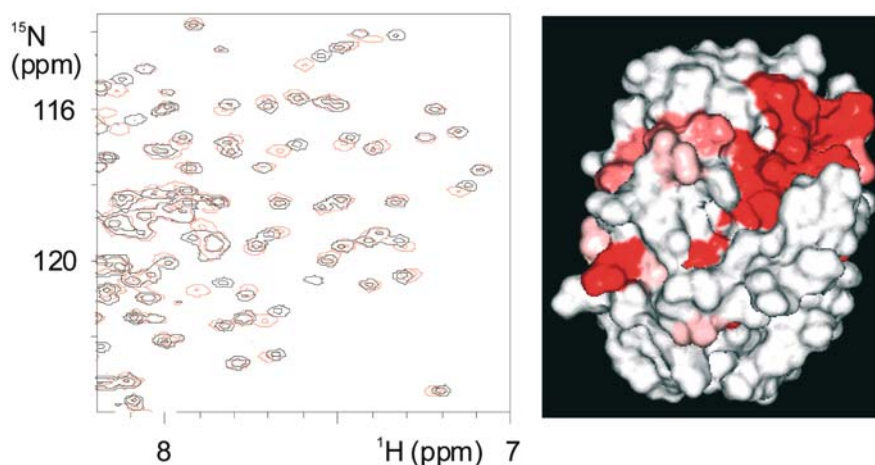
The NMR binding assay can be performed by monitoring either the protein resonances (“protein observation”), or the ligand resonances (“ligand observation”) or the resonances of a suitable reporter ligand (“reporter screening”) (1). The three variations are briefly described and illustrated below. The method of choice depends mainly on the nature and available quantities of the target protein, and the solubility of the test compounds.

Protein Observation

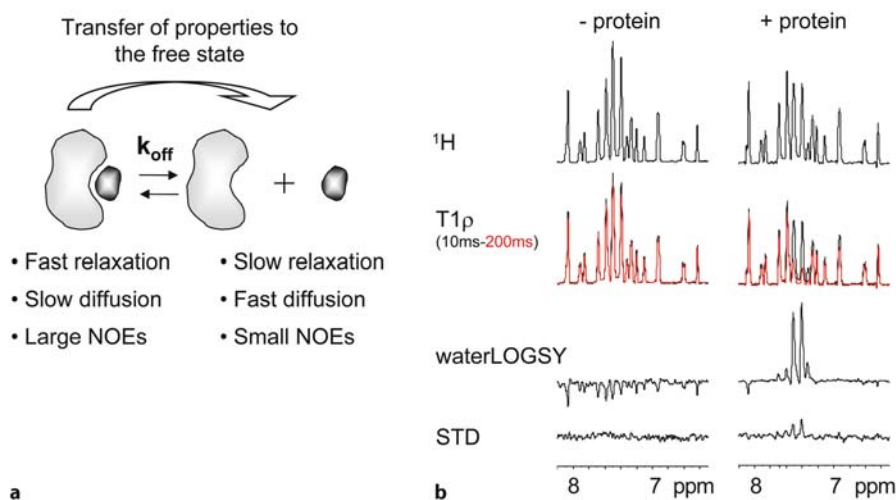
The chemical shifts of protein resonances depend sensitively on the local environment of the respective nuclei. When a protein is complexed by a ligand, the local environment changes and protein residues consequently experience chemical shift perturbations. These can be due to direct contact with the ligand or due to secondary effects such as conformational changes. Since one-dimensional protein NMR spectra are often too crowded, 2D spectra such as ►HSQC are often used to reduce spectral overlap. Chemical shift perturbations indicate the binding event, and also the ligand-binding site, assuming that chemical shift perturbations are mainly due to direct contact with the ligand and provided that the protein resonances have been assigned to their position in the amino acid sequence. An example of the protein observation technique is shown in Fig. 1, which also illustrates the identification of the binding site by this technique. The main advantage of the protein observation method is the high information content. By analysis of the ligand binding site, active-site ligands can be discriminated from non-active-site ligands, and ligands binding to one subsite can potentially be discriminated from ligands binding to another subsite. Disadvantages of the protein observation method are the need for large amounts of isotopically labeled protein (about 1 mg per sample), and the need for high ligand solubility. There is also an upper limit of about 30–50 kD for the target protein, since spectral overlap and peak broadening deteriorate the spectral quality for larger proteins.

Ligand Observation

An unbound ligand is a small molecule and behaves as such, i.e. it has sharp lines (relaxes slowly), small and positive ►NOEs and diffuses fast. A bound ligand, on the other hand, temporarily behaves like and adopts the properties of a large molecule. It shows broad resonances (fast relaxation), large negative NOEs and slow diffusion. Measurements of ligand relaxation, NOE or diffusion properties can therefore be used to detect protein-ligand interactions (1). Due to the high excess of ligand, ligand-observation techniques



NMR-based Screening. Figure 1 Protein observation: A ^{15}N - ^1H HSQC experiment on ^{15}N -labeled target protein is recorded in the absence (black) and presence (red) of compound (left). Chemical shift perturbations indicate binding of the compound and coloring of the residues that experience the greatest chemical shift perturbations indicate the ligand binding site (right).



NMR-based Screening. Figure 2 Ligand observation techniques: (a) Properties of the bound ligand are transferred and detected on the unbound ligand, if the ligand dissociates within the time scale of the NMR experiment, typically 100–1000 ms. (b) Several ligand observation techniques exist, measuring relaxation rates ($T1\rho$), water-ligand NOEs (waterLOGSY) or protein-ligand NOEs (STD).

observe only resonances of the unbound ligand, but they are able to detect the ligand history if the ligand dissociates within the time scale of the experiment. Any effects from a past binding event have been transferred to the unbound ligand and can be detected there (Fig. 2a).

Different ligand properties can be detected by different experiments. Representative spectra are displayed in Fig. 2b. A $T1\rho$ experiment measures ligand relaxation rates. If signal intensity after a 200 ms relaxation period (red spectrum) is not much weaker than signal intensity after a 10 ms relaxation period (black spectrum), the

compound relaxes slowly and is thus unbound. If signals decrease significantly at long relaxation periods, the compound is temporarily bound and is a ligand to the target protein. The waterLOGSY experiment (water-ligand observation with gradient spectroscopy) measures the sign and magnitude of NOEs between water and compound (2). This NOE is small and positive for unbound compounds (leading to negative signals in Fig. 2b), and is large and negative if the compound is bound to protein, provided the binding site also contains water molecules with high residence times. The STD experiment (saturation

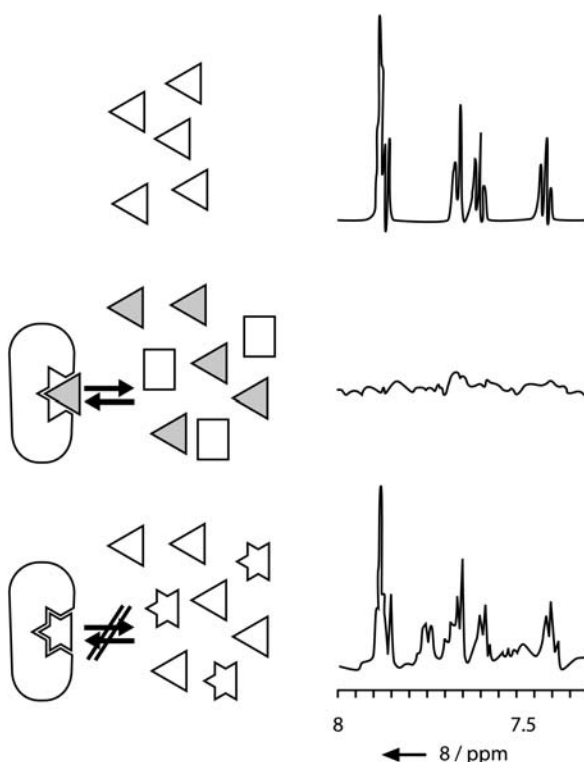
transfer difference) measures transferred NOEs between protein and ligands (4). In this experiment, compound signal intensities are compared with and without protein presaturation. When a compound is bound to the saturated target protein, the saturation is partially transferred from protein to ligand by protein-ligand NOE, so that the binding event can be detected. Literature reviews describe these experiments in more detail (1, 2, 3, 4, 7).

The main advantage of ligand-observation techniques is the significantly lower protein consumption (down to 10 µg per sample) and the lack of need for isotopic labeling. This allows the screening of poorly expressing proteins or of proteins expressed in organisms other than *E.coli*. Disadvantages are the lower information content of the spectra. The protein-binding site cannot be deduced and in fact non-specific binding is not directly discriminated. Hits from ligand-observed NMR techniques should therefore be checked by HSQC, if labeled protein is available, or by competition experiments as shown below. Another major drawback of ligand-observation techniques is the need for significant ligand dissociation during the NMR experiment. This essentially lets tightly binding ligands, or even medium binding affinity ligands with slow binding kinetics, appear non-binding (false negative). In order to alleviate this severe shortcoming, competition-based methods have been developed.

Reporter Screening

Competition assays (“NMR reporter screening”) have been developed to detect tightly binding ligands with small protein amounts (1, 2). For this method, a known, weakly binding ligand (the “spy molecule”, “probe” or “reporter ligand”) is added to the mixture of protein and test compounds. The binding of the reporter ligand is measured, not the binding of the test compounds. If the reporter ligand is bound even in the presence of test compounds, none of the test compounds has comparable or higher binding affinity. In contrast, if the reporter ligand is not bound any more, one or more test compounds must have displaced it and therefore must have comparable or higher binding affinity (Fig. 3). Even poorly soluble test compounds can be detected in this way, if their affinity is high enough.

Advantages of reporter screening are not only the ability to detect tightly binding ligands, but also the elimination of non-active site binders. If the reporter ligand binds to the active site, every compound displacing it should also bind to the active site. Additionally, the binding affinity of test compounds can be calculated based on the known affinity of the reporter ligand. If the reporter ligand is ^{13}C -labeled at one position or is chosen to contain a ^{19}F -atom, signal overlap with test compounds can be completely avoided (2).



NMR-based Screening. Figure 3 The principle of reporter screening. The reporter ligand (triangles) binds to the protein active site with moderate affinity. This leads to severely broadened lines in the presence of protein (middle panel), whereas the reporter ligand in the absence of protein has sharp lines (upper panel). In the presence of non-binding test compounds, the reporter ligand is still bound and shows broad lines (middle panel). If at least one of the test compounds has higher binding affinity, it can displace the reporter ligand; this is made visible by its sharpened lines (lower panel).

Hit Elaboration and Optimization

Primary hits from NMR-based screening have specific and validated, but generally only weak, binding affinity to the target protein and their binding affinities need to be increased. The NMR hit can be followed up by one of the following strategies (1, 7). In a linked-fragment approach, second-site screening is employed to identify a second ligand, which binds simultaneously and in the vicinity of the NMR hit (the “first ligand”). Linking both ligands then results in a significant increase in binding affinity if the linker has been well designed. In a merged-fragment approach, information on binding epitopes from several ligands is used to construct a merged ligand, which exploits binding interactions from two or more distinct ligands. These ligands cannot be linked since they bind at overlapping binding sites. Both the linked-fragment approach and the merged-fragment approach rely heavily on structural characterization of the binding of ligands to the target protein. In

a third strategy, hits from NMR-based screening are used as input to design directed libraries around these hits, perhaps using combinatorial chemistry. Structural information on the complex between the hit and the target protein is desirable, but not mandatory, since NMR screening is essentially used as a filter to select the most promising scaffold for combinatorial chemistry. In the elaboration strategy, hits from NMR-based screening are used to select analogous compounds, e.g. identified by a similarity search, for further follow-up studies such as computational docking and scoring or testing by NMR reporter screening or a biochemical assay. The follow-up strategy of choice depends on the nature of the NMR hits, the nature of the protein-binding site, available chemistry capacity and the amount of available experimental information. Several examples of successful NMR-based screening and subsequent optimization of the primary hits have been described and reviewed (1, 2, 3, 4, 7).

Clinical Applications

NMR-based screening is applied in pre-clinical or genomic research.

Therapeutic Consequences

Fragment-based ligand design including NMR techniques is expected to accelerate the drug discovery process, so that lead compounds as potential drug candidates are expected to be more rapidly identified.

References

1. Jahnke W, Widmer H (2004) Protein NMR in biomedical research. *Cell Mol Life Sci* 61:580–599
2. Stockman BJ, Dalvit C (2002) NMR screening techniques in drug discovery and drug design. *Prog NMR Spectrosc* 41:187–231
3. Diercks T, Coles M, Kessler H (2001) Applications of NMR in drug discovery. *Curr Opin Chem Biol* 5:285–291
4. Meyer B, Peters T (2003) NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angew Chem Int Ed Engl* 42:864–890
5. Shuker SB, Hajduk PJ, Meadows RP et al (1996) Discovering high-affinity ligands for proteins: SAR by NMR. *Science* 274:1531–1534
6. Lepre C (2003) Strategies for NMR screening and library design. In Zerbe O (ed) *BioNMR in drug research*. vol 16, Wiley-VCH, Weinheim, pp 391–415
7. Peng JW, Lepre CA, Fejzo J et al (2001) Nuclear magnetic resonance-based approaches for lead generation in drug discovery. *Methods Enzymol* 338:202–230

N-nitroso-N-ethylurea

► ENU

Nodal

Definition

Nodal proteins are members of the transforming growth factor- β (TGF β) superfamily. They have been identified as important endogenous mesoderm inducers in vertebrates.

► [Axis Formation – Formation and Function of the Dorsal Organizer](#)

Nodes of Ranvier

Definition

Nodes of Ranvier are periodic gaps in the myelin sheath of axons. The node region has a width of up to 1 μm . The passage of chemical ions in and out of axons is restricted to these naturally unmyelinated axon segments, and this is also true for the emergence of action potentials.

► [Glial Cells and Myelination](#)

NOE

Definition

NOE stands for Nuclear Overhauser Effect. It occurs when dipolar coupled spins exchange magnetization through space. In a first approximation, the NOE signal is proportional to the inverse 6th power of the distance between two atoms that are close in space. Therefore, NOE signals can be transformed into distant restraints that serve as experimental input for the structure calculation of molecules by NMR.

► [3D Structure by NMR](#)

► [NMR-Based Screening](#)

► [Protein-Ligand-Interaction studied by NMR](#)

NOESY

Definition

NOESY refers to two-dimensional [NOE](#) spectroscopy. This technology detects, as cross peaks, correlations between nuclei that are close in space, typically up to 5 Å.

► [3D Structure by NMR](#)

Non Rapid-Eye-Movement Sleep

► NonREM (Non Rapid-Eye-Movement) Sleep

Noncoding RNA

Definition

Noncoding RNA designates RNA molecules that function without being translated into protein.

► X-Chromosome Inactivation

Nondisjunction

Definition

Nondisjunction designates an incorrect segregation of a pair of chromosomes or sister chromatids during cell division at anaphase I or II, which results in the formation of daughter cells with an abnormal number of chromosomes (► Aneuploidy).

► Meiosis and Meiotic Recombination

► Prader Willi and Angelman Syndromes

Non-Histone Chromatin Proteins

MONICA FEDELE, ALFREDO FUSCO

Istituto di Endocrinologia ed Oncologia Sperimentale,
University “Federico II”, Naples, Italy
afusco@napoli.com

Synonyms

Chromosomal nonhistone proteins; chromatin-remodeling enzymes; chromatin remodelers; architectural factors; nuclear matrix; nuclear acidic proteins

Definitions

The non-histone chromatin proteins are a heterogeneous group of proteins that act in the eukaryotic nucleus as part of large multisubunit complexes, playing important roles in regulating many processes such as nucleosome remodeling, DNA replication,

RNA synthesis and processing, nuclear transport, steroid hormone action and interphase/mitosis transition. The group includes proteins that regulate chromatin assembly, such as SWI/SNF; proteins that regulate enhanceosome and transcriptosome assembly, such as the high mobility group A proteins; modifying proteins, such as histone acetylase, deacetylase and methylase; proteins that help chromosomes to condense, such as topoisomerase II and many others generally involved in the establishment of functional and dynamic chromatin domains.

Characteristics

All genes are packaged into chromatin structures that can be modulated to generate transcriptionally active or repressed configurations in different cellular contexts and under changing environmental conditions. This remarkable plasticity is governed in part by multi-subunit protein complexes that enzymatically regulate chromosomal structure and activity. These complexes can either chemically modify the histone tails of nucleosomes or disrupt histone-DNA contacts through ATP hydrolysis. Specific combinations of modifying and remodeling complexes often act in a temporal manner to modulate the nucleosomal structure and expression of individual genes.

One important ► chromatin remodeling enzyme is SWI/SNF, which is a 2 MD multisubunit complex that is implicated in regulating critical cellular processes such as differentiation and cell-cycle arrest. Mammalian SWI/SNF complexes are present in biochemically diverse forms, indicating that they may have specialized nuclear functions. Each of these complexes contains one of two ATPases, BRG1 or BRM, and a variable subunit composition of BRG1-associated factors (BAFs). Association of specific protein domains with individual SWI/SNF subunits directs recruitment of the enzymatic complex to nucleosomal sites where stable remodeling occurs. Thus, the functional diversity of mammalian SWI/SNF can result from differential gene targeting of specific complexes by transcription factor interactions with individual SWI/SNF subunits. Another group of ATP-dependent chromatin-remodeling factors contains the nucleosome-dependent ATPase ISWI, which forms several functionally different complexes. ACF1 interacts with the SNF2H isoform of ISWI in the related complexes ACF, WCRF and CHRAC, all of which are conserved in *Drosophila melanogaster*, *Xenopus laevis* and human cells. ISWI is the primary motor of the nucleosome remodeling, whereas ACF1 enhances and directs the process. An interesting family of nonhistone proteins involved in changing the local conformation of DNA and/or nucleosomes and enhancing their accessibility and plasticity, is the high mobility group proteins (HMGs). They can be divided into three

groups that are completely dissimilar from each other as regards sequence and structure but are internally homogeneous, the HMGB (formerly HMG1/2), HMGN (formerly HMG14/17) and HMGA (formerly HMG1/Y) families. Each group is characterized by a distinct functional sequence motif; the "HMG box" for the HMGB proteins, the "nucleosomal binding domain (NDB)" for the HMGN proteins and the "AT-hook" for the HMGA proteins. The HMGA proteins include four members, HMGA1a, HMGA1b and HMGA1c generated by alternative splicing of the same gene, and HMGA2. Each of them contains three AT hook domains through which they bind AT rich sequences in the minor groove of the DNA and induce complex structural alterations. The "HMG-motif" proteins are a similar but distinct family of nonhistone chromatinic proteins. They contain one of the functional motifs of the HMG proteins, but the rest of the protein is different. The mixed lineage leukemia gene (MLL, ALL-1, HRX) is one of the "HMG-motif" proteins. It is homologous to the *Drosophila* protein Trithorax, and, like Trithorax, has been shown to be a positive regulator of homeobox gene expression. The main region of homology with Trithorax includes two small regions in the amino terminal half of the protein, a central zinc finger domain and a carboxy terminal SET domain (a motif that is also found in some mammalian Polycomb homologues, which are negative regulators of Hox gene expression).

Clinical Relevance

The chromatin remodeling complexes are involved in critical aspects of cellular growth and genomic stability. A variety of human malignancies are associated with mutations in BRG1 and INI1/SNF5, suggesting that loss of function of certain SWI/SNF complexes may contribute to tumorigenesis.

MLL and Human Leukemias

Rearrangements of MLL are common in infant acute lymphocytic leukemia, 5–10% of adult and pediatric acute leukemia cases, myelomonocytic leukemias and in leukemias that arise secondarily to DNA topoisomerase II inhibitor therapy. These rearrangements follow chromosomal translocations involving the chromosomal region 11q23. The fusion of MLL may occur with at least 20–30 different genes mapping to different chromosomal regions. The translocations delete the SET domain of MLL and replace it with one of many translocation different partners that in general share sequence homology. To date, two MLL motifs have been identified that are necessary for leukemic transformation. The first of these is the three AT-hook motif homologous to that in HMGA proteins and not present in *Drosophila* Trithorax. These motifs may be relevant to the function of MLL in leukemogenesis

because in all of these translocations the AT hook motifs are retained and are separated from the carboxy terminal portion of the protein. The second region important for transformation is that homologous to the noncatalytic region of DNA methyltransferase.

In some cases there are alterations in the MLL locus not associated with cytogenetic abnormalities; these were found to involve a direct tandem duplication of a portion of MLL spanning from exon 2 to exon 6 or from exons 8 or 4 to exon 6. This second type of alteration may be associated with trisomy of chromosome 11 and occurs in around 5–10% of AML patients.

The HMGA Proteins

HMGA expression is essentially restricted to embryonic development and is negligible in normal adult tissues. However, over-expression of HMGA proteins has been found to be a common feature of experimental and human malignant neoplasias, including thyroid, prostate, uterus, pancreas, breast, colorectal and ovarian carcinomas. Moreover, the expression level of the HMGA proteins is significantly correlated with parameters of a poor prognosis in patients with colorectal cancer. Therefore, an evaluation of the HMGA protein expression by immunohistochemistry can be very helpful in the diagnosis and prognosis of human neoplasias. Finally, several published papers demonstrate that HMGA over-expression is a necessary event in *in vivo* cell transformation and that the suppression of HMGA1 protein synthesis specifically kills malignant cells. In fact, an adenovirus carrying the *HMGA1* gene in an antisense orientation induces programmed cell death in carcinoma cell lines derived from human thyroid, lung, colon and breast cancers. This suggests the fascinating possibility of a new approach to the therapy of human malignant neoplasias based on the suppression of the HMGA protein function. The advantage of such an approach would be that:

- the low or null expression levels of HMGA gene expression in normal adult tissues would account for the specificity and low toxicity of such a kind of therapy.
- since HMGA gene over-expression appears to be a general feature of human malignant neoplasms, a therapy based on the inhibition of HMGA function might be applicable to diverse types of cancer.

HMGA proteins are also involved in the generation of benign tumors of mesenchymal origin, including lipomas, lung hamartomas, uterine leiomyomas, endometrial polyps, fibroadenomas and adenolipomas of the breast. In fact, rearrangements of the *HMGA1* and *HMGA2* gene have been detected in these neoplasias. Chromosomal translocations involving the regions 6p21 and 12q13-15, where the *HMGA1* and *HMGA2* genes respectively are located, account for these

rearrangements. In most of the cases of *HMGA2* rearrangements, breaks occur within the third intron of the gene, resulting in chimeric transcripts containing exons 1 to 3 of *HMGA2* (encoding the AT-hook domains) and ectopic sequences from other genes. The generation of transgenic mice over-expressing a truncated or a wild type *HMGA2* gene has confirmed the critical role of the *HMGA2* rearrangements in the development of human lipomas. From the most recent results, it seems that *HMGA2* expression may have an important role in obesity. In fact, the *HMGA2* gene seems to be necessary for physiological proliferation of adipocytes. Indeed, *HMGA2* knockout mice display a pygmy phenotype with a remarkable reduction in adipose tissue and disruption of the *HMGA2* gene prevents both diet- and gene-induced obesity. A possible therapy for human obesity by blocking of *HMGA2* function cannot be excluded.

References

1. Horn PJ, Peterson CL (2002) Chromatin higher order folding: wrapping up transcription. *Science* 297:1824–1827
2. Kadam S, Emerson BM (2003) Transcriptional specificity of human SWI/SNF BRG1 and BRM chromatin remodeling complexes. *Mol Cell* 11:377–389
3. Reeves R, Beckerbauer L (2001) HMGI/Y proteins: flexible regulators of transcription and chromatin structure. *Biochim et Biophys Acta* 1519:13–29
4. Fedele M, Battista S, Manfioletti G et al (2001) Role of the high mobility group A proteins in human lipomas. *Carcinogenesis* 22:1583–1591
5. Croce CM (1999) Role of TCL1 and ALL1 in human leukemias and development. *Cancer Res* 59:1778s–1783s

Non-Homologous End Joining

Definition

Non-Homologous End Joining (NHEJ) defines a DNA double strand repair pathway in which the ends of the broken DNA molecule are covalently connected, after they have been converted into a ligatable structure.

► [DNA-Repair Mechanisms](#)

NonREM (Non Rapid-Eye-Movement) Sleep

Definition

NonREM sleep is one of two distinct brain states occurring during sleep. nonREM sleep is characterized

by a synchronized EEG. In humans, nonREM sleep is divided into 4 stages according to the amount of delta activity. Stages 3 and 4 are called slow-wave sleep.

► [Narcolepsy](#)

► [REM Sleep](#)

Nonsense Mutation

Definition

Nonsense mutation describes single-base substitutions that change a sense codon coding for an amino acid into a termination codon, for instance UGG (tryptophan) to UGA (stop). Nonsense mutations are deleterious, as at such mutated positions the synthesis of the protein is terminated prematurely and full-length, functional protein is not made. Truncated proteins formed as a consequence of nonsense mutations may impair cell functions. In eukaryotic cells, mRNAs containing premature stop codons are subject to rapid, preferential nucleolytic degradation ► [nonsense-mediated RNA decay](#).

► [Hereditary Nonpolyposis Colorectal Cancer](#)

► [Heritable Skin Disorders](#)

► [Hereditary Spastic Paraplegias](#)

► [Limb Development](#)

► [Ribosomes](#)

Nonsense-Mediated RNA Decay

Definition

mRNAs harbouring premature termination codons (PTC) in their sequence can be selectively degraded by nonsense mediated degradation, or nonsense-mediated RNA decay (NMD).

A termination codon is recognised as premature when it precedes an exon/exon splice junction. PTCs are generated following mutations, errors in fidelity of transcription, random gene rearrangement as occurring in immunoglobulin or T-cell receptor genes. This “quality control” process prevents the synthesis of truncated proteins that may be toxic. The NMD pathway has a direct impact on hundreds of genetic disorders in the human population, where about a quarter of all known mutations are predicted to trigger NMD.

► [Nonsense Mutation](#)

► [Repetitive DNA](#)

► [Ribosomes](#)

► [RNA Polymerase II Transcription](#)

► [RNA Stability](#)

Non-Silent Substitution

Definition

Single base substitutions in a codon are called non-silent, if the corresponding amino acid is changed. In practice, the result might be a protein with a slightly changed structure or function.

► [Sequence Annotation in Evolution](#)

Nonsyndromal/Nonsyndromic Deafness

Definition

Nonsyndromic/nonsyndromal deafness is a hereditary autosomal hearing loss that is not associated with other signs and symptoms. (In contrast, “syndromic deafness” includes forms of hearing loss that are part of genetic syndromes which also affect other parts of the body.) The different types of nonsyndromic deafness are named according to their inheritance patterns. More than 30 genes that, when mutated, can cause nonsyndromic deafness. One form is caused by mutations in the gene GJB2 that encodes for ► [connexin 26](#), which is expressed in the cochlear and vestibular regions of the ear.

► [Intermediate Filaments](#)

Non-Uniform Sampling

Definition

Non-uniform sampling is a technique that is used in multi-dimensional NMR experiments, where the free induction decays with variable number of scans are collected as non-consecutive integers of fixed time increment in the indirect dimension. The data collected are typically processed with maximum entropy methods, since they cannot be processed using conventional Fast Fourier transform algorithms.

► [3D Structure by NMR](#)

Normalization

Definition

Normalization corrects for the different amounts of cDNA in the samples that are compared. Most often

this is achieved by measuring the signal intensity of so-called housekeeping genes, which are usually expressed at a stable level in all cells, independently of external factors. The expression level of genes that are compared between two samples is then mathematically corrected according to the differences found in the expression level of the housekeeping genes.

► [Rheumatism Related Genes, Identification](#)

Normoxia

► [Hypoxia/Normoxia](#)

NORs

► [Nucleolar Organising Regions](#)

Northern Blot

Definition

Northern blot is a technique by which RNA is size-separated on an agarose gel, and transferred to a nitrocellulose or Nylon filter on which it can be hybridized to a complementary DNA.

► [Microarrays in Colorectal Cancer](#)

Notch Pathway

THOMAS KLEIN

Institute for Genetics, University of Cologne, Cologne, Germany

th.klein@uni-koeln.de

Definition

The *Notch* pathway is one of the fundamental signalling mechanisms that mediates cell communication in a wide range of developmental processes

throughout the animal kingdom (1, 2). It mediates communication between cells immediately adjacent to one another. The pathway was discovered and initially characterized in the fruit fly *Drosophila melanogaster*, but has been found in all other animal systems, including nematodes and vertebrates. In vertebrates, Notch signalling is required in many developmental processes, including limb development, segmentation of the mesoderm, vasculogenesis and haematopoiesis. Furthermore, it is involved in several heritable diseases, such as ►Alagille syndrome and ►Cadasil disease (3, 4). In many tissues it also seems to act to promote ►oncogenesis, in some cases as an important mediator of oncogenic Ras. On the other hand, recent work suggests that the pathway acts as a tumour suppressor in the skin (4).

Earlier work in *Drosophila* showed that loss of the function of the pathway during early embryogenesis led to the formation of excess neural tissue at the expense of epidermis. This led to the classification of all genes involved in the *Notch* pathway as ►neurogenic genes. The study of the role of the pathway during neurogenesis in vertebrates and invertebrates revealed that it mediates a binary cell choice among cells of an equivalence group. During this process, which is often termed lateral inhibition, a cell determined to become a neural precursor (e.g. a neuroblast) is thought to emit an inhibitory signal that forces its equipotent neighbours to adopt the epidermal fate (1, 2). This inhibitory signal is mediated through the *Notch*-pathway. Furthermore, the pathway is involved in a variety of inductive signalling events between non-equivalent cells in other developmental situations. One example of this type of signalling is the induction of the expression of genes along the dorsoventral compartment boundary of the *Drosophila* wing through Notch-mediated interactions between dorsal and ventral cells. Notch mediates a similar interaction between dorsal and ventral ectodermal cells during vertebrate limb development leading to the induction of the ►apical ectodermal ridge (AER).

Characteristics

Structure of the Core Components of the Notch Pathway

In *Drosophila*, the pathway has four core elements, two ►ligands, Serrate (►Ser) and Delta (►Dl), the Notch receptor and Suppressor of Hairless (Su(H)), which is a nuclear factor (1, 2). Notch is a large Type 1 transmembrane protein with several distinguishable motifs in its extra- and intracellular domains (Fig. 1). The extra-cellular domain contains a region containing 36 tandem repeats of the ►EGF-like motif, followed by a cysteine-rich region, termed the LNR (for Lin-12/Notch-related region). The intracellular domain contains a RAM 23 domain, adjacent to the transmembrane

domain, six repeats of an ankyrin/cdc10-like domain and a PEST sequence. The six ankyrin-like repeats are necessary and sufficient for the activity of the *Notch* receptor in *Drosophila*.

Both Dl and Ser belong to the DSL (Delta, Serrate, Lag-2) family, which are all ligands for Notch receptors (Fig. 1). They are transmembrane proteins that contain a DSL domain in the N-terminus of their extra-cellular domain and a stretch of EGF-like motifs. In Dl, the stretch comprises nine repeats, whilst Ser contains 14. Ser also has a cysteine-rich domain close to the membrane, which is absent in Dl (Fig. 1).

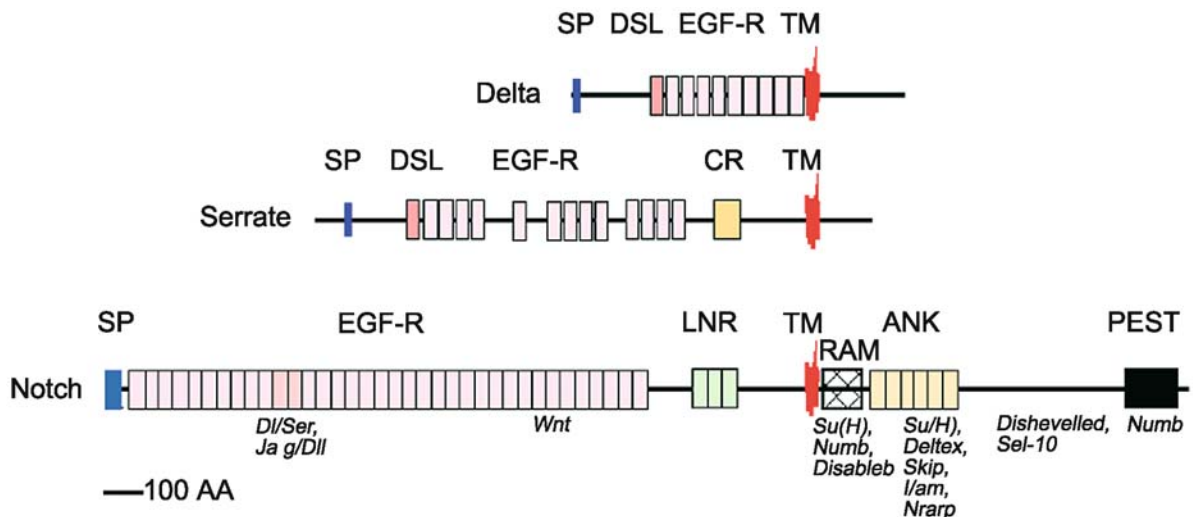
Orthologues in Vertebrates

Studies in various organisms revealed that all the components of the *Notch* pathway found in *Drosophila* are conserved throughout the animal kingdom (1, 2). In mouse and humans three ►orthologues of Dl exist, ►Delta-like (Dll) -1, -3 and -4 and two Ser-orthologues have been found and named ►Jagged (Jag 1 and 2). Four Notch receptors, have been isolated, Notch 1–4. While Notch 1 and 2 have similar overall structures to Notch, Notch 3 and 4 have only 34 and 29 EGF-repeats respectively and contain shorter intracellular domains. The vertebrate Notch receptors require proteolytic cleavage to become functional. A single precursor polypeptide is cleaved in a step referred to as S1 cleavage to produce a heterodimer consisting of a ligand-binding extra-cellular domain and a single pass transmembrane domain (Fig. 2; 1). Both domains associate through Ca^{2+} ion bridges and the LNR region plays an important part in this association. The S1 cleavage occurs in the Golgi apparatus and is performed by a furin protease. For *Drosophila* Notch receptor there is conflicting evidence as to whether S1 cleavage occurs.

Only one orthologue of the Su(H) transcription factor exists in mammals, which is referred to as CBF1 or RBPjk or CSL. In the absence of Notch activation, it seems to be a transcriptional repressor. Su(H)/CBF1 seems to form a complex that contains Hairless (H) in *Drosophila*, which recruits the co-repressors Groucho (Gro) and CtBP (c-terminal binding protein). In vertebrates there is evidence the repressor complex contains SMRT (silencing mediator of retinoid and thyroid hormone receptor), SKIP (ski-related protein) and histone deacetylase (HDAC) (2).

Signal Transduction

The intracellular domains of the Notch receptors are in fact membrane bound transcription factors, which are released through a sequence of two proteolytic cleavages induced by ligand binding (summarized in Fig. 2). Following cleavage the intracellular domain travels to the nucleus and associates with Su(H) to form a transcriptional activator complex. Signal transduction



Notch Pathway. Figure 1 The structure of the founding members of the *Notch*-receptor and DSL- families. All three depicted proteins are type 1 transmembrane proteins, with the N-terminus in the extra-cellular space and their primary amino acid sequence contains a typical signal sequence (SP) in their extreme N-termini. Delta and Serrate are ligands that contain the family specific DSL (Delta, Serrate, Lag-2) domain followed by a stretch of repeats homologous to the epidermal growth factor (EGF). The stretch comprises nine repeats for *Dl* and 14 for *Ser*. *Ser* has a cysteine-rich domain (CR) close to the membrane, which is absent in *Dl*. The intracellular domains of both ligands are relatively small and bear no significant motives.

Notch is the founder of the *Notch*-receptor family. The extracellular domain is large and contains a region of 36 EGF-like repeats followed by a cysteine-rich region, termed the LNR (for Lin-12/*Notch*-related region). In the intracellular domain, the RAM 23 domain is located immediately after the transmembrane domain (TM), followed by six repeats of an ankyrin/cdc10-like domain and a PEST sequence. Some of the factors that interact with different regions of the *Notch* receptor are listed below in italics.

in this pathway is in principle very simple, since it requires no intermediate factors or signal-amplification steps to mediate signal transduction from the receptor to the nucleus. However, the regulation of the release of the intracellular domain is complex and involves additional factors. The detailed events leading to the activation of the receptor are described below.

There is also evidence from several systems that a second, *Su(H)* independent signalling transduction mechanism exists. However, no details are known about this pathway (1, 2).

Molecular Interactions

Ligand Receptor Interactions

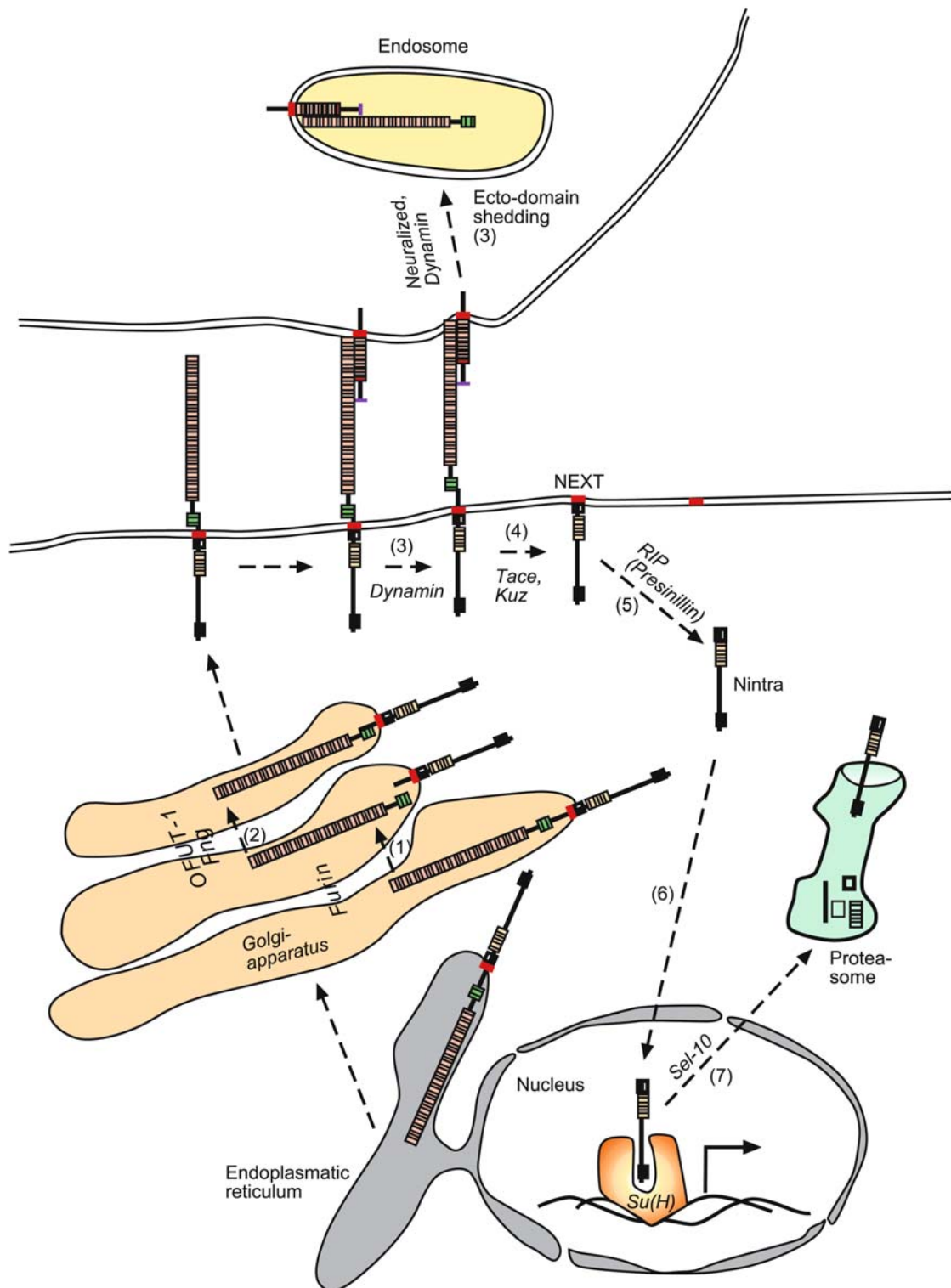
Several studies have revealed structural features of *Notch* and its ligands, which are crucial for their interaction. The ligands require the DSL domain to activate *Notch*, which seems to bind EGF-repeats 10–12 within *Notch*. Furthermore, EGF repeats 24–26 and 17–19 of *Notch* are required for the interaction of *Notch* with its ligands.

Cleavage of the Receptor

The binding of a ligand induces the first cleavage of *Notch*. This cleavage, called S2, is performed by a

protease of the TACE /ADAM metalloprotease family and occurs in the extra-cellular space, close to the membrane (Fig. 2; 4) (1, 2). In *Drosophila* the S2 cleavage is probably carried out by Kuzbanian (Kuz). However, it is unclear whether its orthologue ADAM 10 performs this cleavage in vertebrates (1, 2). The result of S2 cleavage is the shedding of the extra-cellular domain of *Notch*, which is endocytosed together with the ligand in the signal-emitting cell (Fig. 2; 3) (2). The remaining membrane-inserted intracellular domain referred to as **▶NEXT** (*Notch* extra-cellular truncation) is rapidly cleaved by an enzyme complex, the gamma-secretase complex that also cleaves the **▶amyloid precursor protein** (APP) (Fig. 2; 5) (2). This proteolytic complex contains one of the presenilin (Ps) proteases, Nicastrin, APH-1 and PEN-2 (5). **▶NEXT** is cleaved in the trans-membrane region to release the intracellular domain of *Notch* (**▶Nintra**) into the cytoplasm. This phenomenon is an example of regulated intra-membranous proteolysis (**▶RIP**) (2).

The activation of the *Notch* receptor also requires the functioning of the endocytic pathway on both the signal emitting and receiving sides of the pathway (Fig. 2; 3) (1). Recent studies in *Drosophila* revealed that the



Notch Pathway. Figure 2 Notch trafficking and signal transduction through the Notch pathway. After Notch is synthesized in the endoplasmic reticulum, the signal peptide is removed and the receptor is transported to the Golgi apparatus for further maturation. An important event is the cleavage by furin proteases (1) to produce the mature receptor, consisting of the extra-cellular domain connected to the membrane inserted intracellular domain through Ca^{2+} ion bridges. A further modification that occurs in some cells is the decoration of certain EGF-repeats in the

short intra-cellular domain of DI is important for endocytosis. It is probably ubiquitinated by the RING finger E3-ligase Neuralized (Neur), which regulates the endocytosis and functioning of DI. In vertebrates this function seems to be performed by the E3-ubiquitin ligase, Mindbomb (1, 2).

Although endocytosis is also required on the signal-receiving side of the pathway, its function is less clear. Studies in *Drosophila* suggest that endocytosis is not required for S3 cleavage, suggesting that S2 cleavage might be dependent on this process. One attractive hypothesis is that endocytosis is required when the ligand binds to the receptor in order to stretch the receptor and change its configuration. A conformational change could make the S2-cleavage side accessible to the Kuz/Tace-like protease.

Interactions in the Nucleus

Once Nintra is released into the cytoplasm, two nuclear localization signals direct it to the nucleus (Fig. 2; 6). In the nucleus Nintra seems to cause the dissociation of the Su(H)/CBF1 containing repressor complexes with high efficiency (1, 2). After dissociation of the repressor complex, an activator complex that contains SKIP, Mastermind (Mam) and histone acetylases is thought to assemble with the help of Nintra. Nintra has been shown to be able to assemble an activator complex in the absence of Su(H). Hence, Su(H) provides DNA binding specificity and Nintra is responsible for the assembly of the complex.

Studies in flies indicate that the Su(H)/Nintra activator complex alone is not sufficient to activate gene transcription. Additional transcription factors or complexes are also required to activate the transcription of target genes. These inputs also modulate the *Notch*-signal to provide the tissue specific transcriptional response.

Termination of Signalling

At the end of the signalling event, Nintra must be removed from the nucleus to terminate target gene transcription. This process is not very well understood, although recent evidence shows that in mammals and

nematodes Sel-10, a F-box protein, specifically targets nuclear Nintra for destruction by the [proteasome](#) (Fig. 2;) (1, 2).

Ligand Processing

A recent discovery is that during activation of the *Notch*-pathway not only the receptor but also the ligands seem to be cleaved in the same sequence and by the same proteases (6). The binding of at least some of the ligands (tested are Dll-1, Ser and DI) elicits their cleavage by Kuz and consequently by the γ -secretase complex and releases the intracellular domain into the cytoplasm. Interestingly, the intracellular domains travel to the nucleus as Nintra does. The significance of this nuclear translocation is at the moment not clear, but one work reports the interaction of the intracellular domain of Dll-1 with the AP-1 transcriptional complex. It is also possible that processing of the ligands is part of the degradation process of the ligand/receptor complex after termination of the signalling event.

Ligand Specificity

In vertebrates not only four Notch receptors, but also five different ligands have been discovered. These ligands and receptors are often expressed in strongly overlapping domains. This raises the interesting question as to whether specific interactions between certain ligands and receptors exist to perform specific functions. Whether such specific interactions exist is one of the most pressing questions.

Regulatory Mechanisms

Given the importance of the pathway and its involvement in so many processes, its activity must be regulated precisely. Several mechanisms have been discovered. One major mode of regulation is transcriptional regulation of the ligands and the Notch receptors. In *Drosophila*, Notch is ubiquitously expressed. Nevertheless, it is expressed at different levels, suggesting that transcriptional regulation is important. Overall, one major regulation of its activity occurs through the differential expression of the ligands.

extra-cellular domain with sugar chains by OFUT-1 and Fng (2, 8). The mature receptor is then transported to the cell membrane. The binding of a ligand such as DI activates the receptor. An important step in the activation of the receptor is endocytosis (3), which is required on the signal emitting side as well as the receiving side of the pathway. Dynamin and Neuralized are important for this process. The extracellular domain of Notch is endocytosed together with DI by the signal-emitting cell. This event is named ecto-domain shedding. The binding elicits the first cleavage of the receptor by a member of the ADAM/TACE family, which is encoded by *kuz* in *Drosophila* (4). The cleavage produces the intermediate NEXT (for *Notch* *Extra-cellular* *Truncation*), which is immediately subjected to RIP (for *regulated* *intra-membranous* *proteolysis* *release*) by the presenilin-containing gamma-secretase complex (5). RIP releases the intracellular domain (Nintra) from the membrane (6) and this travels into the nucleus (6) and combines with Su(H) to generate a transcriptional activator that controls the expression of the target genes. At the end of the signalling event the activator complex is targeted for protein degradation by the proteasome (7). Sel-10 is involved in this process.

In vertebrates, regulation of expression of the ligands and the receptors is much more complex, indicating a tight regulation. Although not tested, it is likely that this precise regulation of expression is used to tightly control the activity of the pathway.

► Glycosylation

In several situations certain EGF-repeats of Notch become modified by glycosylation (Fig. 2) (1, 2, 8). The sugar chain is added in the Golgi apparatus. Initially, a fucose residue is added by the *O*-fucosyltransferase (►OFUT1), encoded by the gene *neurotic (nti)* in *Drosophila*. The next step is the addition of N-acetylglucosamine to the fucose residue by the action of Fringe (Fng). The glycosylation of Notch enhances activation by D1 and suppresses activation through Ser. In several developmental situations, the modification leads to the activation of the *Notch*-pathway in a thin stripe along the boundary of Fng-expressing and non-expressing cells. Examples are the activation of Notch target genes along the dorsoventral boundary during *Drosophila* wing development or induction of the apical ectodermal ridge at the dorsoventral boundary during limb formation in vertebrates. Three variants of Fng have been found in vertebrates, named Radical fringe, Manic fringe and Lunatic fringe.

Ligand Concentration

Another means regulating Notch activity has been found by investigating the role of the *Notch*-pathway during *Drosophila* wing development: In cells that express high levels of ligands, the *Notch*-pathway cannot be activated. However, the cells with high levels of ligands can activate the pathway in the adjacent cells that express lower levels of the ligands (1). This cell-autonomous suppression occurs only if the levels of ligands increase above a certain threshold. The suppressive effect is used during wing development for regulation of the activity of Notch. The underlying mechanism of suppression of the activation of Notch is not understood. It is also not known, whether the vertebrate orthologues of the ligands have similar properties.

Numb and Asymmetric Cell Divisions

Evidence from *Drosophila* and more recently from mammals indicates that the *Notch* pathway plays an important role in promoting asymmetric cell division (7). In this process a cell divides to give rise to two different daughter cells. For example this occurs during neurogenesis to generate two different neurons from a single precursor cell. An important factor in this process is Numb, a protein that can prevent the activation of Notch in one of the two daughter cells. Numb is a membrane-associated cytoplasmic protein

that contains a phosphotyrosine-binding (PTB) domain and associates with the intracellular domain of Notch. During cell division Numb is asymmetrically distributed to one of the daughter cells and represses the activation of Notch in this cell. As a result the *Notch* pathway will be activated only in the sister cell, which adopts another developmental fate. Numb binds to alpha adaptin, a protein essential for the formation of clathrin coated endocytic vesicles (1), suggesting that Numb might target Notch for endocytosis and thereby remove the receptor from the cell surface. However, recently it was discovered that Notch requires the activity of the four pass transmembrane protein Sanpodo (Spdo) to promote cell asymmetry in *Drosophila* (7). Spdo associates with Notch and Numb and the interaction with Notch is required for the function of the receptor. It seems that Numb prevents the transport of Spdo to the membrane and thus prevents the activation of Notch in one of the daughter cells.

Suppressor of Deltex

The E3 ubiquitin ligase, named Suppressor of Deltex (Su(dx)) in flies and Itch in mammals, is a negative regulator of Notch when inserted in the plasma membrane. However, its function is so far not understood.

Notch Regulated Ankyrin Repeat Protein (Nrarp)

As in many other important signalling pathways, a feedback loop exists that counteracts the activity of the *Notch*-pathway. Activation of the *Notch*-pathway induces the expression of a small protein termed Nrarp (for Notch regulated ankyrin repeat protein) that contains two cdc10/ankyrin repeats (1). Nrarp can bind to the CBF/Nintra ternary complex and inhibit its activity. This feedback mechanism seems to be specific to vertebrates, since no equivalent gene exists in the genome of *Drosophila* or the nematode *C. elegans*.

Interactions with Other Signalling Pathways

Several important interactions with two other pathways have been reported. Various publications report negative interactions of the *Notch*- and Wnt- pathways at different levels (1, 2). For example the Wnt transducer Dishevelled (Dsh) can bind to the intracellular domain of Notch and prevent the activation of the receptor (Fig. 1). Furthermore, interactions are reported between Armadillo/ β -catenin, the major Wnt transducer, and Psn, which seem to promote the degradation of β -catenin.

Moreover, a strong antagonism between the EGF-pathway and Notch exists in several processes during *Drosophila* development. Whether this is also the case in the vertebrates remains to be determined.

► Wnt/ β -Catenin Signaling Pathway

References

1. Baron M, Aslam H, Flasz M et al (2002) Multiple levels of Notch regulation. *Mol Membr Biol* 19:27–38
2. Kopan R (2002) Notch: a membrane-bound transcription factor. *J Cell Sci* 115:1095–1097
3. Gridley T (2002) Notch signalling and inherited disease syndromes. *Hum Mol Genet* 12:R9–R13
4. Pear WS (2003) Notch and Cancer: Best to avoid the ups and downs. *Cancer Cell* 3 (3):203–205
5. Edbauer D, Winkler E, Regula JT et al (2003) Reconstitution of γ -secretase activity. *Nat Cell Biol* 5:486–488
6. La Voie MJ, Selkoe DJ (2003) The Notch ligands, Jagged and Delta, are sequentially processed by alpha-secretase and presenilin/gamma-secretase and release signaling fragments. *J Biol Chem* 278:34427–34437
7. O'Connor-Giles KM (2003) Numb inhibits membrane localization of Sanpodo, a four-pass transmembrane protein, to promote asymmetric divisions in *Drosophila*. *Dev Cell* 5:231–243
8. Haines N, Irvine KD (2003) Glycosylation regulates notch-signaling. *Nat Rev Mol Cell Biol* 4:786–797

Notochord

Definition

Notochord refers to the transient mesodermal “backbone” of the embryo located ventral of the neural tube, which is essential for initiating the differentiation of the nervous system and the somites.

- Bone Disease and Skeletal Disorders, Genetics
- Neural Development

NPC

- Nuclear Pore Complex

NPxY/F Motif

Definition

NpxY/F motif designates a conserved amino acid motif of the cytoplasmic tail of the β integrin chains. It is a docking sequence for talin. Talin-integrin interactions mediate integrin activation. It is likely, but not proven, that this occurs by spatial separation of the cytoplasmic integrin chains, allowing unfolding of the extracellular part into a ligand binding integrin.

- Focal Complexes/Focal Contacts
- Integrin Signalling

NSF

Definition

NSF (N-ethylmaleimide-sensitive-factor) was one of the first proteins of the cellular trafficking machinery identified as necessary for vesicle fusion in an *in vitro* assay, which reconstitutes intra-Golgi transport (“Rothman assay”). NSF is a hexameric ATPase of the AAA family that is essential to prime ►SNAREs for membrane fusion and to untangle v-/t-SNARE complexes after membrane fusion has occurred. Sec18p, the orthologue of NSF in yeast, is encoded by the essential SEC18 gene.

- Vesicular Traffic

N-TAD

Definition

N-TAD denotes a (or the) amino-terminal transactivation domain. HIF- α (Hypoxia inducible transcription factor) has two independent transactivation domains (TADs).

- Hypoxia Inducible Factors

N-Terminal Myristoylation

- Fatty Acid Acylation of Proteins

NTPase

Definition

NTPases comprise of a group of enzymes that break the high-energy phosphodiester bonds in nucleotide triphosphate (NTP), during which the energy is transferred to energy-dependent cellular activities.

- DNA Helicases

Nuclear Acidic Proteins

►Non-Histone Chromatin Proteins

Nuclear Compartments

ALISTAIR E. T. NEWALL, JAYSON WANG,
DENISE SHEER

Human Cytogenetics Laboratory, Cancer Research UK
London Research Institute, London, UK
denise.sheer@cancer.org.uk

Definition

The genome of every eukaryotic cell is packaged within the cell nucleus. Conceptually, the role of the nucleus is to provide an environment suitable for both expression and maintenance of the genome. These two functions can be sub-divided into a broad range of activities, including such diverse processes as transcription, RNA processing and export, DNA replication and DNA damage sensing and repair. In addition the genome must be tightly packaged, yet large parts of it nonetheless must be readily available as a transcriptional substrate either for basal metabolism or in response to specific stimuli.

The subdivision of biological processes into spatially and/or temporally discrete compartments confers numerous advantages. It permits the segregation of otherwise mutually exclusive processes, creates locally high concentrations of factors necessary for specific events/pathways and allows the coupling of energetically favourable and unfavourable reactions. Thus compartmentalisation allows for a much greater degree of complexity and efficiency than a simple system consisting of a homogenous mixture of its components. Unlike the compartments (organelles) found within the cytosol, those of the nucleus are not bounded by membranes. However, they are generally cytologically visible and can be defined by their constituent sub-units. While many of these compartments remain poorly described, they are the subject of intensive study and their roles in normal and aberrant nuclear function are slowly being understood.

Characteristics

Nuclear compartments can be divided into two categories, ►chromatin and the nuclear bodies that are found in the inter-chromatin space.

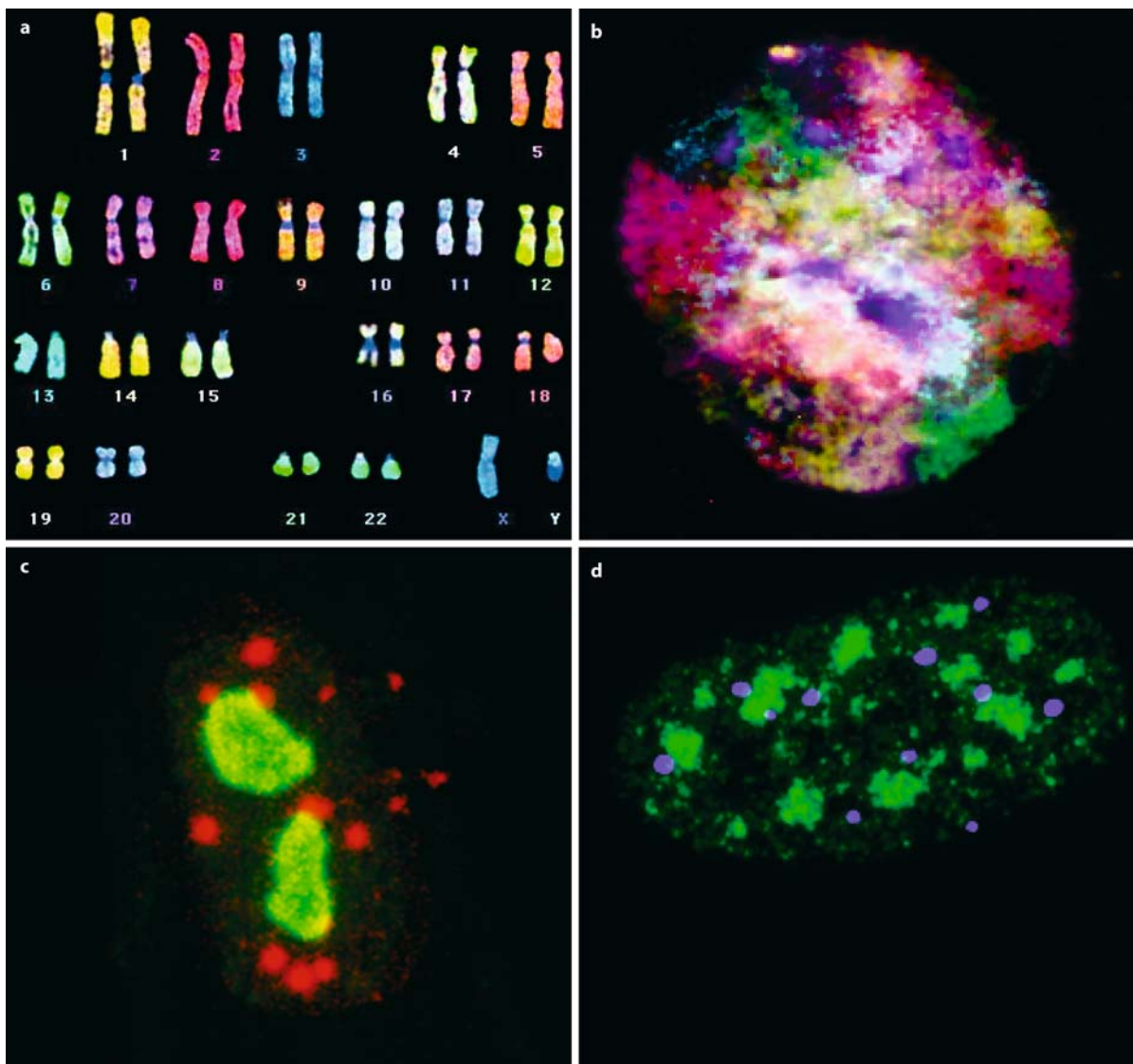
Chromatin

Unlike the compacted chromosomes seen in ►metaphase cells, which have a clear higher-order structure and exist distinct from one another, the conformation and position of decondensed chromatin in ►interphase nuclei is much less well defined. With the development of ►fluorescence in-situ hybridisation (FISH) methods, it has been shown that each chromosome occupies its own discrete territory in the nucleus (Fig. 1a,b). Between and even within these chromosome territories (CTs) is a chromatin-free space of varying width, termed the interchromatin domain (ICD). The ICD forms a network of channels extending to the nuclear pores, and contains nuclear bodies (see below) and smaller diffusible nuclear factors. It has been proposed that the interface between the surface of the CTs and the ICD is the major site of nuclear activity.

The subject of CT positioning in the interphase nucleus remains controversial. It has been suggested that smaller or more gene-dense chromosomes tend to be positioned more towards the nuclear interior. However it may be that differences in arrangements exist in different cell types, and indeed different species.

Within each CT the positions of specific genes and genomic regions also suggest some form of functional organisation. It has been shown that late-replicating, gene-poor regions are more likely to be located at the nuclear periphery and around the nucleolus than early-replicating, gene-rich regions (corresponding to R/T ►chromosome bands in metaphase chromosomes), which are found throughout the nuclear interior. It has been proposed therefore that chromosomes may be “polar”. Numerous studies have demonstrated the positioning of active genes on the CT surface, although the observation of transcription foci in the territory interior argues against such a straightforward relationship. This discrepancy may be explained by the fact that CTs have an irregular surface topology, with regions of invagination. In addition CTs may not form impenetrable barriers since diffusible molecules have been shown within them. Genes can also lie at the periphery of their chromosome territory, in extreme cases forming large-scale extensions or “loops”. In addition to the radial position of genes, studies have also shown that certain genes can associate, depending on their transcription status, with specific compartments, such as nucleoli, Cajal bodies, PML bodies, speckles and centromeric heterochromatin clusters (see below).

The studies thus far have largely been performed in fixed cells and so are uninformative about dynamic movements *in vivo*. Recently, live cell studies have shown that while CTs may have limited mobility, genes can also move up to 1–2 micrometres. It remains to be seen, however, if specific, directed movements exist.



Nuclear Compartments. Figure 1 Fluorescent staining of nuclear compartments (a) Human metaphase chromosomes visualised by M-FISH. (b) M-FISH image of a human B cell interphase nucleus, showing discrete chromosome territories. (c) Human fibroblast nucleus immunostained for nucleoli (green) and PML bodies (red). (d) Human fibroblast nucleus immunostained for PML bodies (blue) and SFC (green). Images (a) and (b) courtesy of R.Vatcheva

Nuclear Bodies

The dynamic complexity of the nucleus, together with the small size and inherent difficulty of biochemical purification of nuclear sub-domains, have rendered the study of nuclear bodies problematic. To date only a few have been extensively characterised, the remainder having only a simple morphological or behavioural description (Table 1 and Fig. 1c,d).

The Nucleolus

This is the most prominent nuclear structure. It is dynamic, its size and organisation depending on cell type and status.

The chief function of the nucleolus is in ribosomal biogenesis. It is assembled around tandemly repeated clusters of **▶ribosomal RNA genes** (rDNA) found on human chromosomes 13, 14, 15, 21 and 22. These clusters are termed the “nucleolar-organising regions” (NORs). Although nucleolar integrity is dependent on ongoing rDNA transcription, only a fraction of rDNA genes are active. It is believed that heterochromatin at the NORs is important in nucleolar assembly, possibly *via* clustering of silent regions adjacent to active rDNA genes.

The nucleolus consists of three distinct morphological regions through which rRNA biogenesis occurs in a

Nuclear Compartments. Table 1 Nuclear bodies

Body	Appearance/behaviour	Postulated function(s)
Nucleolus	Large	rRNA biogenesis, RNA processing, nuclear regulation
Cajal body/ GEM	Small, multiple, dynamic	RNP maturation
PML body	Small, multiple, generally static	Multiple postulated functions
Splicing factor compartment	Multiple “speckles” of variable size, close to active genes	Splicing factor “depots”
OPT domain	Small, associate with chromosomes 6 and 7	Transcription of specific genes
Perinucleolar compartment	Touching the nucleolus	Transcription/RNA processing
Cleavage body	Small foci	RNA processing
GATA-1 body	Small foci	Transcription
PcG domain	Pericentromeric foci	Gene silencing

For further details see (6) and references therein.

vectorial manner: the “fibrillar centre” (*fc*), the surrounding “dense fibrillar component” (*dfc*), and the external “granular component” (*gc*). The *fc* consists of rDNA, active genes having a more peripheral localisation that extends into the *dfc*. rDNA primary transcripts transiently accumulate in the *dfc* and undergo a complex series of ►[rRNA processing events](#) as they radiate out into the *gc*. The final rRNA processing occurs here, following which the pre-ribosomal particles are translocated to the cytoplasm. In addition to rRNA synthesis and processing, the nucleolus is believed to be the site of assembly and action of other ►[RNP complexes](#), many of these having a role in protein synthesis. The nucleolus may also have a role in the processing of small transcripts generated by RNA polymerase III. Finally, aside from its role in RNA processing, there is increasing evidence that the nucleolus is important as a site of sequestration of regulatory complexes away from their target molecules. In particular it seems that confinement within and release from the nucleolus of cell-cycle and mitogenic growth factors has a critical role in the regulation of cell division and proliferation.

Cajal Bodies

Also known as “coiled bodies”, Cajal bodies (CBs) were initially described as “nucleolar accessory bodies”. A typical nucleus has 1–5 CBs although this number varies between cell types and within the cell cycle. They are spherical in shape, having a diameter of 0.1–1 µm, and are highly dynamic, most likely moving by simple diffusion.

By electron microscopy CBs appear as a tangle of coiled fibrillar strands. They are specifically detected by antibodies to the protein p80-coilin, although only a minor fraction of nuclear coilin is present in CBs at any given time. CBs have a large number of other components, being enriched in spliceosomal snRNPs, several nucleolar proteins and snoRNPs, RNA polymerase II transcription factors and a novel class of CB-specific small RNAs (scaRNAs).

The function of CBs remains unclear. They are frequently found associated with chromosomal loci encoding histones and several sn- and sno-RNAs. This association is apparently dependent on RNA polymerase, although CBs themselves are unlikely to be sites of transcription or splicing. Instead there is increasing evidence that they are functionally connected to the nucleolus and have a role in co-ordinating the assembly and maturation of RNPs.

PML Bodies

The existence of this class of nuclear body was first shown in 1991, when it was found to be targeted by autoantibodies against the Sp100 protein in patients with primary biliary cirrhosis. Its main clinical interest, however, came in the mid-1990s when one of its main constituents was found to be the promyelocytic leukaemia (PML) protein (see Clinical Relevance, below). In normal cells there are usually 5–30 PML bodies, depending on cell type and the stage in the cell-cycle. As with Cajal bodies, they are spherical, having a diameter of 0.25–1 µm. However, unlike CBs, they are relatively immobile.

Nuclear Compartments. Table 2 Components of PML bodies

Constitutive proteins	Conditionally present components
PML	HP1
Sp100	CBP
Daxx	Rb
BLM	GAPDH
SUMO	P53
NDP55, NDP65	HMG1/2
	Mutant ataxin
	BRCA1
	PA28
	RAD51
	RAD52
	ICPO (HSV-1)
	E4 ORF33 (Adenovirus-5)
	E1A (Adenovirus-5)

For a more comprehensive list see (6) and (7) and references therein.

The PML protein is the main structural component of PML bodies, being essential for the recruitment of other proteins. The proteins that have been detected in PML bodies form a diverse and (at present) expanding list (Table 2). These proteins have implicated PML bodies in gene transcription, repair, cell cycle control and apoptosis. In addition, PML bodies have been found to associate physically with indicators of transcription (including nascent RNA and transcription factors), sites of DNA repair and viral replication. It has also been proposed that PML bodies may simply be depot sites in which disparate proteins are sequestered. For example, the tumour suppressor protein, BRCA1, colocalizes with these bodies only when over-expressed. To elucidate the function(s) of these bodies, mutant PML models have been analysed. PML $-/-$ mouse embryonic fibroblasts (MEFs), were found to grow faster than PML $+/+$ cells. Conversely, HeLa cells over-expressing PML showed an accumulation of cells in G1 phase and a delay into S phase. PML $-/-$ knockout mice were also more likely to succumb to infections, as well as developing tumours when exposed to carcinogens or ionising radiation.

Other Nuclear Bodies

Less well-characterised nuclear structures include:

- The splicing factor compartment (SFC), found as numerous “speckles” enriched in pre-mRNA splicing factors that are not directly associated with sites of transcription. The SFC consists of perichromatin fibrils, thought to represent nascent transcripts, and interchromatin granule clusters. It is believed that the SFC represents sites of deposition and assembly of spliceosomal factors.
- GEMS (Gemini of coiled bodies) are closely related to/overlap with CBs. They contain the SMN (survival of motor neurones) protein involved in snRNP assembly.
- Oct1, PTF transcription (OPT) domains are found as one or few bodies with a diameter of up to 1.5 μ m that are only present during the G1 phase of the cell cycle. They are transcriptionally active, their preferential association with chromosomes 6 and 7 suggesting that they may have a role in transcription of specific genes from these chromosomes
- The perinucleolar compartment (PNC), an electron-dense region has multiple strands that directly contact the nucleolus. It contains several RNA polymerase III snRNAs and a high concentration of certain RNA binding proteins, suggesting a possible role in transcription and RNA processing.

Other Nuclear Compartments

Aside from the structures outlined above there are other nuclear regions that may also be considered as functional compartments. There is evidence that DNA replication and mRNA production may each occur in a large number of discrete immobilised foci through which the template DNA is postulated to pass. These are termed replication and transcription factories respectively. Structural nuclear elements may also play an important role in gene regulation. For example in some cell types chromatin recruitment to domains such as the nuclear periphery and centromeric clusters is an important mechanism of gene silencing. The nuclear lamina, lining the inner surface of the nuclear envelope serves several roles. It determines the size and shape of the nucleus, supports the nuclear pore complexes (NPCs) necessary for transport of macromolecules between the nucleus and cytoplasm and may have a role in tethering chromosome territories.

Clinical Relevance

Although the functions of many of the above bodies have not yet been defined, there is evidence that changes in the distribution and assembly of these structures, and in the interactions between them, are relevant in a variety of human diseases.

Of the non-chromatin bodies, PML bodies must arguably have the most diverse clinical links. Their role in an autoimmune disease has been mentioned above. Their link to malignant disease is well

established in acute promyelocytic ▶**leukemia** (APL), in which PML is found as a chimeric protein, fused with the retinoic acid receptor- α (RAR α). APL patients have an abnormal PML body distribution. Treatment with all-transretinoic acid causes PML body reformation, induces differentiation of the leukaemic cells and results in clinical remission. PML bodies also appear to be involved in cellular immunity against infections, although this function remains controversial. While interferon causes an increase in their number, suggesting an antiviral action, certain viruses such as cytomegalovirus (CMV) and herpes simplex virus type 1 (HSV-1) disrupt PML bodies. Finally, in spinocerebellar ataxia type I (SCA I), a neurodegenerative disease, the gene encoding ataxin-1 (a protein normally found in PML bodies) becomes mutated and disrupts the normal distribution of PML bodies. Changes in Cajal body distribution and/or function are found in a number of diseases, which may be linked by a failure to process RNAs correctly. Examples include the association of CBs with the neuronal nuclear inclusions characteristic of poly-glutamate expansion diseases such as Huntington's and the loss of functional SMN protein (involved in RNP assembly) from GEMs/CBs in patients with the severe neurodegenerative disorder spinal muscular atrophy (SMA).

As mentioned above, the nucleolus is thought to play a role in the regulation of cell proliferation. In particular, nucleolar compartmentalisation of the oncoprotein MDM2 and its antagonist, the tumour-suppressor protein ARF, are believed to be important in the control of cell growth arrest *via* MDM2-mediated degradation of the tumour suppressor p53. In addition, the presence of a perinucleolar compartment closely correlates with oncogenic transformation. Clues as to other nucleolar roles may come from the finding that the WRN protein, which normally accumulates in the nucleolus, is lost in patients with the accelerated ageing disease Werner syndrome. WRN is a DNA helicase and is postulated to have a role in the maintenance of nucleolar integrity.

Chromatin positioning too has been implicated in the pathogenesis of certain cancers. The specificity with which some malignancies have chromosomal translocations has prompted investigators to examine translocation breakpoints in interphase cells. The results suggest that genes are often found in proximity to their translocation partners. Examples include the *ABL* and *BCR* genes on chromosomes 9 and 22 (translocated in chronic myeloid leukaemia) and *MYC* and Ig-heavy and light chain genes on chromosomes 8 and 2, 14 or 22 (translocated in Burkitt's lymphoma and B-cell leukaemia).

In conclusion, the study of individual nuclear compartments has revealed much about the distinct processes occurring within the nucleus. The greater challenge

now, however, lies in understanding the regulatory networks that link these functional compartments so that we may derive an integrated model for how the nucleus functions as a whole.

References

1. Carmo-Fonseca M (2002) The contribution of nuclear compartmentalization to gene regulation. *Cell* 108:513–521
2. Chevret E, Volpi EV, Sheer D (2000) Mini review: form and function in the human interphase chromosome. *Cytogenet Cell Genet* 90:13–21
3. Cremer T, Cremer C (2001) Chromosome territories, nuclear architecture and gene regulation in mammalian cells. *Nat Rev Genet* 2:292–301
4. Francastel C, Schubeler D, Martin DI et al (2000) Nuclear compartmentalization and gene activity. *Nat Rev Mol Cell Biol* 1:137–143
5. Dundr M, Misteli T (2001) Functional architecture in the cell nucleus. *Biochem J* 356:297–310
6. Matera AG (1999) Nuclear bodies: multifaceted subdomains of the interchromatin space. *Trends Cell Biol* 9:302–309
7. Negorev D, Maul GG (2001) Cellular proteins localized at and interacting within ND10/PML nuclear bodies/PODs suggest functions of a nuclear depot. *Oncogene* 20:7234–7242
8. Ogg SC, Lamond AI (2002) Cajal bodies and coilin-moving towards function. *J Cell Biol* 159:17–21

Nuclear Envelope

N

Definition

Nuclear envelope defines a double bilayer membrane that encircles and encloses the genome, and regulates the movement of molecules into and out of the nucleus through special nuclear pores.

- ▶ Centromeres
- ▶ RNA Export

Nuclear Export Factor 1

- ▶ NXF1

Nuclear Export Signal

Definition

Nuclear export signal (NES) refers to a short hydrophobic stretch of amino acids that is rich in leucine

residues, and which is present on export adapter proteins that interact with ►[karyopherin/exportin receptors](#). For example, LR-NESs' are enriched in the amino acid leucine and recognized by the ►[exportin Crm1](#).

- [NFκB Pathway](#)
- [Nuclear Import and Export](#)
- [Nuclear Pore Complex](#)
- [RNA Export](#)

Nuclear Hormone Receptors: Regulators of Gene Transcription and Cellular Signaling

ANDREW C.B. CATO, LIUBOV SHATKINA
Forschungszentrum Karlsruhe, Institute of Toxicology and Genetics, Karlsruhe, Germany
andrew.cato@itg.fzk.de

Definition

Nuclear hormone receptors form a large class of ►[ligand](#)-activated proteins. In the activated state, they bind specifically to distinct DNA sequences to regulate the transcription of a number of genes that control cellular proliferation, differentiation, body homeostasis and reproduction. The ligands that activate the nuclear receptors are ►[steroids](#), ►[retinoic acids](#) (all-trans and 9-cis isoforms), vitamin D3, ►[ecdysone](#), oxysterols, ►[bile acids](#), thyroid hormones, fatty acids, leukotrienes and ►[prostaglandins](#). In addition, a number of receptors have been identified for which no ligands exist and these are referred to as orphan receptors.

Characteristics

The nuclear receptors share a common structural organization. They possess a centrally conserved DNA binding domain (DBD, also termed C-domain) with a short motif termed the P-box, which is responsible for DNA binding. In addition they contain an N-terminal domain (A/B domain) variable in length from 50 to 500 amino acids with a ►[transactivation function](#) (AF-1) and a carboxy-terminus moderately conserved ligand-binding domain (E-domain) (Fig. 1). Additionally, the E-domain contains a further transactivation domain (AF-2), a strong dimerisation interface, a nuclear localization sequence (NLS) and often a repression function. Between the C and E domains, is a non-conserved region (D-domain) that contains a ►[bipartite NLS](#). Some nuclear receptors contain a further domain at the C-terminus of the E-domain

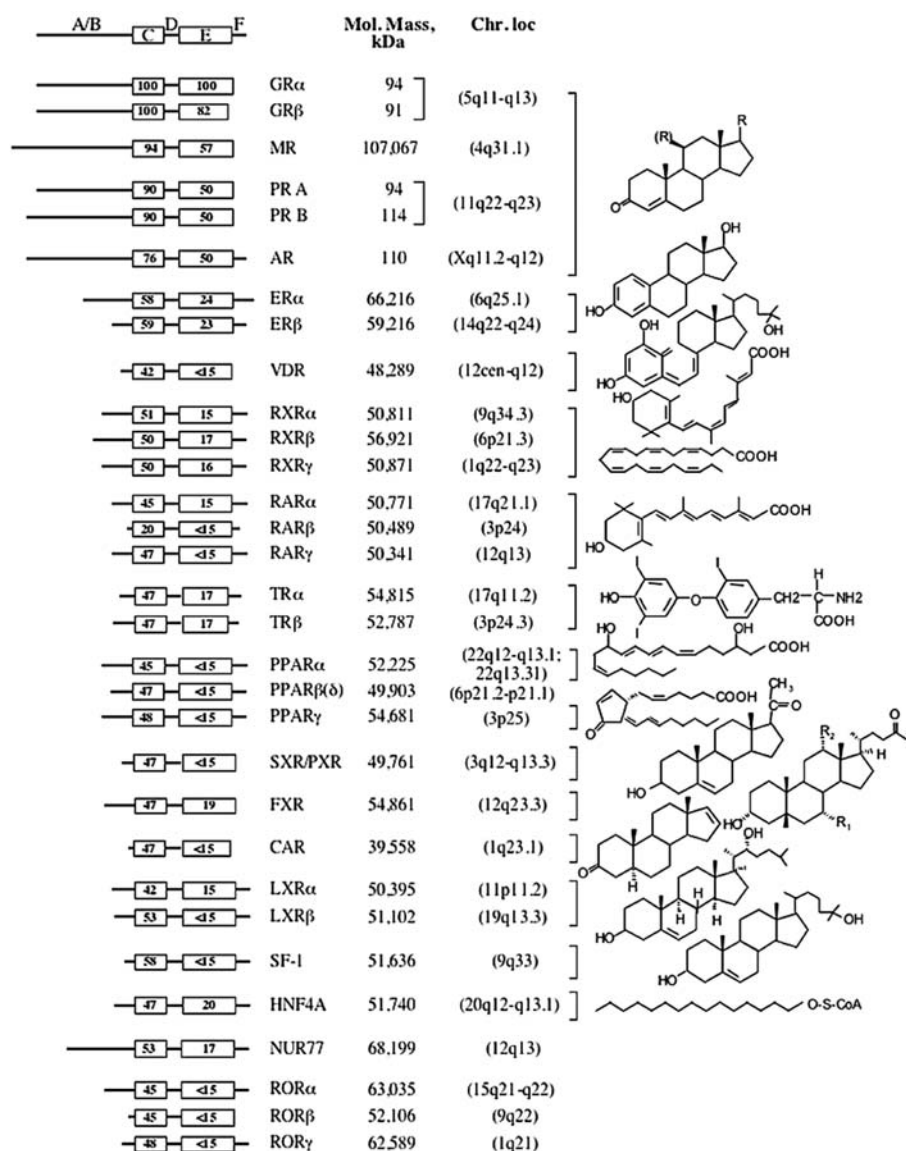
(F-domain), whose structure and function are not yet known (Fig. 1).

Nuclear receptors are organized in a ►[phylogeny](#)-based nomenclature (Nuclear Receptor Nomenclature Committee 1999) of the NR_{xyz} where X is the subfamily, Y is the group and Z the genes of receptors that possess both DNA binding and ligand binding domains (1) (Table 1). Some of the nuclear receptor genes code for different isoforms originating from alternative splicing of the primary transcript or may originate from different genes at distinct chromosomal loci (Fig. 1).

Molecular Interactions Response Elements

Nuclear receptors regulate transcription of target genes by binding DNA as homo- or hetero-dimers to hormone ►[response elements](#), which may be formed by a repetition of the consensus sequence motif PuGGTCA. Two such core sequences separated by three nucleotides arranged as ►[palindromes](#) are recognized by most steroid receptors as homodimers. Androgen receptor (AR), a member of the steroid receptor family, can also bind as homodimer to direct repeats of this motif and retinoic acid receptors (RARs), thyroid receptors (TRs) and some orphan receptors heterodimerize with retinoid X receptors (RXRs) and bind to direct repeats of the core motif. The spacing between the two halves determines the type of heterodimer that will be formed. For example direct repeats separated by 5 nucleotides (DR5) will be recognized by RXR-RAR heterodimer whereas a DR4 will be bound by RXR-TR. Some orphan receptors such as steroidogenic factor 1 (SF1), RAR-related orphan receptor (ROR), Reverse-erb A (Rev-erb) or Nur77 can bind DNA as monomers through a single core sequence. In this case, an A/T-rich region 5' to the core element controls the binding specificity. The X-ray structure of sequence with which the nuclear receptor binds DNA has been solved. The DBD contains a ►[zinc finger](#) motif with two subdomains each containing a zinc ion coordinated by four cysteine residues, followed by an alpha helix of the general sequence -C-X₂-C-X₁₃-C-X₂-C- and -C-X₅-C-X₉-C-X₂-C-. The four cysteines of each finger chelate a Zn²⁺ ion.

Response elements for nuclear receptors for the activation of gene expression (also known as transactivation) and elements for negative regulation of gene expression (transrepression) have been described. There are at least three different types of response elements for transactivation using the glucocorticoid receptor as a classic example (although other receptors can also function in the same way). First, there are simple response elements that are usually of imperfect palindromic nature where the nuclear receptor is the sole sequence-specific DNA binding protein. Secondly,



Nuclear Hormone Receptors: Regulators of Gene Transcription and Cellular Signaling. Figure 1 Schematic diagram of the domain structure of several members in the superfamily of nuclear receptors showing the amino acid identity of the individual domains with reference to the human glucocorticoid receptor. Indicated are the molecular mass of the receptors, their chromosomal localization and the ligands they bind to. For simplification only human nuclear receptors are presented. In the case of CAR, ligand binding has only been described for the mouse receptor.

there are composite response elements that specify binding sites for the nuclear receptor and one or two non-receptor proteins. Thirdly, there are tethering response elements where the receptor is recruited through protein-protein interaction to another DNA-bound factor (2). In negative regulation of transcription, the receptors can bind negative response elements described for the glucocorticoid receptor as the negative glucocorticoid response element. The interaction of the receptor with this element leads to the

repression of gene expression possibly due to the replacement of or interference with binding of essential transcription factors required for the activity of the responsive gene. The most important mechanism for transrepression is through tethering response elements. In this case the activated receptor is tethered to already bound transcription factors, mostly the activator protein-1 (▶AP-1) or nuclear factor-κB (▶NF-κB) to inhibit their activity. As these transcription factors control the expression of a number of pro-inflammatory

Nuclear Hormone Receptors: Regulators of Gene Transcription and Cellular Signaling. Table 1 A subset of the phylogenic-based nomenclature of nuclear receptors and the ligands that bind to the different classes of receptors. Detailed information on the classification of the nuclear receptors can be found in NUREBASE (► <http://www.ens-lyon.fr/LBMC/laudet/nurebase/nurebase.html>)

Group	Nomenclature	Receptor	Denomination	Ligand
I	NR1A1, NR1A2	TR α , β	Thyroid hormone receptor	Triiodothyronine
	NR1B1, NR1B2, NR1B3	RAR α , β , γ	Retinoic acid receptor	All-trans retinoic acid
	NR1C1, NR1C2, NR1C3	PPAR α , β , γ	Peroxisome proliferator activated receptor	Fatty acids (prostaglandin J2, leukotriene B4)
	NR1D1, NR1D2	Rev-erb α , β	Reverse ErbA	
	NR1D3	E75*		
	NR1E1	E78*		
	NR1F1, NR1F2, NR1F3	ROR α , β , γ	Retinoic acid-related orphan receptor	
	NR1F4	DHR3*	<i>Drosophila</i> homolog of orphan nuclear receptors 3	
	NR1H1	ECR*	Ecdysone receptor	20-OH-Ecdysone
	NR1H2, NR1H3	LXR α , β	Liver X receptor	22(R)-Hydroxycholesterol
	NR1H4, NR1H5	FXR α , β **	Farnesoid X receptor	CDCA (bile acids)
	NR1I1	VDR	Vitamin D receptor	1 α ,25-Dihydroxy-vitamin D3
	NR1I2	PXR	Pregnane X receptor	Pregnenolone
	NR1I3	CAR1	Constitutive androstane receptor	Androstenol
	NR1J1	DHR96*	<i>Drosophila</i> homolog of orphan nuclear receptors 96	
II	NR2A1, NR2A2	HNF4 α , γ	Hepatocyte nuclear factor 4	Palmitoyl-CoA
	NR2A4	HNF4*		
	NR2B1, NR2B2, NR2B3	RXR α , β , γ	Retinoid X receptor	9- <i>cis</i> -Retinoic Acid, Docosahexaenoic Acid
	NR2B4	USP*	Ultraspiracle	
	NR2C1, NR2C2	TR2,4	Testis receptor	
	NR2D1	DHR78*	<i>Drosophila</i> homolog of orphan nuclear receptors 78	
	NR2E1	TLX	Tailless-related receptor	
	NR2E2	TLL*	Tailless	
	NR2E3	PNR/PNR*	Photoreceptor specific nuclear receptor	
	NR2E4	Dissatisfaction*		
	NR2F1, NR2F2	COUP α , β	Chicken ovalbumin upstream promoter transcription factor	
	NR2F3	SVP*	Seven up	
	NR2F6	EAR2	ERBA-related gene	

Nuclear Hormone Receptors: Regulators of Gene Transcription and Cellular Signaling. Table 1 A subset of the phylogenetic-based nomenclature of nuclear receptors and the ligands that bind to the different classes of receptors. Detailed information on the classification of the nuclear receptors can be found in NUREBASE (► <http://www.ens-lyon.fr/LBMC/laudet/nurebase/nurebase.html>) (Continued)

Group	Nomenclature	Receptor	Denomination	Ligand
III	NR3A1, NR3A2	ER α , β	Estrogen receptor	17 β -Estradiol
	NR3B1, NR3B2, NR3B3	ERR α , β , γ	Estrogen-related receptor	
	NR3B4	ERR*		
	NR3C1	GR	Glucocorticoid receptor	Cortisol
	NR3C2	MR	Mineralocorticoid receptor	Aldosterone
	NR3C3	PR	Progesterone receptor	Progesterone
	NR3C4	AR	Androgen receptor	Dihydrotestosterone
IV	NR4A1	NGFIB	NGF-induced clone B	
	NR4A2	NURR1	Nur-related protein 1	
	NR4A3	NOR1 (NUR77)	Neuron-derived orphan receptor 1	
	NR4A4	DHR38*	<i>Drosophila</i> homolog of orphan nuclear receptors 38	
V	NR5A1	SF-1	Steroidogenic factor 1	25-Hydroxycholesterol
	NR5A2	LRH-1	liver receptor homolog 1	
	NR5A3	FTZF-1*	Fushi Tarazu factor 1	
	NR5B1	DHR39*	<i>Drosophila</i> homolog of orphan nuclear receptors 39	
VI	NR6A1	GCNF1	Germ cell nuclear factor 1	
	NR6A2	THR4*	<i>Tenebrio molitor</i> hormone receptors 4	

* - *Drosophila melanogaster* proteins

** - present in the mouse only

genes, the negative regulation of their activity has become a paradigm for the anti-inflammatory action of a number of nuclear receptors, including glucocorticoid receptors. Transrepression by nuclear receptor (also described as cross-talk of nuclear receptors and AP-1 or NF- κ B) has been described not only for the glucocorticoid receptor but also for a number of family members including the estrogen, thyroid, vitamin D3 and retinoid receptors. Transrepression of AP-1 by nuclear receptors has also been described in *Drosophila*, indicating that the mechanism involved in this action is conserved in the fly as well.

The Ligand-Binding Domain

The ligand-binding domain (LBD) of nuclear receptors is structured as a three layered α -helical anti-parallel sandwich of 12 helices forming a ►hydrophobic

pocket. In the presence of ligand, contacts are established between the ligand and amino acid residues in the LBD, producing a conformational change that closes the lid (helix 12) of the hydrophobic pocket (3). This conformational change is important for transcriptional activation since it allows the AF2 that is situated on helix 12 to interact with coregulatory proteins (►coactivators). The recruitment of coactivators to helix 12 is usually mediated by a conserved short LXXLL helical motif present in the majority of coactivators. Intramolecular protein-protein contact of the nuclear receptors is also essential for transactivation. This is established between the AF1 at the N-terminus and the AF2 at the C-terminus and may be necessary as a platform for the recruitment of coactivators to the AF1 or for the stabilization of helix 12. Mutational studies have identified key features at

the N-terminus for intramolecular contact. In the androgen receptor a core sequence FXXLF binds the AF2 while another motif WXXLF binds to the LBD outside AF2 (where F, W and L are the single letter code for the amino acids phenylalanine, tryptophan and leucine, respectively, and X stands for any amino acid). In addition to coactivators, **corepressors** such as nuclear receptor corepressor (NCoR) and silencing mediator of retinoid and thyroid receptors (SMRT) interact with the nuclear receptors in the unliganded state. Their interaction is dependent on a conserved motif referred to as the CoRNR box or as LXXI/HXXXI/L.

Regulatory Mechanisms

Regulation of Gene Expression

In the absence of ligand, most nuclear receptors are present in the nucleus or cytoplasm or in complex with **molecular chaperones and cochaperones**, **immunophilins** or **cyclophilins**. Upon ligand binding, the receptor dissociates from the non-receptor proteins and acquires the ability to bind to its response elements. Those receptors that reside in the cytoplasm migrate to the nucleus to bind DNA. One of the primary changes that take place after ligand binding is **phosphorylation** of the receptor. Most phosphorylation occurs on serine and threonine residues but the estrogen receptor (ER) is additionally phosphorylated on tyrosine and this is used to bind to Src-homology domain 2 (SH2) of the oncoprotein c-Src in rapid non-genomic signaling by this receptor. Those receptors that are bound to molecular chaperones have basal phosphorylation that is increased upon ligand binding. Phosphorylation is quite essential for receptor action as it has been implicated in DNA binding, transcriptional activation and stability of the receptors.

Binding of ligand to the LBD of nuclear receptor also leads to the recruitment of coactivators. Among the most well characterized coactivators are the p160 steroid receptor coactivators (SRC), that serve as a platform for the recruitment of histone-modifying enzymes including CREB binding protein (CBP/p300) and methyltransferases that acetylate or methylate histones. These modifications are responsible for the decondensation of histones, **chromatin remodeling**, assembly of general transcription factors and transcription of target genes. The chromatin remodeling is carried out by complexes such as mating type switch/sucrose nonfermenting complex (SWI/SNF), polybromo- and brahma-related gene 1-associated factors (PBAF) or Mediator/thyroid hormone receptor-associated protein (TRAP)/Vitamin D receptor-interacting proteins (DRIP) (4). In the case of thyroid receptors, **corepressors** with histone deacetylase activity repress transcription in the absence of ligand. Binding of the ligand results in the dissociation of the

corepressors and recruitment of coactivators leading to the activation of **gene expression**. A number of the corepressor proteins such as NCoR, SMRT, receptor-interacting protein 140 (RIP140) and ligand-dependent nuclear receptor corepressor (LCoR) may be recruited to the surface of some nuclear receptors depending on the ligand bound to the LBD. They too function as platforms to recruit enzymes such as histone deacetylases or redox-sensing cofactors including C-terminal binding protein (CtBP) to repress transcription (5). A different type of regulation is demonstrated by the orphan receptor, constitutive androstane receptor β (CAR β) for which the steroid like compound (androstanol) represses transcription upon binding. In the unliganded conformation, CAR β interacts with coactivators such as SRC-1 whereas androstanol binding reverses this interaction.

Although ligands that bind some orphan receptors have been identified, not all the orphan receptors may bind ligand. An example of such a receptor is Nurr1 (NR4A2, Nur-related protein 1) that possesses a LBD with constitutive and cell-type specific activity. The crystal structure of its LBD shows that it adopts a canonical protein fold resembling that of the agonist-bound transcriptionally active LBD of nuclear receptors. Nurr1 LBD contains no cavity for ligand binding. The NGF1-B subfamily (NR4A1) and Nor1 (NR4A3, neuron-derived orphan receptor 1) as well as the Drosophila DHR 38 (NR4A4) have a similar LBD and are therefore predicted to be ligand-independent transcription factors.

Ligand Independent Activation of Nuclear Receptors

In general nuclear receptors are conventional transcription factors regulated by **allostery**. A signal that can trigger a conformational change in the receptor similar to that produced by the ligand should be able to activate the receptor (Table 2). Indeed epidermal growth factor-mediated phosphorylation at serine 118 of ER α leads to transcriptional activation of this receptor in the absence of ligand (6). Similarly progesterone receptor can be activated in a ligand independent manner by adrenocorticotrophic hormone (ACTH). Examples of other nuclear receptors whose activity can be controlled by phosphorylation include peroxisome proliferator activated receptor PPAR γ and orphan receptor NGFI-B where phosphorylation reduces their DNA binding activity. Another way of regulating nuclear receptor action in the absence of ligand is the interaction of the nuclear receptor with non-receptor proteins through its LBD. This converts the nuclear receptor into an appropriate conformation for its interaction with the transcriptional machinery in the absence of ligand. An example of this type of regulation is the activation of the orphan receptor SF-1 through an interaction with the bicoid-related homeobox transcription factor

Nuclear Hormone Receptors: Regulators of Gene Transcription and Cellular Signaling. Table 2 Ligand-independent regulation of the activity of selected members of nuclear receptors and the signaling pathways involved in this regulation

Receptor	Regulation	Activator/Repressor	Signaling pathway
ER (NR3A)	Activation	Epidermal growth factor	Phosphorylation of Ser118 (Mitogen-activated protein kinase pathway)
		Substitution of Tyr537 by Ala or Ser	
		Human X box-binding protein 1	Protein-protein interaction leads to alteration of LBD folding
		Cyclin D1	
		Cadmium	Metalloestrogen-Cd ²⁺ binds to the LBD
PR _β (NR3C3)	Activation	Dopamine	
		8-bromo cAMP	
		Gonadotropin-releasing hormone antagonist	
AR (NR3C4)	Activation	Phorbol ester	Growth factors: Phosphorylation of Ser515 (Mitogen-activated protein kinase pathway)
		Epidermal growth factor	
		Insulin-like growth factor-1	Akt phosphorylation of Ser213 and Ser791 (Phosphatidylinositol-3'-kinase pathway)
		Interleukin 6	JAK (Janus activated kinase) signaling through STAT3 (signal transducer and activator of transcription 3)
		Sphingosine-1-phosphate	Rho (ras homolog gene) family of proteins signaling pathway; FHL2 (four and a half Lim-domain protein 2)
		Protein Kinase A	
Nur77 (NR4A3)	Repression	Adrenocorticotrophic hormone (ACTH)	Alters phosphorylation at Ser354 and DNA binding activity
SF-1 (NR5A1)	Activation	Pituitary homeobox 1 (Ptx1)	Protein-protein interaction leading to alteration of LBD folding
		Phosphorylation	Phosphorylation of Ser203 (Mitogen-activated protein kinase pathway)

pituitary homeobox 1 (Ptx1). Similarly, cyclin D1 binds to the ER α LBD and activates this receptor in the absence of estrogen (Table 2).

Rapid Action of Nuclear Receptor

Nuclear receptors activate a number of signaling molecules within seconds and minutes. These effects are too rapid to arise from the genomic effects of the receptors that are usually manifested within minutes to hours. This rapid action of nuclear receptors is insensitive to inhibitors of transcription and translation and ranges from activation of **▶mitogen-activated protein kinases** (MAPKs) and phosphatidylinositol 3-kinase (**▶PI3K**) to activation of protein kinase C (Table 3) (7). The mechanism of activation of these pathways is not fully understood but it involves the interaction of the

receptors with Src-homology region 2 or 3 domains on a number of signaling molecules. Following these interactions, the kinase domains of the signaling molecules are activated leading to the transmission of signals that turn on a cascade of kinases finally resulting in activation of MAPK and PI3K and an increase in intracellular calcium. The proto-oncogene product c-Src has been identified as a key protein in this rapid action of nuclear receptors since most of the receptors analyzed so far interact with it and it is thought that binding of the receptor to this protein is essential for triggering the rapid action of the nuclear receptors. The physiological significance of the rapid action of nuclear receptors is diverse, ranging from bone protection, proliferation of mammary epithelial and prostate cancer cells and vasodilation to neuroprotection and oocyte maturation.

Nuclear Import and Export. Table 3 Different signaling pathways activated by the rapid action of steroid receptors

Rapid action	Receptor(s)
Ca ²⁺ mobilization	ER α
Activation of ERK1 and ERK2	ER α , ER β , PR, AR, GR
Activation of G proteins	ER α , ER β
Inositol phosphate production	ER α , ER β
Stimulation of adenylyl cyclase	ER α , ER β
Increase in intracellular cAMP level	MR
Activation of JNK	ER β , GR
Inhibition of JNK	ER α , GR
Activation of p38	GR
Activation of PI3K	ER α , PR, AR
Inhibition of PI3K	GR
Activation of PKC	GR, MR
Activation of Akt	ER α , AR
Inhibition of Akt	GR
Induction of Ca ²⁺ influx	PR, MR
Inhibition of EGF receptor signaling	GR
Activation of IGF-1	ER α
Activation of Na ⁺ -H ⁺ antiporter	MR
Activation of eNOS	ER α
Inhibition of Na ⁺ ,K ⁺ -ATPase channel	MR
Activation of Na ⁺ /K ⁺ /2Cl ⁻ cotransporter	MR

References

1. Escriva H, Delaunay F, Laudet V (2000) Ligand binding and nuclear receptor evolution. *BioEssays* 22:717–727
2. Lefstin JA, Yamamoto KR (1998) Allosteric effects of DNA on transcriptional regulators. *Nature* 392:885–888
3. Wurtz J-M, Bourguet W, Renaud J-P et al (1996) A conical structure for the ligand-binding domain of nuclear receptors. *Nat Struct Biol* 3:87–94
4. Glass CK, Rosenfeld MG (2000) The coregulator exchange in transcriptional functions of nuclear receptors. *Genes Dev* 14:121–141
5. Belandia B, Parker MG (2003) Nuclear receptors: a rendezvous for chromatin remodelling factors. *Cell* 114:277–280

6. Tremblay A, Tremblay GB, Labrie F et al (1999) Ligand-independent recruitment of SRC-1 to estrogen receptor β through phosphorylation of activation function AF-1. *Mol Cell* 3:513–519
7. Cato ACB, Nestl A, Mink S (2002) Rapid actions of steroid receptors in cellular signalling pathways. *Science's STKE* 138:r8

Nuclear Import and Export

ANJA PANNEK, KATRIN STADE

Max-Delbrück-Center for Molecular Medicine,
Berlin, Germany

kstade@mdc-berlin.de

Definition

In eukaryotic cells, DNA replication and RNA transcription occur in the cell nucleus, whereas protein synthesis occurs in the cytoplasm. This spatial separation of essential cellular processes can only be overcome by allowing macromolecules to traverse the nuclear membrane in both directions. Proteins as well as several classes of RNAs, which in the following will be collectively referred to as cargo, need to be imported into and exported from the nucleus, respectively. The most prominent cellular cargoes for nuclear import are transcription factors, ribosomal proteins and RNA binding proteins. In addition, in virally infected cells, whole viral genomes need to be imported into the nucleus. Cellular export cargo largely consists of proteins, messenger RNAs (mRNAs), transfer RNAs (tRNAs), ribosomal RNAs (rRNAs) and other small nuclear RNAs (snRNAs). All classes of RNAs are found in complexes with specific proteins and hence leave the nucleus as ribonucleoprotein particles (RNPs).

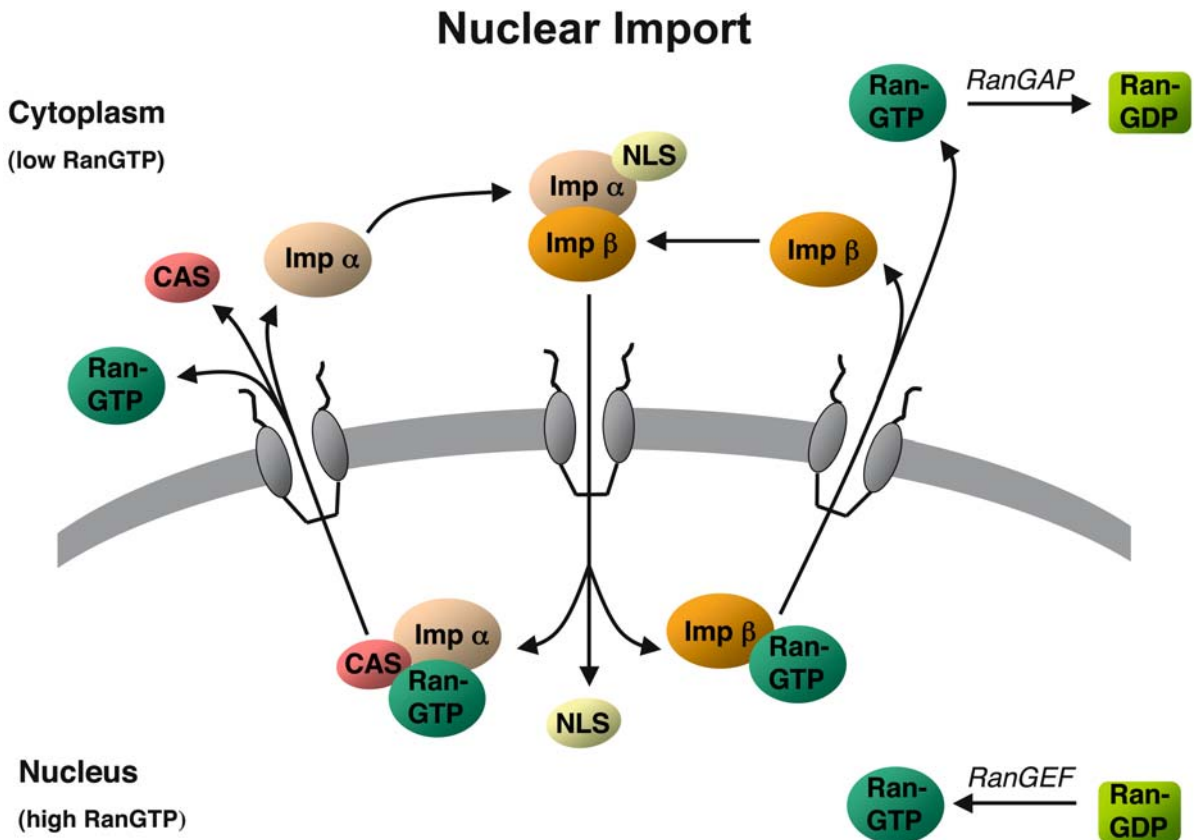
Characteristics

In general, nuclear transport relies on signals, receptors and energy (1). Viewed at the molecular level, nucleocytoplasmic trafficking can be described as follows. First, individual transport cargo is recognized by a specific transport receptor by virtue of either a ►nuclear localization signal (NLS) or a ►nuclear export signal (NES), respectively. Then, cargo-receptor complexes traverse the nuclear membrane through large proteinaceous structures called ►nuclear pore complexes (NPCs) that mediate all traffic between the nucleus and cytoplasm. Facilitated diffusion has been proposed to be responsible for this translocation step.

Finally, the release of individual transport cargo into the recipient compartment is achieved through interaction of the transport receptor with the guanine-nucleotide binding protein ►**Ran**. As a GTPase, Ran constantly switches between a GTP and a GDP bound form, although at a low intrinsic rate. *In vivo*, this reaction is enzymatically catalyzed by so-called effector proteins. The cytoplasmic localization of ►**RanGAP**, a protein that induces hydrolysis of RanGTP to RanGDP and the nuclear localization of ►**RanGEF**, which exchanges GDP for GTP on Ran, result in a low cytoplasmic concentration of RanGTP *versus* a high concentration of RanGTP in the nucleus. It is in fact this asymmetric distribution of RanGTP, also known as the Ran gradient, which defines compartment identity and confers directionality on transport processes by affecting the binding status of a transport receptor and its cargo.

Molecular Interactions Nuclear Import Pathways

Proteins destined for the nucleus contain a short amino acid motif generally termed a nuclear localization signal (NLS). Since NLSs can vary in amino acid composition and size, they are grouped into distinct classes. A particular NLS is recognized by a specific nuclear transport receptor or ►**karyopherin**. Karyopherins are a family of proteins of which the founding members were ►**importins** α and β . In Fig. 1, nuclear import is exemplified by the importin α/β route. In the cytoplasm, importin α binds to a cognate NLS on the cargo protein, associates with importin β and subsequently this trimeric complex docks at the nuclear pore complex. After NPC targeting is achieved, the complex travels through the nuclear pore, a process that is still not fully understood in its mechanistic details. Following translocation, the import complex dissociates in the



Nuclear Import and Export. Figure 1 Nuclear import. In the cytoplasm, the NLS-containing import cargo binds to the import receptor importin α which then associates with importin β . This trimeric complex translocates through the NPC into the cell nucleus, where importin β binds to RanGTP. This induces the release of NLS-cargo and importin α which binds to its own export receptor CAS in the presence of RanGTP. The importin subunits then travel back to the cytoplasm individually. On arrival in the cytoplasm, the two molecules of RanGTP utilized for importin export are converted to RanGDP by the RanGAP protein. This finally results in the release of the importins which are now free to engage in another round of import.

nucleoplasm releasing NLS-cargo and the two importin subunits. Release of cargo is achieved through binding of importin β to RanGTP whereas importin α is bound by its own export receptor, CAS, also in the presence of RanGTP. Both importins are then exported individually. Since two molecules of RanGTP are lost from the nucleus during importin subunit export, nuclear Ran needs to be constantly replenished. This goal is achieved by a specific Ran import pathway (1). Although the importin α/β route probably represents the best-studied example for nuclear protein import, additional import pathways exist. For example, nuclear import of the hnRNA binding protein A1 is mediated by an importin β -like transport receptor called transportin. In this case, a transportin-specific NLS is recognized on the A1 protein which is termed the M9 signal and which is different from the NLS recognized by importin α . Nuclear import of ribosomal proteins is mediated by several importins such as transportin, importin β , importin 5 and importin 7, which probably reflects the fact that such abundant cargo must enter the nucleus *via* parallel pathways to be effectively imported under conditions when ribosome production becomes a major cellular task. Again, the NLS signals used in this case differ from the so-called classical NLS recognized by importin α . Apart from nuclear protein import, nucleic acids are imported as well. For example, cellular uridine-rich small nuclear RNAs (UsnRNAs) are transcribed in the nucleoplasm but need to traverse the cytoplasm for maturation before returning to the nucleus. In addition, in virally infected cells, the viral genome needs to traverse the nuclear membrane to integrate into the host DNA for replication. However, in these cases it is not yet clear whether proteins with NLSs mediate nuclear import of these nucleic acids or whether alternative import pathways exist (2).

Nuclear Export Pathways

Nuclear export is fuelled by individual proteins as well as several classes of RNAs, which during their biogenesis in the nucleus are packaged into compact RNP particles by a diverse set of RNA binding proteins. As is the case for nuclear import, a specific signal on the cargo molecule is recognized by a nuclear export receptor or **▶exportin**, which in this complex travels through the NPC and releases its substrate into the cytoplasm (Fig. 2). However, unlike import complexes, export complexes can only be formed in the presence of RanGTP, i.e. in the nucleus and will dissociate in the absence of RanGTP, i.e. in the cytoplasm. Nuclear export signals of the so-called leucine-rich type (LR-NESs) have been identified on proteins such as protein kinase inhibitor (PKI) and are recognized by a karyopherin known as Crm1. Crm1 is also responsible for the nuclear export of other

important cellular regulators such as p53, I κ B α and cyclins. In addition, it mediates export of small nuclear RNAs, ribosomal RNAs and unprocessed viral messenger RNAs such as the precursor-mRNA of HIV. However in this case, additional adaptor proteins are needed since Crm1 does not bind directly to RNA. In addition to Crm1, several other exportins have been identified. For example, CAS is the exportin for the importin α subunit of the α/β import receptor (Fig. 1). It is noteworthy, that not all nuclear export is mediated by karyopherin family members. During their biogenesis in the nucleus, cellular mRNA transcripts become tightly packaged with specific sets of proteins. However, only after **▶mRNA processing** events such as 5'-capping, splicing and polyadenylation are complete, is an mRNA particle considered export competent and transported to the cytoplasm. Although some of the mRNA binding proteins have been proposed to contain NESs, no single karyopherin of the importin- β type has been shown to be responsible for bulk cellular mRNA export. Rather, an evolutionarily highly conserved complex of two proteins, TAP and p15, has been shown to mediate export of most cellular mRNAs. Since TAP/p15 can directly interact with the NPC, no additional adaptor protein is needed for nuclear export of cellular mRNAs.

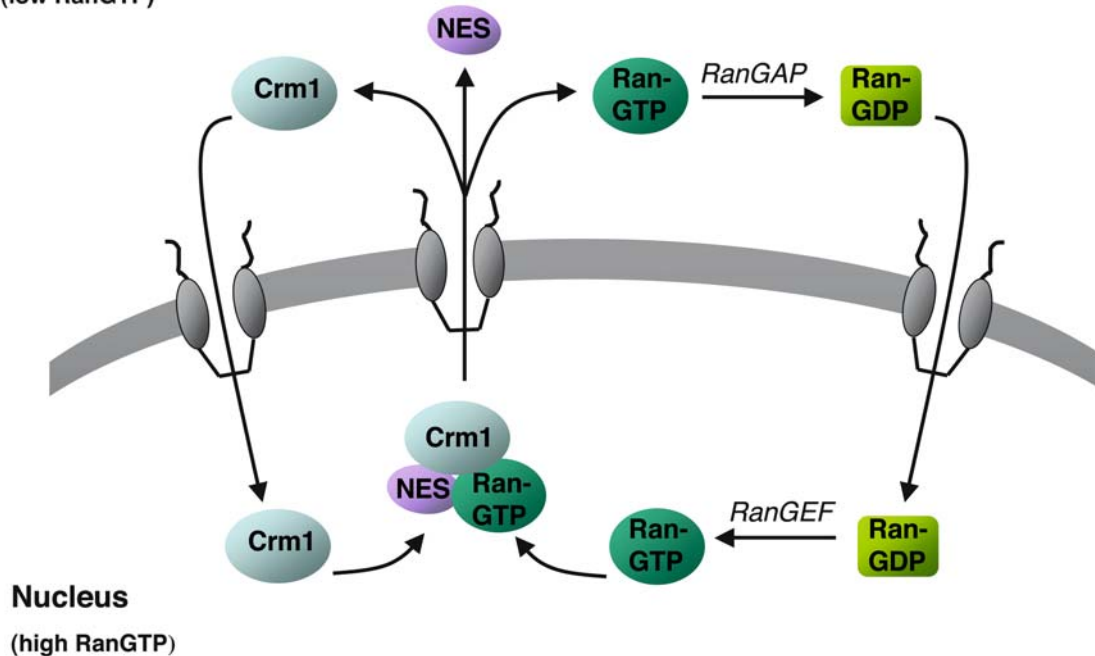
Regulatory Mechanisms

Nucleocytoplasmic trafficking is a major cellular activity, in terms of both number of individual particles involved and energy consumption. Despite a high degree of complexity, nuclear transport pathways offer an exquisite opportunity for regulating gene expression and cell-cycle progression at various levels (3). There are at least three ways in which nuclear transport can be regulated. First, the cargo itself can be modified in a way that affects its ability to bind to a karyopherin. For example, masking and unmasking of nuclear import signals by phosphorylation has been shown to control transcription factors such as Pho4 and NF-AT. An interesting example of masking of a nuclear export signal was demonstrated recently for the transcription factor Yap1 in budding yeast. Yap1 is a shuttling protein, which constantly travels between the nucleus and the cytoplasm. However under steady state conditions, Yap1 is mainly found in the cytoplasm due to a LR-NES recognized by yeast Crm1. This specialized NES also contains cysteine residues, which are proposed to sense the redox state of the cell. In a situation of oxidative stress, the cysteines in the NES region are oxidized and Crm1 can no longer recognize the NES on nuclear Yap1. This lack of export in the presence of functional import finally leads to nuclear accumulation of Yap1 and results in increased transcription of anti-oxidative stress genes.

Nuclear Export

Cytoplasm

(low RanGTP)



Nucleus

(high RanGTP)

Nuclear Import and Export. Figure 2 Nuclear export. In the nucleus, an NES-containing export substrate binds to the export receptor Crm1 in the presence of RanGTP. This trimeric complex travels through the NPC into the cytoplasm, where RanGTP is converted to RanGDP by RanGAP. This induces the release of export cargo from Crm1. Crm1 then travels back to the nucleus to engage in another round of export. Since during export RanGTP is lost from the nucleus, nuclear Ran needs to be replenished. This is achieved by reimporting RanGDP into the nucleus where it is converted to RanGTP by RanGEF.

As a second possibility, the activity of the soluble transport machinery can be regulated. Some karyopherins are expressed only in certain tissues or cell types or at specific stages of the cell-cycle. For example, in the fruit fly *Drosophila*, the importin α subunit enters the nucleus at the onset of the mitotic phase of the cell-cycle. Finally, in a third scenario, the individual proteins of the NPC and their unique functions in protein import and RNA export may be utilized to control nuclear transport routes (4). Since all import and export pathways converge at the nuclear pore complex, the NPC can be considered as the central gate for all traffic across the nuclear membrane. Nuclear pore proteins have been shown to interact with karyopherins during translocation although the precise mode of their interaction is not yet fully understood. Very recently, it was shown that the diameter of the NPC could vary during the cell-cycle which allows regulated access of cargo-receptor complexes to the nuclear interior.

Nuclear Transport and Human Disease

It is clear from the study of nuclear transport pathways that perturbations at any of the regulatory levels involved could potentially lead to abnormal cellular function and ultimately cell death (5). As an example, it is well established that cytoplasmic mislocalization of the transcription factor p53 results in a failure to activate its nuclear target genes, which has a severe impact on a cell's ability to respond to stress situations. In the same line of thought, it can be easily imagined that over- or under-representation of any of the karyopherins involved in **p53** import or export could potentially have the same cellular effects. In addition, any mutations in the general transport machinery such as NPC proteins or the Ran GTPase system could severely affect multiple nuclear import and export pathways. Indeed, several proteins of the nuclear pore complex have been found to be mutated in different types of leukemia (5). Hopefully, in the near future the ongoing study of nuclear transport pathways will help

to shed some light on the development of human cancer.

- [Nuclear Pore Complex](#)
- [Sumoylation](#)

References

1. Görlich D, Kutay U (1999) Transport between the Cell Nucleus and the Cytoplasm. *Annu Rev Cell Dev Biol* 15:607–660
2. Greber UF, Fassati A (2003) Nuclear Import of Viral DNA Genomes. *Traffic* 4:136–143
3. Kaffman A, O'Shea EK (1999) Regulation of nuclear localization: a key to a door. *Annu Rev Cell Dev Biol* 15:291–339
4. Fahrenkrog B, Aeby U (2003) The nuclear pore complex: Nucleocytoplasmic transport and beyond. *Nat Rev Mol Cell Biol* 4:757–766
5. Kau RT, Way JC, Silver PA (2004) Nuclear transport and cancer: from mechanism to intervention. *Nat Rev Cancer* 4:106–117

Nuclear Lamina

Definition

The nuclear lamina is the dense mat of protein lying immediately internal to the nuclear membrane, to which chromatin attaches, and which is punctuated by nuclear pore channels. It is mainly composed of lamin intermediate filaments, lamins A, B and C.

- [Intermediate Filaments](#)

Nuclear Localization Signal

Definition

Nuclear Localization Signal (NLS) designates a short sequence motif of SV40 type or bipartite type that is recognized by importins, and is required for import through the ► [nuclear pore complexes](#)

- [NFκB Pathway](#)
- [Nuclear Import and Export](#)
- [Nuclear Pore Complex](#)
- [Two-Hybrid System](#)

Nuclear Magnetic Resonance

- [NMR](#)

Nuclear Matrix

- [Non-Histone Chromatin Proteins](#)

Nuclear Medicine

Definition

Nuclear medicine is a medical field that uses radioactive substances to image the body and treat disease. Also known as radionuclide imaging or nuclear scintigraphy, these radiological techniques are primarily used diagnostically to investigate physiological function and anatomy of the body.

- [Molecular Imaging](#)

Nuclear Poly(A) Binding Protein 1

Definition

Nuclear poly(A) binding protein 1 (PABPN1) designates a protein that binds to the growing poly(A) tails of mRNA precursors during nuclear processing. The protein is required for efficient poly(A) synthesis, and for limiting the length of the poly(A) tail in a mammalian cell to approximately 250 nucleotides. At some point during mRNA export, the protein is exchanged for the cytoplasmic poly(A) binding protein.

- [Polyadenylation](#)
- [RNA Export](#)

Nuclear Pore Complex

BIRTHE FAHRENKROG, UELI AEBI
M.E. Müller Institute for Structural Biology,
Biozentrum, University of Basel, Basel, Switzerland
birthe.fahrenkrog@unibas.ch
ueli.aebi@unibas.ch

Definition

Nuclear pore complexes (NPCs) are large supramolecular assemblies that perforate the double-membraned

nuclear envelope (NE). Being composed of ~ 30 different proteins, termed **nucleoporins** (Nups), NPCs have an estimated molecular mass of ~ 125 MD in vertebrates and ~ 60 MD in the yeast *S. cerevisiae*. NPCs are the exclusive gateways for all traffic between the cell nucleus and the cytoplasm. Traffic through the NPC includes diffusion of ions and small molecules, as well as signal-dependent, receptor-mediated transport of proteins, RNAs and ribonucleotide protein (RNP) particles.

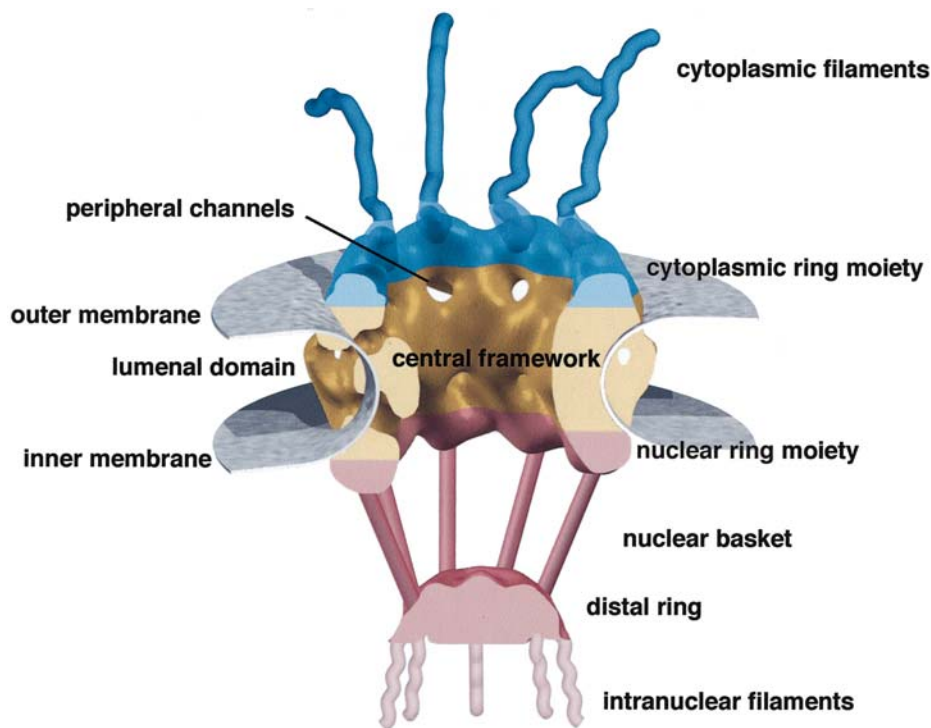
Characteristics

Electron microscopy studies, including 3D reconstructions, both in yeast and vertebrates, have led to a consensus model of the overall 3D NPC architecture (1, 2, 3, 4). Accordingly, the vertebrate NPC has a tripartite architecture with the central framework being embedded in the NE and the cytoplasmic filaments and the nuclear basket emanating towards the cytoplasm and the nucleus, respectively (Fig. 1). In the plane of the NE, the vertebrate NPC has a diameter of ~ 100 nm. The eight-fold symmetrical central framework is composed of eight spokes, each consisting of two approximately identical halves that are placed back to back in the midplane of the NE. The central framework therefore appears also roughly two-fold symmetrical in

the plane of the NE. The cytoplasmic ring moiety of the central framework is decorated by eight, ~ 50 nm-long kinky filaments, whereas the nuclear ring moiety is topped by the nuclear basket, an assembly of eight ~ 50 – 100 nm-long filaments that join into a distal ring (Fig. 1).

The central framework embraces the central pore, which has a length of ~ 90 nm and is narrowest in the midplane of the NE with a diameter of 45–50 nm widening towards the cytoplasmic and nuclear peripheries to ~ 70 nm (1, 2, 3, 4). It represents the sole gateway for all traffic through the NPC, allowing particles up to ~ 40 nm in diameter to traverse. In 2D projection images the central pore often appears obstructed by a particle, termed the central plug, which is highly variable both in size and shape. Based on 2D projection images and earlier 3D reconstructions (1, 2, 3), the central plug was thought to be a stationary component of the NPC. However, a more recent 3D reconstruction of native NPCs as well as nucleocytoplasmic transport studies using **atomic force microscopy** (AFM) suggest that the central plug most probably represents a major part the distal ring of the nuclear basket and, to a minor extent, cargo arrested in transit (3, 4).

3D reconstructions of the central framework have further unveiled eight peripheral channels or holes that



Nuclear Pore Complex. Figure 1 Nuclear pore complex architecture. The reconstructions of the central framework and the distal ring of the nuclear basket have been adapted from a tomographic reconstruction of native NPCs embedded in thick amorphous ice (3). The figure was modeled with a visual programming environment (ViPER) at the Scripps Research Institute (<http://www.scripps.edu>).

perforate the central framework (1, 2, 3). These peripheral channels have been proposed to mediate diffusion of ions and small molecules, to participate in the import of inner nuclear membrane proteins and/or to maintain the NE electrical conductance (1, 2). However, more recent data support the idea that the peripheral channels might act as voids or buffer zones to achieve the plasticity of the central framework needed for translocation of large cargoes, such as ►**Balbani ring mRNP** particles through the central pore or in response to chemical or physical effectors, such as Ca^{2+} or ATP, rather than as transport channels (1, 2, 3).

Molecular Interactions

The molecular building blocks of the NPC are a set of proteins called nucleoporins. Based on proteomic analyses in yeast and in vertebrates, the NPC is composed of ~30 different nucleoporins, each of which is present in 8 to 56 copies per NPC (1, 2). Based on immunolocalization studies by electron microscopy (EM), the majority of the nucleoporins are distributed fairly symmetrically on the cytoplasmic and the nuclear sides of the NPC, with only a few nucleoporins being located asymmetrically to either the cytoplasmic or the nuclear periphery (5, 6). Molecular interactions of nucleoporins involve intra-NPC interactions between distinct nucleoporins, as well as interaction of a subset of nucleoporins with soluble transport receptors that mediate nucleocytoplasmic transport of signal-bearing cargoes (1, 2, 7).

Interactions Between Nucleoporins

Based on biochemical and genetic interactions, nucleoporins are often organized in subcomplexes within the NPC (8, 9). The first identified and molecularly characterized NPC subcomplex was the vertebrate p62 complex, consisting of the nucleoporins p62, p58, p54 and p45 (Table 1). Based on immuno-EM data this complex localizes to the cytoplasmic and nuclear peripheries of the NPC's central pore and is involved in nuclear protein import (1, 2, 5, 8, 9). Additionally, p62 interacts with the nucleoporins CAN/Nup214 and Nup88, both located at the cytoplasmic ring moiety of the NPC. The example of p62 shows that nucleoporins can be organized in multiple subcomplexes. The Nsp1p complex represents the yeast homologue of the p62 complex. The Nsp1p complex is composed of the nucleoporins Nsp1p, Nup57p, Nup49p and Nic96p (Table 1). Like the p62 complex, the Nsp1p complex is implicated in nuclear protein import and is located at the cytoplasmic and nuclear peripheries of the central pore. Nsp1p and Nic96p are additionally found in a subcomplex that resides at the distal ring of the NPC's nuclear basket.

Another extensively studied subcomplex of the yeast NPC is the Nup84p complex, which consists of the

nucleoporins Nup84p, Nup85p, Nup120p, Nup145p, Seh1p and a fraction of Sec13p. Sec13 is also a subunit of the ►**COPII complex**, which is involved in vesicle transport from the ER to the Golgi apparatus. Based on immuno-EM data, the Nup84p complex is distributed symmetrically on both faces of the NPC and it participates exclusively in mRNA export. The vertebrate homolog of the Nup84p complex is the Nup107-160 complex, which is composed of the nucleoporins Nup107, Nup160, Nup96, Nup133, Nup37, Seh1 and Sec13 (1, 2, 8, 9). Besides having a role in mRNA export, this vertebrate NPC subcomplex is critical for nuclear pore assembly (1). The Nup107-160 complex has also been identified by its interaction with the vertebrate nucleoporin Nup153, a component of the nuclear basket, suggesting that Nup153 is also part of this subcomplex of the NPC. However, whereas the molecular composition of the distinct NPC subcomplexes is well described, how the individual nucleoporins interact with one another within as well as between subcomplexes has remained elusive.

Interactions Between Nucleoporins and Transport Receptors

A common feature of about one third of the nucleoporins is the presence of hydrophobic FG (phenylalanine-glycine) repeat motifs within their amino acid sequence (1, 2, 5, 6). ►**FG repeats** are often embedded in larger repeat motifs, such as FXFG (where X is any amino acid) or GLFG (where L is leucine). FG-nucleoporins predominantly line the cytoplasmic and the nuclear peripheries of the central pore and FG-repeat domains appear highly mobile – they can reach into and even through the central pore of the NPC (1, 2, 5, 6). FG-repeat nucleoporins might attract each other through weak hydrophobic interactions. Such weak hydrophobic interactions may give rise to a transient barrier within the central pore, which allows only the translocation of cargo that can penetrate into this hydrophobic meshwork (1, 2, 5, 6).

The interaction of FG nucleoporins with soluble transport receptors, which, in turn, associate strongly and specifically with signal-bearing cargo, is key for the translocation of cargo-receptor complexes through the NPC. Crystal structures of FG-repeat peptides in complex with different transport receptors, for example the protein import receptor ►**importin** β or the mRNA export receptor ►**TAP**, have unveiled that the interaction between the nucleoporin and the receptor primarily involves the phenylalanine ring of the FG core and hydrophobic residues on the surface of the receptor (1, 2). However, the strength and specificity of the interaction vary between different transport receptors and different nucleoporins.

Although the exact mechanism by which cargo-receptor complexes traverse the NPC has remained

elusive, different models for NPC translocation have been proposed based on the affinity and nature of the interactions between nucleoporins and transport receptors (1, 5, 7). However, there are nucleoporins, such as the yeast nucleoporin Nic96p, that exhibit no FG repeats within their amino acid sequence and yet are able to interact with transport receptors (1, 8),

indicating that other interactions between nucleoporins and transport receptors might play a role in nucleocytoplasmic transport.

Regulatory Mechanisms

The regulation of binding of cargo-receptor complexes to and their release from the NPC is best understood for

Nuclear Pore Complex. Table 1 Vertebrate and yeast nucleoporins

Nucleoporin	Localization	Motifs	Potential functional yeast homologue
Nup153	Nuclear basket	FG, ZF	Nup1p
Nup50	Nuclear basket	FG	Nup2p
Nup98	both sides	GLFG	C-Nup145p, Nup116p, Nup100p
Nup160	both sides		Nup120p
Nup133	both sides		Nup133p
Nup107	both sides	LZ	Nup84p
Nup75	both sides		Nup85p
Nup96	both sides		N-Nup145p
Sec13	both sides		Sec13p
Seh1	both sides	WD	Seh1p
Gle2/RAE1	?	WD	Gle2p
Nup155	both sides		Nup157p, Nup170p
Nup188	both sides		Nup188p
Nup205	nuclear	LZ	Nup192p
Nup93	nuclear	CC	Nic96p
p62	both sides	FG, CC	Nsp1p
p58	both sides	FG, CC	Nup49p
p54	both sides	FG, CC	Nup57p
p45	both sides	FG, CC	Nup49p
Nup88	cytoplasmic	CC	Nup82p
Nup214/CAN	cytoplasmic	FG, CC, LZ	Nup159p
Gle1	cytoplasmic		Gle1p
Nup358/RanBP2	cytoplasmic	FG, ZF, LZ	-
NLP1	cytoplasmic	FG, ZF	Nup42p
ALADIN	cytoplasmic	WD	-
Tpr	nuclear	CC, LZ	Mlp1p, Mlp2p
Nup37		WD	-
Nup35	both sides?		Nup53p, Nup59p
POM121	Integral membrane	FG	-
gp210	Integral membrane		-
Nup43		WD, CC	-

nucleocytoplasmic transport pathways that involve transport receptors of the importin family (also called ►**karyopherins**). Karyopherins collectively bind the small GTPase Ran, and can be either import receptors (importins) or export receptors (exportins) (7).

Ran, like all small GTPases, cycles between a GDP-bound and a GTP-bound state. The conversion between these two forms is achieved by two accessory factors, RanGAP, the GTPase-activating protein and RanGEF, the guanine-nucleotide exchange factor, termed RCC1. RanGAP is a cytoplasmic protein and catalyzes GTP hydrolysis by ►**Ran** with the help of two other cytoplasmic, Ran-binding proteins, RanBP1 and RanBP2. RanBP2, in turn, is a constituent of the cytoplasmic filaments of the NPC (Fig. 1), and is also referred to as Nup358 (Table 1). RCC1, the other Ran accessory factor, is, in contrast, a nuclear protein that catalyzes the loading of Ran with GTP. Due to this compartmentalization of the Ran accessory factors, a Ran gradient exists between the cytoplasmic and the nuclear NPC periphery, a high concentration of RanGTP within the nuclear nearfield of the NPC, and a high concentration of RanGDP in the cytoplasmic NPC nearfield. Ultimately, it is this RanGDP/GTP gradient between the cytoplasmic and the nuclear nearfield of the NPC, which determines the directionality of nucleocytoplasmic transport (6).

All karyopherins interact with RanGTP through a conserved amino-terminal domain. Importins bind their cargo in the cytoplasm in the absence of RanGTP. The binding of RanGTP to the importin causes a conformational switch within the importin molecule, so that its affinity for the cargo and the NPC is strongly reduced. As a consequence the cargo is released into the nucleus, whereas the importin-RanGTP complex is recycled back into the cytoplasm in a yet uncharacterized manner. In contrast, exportins can bind their nuclear export cargoes only in complex with RanGTP in the nucleus. Once on the cytoplasmic face of the NPC, the heterotrimeric cargo-exportin-RanGTP complex becomes dissociated upon RanGTP hydrolysis, which is induced by the orchestrated action of RanGAP, RanBP1 and RanBP2. RanGDP is imported back into the nucleus by its own import-receptor, called NFT2, which dissociates from Ran in the nucleus upon exchanging its bound GDP to GTP with the help of RCC1 (7).

Other regulatory mechanisms that might play a role in nucleocytoplasmic transport are, for example, phosphorylation of nucleoporins and/or the cargo, masking of ►**nuclear localization signals** or ►**nuclear export signals**, or the presence of nuclear retention signals on immature RNPs/RNAs. However, at the present time these latter mechanisms are not as well understood as the Ran cycle and its regulation of nucleocytoplasmic transport.

►Nuclear Import and Export

►RNA Export

►Sumoylation

References

1. Fahrenkrog B, Köser J, Aebersold U (2004) The nuclear pore complex: a jack of all trades? *Trends Cell Biol* 29:175–182
2. Fahrenkrog B, Aebersold U (2003) The nuclear pore complex: Nucleocytoplasmic transport and beyond. *Nat Rev Mol Cell Biol* 4:757–766
3. Stöffler D, Feja B, Fahrenkrog B et al (2003) Cryo-electron tomography provides novel insights into nuclear pore complex architecture—Implications for nucleocytoplasmic transport. *J Mol Biol* 328:119–130
4. Beck M, Forster F, Ecker M et al (2004) Nuclear pore complex structure and dynamics revealed by cryoelectron tomography. *Science* 306:1387–1390
5. Rout MP, Aitchison JD (2001) The nuclear pore complex as a transport machine. *J Biol Chem* 276:16593–16596
6. Rout MP, Aitchison JD, Magnasco MO et al (2003) Virtual gating and nuclear transport: the hole picture. *Trends Cell Biol* 13:622–628
7. Fried H, Kutay U (2003) Nucleocytoplasmic transport: taking an inventory. *Cell Mol Life Sci* 60:1659–1688
8. Ryan KJ, Wenthe SR (2000) The nuclear pore complex: a protein machinery bridging the nucleus and the cytoplasm. *Curr Opin Cell Biol* 12:361–371
9. Vasu SK, Forbes DJ (2001) Nuclear pores and nuclear assembly. *Curr Opin Cell Biol* 13:363–375

Nuclear Receptor Co-Activators

Definition

Nuclear receptor co-activators are nuclear molecules interacting with the glucocorticoid or mineralocorticoid receptors and components of the transcription machinery, potentiating the transactivation effects of glucocorticoids or mineralocorticoids on their responsive genes.

►Nuclear Hormone Receptors: Regulators of Gene Transcription and Cellular Signaling

►Steroid Hormone Receptor Defects, Molecular Basis

Nuclear Receptors

Definition

Intracellular or nuclear receptors bind lipophilic hormones (e.g. steroid hormones) specifically, and are

thereby activated to function as transcription factors; also sometimes referred to as nuclear steroid receptor.

- Methylation of Proteins
- Nuclear Hormone Receptors: Regulators of Gene Transcription and Cellular Signaling
- Protein/DNA Interaction
- Transcription Factors and Regulation of Gene Expression

Nuclease

Definition

Nuclease is an enzyme that degrades nucleotide strands, as for example DNase or RNase.

- Morpholino Oligonucleotides, Functional Genomics by Gene ‘Knock-down’

Nucleic Acid

Definition

Nucleic acids are polynucleotides that are composed of heterocyclic purin- (adenine, guanine) and pyrimidine- (cytosine, thymine) bases, carbohydrates (2-deoxyribose (DNA) or ribose (RNA) and phosphoric acid. A strand of nucleotides carries within its sequence the genetic information.

- DNA Structure
- Morpholino Oligonucleotides, Functional Genomics by Gene ‘Knock-down’

Nucleobase

Definition

Nucleobases are heterocyclic compounds that are constituents of nucleic acid (DNA, RNA). They include the purines adenine (A) and guanine (G) and the pyrimidines cytosine (C), thymine (T) and uracil (U; replaces T in RNA). They are involved in base pairing. A nucleobase that is bound covalently to the 1' carbon of deoxyribose or ribose is called a nucleoside; when a phosphate group gets coupled to the 5' carbon of a nucleoside, it is called nucleotide.

- SNP Detection and Mass Spectrometry

Nucleocapsid

Definition

Nucleocapsid is a proteinaceous structure encapsulating the genomic nucleic acid (RNA or DNA) of a virus, which may or may not be surrounded by a membranous structure.

- Translational Frameshifting, Non-standard Reading of the Genetic Code

Nucleolar Organising Regions

Definition

The NOR is the chromosomal domain that contains the rDNA repeats around which a nucleolus can form.

- Nuclear Compartments
- RNA Polymerase I

Nucleolus

Definition

The nucleolus is a cytologically and functionally distinct non-membrane-bound subcompartment of the nucleus, where RNA Polymerase I transcription of the rRNA genes occurs

- RNA Polymerase I
- Nuclear Compartments

Nucleophile/Nucleophilic

Definition

“Nucleus-loving”. Nucleophile designates a substance that becomes an electron donor in bonding during a chemical reaction. For instance, nucleophilic amino acid side chains are amino acid side chains containing chemical functional groups that are Lewis bases, which contribute electron pairs in covalent bond formation. The more nucleophilic a species is, the faster it will react.

- Proteases and Inhibitors
- Protein Interaction Analysis: Chemical Cross-Linking
- Protein Prenylation

Nucleoporins

Definition

Nucleoporins are proteins that line the nuclear pore complex, and which contain phenylalanine-glycine (FG) dipeptide repeats.

- ▶ Nuclear Import and Export
- ▶ Nuclear Pore Complex
- ▶ RNA Export

Nucleoside, Nucleotide

Definition

Nucleotides are the subunits forming the DNA and RNA molecules. Nucleotides are composed of a nucleobase (purin- or pyrimidin base), which is linked via a glycosidic bond to a ribose or 2'-deoxyribose (together a nucleoside), with 1, 2, or 3 phosphate groups attached to the 5'-position of the ribose.

- ▶ DNA Structure
- ▶ Nucleotide Biosynthesis
- ▶ SRY – Sex Reversal

Nucleosome Remodelling (Enzymes/Factors)

- ▶ Chromatin Remodelling

Nucleosome Repositioning

Definition

Nucleosome repositioning designates the change in location of a nucleosome particle relative to the DNA sequence.

- ▶ Nucleosomes

Nucleosomes

LOREDANA VERDONE¹, MICAELA CASERTA²

¹Dipartimento di Genetica e Biologia Molecolare, University La Sapienza, Rome, Italy

²Istituto di Biologia e Patologia Molecolari, CNR, Rome, Italy

micaela.caserta@uniroma1.it

Definitions

In the nucleus of all eukaryotic cells the DNA molecules, which carry the genetic information, wrap around well conserved protein particles forming the nucleosomes (1). They represent the basic units of ▶ [chromatin](#), which constitutes the fundamental material of chromosomes. Chromatin structure is strongly linked to the function of the eukaryotic genome. Significant progress has been made in understanding the organization and the dynamics of nucleosomes and of chromatin fibers and how these structures influence nuclear processes.

Characteristics

The DNA of a single eukaryotic cell is packed in the nucleus, according to a hierarchical scheme of folding. At the first level of organization, almost two tight superhelical turns (▶ [Supercoiled DNA](#)) of DNA (146 base pairs) are wrapped around an octamer of two copies each of the four ▶ [histone proteins](#), H2A, H2B, H3 and H4. This unit, the nucleosome core particle (NCP), represents the basic repeating structure in chromatin and is invariant over the whole eukaryotic kingdom. Histones are characterized by their abundance, relatively small size and strongly basic charge due to their high content of lysine and arginine residues. The charged amino acids are localized on the N-terminal tails that extend from the histone molecules. Histones are easily extracted by exposing chromatin to acids or salts at elevated concentrations. This kind of removal demonstrates that histones are held in chromatin by electrostatic attraction (between their positively charged residues and the negatively charged phosphate groups present on the surface of DNA helix) rather than by covalent bonds.

According to analysis of the crystal structure of a single nucleosome, DNA lies on the surface of the ▶ [histone octamer](#) and makes about 1.67 left-hand superhelical turns around it (2, 3). Nucleosomes are connected to each other by 10–90 base pairs of ▶ [linker DNA](#) depending on cell type, organism and physiological status. This DNA tract interacts with the histone H1, also called the linker histone because it binds to the

exterior of nucleosomes and stabilizes the highly condensed states of chromatin fibers. The simple arrays of nucleosomes along the DNA molecule represent the first level of chromatin structure (the 10 nm fiber, also called “the beads-on-a-string” structure). Most of the chromatin in the nucleus is even more highly compacted. The next stage of packing involves folding the beaded structure into a 30 nm fiber. These fibers may be further folded on themselves to make the thicker fibers visible in both metaphase chromosomes and nuclei of non-dividing (interphase) cells. The highly condensed regions of chromosomes are called ►**heterochromatin**, the more open chromatin regions are called ►**euchromatin**; in these relaxed domains transcription may occur.

Histone Variants

Histone variants within any species are highly conserved specialized histones that co-exist with the major histone types in the nucleus and have the potential to alter chromatin structure locally. H2A and H3 variants have diversified to assume roles in ►**epigenetic** silencing, gene expression and centromere function (4).

One of the best-characterized variants is H2A.Z; it has been shown to be important for viability in many organisms. Its function is thought to be different from that of the major H2A protein; H2A.Z appears to alter nucleosome stability and is involved in gene regulation by maintaining a transcriptionally permissive “open” state and by protecting euchromatin from encroachment by “silent” heterochromatin. Another H2A variant is H2A.X, whose defining feature is a C-terminal extension containing a motif that is crucial for chromatin compaction and DNA repair.

Cse-4 and CENP-A are yeast and human, respectively, histone H3 variants that appear to be essential structural components of centromeric nucleosomes. Cse-4 can heterodimerize with H4 *in vivo*; CENP-A can be reconstituted into a *bona fide* nucleosome *in vitro*. H4 and H2B are invariant.

Histone Modifications

The structure and functions of nucleosomes are influenced by chemical alterations in the histone molecules. Covalent modifications of the core histones have important roles in gene regulation. In particular, the amino-terminal tails, which protrude out of the histone octamer, are subject to various covalent alterations, acetylation, phosphorylation, methylation and ubiquitination.

A histone modification affects chromosome function by two possible distinct mechanisms. Either it changes the electrostatic charge and as a consequence the structural properties of the histone or its binding to DNA, or it can create binding surfaces for protein

recognition modules, thus recruiting specific functional complexes to their proper sites of action (for instance, acetylated lysines are recognized by protein ►**bromodomains**, methylated lysines by protein ►**chromodomains**).

This multiplicity of modifications and their occurrence in defined patterns of combinations argues for the existence of a “histone code”(5).

Acetylation

This consists of the addition of acetyl groups to conserved lysine residues in histone tails. This process is catalyzed by ►**histone acetyltransferases** (HATs); the inverse process is catalyzed by ►**histone deacetylases** (HDACs). As a consequence of histone amino-terminal tail acetylation, a decrease in their overall positive charge occurs; as these tails are highly basic, it has long been postulated that acetylation decreases their affinity for the negatively charged DNA and facilitates the binding of proteins that regulate transcription to chromatin templates. HATs utilize acetyl-coenzyme A as a cofactor. Many of the HATs are components of large multisubunit complexes, recruited to promoters by interaction with DNA-bound activator proteins. Two major families have been discerned on the basis of structural similarities: GCN5-related N-acetyltransferases (GNAT) and MOZ, Ybf2/Sas3, Sas2, Tip60 (MYST). Although most nuclear acetyltransferases have been linked to transcriptional regulation there is evidence that they play a role in other chromatin-based processes. TAFII250 and Esa1 HATs regulate cell-cycle progression, Tip60 HAT complex is involved in DNA repair and apoptosis and Elp3 HAT activity helps the RNA polymerase II elongation complex. In addition, histone acetylation promotes cell differentiation and modulates cell survival, in fact MyoD, the key regulator of myogenesis, is associated with both p300/CBP and PCAF (human-HATs complexes).

Phosphorylation

This consists of the addition of phosphate groups to conserved serine residues in histone tails. The phosphorylation involving Ser-10 of histone H3 has emerged as an important modification, both in transcriptional activation and in chromosome condensation during mitosis. The identity of histone kinases (HKs) responsible for Ser-10 phosphorylation with previously known transcription-associated factors suggests that they may be recruited to specific promoters as coactivators just like the HATs.

Methylation

This consists of the addition of methyl groups. There are two types of histone methylation, targeting either arginine or lysine residues. In contrast to the previous modifications, methylation does not alter the overall

charge of the histone tails. By increasing methyl addition (mono, di or tri) the histone gets more and more basic and hydrophobic. These observations suggest a tight association between methylated histone tails and DNA and/or chromatin. Methylation of histones is catalyzed by histone methyltransferases (HMTs), which use S-adenosyl methionine (SAM) as a cofactor.

Unlike histone acetylation, which occurs throughout the cell-cycle, histone methylation peaks in the G2 phase, subsequent to DNA replication and histone synthesis and during heterochromatin assembly. Different histone methylations are associated with different chromatin functions, for example, H3 K4 methylation is linked to active genes, whereas H3 K9 methylation is linked to inactive genes.

Ubiquitination

Histone ubiquitination is still poorly understood. In the yeast *Saccharomyces cerevisiae* it has been shown that the ubiquitin-conjugating enzyme Rad6 (Ubc2) mediates methylation of histone H3 at lysine 4 through the ubiquitination of H2B at Lys 123. This modification is critical to mitotic and meiotic growth, although it is not yet clear whether it is involved in transcription.

► Chromatin Remodeling

Nucleosome structure is dynamic. This feature represents a fundamental requisite for all the processes taking place in the nucleus. In fact, transcription and replication, as well as DNA recombination and repair can occur only when specific protein complexes overcome the nucleosome barrier and reach the relevant DNA sites of action, e.g. promoters and replication origins. Nevertheless, nucleosomes do not only play a repressive role; they also contribute to the assembly of productive protein complexes on their surface.

The dynamic behavior of chromatin structure is largely due to enzymatic covalent modifications of the N-terminal amino acids of the core histones, but it is also due to the function of ATP-dependent chromatin remodeling complexes (6). These are able to alter the folding, fluidity and basic structure of chromatin by using the chemical energy of ATP hydrolysis. The results are mobilization and ►nucleosome repositioning.

Protein purification has revealed different classes of multi-component protein complexes that require ATP hydrolysis for their action, the *Drosophila* NURF, CHRAC and ACF, the yeast SWI/SNF, RSC, ISW1 and ISW2 complexes and the vertebrate SWI/SNF, ERC1, RSF and Mi-2/NRD/NURD complexes.

Although these complexes are compositionally and functionally different, they all share a motor subunit that belongs to the Snf2-like family of ATPases. The ATPase polypeptides of these complexes possess an

intrinsic chromatin-remodeling function, while the non-ATPase subunits may enhance or regulate the motor activity of the ATPase subunits. The non-ATPase subunits can mediate other specialized functions, for example both human SWI/SNF complex and its yeast counterpart can be recruited to target genes through direct interactions with gene-specific regulatory proteins.

As far as the remodeling mechanism is concerned, several models have been suggested. The first model for SWI/SNF-like enzyme activity is based on the idea that the energy coming from ATP hydrolysis is used to drive removal of one or both of the histone H2A-H2B dimers, resulting in DNA wrapping onto a tetramer of histones H3 and H4. A model was later proposed in which dimers are not lost but undergo dramatic rearrangements, giving rise to a novel nucleosome conformation and to its relocation to different positions relative to the DNA template. Most of the proposed mechanisms involve changes in the topology of nucleosomal DNA. ATP hydrolysis is used to generate superhelical torsion in nucleosomal DNA. Superhelical stress might then be relieved on the nucleosome surface by random dissociation (bulging) of small stretches of nucleosomal DNA. In these models the remodeling enzyme must continually inter-convert the distribution of bulged species, producing a window of opportunity for factor binding at essentially any position within the nucleosome. Although there are many data supporting these models, the molecular mechanisms of chromatin remodeling still remain to be determined.

Clinical Relevance

In the past few years the importance of chromatin structure in human disease has become increasingly clear; in fact, the mutation of genes whose products are predicted to regulate nucleosome structure and location is often the cause of complex multi-system disorders or neoplasias. Inherited diseases have been associated so far with mutations in several nucleosome remodeling ATPases, such as SWI/SNF, as well as in histone acetyltransferases (7).

The best example of a chromatin-remodeling disorder is the α -thalassemia/mental retardation syndrome (ATR-X, where the X indicates that it is linked to the X chromosome). It is due to mutations of the ATRX gene, coding for a nuclear protein that belongs to the family of SWI/SNF ATPases. Further examples of chromatin-remodeling disorders include mutations in other SWI/SNF related ATPases involved in nucleosome repositioning. Human diseases are also determined by mutations of genes coding for histone covalently modifying enzymes, as for example the Rubinstein Taybi syndrome caused by mutations in the CREB binding protein (CBP), a histone acetyltransferase. In addition, chromosomal translocations resulting

in the fusion of CBP or EP300, both involved in tumor suppression, with the ZNF220 gene, another histone acetyltransferase, are associated with aggressive human acute myeloid leukemia. A few diseases have been associated with aberrant histone deacetylases. The best example of this group is ►[Rett syndrome](#) caused by mutations in the MECP2 gene, encoding a protein that binds methyl-CpG in DNA. The current hypothesis is that the MECP2 protein recruits HDACs to methylated DNA and induces transcriptional silencing.

- [Protein/DNA Interaction](#)
- [Transcriptional Repression](#)
- [Transgene Silencing](#)

References

1. Kornberg RD, Thomas JO (1974) Chromatin structure; oligomers of the histones. *Science* 184:865–868
2. Luger K, Mader AW, Richmond RK et al (1997) Crystal structure of the nucleosome core particle at 2.8 Å resolution. *Nature* 389:251–260
3. Richmond TJ, Davey CA (2003) The structure of DNA in the nucleosome core. *Nature* 423:145–150
4. Malik HS, Henikoff S (2003) Phylogenomics of the nucleosome. *Nat Struct Biol* 10:882–891
5. Fischle W, Wang Y, Allis CD (2003) Histone and chromatin cross-talk. *Curr Opin Cell Biol* 15:172–183
6. Lusser A, Kadonaga JT (2003) Chromatin remodeling by ATP-dependent molecular machines. *Bioessays* 25:1192–1200
7. Huang C, Sloan EA, Boerkoel CF (2003) Chromatin remodeling and human diseases. *Curr Opin Genet Dev* 13:246–252

Nucleotide

- [Nucleoside, Nucleotide](#)
- [Nucleotide Biosynthesis](#)

Nucleotide Binding Fold

Definition

Nucleotide binding fold refers to a binding site of ATP or ADP. Nucleotide binding folds consist of Walker A and B motif and canonical signature motif (S-G-G). Nucleotide binding folds contain a highly conserved phosphate binding loop.

- [High-HDL Syndrome](#)

Nucleotide Biosynthesis

RICHARD I. CHRISTOPHERSON

School of Molecular and Microbial Biosciences,
University of Sydney, Sydney, Australia
ric@mmb.usyd.edu.au

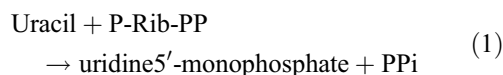
Synonyms

►[Nucleotides](#) are the building blocks of ►[DNA](#) and ►[RNA](#) and contain either a ►[pyrimidine](#) or a ►[purine](#) base. The pyrimidine bases found in DNA are cytosine and thymine; cytosine and uracil are found in RNA. The purine bases found in both DNA and RNA are adenine and guanine.

Definitions

Nucleotides consist of three chemical components, one of the 5 bases mentioned above, a ribose moiety in RNA and a deoxyribose in DNA and one to three phosphate groups linked together and to the ribose. There are two routes for nucleotide biosynthesis in living cells, the *de novo* and ►[salvage pathways](#).

The salvage synthesis of nucleotides may occur from a base such as uracil using a reaction catalyzed by a phosphoribosyltransferase enzyme.



All bases listed under Synonyms may participate in such a reaction. The alternative salvage reaction involves a ►[nucleoside](#) kinase enzyme.



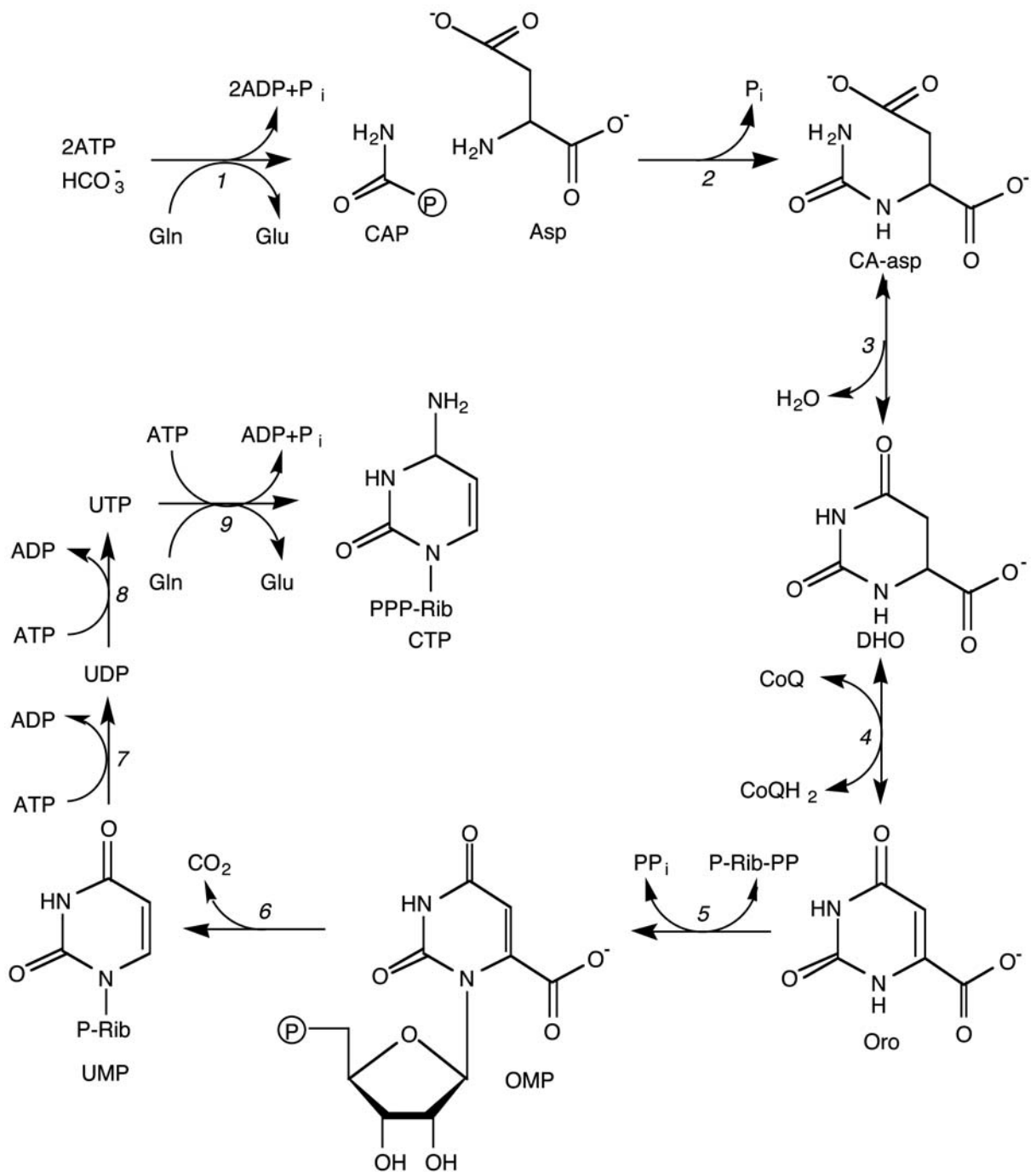
Nucleosides that may participate in such a reaction are uridine, cytidine, (deoxy)thymidine, deoxycytidine, adenosine, guanosine, deoxyadenosine and deoxyguanosine.

The pathway for *de novo* biosynthesis of pyrimidine nucleotides consists of 9 enzyme-catalyzed reactions involving combination of the precursors, bicarbonate, glutamine, ATP, aspartate and P-Rib-PP to yield cytidine 5'-triphosphate (Fig. 1, 1). The pathway for *de novo* biosynthesis of purine nucleotides consists of 14 enzyme-catalyzed reactions from the precursors, P-Rib-PP, glutamine, ATP, glycine, N¹⁰-formyl tetrahydrofolate and aspartate (Fig. 2, 2).

Characteristics

De Novo Pyrimidine Nucleotide Biosynthesis

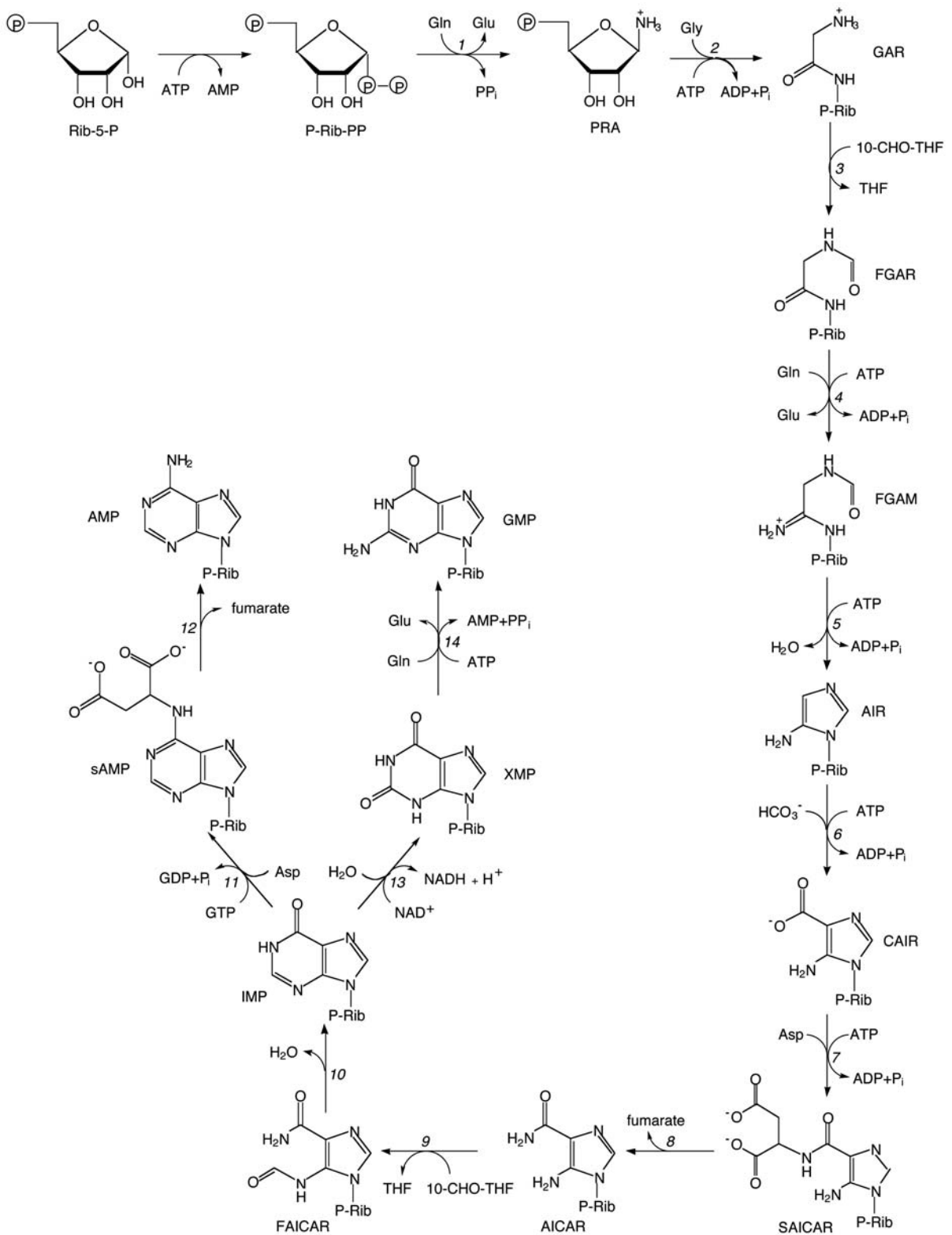
The pathway is shown in Fig. 1 with the reactions numbered in sequence. All of the reactions occur in the



Nucleotide Biosynthesis. Figure 1 The pathway for *de novo* biosynthesis of pyrimidine nucleotides.

►cytoplasm of human cells, except for dihydroorotate dehydrogenase (reaction 4), which is located on the outer side of the inner ►mitochondrial membrane (1). In mammals, the enzymes catalyzing the first three reactions of the pathway, carbamyl phosphatase, aspartate transcarbamylase and dihydroorotase,

are contained in a trifunctional protein called CAD (or DHO synthetase). The dihydroorotate produced must diffuse across the outer mitochondrial membrane to be oxidized by dihydroorotate dehydrogenase, which is coupled directly to the electron transport chain in mitochondria via ►coenzyme Q. The orotate formed



Nucleotide Biosynthesis. Figure 2 The pathway for *de novo* biosynthesis of purine nucleotides.

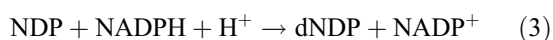
then diffuses back into the cytoplasm where it is further converted by UMP synthase. The enzymic activities catalyzing reactions 5 and 6, orotate phosphoribosyltransferase and OMP decarboxylase, are found in mammals as a bifunctional protein called UMP synthase. The uridine 5'-monophosphate (UMP) produced by reaction 6 is sequentially phosphorylated, $\text{UMP} \rightarrow \text{UDP} \rightarrow \text{UTP}$, the UTP is then converted to CTP by the enzyme, CTP synthetase. In mammals, the pyrimidine pathway is regulated at the first reaction, catalyzed by carbamyl phosphate synthetase II, by UTP, which is a feedback inhibitor and **►P-Rib-PP**, which is an activator. Thus when P-Rib-PP is available, pyrimidine biosynthesis proceeds (1).

De Novo Purine Nucleotide Biosynthesis

The pathway is shown in Fig. 2 with the reactions numbered in sequence. All of the reactions occur in the cytoplasm of human cells. The first reaction of the pathway is catalyzed by the enzyme, amido phosphoribosyltransferase, which is a very unstable enzyme with an iron-sulfur center. The unbranched portion of the pathway consists of 10 reactions yielding IMP; there is then a branch to AMP and another branch to GMP (Fig. 2). GMP and AMP are then phosphorylated ($\text{NMP} \rightarrow \text{NDP} \rightarrow \text{NTP}$). In mammals, purine biosynthesis is controlled by feedback inhibition of amido phosphoribosyltransferase by the end products, AMP, GMP and IMP (Fig. 2). This enzyme is also subject to inhibition by dihydrofolate (DHF) polyglutamates, which may regulate the pathway according to the state of reduction of metabolic cofactors (3). The enzymes catalyzing reactions 2, 3 and 5 of the pathway are contained in a trifunctional protein; those catalyzing reactions 6 and 7 and reactions 9 and 10 are contained in bifunctional proteins. There have been proposals that the intermediates of this pathway are "channeled", that is preferentially utilized by the next enzyme of the pathway without diffusing into the bulk solvent. There is some evidence for association of the **►multifunctional enzymes** of this pathway into a "**►pathway particle**" which could mediate substrate channeling and provide coordinate control of enzyme activity. However, more evidence is required to support these proposals.

Deoxynucleotide Biosynthesis

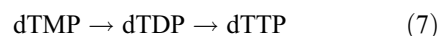
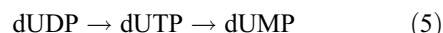
The deoxynucleoside 5'-triphosphates, dCTP, dTTP, dATP and dGTP, are formed in a series of reactions starting with a nucleoside 5'-diphosphate precursor (NDP).



For dCTP, dATP and dGTP, there is then a simple phosphorylation



For dTTP, the sequence of reactions is more complex



The 4 **►dNTPs** are then polymerized into a growing strand of DNA where the nucleotide (NMP) sequence is copied from an existing strand using the enzyme, **►DNA polymerase**.

Clinical Relevance

The *de novo* and salvage pathways described above for biosynthesis of pyrimidine and purine nucleotides may be considered universal, being found in organisms ranging from bacteria to humans. For cell division to occur, there must be two copies of each chromosome, for humans 92 chromosomes rather than the diploid number of 46. These chromosomes consist of DNA containing the genetic code or "blueprint for life". As stated above, DNA is synthesized from the 4 dNTPs and therefore active nucleotide biosynthesis is required for DNA synthesis to occur. Thus drugs for the treatment of cancer, bacterial or parasitic infections could selectively block nucleotide biosynthesis in the invading cells.

Cancer

There are a number of drugs that block pathways for nucleotide biosynthesis (4), the best known is probably methotrexate (MTX). MTX is a so-called antifolate drug, which inhibits the enzyme, **►dihydrofolate reductase**. The consequence of this inhibition is the accumulation of dihydrofolate polyglutamates in cells, which then block the pathway for *de novo* biosynthesis of purines (Fig. 2) and the synthesis of dTMP (Eq. 6). The result is major decreases in the cellular concentrations of dATP, dGTP and dTTP, while dCTP is unchanged. This imbalance in the pools of dNTPs results in genetic miscoding during replication of DNA and cell death. The basis of the selective toxicity of MTX on cancer cells is due to the more rapid growth of some cancers compared with normal cells of the body or to the slow but continuous growth of some cancers. It should be clear from this explanation that such cytotoxic drugs have side effects on the patient.

Bacterial and Parasitic Infections

While the differences between cancer and normal cells are subtle, there is considerable variation between the pathways and enzymes of pathogenic organisms and those of human cells. Thus, there is more scope for drugs with selective toxicity (4). For example, there are antifolate drugs that are more toxic to the pathogen than human cells and sulfa drugs block the synthesis of

folate in some bacteria, while humans must obtain it as a vitamin in the diet. The malarial parasite, *Plasmodium falciparum*, has the ►*de novo* pathway for biosynthesis of pyrimidine nucleotides (Fig. 1) but is unable to salvage pyrimidine nucleosides (Eq. 2). Conversely, the parasite lacks the *de novo* pathway for purine nucleotides (Fig. 2), but salvages the purine nucleosides, adenosine and guanosine (substitute these nucleosides into Eq. 2). This difference in pathways between malaria and the human host can be exploited in chemotherapy. Atovaquone is an antimalarial drug that blocks the electron transport chain, thus inhibiting dihydroorotate dehydrogenase (reaction 4) of the pyrimidine pathway (Fig. 1). The parasite cannot synthesize pyrimidine nucleotides required for DNA synthesis and cell division, while the patient can synthesize pyrimidine nucleotides *via* the alternative salvage pathway. Thus the malarial parasite dies and the patient survives. The basis for selective toxicity here is very strong, involving differences in pathways. Other enzymes of the pyrimidine pathway in *P. falciparum* are also potential targets for development of antimalarial drugs.

References

1. Jones ME (1980) Pyrimidine nucleotide biosynthesis in animals: genes, enzymes, and regulation of UMP biosynthesis. *Annu Rev Biochem* 49:253–279
2. Zalkin H, Dixon JE (1992) *De novo* purine nucleotide biosynthesis. *Prog Nucleic Acid Res Mol Biol* 42:259–287
3. Sant ME, Lyons SD, Phillips L et al (1992) Antifolates induce inhibition of amido phosphoribosyltransferase in leukemia cells. *J Biol Chem* 267:11038–11045
4. Christopherson RI, Lyons SD, Wilson PK (2002) Inhibitors of *de novo* nucleotide biosynthesis as drugs. *Acc Chem Res* 35:961–971

Nucleotide Excision Repair

JAMES E. CLEAVER

Auerback Melanoma Laboratory, UCSF Cancer Center, University of California, San Francisco, CA, USA
jcleaver@cc.ucsf.edu

Definitions

DNA repair encompasses a large and important set of cellular processes that maintain the integrity of the genome under onslaught from external and internal radiations, chemicals and oxidation. Every sequenced genome contains an extensive series of genes that

encode proteins for detection, modification, removal, replacement and bypass of damaged and non-informative DNA bases (1). These function in a number of different DNA repair mechanisms, each with a range of substrates and distinctive mechanisms. The mechanism by which large DNA adducts especially those produced by ultraviolet (UV) light are repaired is known as ►*nucleotide excision repair* (NER) (Fig. 1) (2).

Transcribed genes and their transcribed strands are more rapidly repaired by NER than other regions of DNA (3). A specific branch of NER has therefore been named “►*transcription coupled repair* (TCR)” to distinguish it from repair of the remainder of the genome, called “►*global genome repair*” (GGR). Many components of the basal transcription factor IIIH (TFIIH) are also involved in NER, contributing to its linkage to transcription.

Characteristics

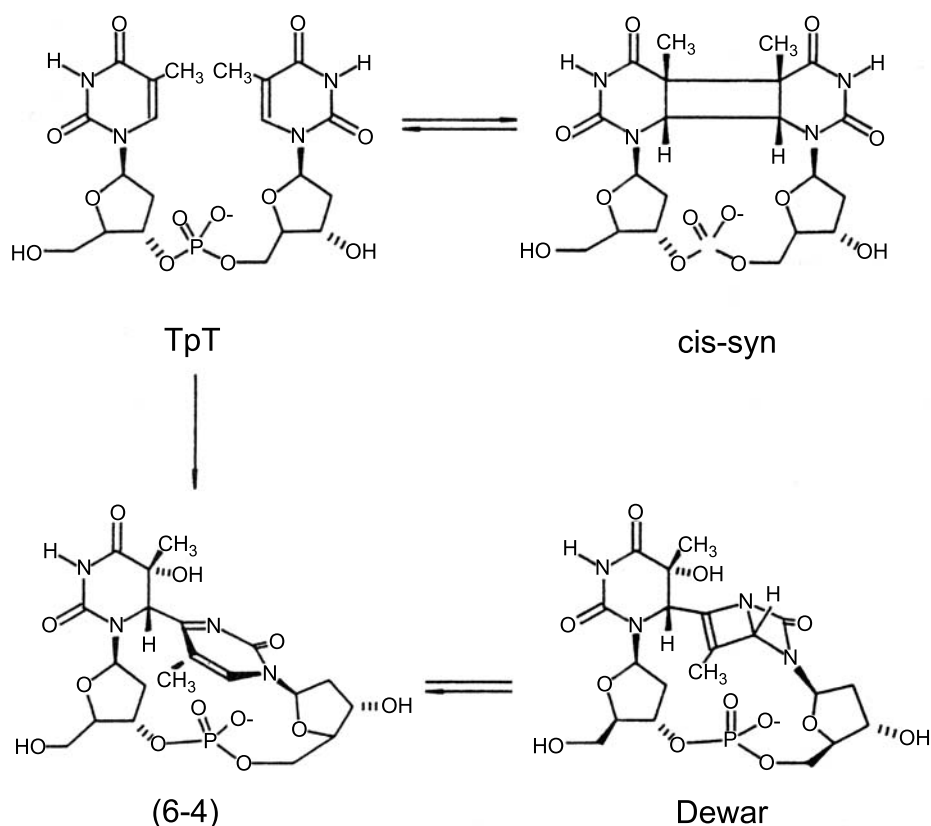
Types of Damage Repaired by NER

The NER system recognizes and repairs DNA damage that consists of photoproducts produced by UV light and large DNA adducts produced by carcinogenic chemicals (Fig. 2). The most important wavelengths of UV are those in the UVC (240–280) and the UVB (280–320 nm) ranges that are strongly absorbed by nucleic acids. The photoproducts include the cyclobutane [5–5], [6–6] pyrimidine dimers (CPDs) and the [6–4] pyrimidine pyrimidinone dimers ([6–4]PDs) that involve both T and C pyrimidines (Fig. 1). The [6–4]PD can further photoisomerize to the Dewar photoproduct at longer UV wavelengths. Cytosine in dimers has an increased rate of deamination leading to cytosine to thymine (C to T) mutagenesis. Chemical adducts include those produced by N-acetoxy-N-acetyl amino-fluorene (AAAF), benzo(a)pyrene, aflatoxin, photo-activated psoralens, and cis-platinum. An oxidative purine product, 5',8-purine cyclodeoxynucleoside, that may accumulate in neurological tissue also requires the NER system for repair. The NER system can even recognize DNA triplexes formed by the binding of short oligonucleotides to double stranded DNA.

Other kinds of damage involving smaller modifications to DNA bases (alkylation) or DNA breakage require a different suite of enzymes, many of which are involved with immunoglobulin rearrangements and neurodegeneration. The distinctions between the various repair systems are not absolute, however, and there are overlaps in the substrate specificity of these various repair systems and variations in the efficiencies and sites of action on DNA in differing metabolic states.

Recognition of Damage in Non-Transcribing Regions of the Genome

The initial damage recognition factors uniquely required for GGR are the XPC and XPE DNA binding



Nucleotide Excision Repair. Figure 1 The sequence of steps involved in nucleotide excision repair from damage recognition by GGR or TCR mechanisms (I), DNA unwinding (II), verification (III), excision (III, IV), polymerization and ligation (V). (reproduced from Hoeijmakers JH (2001) *Genome maintenance mechanisms for preventing cancer*. Nature 411:366-374, with permission from Macmillan Publishers)

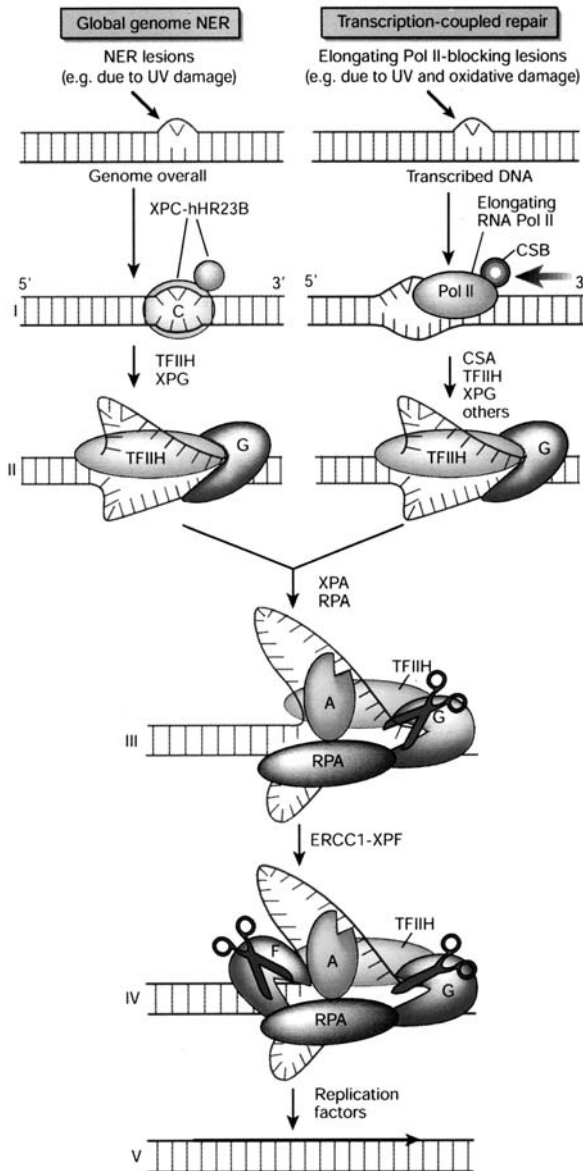
proteins (Table 1) (4). XPC is found in association with a cofactor HR23B, one of the two human homologs of the yeast Rad23 gene product. The XPC/HR23B complex is the earliest damage detector that initiates NER in nontranscribed DNA with highest affinity for the [6-4]PD. This complex stabilizes subsequent binding of the XPA-RPA proteins to the damaged site. In cells that lack the XPC/HR23B complex, repair is confined to small transcriptionally active islands of the genome, where repair occurs at normal rates, surrounded by oceans of unrepaired transcriptionally silent DNA. The constitutive level of expression of XPC is controlled by p53, and the protein can be induced further by UV irradiation. If a damaged site already contains a significant degree of unpaired bases the requirement for XPC/HR23B is relaxed.

The XPE protein is a ▶heterodimer having subunits of 127 kD (DNA Damage Binding Protein, DDB1) and 48 kD (DDB2) and mutations that are diagnostic for XP-E patients are located in the DDB2 subunit. The DDB1/2 heterodimer is involved in the recognition of damaged DNA but reconstitution of excision repair

using purified proteins and synthetic substrates lacking chromatin structure shows no requirement for DDB1/2 for damage removal and repair synthesis. The XPE protein is mainly involved in repair of mutagenic UV-induced photoproducts.

DDB1 (p127) is present in excess over its partner, and is predominantly cytoplasmic but translocates to the nucleus after UV irradiation. Mutations in DDB2 also prevent the accumulation of DDB1 in the nucleus. UV-induced expression of DDB2 involves p53 transactivation in human, but not rodent, cells. The consensus DNA-binding site for p53 is present in the human p48 promoter region but has mutated in the mouse homolog. These observations provide a partial explanation for early observations that excision repair can be low in some mouse cell strains.

By *in vitro* reconstruction using recombinant proteins, the complete excision process on metabolically inactive (nontranscribed) DNA can be carried out with a minimal set of components, replication protein A (RPA), XPA, TFIIH (6 sub-unit core of XPB, XPD, p44, p34, p52, p62 and 3 sub-unit kinase, CAK), XPC,



Nucleotide Excision Repair. Figure 2 Photochemical reactions in a dipyrimidine DNA sequence (top left, TpT) leading to the formation of CPDs (top right, cis-syn) or a [6-4]PD (bottom left) and its photolytic derivative, the Dewar pyrimidine (bottom right). (redrawn from Taylor, J. S., Cohrs, M. P. (1987). DNA, light. and Dewar pyrimidines: the structure and significance of TpT3. *J. Amer. Chem. Soc.* 109, 2834-2835, and reproduced from Ultraviolet radiation carcinogenesis, in *Cancer Medicine*, 6th edition, volume 1 edited by Kufe et al. 2003, with permission from B C Dekker, Hamilton, Canada)

HR23B, XPG, ERCC1 and XPF. CAK appears to act as a negative regulator and inhibition of phosphorylation stimulates repair. These individual factors of NER associate sequentially and independently on UV

photoproducts *in vivo*, without preassembly of a "repairosome" complex.

Recognition of Damage in Transcriptionally Active Regions of the Genome

CPDs are excised more rapidly from actively transcribed genes, especially from the DNA strand used as the template for transcription. Excision can be slow where binding proteins interact with the promoter but increases immediately after the ATG start site for transcription. The initial damage recognition mechanism for TCR may be the stalled RNA polymerase (Pol II) itself, although a potentially large number of proteins play a role in mediating the increased rate of repair in transcriptionally active genes. The blocked RNA Pol II masks damaged sites and must be removed, or backed off, to alleviate the arrested transcription and provide access for repair enzymes. The processes that alleviate the arrested RNA Pol II enhance the overall accessibility of the damaged site and permit more rapid repair than in non-transcribed regions of the genome. Persistently blocked RNA Pol II can be a signal for UV-induced apoptosis.

The Core Excision Process Common to GGR and TCR

DNA Unwinding

The damaged site is unwound by the 3'-5' (XPB) and 5'-3' (XPD) helicases of TFIID in concert with XPG, and then stabilized by the XPA-RPA binding complex. The basal transcription factor, TFIID that is involved in transcription initiation and elongation plays a major role in remodeling the damaged regions for excision to occur after initiation of damage recognition by TCR and GGR. The XPB and D components of TFIID and CSA interact with p53, suggesting a role for p53 in regulation of TCR. But this role is not as prominent as regulation of GGR by p53 in the overall repair reaction.

Damage Verification

After damage recognition, a process of damage verification occurs in which the initial proteins are replaced by another damage binding complex, the heterodimer of XPA/RPA. XPA binds to both the damaged and the undamaged strand, but RPA binds mainly to the undamaged strand. XPA acts as a foundation on which many of the other components of the NER process subsequently assemble. XPA binds to the p34 subunit of RPA through its N-terminal region and to the p70 subunit through a central portion in exon IV, as well as to many other NER components including ERCC1, TFIID, XPC, XPG, and DNA itself. The final stable complex that remains at the damaged site is unclear, but a subset of these components acts as an assembly point for the XPG and ERCC1/XPF nucleases that cut the damaged strand 3' and 5' to the damaged site.

Nucleotide Excision Repair. Table 1 The main genes and functions of the NER system

Gene	Chrom location ¹	Protein	Function	Partner(s)
XPA	9q34.1	273aa	DNA damage binding	RPA
XPB	2q21	782aa	3'-5' helicase	TFIIH
XPC	3p25.1	940aa	DNA damage binding	HR23B
XPD	19q13.2	760aa	5'-3' helicase	TFIIH
XPE	11p11-12	427aa	DNA damage binding (DDB2)	DBB1(p127)
XPF	16p13.3	916aa	Endonuclease 5' to damage	ERCC1
XPG	13q32-33	1186aa	Endonuclease 3' to damage	Thymine glycosylase (nth)
XPV	6p21	713aa	Low fidelity DNA polymerase (Pol η)	PCNA
CSA	5q11.2	396aa	WD protein, ubiquitination of RNA pol II	
CSB	10q23	1493aa	DNA dependent ATPase, ubiquitination of RNA pol II	

¹. For further details on chromosome locations and sequence see ► www.ensembl.org and for mutations in the genes see ► www.xpmutations.org.

Excision of the Damaged Strand

The excision process involves removal of an oligonucleotide, 27–29 nucleotides long, containing the photoproduct as a result of cleavages 5 nucleotides on the 3' side of the photoproduct, and 24 nucleotides on the 5' side. The structure-specific cleavage pattern is determined by binding of RPA to the unwound damaged site, and the excised fragment is close in size to the footprint of RPA on DNA. Slight variations in the precise sites of cleavage result in the removal of variable fragments between 27 and 29 nucleotides. The XPG nuclease cleaves first on the 3' side of the damage and interacts with the XPC-HR23B complex and with TFIIH; the XPF-ERCC1 heterodimer cleaves subsequently on the 5' side of the damaged site. XPA serves as an anchor for the 5' nuclease through binding to ERCC1.

Synthesis of the Repair Patch

The excised region is replaced by the action of a complex similar to that involved in normal DNA replication. Proliferating cell nuclear antigen (PCNA) is loaded onto the DNA by the 5 subunit replication factor C (RFC) complex which then anchors the replicative DNA polymerases, Pol δ or Pol ϵ . The final closure of the repaired site occurs with DNA ligase I.

Replication of Damaged DNA

NER can remove DNA damage before DNA replication begins, and consequently plays a major role in reducing the amount of damage that becomes fixed as mutations during replication. DNA photoproducts,

however, are blocks to the replicative DNA polymerases, alpha, delta and epsilon (Pol α , δ , ϵ) that cannot accommodate large distortions such as DNA photoproducts or adducts in their active sites. Replicative bypass of these photoproducts is achieved instead by damage-specific polymerases with relaxed substrate specificity, known as class Y polymerases (5). These polymerases have larger active sites that allow them to read-through noninformative sequence information resulting from DNA damage. Consequently these polymerases have high error rates of the order of 1% when assayed *in vitro*, and this property must be controlled *in vivo* otherwise the results would be catastrophic to the cell.

Three class Y polymerase genes have been identified in the mammalian genome, Pol η , ι and κ . Pol η and ι are close homologs, unique to mammalian cells, only a single Pol η gene is found in yeast. Mutations in Pol η are found in one group of XP (Xeroderma pigmentosum) patients (XP variants) with clinical symptoms generally similar to those seen in the NER-defective patients. Pol η has the highest efficiency for replication of UV damage, followed by Pol ι with Pol κ having the poorest. Their relative efficiencies may differ for other kinds of chemical adducts. Pol η and Pol ι exhibit different base-specificities when replicating mis-coding lesions. Pol η preferentially inserts adenines, thus accurately replicating T-containing photoproducts. Pol ι preferentially inserts guanines, thus accurately replicating C-containing photoproducts. Pol η can therefore accurately replicate a T-T CPD. Pol ι can accurately replicate a T-C CPD or [6-4]PD. Pol η acts

distributively to extend nascent DNA chains by one or two bases across from CPDs, and the inserted bases may be edited by a separate exonuclease. Chain extension then requires the activity of Pol ζ or Pol κ that can extend the nascent chain from a terminal mis-paired base and then restore normal DNA replication. Pol η is uniformly distributed in the nucleus, and excluded from the replication fork until replication is stalled by UV damage. Pol η then traffics into the nucleus and accumulates with a large number of other proteins in microscopically visible foci at the replication fork. This requires specific sequence motifs at the C-terminal end of the protein for translocation and for binding to PCNA.

Clinical Relevance

► **Xeroderma pigmentosum** (XP) is a human, autosomally inherited, skin and neurodegenerative disease in which exposure to normal sunlight can result in a very high incidence of all kinds of skin cancer (6). XP is predominantly due to a failure of NER in cells that contain UV-induced photoproducts in their DNA. When these photoproducts are unrepaired, due to NER deficits, subsequent replication errors and cytosine deamination lead to characteristic C to T mutations, especially CC to TT mutations, that are found in the *p53* gene in sun- and UVB-induced skin cancers.

Mutations in NER genes are found in XP and in other cancer, developmental or neurodegenerative diseases such as Cockayne syndrome (CS), trichothiodystrophy (TTD), and Cerebro-Oculo-Facio-Skeletal (COFS) syndrome and have been summarized in the web-site ► www.xpmutations.org and in Table 1. Patients with mutations in those genes predominantly responsible for global genome repair, *XPC*, *XPE*, *XPF*, or replicative bypass, *XPV*, are at high risk for solar-induced skin cancer, but have few other complications. Patients with mutations in genes linked to transcription, *XPA*, *XPB*, *XPD*, *XPG* and the *CSA* and *CSB* genes, show high risks for neurodegenerative disease associated with diagnoses of CS or TTD.

Cells from CS patients are specifically defective in TCR. The excision of DNA photoproducts from total genomic DNA is normal in CS cells, but repair of transcriptionally active genes is reduced. Two CS genes, *CSA* & *CSB*, are involved specifically in TCR. *CSA* contains WD-repeat motifs that are important for protein-protein interactions. *CSB* contains an ATPase, helicase motifs and a nucleotide-binding domain, but only the last is essential for TCR and *CSB* does not function as a helicase. Cells lacking either *CSA* or *CSB* are unable to ► **ubiquitinate** the C-terminal domain (CTD) of RNA Pol II.

The XPG nuclease has a particularly complex range of activities, and mutations in XPG are associated with

both XP and CS diseases. XPG is an endonuclease in the FEN-1 family that is capable of strand-specific cleavage of a range of branched DNA substrates that may arise during DNA replication, repair and recombination. The XPG gene product also interacts with RNA pol II and facilitates efficient transcription elongation, thereby providing an explanation for the complex symptoms of some XPG patients who show both XP and CS symptoms. XPG is also a cofactor required for the activity of thymine glycosylase (*nth* gene) and is thereby linked to repair of oxidative damage that may also be important in maintaining neural functions that fail in CS.

Cancer risks are low or negligible in CS and TTD patients however. CS patients are photosensitive but do not show increased skin carcinogenesis, unlike mice with inactivated *Csa* or *Csb* genes that show increased skin carcinogenesis from UV light or chemical carcinogens (7). Mice however, also have inactive *XPE* genes, due to mutations in the *p53* response elements in the promoter region of the gene. Consequently the increased cancer rates in Cockayne mice might be due to their being double mutants in *Cs* and *Xpe*.

► **Polymorphisms** that are not directly associated with the high penetrance XP disease phenotype have been described for many of the NER and other repair genes. Polymorphisms have been found in the *XPA*, *XPC* and *D* genes, and others are listed in ► <http://greengenes.lbl.gov/dpublic/secure/reseq/>. Many of these polymorphisms result in non-conservative amino acid changes that could have an impact on NER function and disease. A large number of alleles present at low frequency individually could combine to create variation in disease frequency in the overall population.

► DNA Repair Mechanisms

References

1. Wood RD, Mitchell M, Sgouros J et al (2001) Human DNA repair genes. *Science* 291:1284–1289
2. Sancar A (1996) DNA excision repair. *Annu Rev Biochem* 65:43–81
3. Bohr VA (1991) Gene specific DNA repair. *Carcinogen* 12:1983–1992
4. Batty DP, Wood RW (2000) Damage recognition in nucleotide excision repair of DNA. *Gene* 241:193–204
5. Trincão J, Johnson RE, Escalante CR et al (2001) Structure of the catalytic core of *S. cerevisiae* DNA polymerase η : implications for translesion synthesis. *Mol Cell* 8:417–426
6. Kraemer KH, Lee MM, Scotto J (1987) Xeroderma pigmentosum: cutaneous, ocular and neurological abnormalities in 830 published cases. *Arch Dermatol* 123:241–250
7. Berg RJ, Rebel H, Horst GT van der et al (2000) Impact of global genome repair versus transcription-coupled repair on ultraviolet carcinogenesis in hairless mice. *Cancer Res* 60:2858–2863

Nucleotide Microarrays

- ▶ Automated High-Throughput Functional Characterization of Human Proteins
- ▶ DNA Chips
- ▶ Microarray Technology

Nucleus

Definition

Nucleus designates a membrane-enclosed cellular compartment of eukaryotic cells that contains chromosomes.

- ▶ Cap-Independent Translational Control
- ▶ Nuclear Compartments

Null Allele/Null Phenotype

Definition

Null allele denotes a mutation in a gene that results in total loss of the normal function of the gene product encoded by that gene, resulting in a Null Phenotype.

- ▶ Large-Scale Homologous Recombination Approaches in Mice
- ▶ Mouse Genomics

Null Mice

- ▶ Knock-Out Mice

Null Phenotype

- ▶ Null Allele/Null Phenotype

NXF1

Definition

NXF1 is an export receptor for mRNA; via molecular adapters, mRNA interacts with the export receptor (NXF1).

- ▶ Nuclear Import and Export

06-Alkylguanine-DNA Alkyltransferases

- ▶ Alkyltransferases

- ▶ Leptin
- ▶ Prader Willi and Angelman Syndromes

06-Methylguanine-DNA Methyltransferase

- ▶ Alkyltransferases

Oculopharyngeal Muscular Dystrophy

Definition

Oculopharyngeal Muscular Dystrophy is a heritable disease caused by mutations in the gene encoding the nuclear poly(A) binding protein (PABPN1). The mutations cause moderate expansions of a naturally occurring polyalanine tract at the N-terminus of PABPN1. The mutant protein forms filamentous aggregates in the nuclei of muscle cells.

- ▶ Polyadenylation
- ▶ Repeat Expansion Diseases

Obesity

Definition

Obesity is a condition with a markedly increased amount of body fat or adipose tissue in relation to the lean body mass, caused by enhanced uptake of calories or reduced calory expenditure, and/or genetic factors. In the recent years, the body mass index (BMI) has become the medical standard used to measure overweight and obesity. Obesity is an important risk factor for a number of diseases like diabetes mellitus type 2, high blood pressure and has been implicated in the high prevalence of atherosclerosis. Obesity in children and adolescents is a serious issue with many health and social consequences that often continue into adulthood.

- ▶ Adipogenesis
- ▶ Adiponectin

ODD

Definition

ODD (oxygen dependent degradation) domain is a polypeptide segment within HIF- α subunits, which confers hypoxic protein stabilization.

- ▶ Hypoxia Inducible Factors

OFC

- ▶ Orofacial Clefts
- ▶ Cleft Lip Palate

OFUT-1

Definition

OFUT-1 stands for O-fucosyltransferase. It is an enzyme that initiates the addition of a sugar chain with fucose. This type of glycosylation is a rare modification of a protein.

► [Notch Pathway](#)

Okazaki Fragment

Definition

Okazaki fragment designates short DNA pieces at the lagging-strand of the replication fork, which are formed discontinuously during DNA replication, because all known DNA polymerases have a 5' – 3' directionality in synthesizing DNA. They are rapidly joined by DNA ligase to form a continuous DNA strand.

► [Replication Fork](#)

Oligo Microarray

Definition

Oligo microarray refers to a cDNA library in which each clone is represented by an oligo-nucleotide (30–70–mer). These oligo-nucleotides are arranged on a glass slide in a defined pattern, by which each oligo-nucleotide is defined by its position on the slide.

► [Medaka as a Model Organism for Functional Genomics](#)

► [Microarray Technology](#)

Oligoarticular

Definition

Oligoarticular describes a state when five or fewer joints are affected, e.g. in arthritis.

► [Rheumatoid Arthritis](#)

Oligodendrocyte

Definition

Oligodendrocyte is a glial cell which enwraps axons with a lipid-rich myelin sheath providing insulation to neurons in the central nervous system. One oligodendrocyte typically myelinates several axon processes.

► [Glial Cells and Myelination](#)

► [In Vivo Imaging of Transgenic Mice with Fluorescent Protein Expression](#)

Oligogenic

Definition

Oligogenic designates a mode of disease transmission that is determined by a small number of genes, each with a moderate effect.

► [Schizophrenia Genetics](#)

Oligomerization

Definition

Oligomerization defines a state in which a protein associates with itself to form high-order forms, such as dimers or multimers, in order to carry out its functions.

► [DNA Helicases](#)

Oligonucleotides

Definition

Oligonucleotides are short sequences of single-stranded DNA or RNA. Oligonucleotides are often used as probes for detecting complementary DNA or RNA because they bind readily to their complements.

► [DNA Chips](#)

Oligosaccharide

Definition

Oligosaccharide (polysaccharide) consists of a linear or branched chain of monosaccharides attached to one

another via glycosidic linkage. The number of monosaccharide units can vary. The term polysaccharide is usually reserved for large glycans with repeating units.

► [Glycosylation of Proteins](#)

Oligosaccharyltransferase

Definition

Oligosaccharyltransferase (OST) refers to an enzyme complex in the endoplasmic reticulum, which transfers Glc₃Man₉GlcNAc₂ to Asn residues from the donor Glc₃Man₉GlcNAc₂-P-P-(pyrophosphoryl)dolichol.

► [Glycosylation of Proteins](#)

OMIM (TM)

► [Online Mendelian Inheritance in Man](#)

Oncogen Activation

► [Insertional \(Oncogen\) Activation](#)

► [Oncogene](#)

Oncogene

MANFRED SCHWAB
DKFZ, Heidelberg, Germany
m.schwab@dkfz.de

Definition

An oncogene is a derivative of any gene that has the ability to stimulate cellular growth. In experimental assays, oncogene products can, alone or in cooperation with another gene, transform eukaryotic cells so that they grow in a way analogous to tumor cells. The definition was originally applied to the transforming genes acquired by RNA tumor viruses through the

transduction of cellular genes. Today, the term is used rather broadly. Oncogenes contribute to tumorigenesis by any positive modulation of cellular growth; they act by their presence (this in contrast to tumor suppressor genes), an activity that is often referred to as “dominant”. Tumorigenic activation of oncogenes can result from mutational/structural/numeric changes in a gene and possibly from regulatory enhancement of gene expression.

Characteristics

Oncogenes were originally isolated from RNA tumor viruses, where they are responsible for the rapid tumor induction after infection of an animal host. In the viral genome, the oncogene was referred to as a viral oncogene or *v-onc* (1, 2).

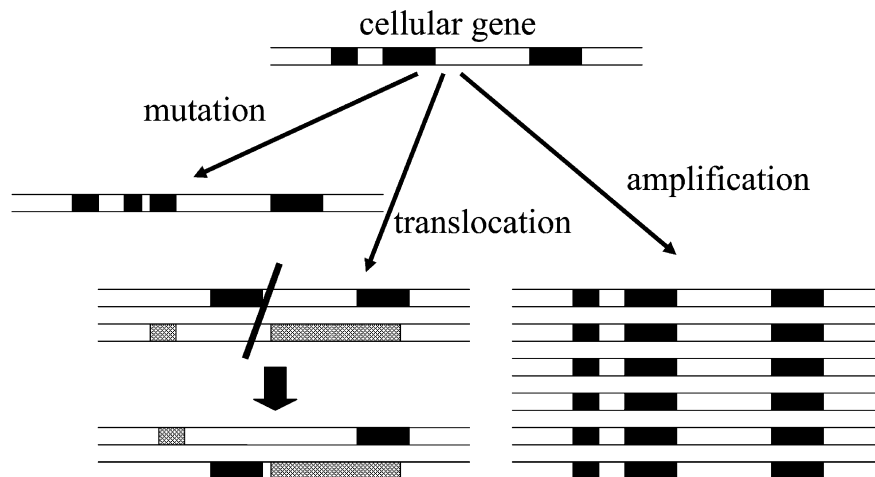
It was soon established that the *v-oncs* are actually derived from the genome of the host cell. They have been captured by the virus after infection of the cell by a process called transduction. Transduction appears in a range of animal species from chickens to monkeys; it has not been observed in humans. The cellular counterparts, from which the *v-oncs* are derived, are referred to as proto-oncogenes, or cellular oncogenes (*c-onc*). Proto-oncogenes are normal constituents of the cellular genome and are highly conserved among all eukaryotic organisms.

This original rigid definition has softened in subsequent years. Broadly speaking, the term oncogene now includes any gene that has a growth stimulatory effect on cells, by means of:

- conferring sustained cellular multiplication
- advancement of cell-cycle progression
- decreased requirement for growth factors
- focus formation under conditions of cell culture
- enabling cells to grow under more restricted experimental conditions, such as in soft agar
- tumorigenic conversion, such as in experimental animals
- conversion of cells to form tumors that undergo metastasis
- escape from apoptosis

The precursors of oncogenes (► [proto-oncogenes](#)) are present with their normal structures and expression activities in at least all higher animal cells. They represent what has been referred to as ‘enemies within’, but perform normal, usually vital, functions. Their oncogenic potential can be activated by any one of the following genetic changes (Fig. 1):

- Point mutation, resulting in an exchange of a base with the consequence of an amino acid change. The first example was the *RAS*^H oncogene from a human gastric cancer that can convert *in vitro* established cells to tumorigenicity.

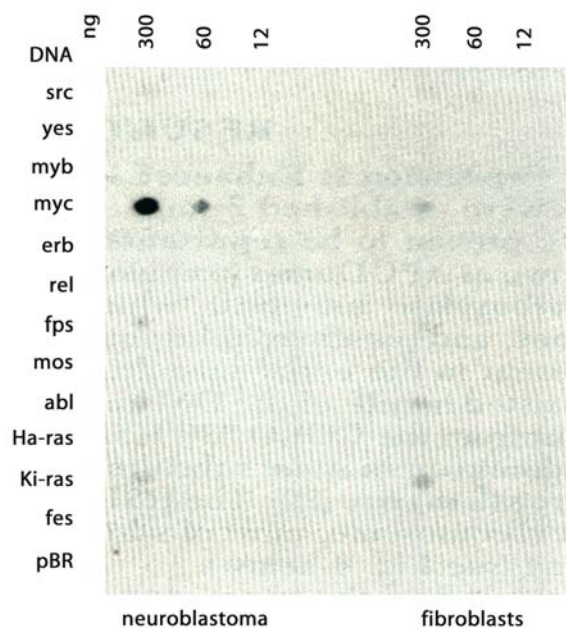


Oncogene. Figure 1 Molecular pathways for activating the oncogenic potential of cellular oncogenes. Translocation is the exchange of genetic material between two non-homologous chromosomes.

- Translocation, in which DNA from two genes on two different chromosomes can recombine, resulting in a fusion gene and a fusion protein (e.g. *BCR/ABL* [*BCR-ABL1*]) in chronic myelogenous leukemia or deregulated gene expression, when normal gene expression signals are replaced by other DNA sequences, for example the juxtaposition of *MYC* and active *Ig* promoters in different types of lymphomas.
- Amplification, where the number of gene copies multiplies, with consequent enhanced gene expression; prototypic for this is the *MYCN* gene in neuroblastomas, which has emerged as the first molecular marker for patient prognosis and is a paradigm for the clinical usefulness of an oncogene alteration (3, 4).

Additionally, activation may be achieved through regulatory enhancement of gene expression, although such a type of activity is more difficult to establish, due to the lack of suitable counterpart cells for a faithful comparison of expression signatures.

Historically, the activity of point mutations was discovered through transfection assays, where total DNA from cancer cells was introduced into non-tumorigenic tester cells, either of murine or human origin. Cells that have incorporated the mutationally activated oncogene have acquired the ability to form tumors when transplanted into an animal host. Non-random translocations had been known for many years, particularly in lymphoblastoid cells. The availability of cloned retroviral oncogenes has allowed rapid determination of the molecular identity of the genetic information altered through the translocation event. The significance of oncogene amplification was



Oncogene. Figure 2 Early low-tech oncogene array for expression profiling of tumor cells. Retroviral oncogenes were spotted as a template onto a filter membrane (Ref.5; note that the cellular homologs had not been isolated at that time), poly-adenylated RNA was extracted from tumor cells, reverse transcribed into cDNA, which then was radioactively labeled and hybridized under conditions of reduced stringency to allow for interspecies cross-hybridization. These principles of array technology involving, as a hallmark, the reverse hybridization of a complex cDNA probe to a gene array, were first described in Ref.5, they are the same as those used in today's high volume arrays with thousands of genes (Ref.6).

established by the first application of array technology, using what in today's terminology is referred to as an "oncochip" (Fig. 2). Oncogene expression profiling of human and animal tumor cells carrying conspicuous cytogenetic manifestations of amplified DNA, "double minutes" (DMs) or "homogeneously staining chromosomal regions" (HSR) revealed the enhanced expression of an individual oncogene, such as *RAS*^K in a murine cell line or, more significantly, *MYCN* in human neuroblastoma. Subsequent analysis of the gene copy number showed that the enhanced expression is the consequence of the amplification (up to 500 additional gene copies) of the otherwise unchanged gene.

Oncogene Cooperation

Experimental approaches have shown that the expression of a single activated oncogene is insufficient to achieve full tumorigenic conversion of a normal cell. Only when at least two altered oncogenes or alternatively, a single altered oncogene under the control of a strong heterologous promoter are introduced can a normal cell assume a tumorigenic phenotype. Oncogene cooperation is well in line with the multiple genetic changes that a tumor acquires during its evolution to metastatic disease (multistep development). In the development of naturally developing tumors, oncogenes can also cooperate with tumor suppressor genes during cellular evolution towards the malignant phenotype.

- Breast Cancer
- Cap-Independent Translational Control
- Colorectal Cancer
- Ras Signalling
- Hedgehog Signalling
- Translational Control in Eukaryotes

References

- 1) Bishop JM (1987) The molecular genetics of cancer. *Science* 235:305–311
- 2) Bishop JM (1982) Retroviruses and cancer genes. *Adv Cancer Res* 37:1–32
- 3) Schwab M, Westermann F, Hero B et al (2003) Neuroblastoma: biology and molecular and chromosomal pathology. *Lancet Oncol* 4:472–480
- 4) Savelyeva L, Schwab M (2001) Amplification of oncogenes revisited: From expression profiling to clinical application. *Cancer Lett* 167:115–123
- 5) Schwab M, Alitalo K, Varmus HE et al (1983) A cellular oncogene (c-Ki-ras) is amplified, overexpressed, and located within karyotypic abnormalities in mouse adrenocortical tumour cells. *Nature* 303:497–501
- 6) Khan J, Wei JS, Ringnér M et al (2001) Classification and diagnostic prediction of cancers using gene expression profiling and artificial neural networks. *Nat Med* 7:673–679

Oncogenesis

Definition

Oncogenesis stands for tumorigenesis (the formation of a malignancy or cancerous growth).

- Mouse Genomics
- Notch Pathway

Oncoprotein

Definition

Oncoprotein is a product of an ► **oncogene** that is able to transform cells and induce tumor formation in animals or humans.

- Leucine Zipper Transcription Factors: bZIP Proteins

Oncostatin M

Definition

Oncostatin M is a member of the Interleukin-6 related cytokine subfamily, which is primarily known for its effect on cell growth. It is produced by monocytes and adherent macrophages, by activated T-lymphocytes and various T-cell lines.

- Tangier Disease

Onionskin Mechanism

Definition

Onionskin mechanism refers to structures resembling onion skins, and arises when DNA replication is initiated from the same origin of replication several times during one S phase.

- DNA Amplification

Online Mendelian Inheritance in Man

Definition

OMIM™ (Online Mendelian Inheritance in Man) refers to the online knowledge base of human genes and genetic disorders authored and edited by Dr. Victor A. McKusick and his colleagues at John Hopkins University (Baltimore, MD) and elsewhere, and developed for the World Wide Web by the National Center for Biotechnology Information (NCBI). The database comprises over 14,000 inherited diseases and genetic traits in humans, with ample links to genomic and mutation databases and to the free MedLine data bank of scientific literature. The database at <http://www.ncbi.nlm.nih.gov/Omim/> is constantly updated and contains textual information and references.

DNA sequence, an open reading frame is characterized by the sequence of nucleotides, which when transcribed into mRNA, results in a series of triplet codons that is not interrupted by a stop codon. It begins with a start codon (ATG), terminates by a stop codons (TAA, TAG, TGA).

- ▶ [Full Length cDNA Sequencing](#)
- ▶ [Functional Assays](#)
- ▶ [High-Throughput Approaches to the Analysis of Gene Function in Mammalian Cells](#)
- ▶ [Protein Databases](#)
- ▶ [Recombinant Protein Expression in Bacteria](#)
- ▶ [Recombinant Protein Production in Mammalian Cell Culture](#)
- ▶ [Translational Frameshifting, Non-standard Reading of the Genetic Code](#)
- ▶ [Two-Hybrid System](#)

Oogenesis

Definition

Oogenesis describes a sequential process involving mitosis, meiosis, and cellular differentiation. It begins during female fetal development, ends during adulthood, and results in formation of unfertilized eggs from primordial germ cells.

- ▶ [Mammalian Fertilization](#)

Operons

Definition

Operons are a form of gene organisation commonly seen in eubacteria, archaea and nematodes. They are polycistronic mRNA units containing two or more genes under the influence of the same promoter. The 5' ends of the mRNA's are transspliced followed by polyadenylation of their 3' ends using different spliced leader sequences.

- ▶ [C. elegans Genome, Comparative Sequencing](#)

Oogonia

Definition

Oogonia are the female germ cells containing 46 chromosomes. During fetal life, the oogonia give rise to a limited number of primary oocytes by mitosis before entering meiosis to become egg cells.

- ▶ [Fragile X Syndrome](#)

Opiates

- ▶ [Heroin](#)

Open Reading Frame

Definition

Open reading frame (ORF) refers to one of three possible reading frames in which an mRNA is potentially translated into protein. In analysis of a

Opsonization

Definition

Opsonization is the process of coating invading microorganisms with several plasma proteins (e.g. IgG antibody, Complement 3 component), thereby making them more susceptible to receptor-mediated

- ▶ [phagocytosis](#) by macrophages.
- ▶ [Epistasis in Cystic Fibrosis](#)

ORC

► [Origin Recognition Complex](#)

Orexins

Definition

Orexin-A and orexin-B are two polypeptides produced within the lateral hypothalamus from one common precursor protein (preproorexin). They are also called hypocretins. Orexins act through two different receptors (ORX-1 and ORX-2 or Hcrt-1 and Hcrt-2 receptor) that are expressed throughout the brain. They are important for the regulation of appetite, metabolism and arousal. Genetic defects in this neuropeptide system lead to animal models of narcolepsy. In human narcolepsy, patients suffer from an acquired orexin deficiency of unknown origin.

► [Narcolepsy](#)

ORF

► [Open Reading Frame](#)

Organelle

Definition

A structurally discrete component of a cell, as for example golgi complex, endoplasmic reticulum, lysosomes, mitochondria etc.

Organizer

Definition

Organizer refers to conceptual domains that produce particular organs or body structures, both cell autonomously and non-cell autonomously, in the field of developmental biology.

► [Axis Formation – Formation and Function of the Dorsal Organizer](#)

Organogenesis

Definition

Organogenesis refers to the development of a body part that exhibits self-contained functions, such as the heart, brain or skin.

► [Drosophila Model of Cardiac Disease](#)

Ori

► [Replication Origin](#)

Origin of Bidirectional Synthesis

► [Replication Origins](#)

Origin Recognition Complex

Definition

Origin Recognition Complex (ORC) is a complex of six proteins, consisting of the subunits ORC1 to 6, that recognizes and binds to the replication origin. ORC was initially identified as binding specifically and ATP-dependently to Autonomously Replicating Sequences (ARSSs) in budding yeast. Further studies have demonstrated ORC to be conserved throughout eukaryotes. ATP-binding and -hydrolyzing activities of ORC depend on the ORC1 subunit. While DNA binding specificity of ORC is clear in budding yeast, it has yet to be clarified how mammalian ORC determines the DNA binding site.

► [DNA Replication Initiation](#)

► [Replication Origins](#)

Orofacial Clefts

Definition

Orofacial clefts result from any failure to fuse the structures of the face and oral cavity. Clefts of the lip and palate are by far the most common such anomalies, although clefts of other parts of the face may occur.

► [Cleft Lip Palate](#)

Ortholog/Orthologous Genes/Orthologous Proteins

Definition

Orthologs/orthologous genes have evolved directly from the same ancestral locus (speciation) and are the corresponding genes in another species. In analogy, orthologous proteins share one common ancestor protein. In most cases, orthologs perform the same function. This assumption does not have to be true, as the definition itself is based on evolution and not function. Cases of orthologs with differing function as well as non-orthologous replacement have been reported. For a comparison see ►[paralog](#).

- [Hereditary Spastic Paraplegias](#)
- [Homeodomain Transcription Factors](#)
- [Medaka as a Model Organism for Functional Genomics](#)
- [Notch Pathway](#)
- [Protein Domains](#)
- [Sequence Annotation in Evolution](#)

Orthostatic Hypotension

Definition

Orthostatic hypotension defines a dramatic lowering of the blood pressure, which occurs when going from a lying or sitting position to a standing position, which causes dizziness and can cause syncope. It is a type of dysautonomia that can be seen in certain diseases including Multiple System Atrophy.

- [Parkinson's Disease: Insights from Genetic Cause](#)

Oscillator

Definition

Oscillator designates a system of components that interact to generate a variable changing between alternative extremes with a defined period length.

- [Circadian Clocks](#)

OST

- [Oligosaccharyltransferase](#)

Osteitis

Definition

Osteitis is an inflammation of the bone, usually caused by secondary infection, after a trauma of the bone; can also be caused by radiation therapy.

- [Paget's Disease](#)
- [Recklinghausen Disease](#)

Osteitis Fibrosa Cystica

Definition

Osteitis fibrosa cystica describes an increased osteoclastic resorption of calcified bone with replacement by fibrous tissue. Synonyms are parathyroid osteosis or Recklinghausen disease of bone.

- [Hyper- and Hypoparathyroidism](#)

Osteoblastogenesis

Definition

Osteoblastogenesis describes a process of osteoblast development encompassing initiation, proliferation, differentiation, and maturation.

- [Bone and Cartilage](#)

Osteoblasts

Osteoblasts are cells that differentiate from mesenchymal cells, and produce a special protein mixture (collagen and enzymes that facilitate mineral deposition) to form the internal structure of bones.

- [Bone Disease and Skeletal Disorders, Genetics](#)
- [Jun/Fos](#)

Osteochondrodysplasias

Definition

Osteochondrodysplasias comprise of a group of disorders that primarily show abnormalities of bone and/or

cartilage. Dysplasias affect bone and/or cartilage as a tissue.

► Bone Disease and Skeletal Disorders, Genetics

Osteoclast(s)

Definition

Osteoclast(s) are large multinuclear cells, differentiated from the monocyte/macrophage lineage, which reabsorb old bone so that osteoblasts can replace it with new bone.

► Bone and Cartilage

► Bone Disease and Skeletal Disorders, Genetics

Osteoclastogenesis

Definition

Osteoclastogenesis refers to a process of osteoclast development that encompasses initiation, proliferation differentiation, and maturation.

► Bone and Cartilage

Osteocyte

Definition

Osteocyte is a mature (“retired”) osteoblast located within the bone matrix, which produces and resorbs bone matrix.

► Bone and Cartilage

Osteogenesis Imperfecta

Definition

Osteogenesis imperfecta is a dominant condition with decreased bone density, recurrent fractures and bowing, and other variable clinical characteristics. It is also called “brittle bone disease”. The disease results from mutations in either COL1A1 or COL1A2 genes, which code for the alpha1 and alpha2 chains of Type I collagen.

► Bone and Cartilage

► Bone Disease and Skeletal Disorders, Genetics

Osteoid

Definition

Osteoid designates a collagen 10A1 rich, weakly calcified bone matrix.

► Bone and Cartilage

Osteomalacia

Definition

Osteomalacia designates an abnormality (“soft bones”) of the mineralization of bone matrix in adults, caused by a vitamin D deficiency or problems in the metabolism of this vitamin. The effects of gravity and muscle tension are less than in the ► rickets of children, and may result in only bowing abnormalities of the weight-bearing long bones.

► Bone and Cartilage

► Hyper- and Hypoparathyroidism

Osteopenia

Definition

Osteopenia is characterized by a less than normal amount of bone, but not as severe as ► osteoporosis.

► Bone and Cartilage

Osteopetrosis

Definition

Osteopetrosis or “dense bone disease” is an inherited failure of bone formation, frequently caused by abnormal remodeling due to infunctional osteoclasts. Since the old bone tissue is not reabsorbed, the bones in the body overgrow. This results in sclerotic, thickened bones with poor mechanical properties. In the bones where bone marrow usually forms, the overgrowth expels the marrow. In the head and spine, this

overgrowth exerts pressure on nerves and causes neurological problems.

- ▶ [Bone and Cartilage](#)
- ▶ [Bone Disease and Skeletal Disorders, Genetics](#)

Osteoporosis

Definition

Osteoporosis is a degenerative bone disease that primarily affects post-menopausal women. It is characterized by a decrease in normal bone density due to the loss of calcium and collagen. As a consequence of decreased bone density, bones become brittle resulting in frequent fractures and other serious effects.

- ▶ [Bone and Cartilage](#)
- ▶ [Bone Disease and Skeletal Disorders, Genetics](#)

Osteoprogenitor

Definition

Osteoprogenitor designates a precursor of osteoblasts and osteocytes.

- ▶ [Bone and Cartilage](#)

Osteosclerosis

Definition

Osteosclerosis describes an abnormal hardening and density of bone.

- ▶ [Hyper- and Hypoparathyroidism](#)

Overexpression/Misexpression

Definition

Overexpression/misexpression describes the expression of a gene product (RNA or protein) in greater-than-normal amounts (overexpression), or at a time or place different from that of wild-type expression

(misexpression). This may be achieved through the microinjection of mRNA, injection or transfection of an expression vector, through the generation of transgenic animals, etc. In classical genetic nomenclature, over-expression is analogous to a hypermorphic allele while misexpression is analogous to a neomorphic allele, although in fact overexpression may also produce a neomorphic phenotype.

- ▶ [Xenopus as a Model Organism for Functional Genomics](#)

Overfitting

Definition

Overfitting refers to the core problem of machine learning: perfect separation of the study data does not necessarily lead to good diagnostic performance in the future.

- ▶ [Computational Diagnostics](#)

OXA1 Complex

Definition

OXA1 complex refers to a translocase of the inner membrane in mitochondria that is involved in the assembly of cytochrome c oxidase.

- ▶ [Mitochondria – Biogenesis and Structural Organization](#)

Oxidases

- ▶ [Dioxygenases, Monooxygenases and Oxidases](#)

Oxidation

- ▶ [Alpha-Oxidation](#)
- ▶ [Beta-Oxidation](#)

Oxidative Phosphorylation Disease

► Mitochondrial Myopathy

Oxidative Stress

Definition

Oxidative stress comprises of a disturbance in the prooxidant-antioxidant balance of a cell in favor of the former. As a consequence, highly reactive molecules known as free radicals (reactive oxygen species; ROS) are produced, which can interact with other molecules in the cell and cause oxidative damage. Indicators of oxidative stress include damaged DNA bases, protein oxidation and lipid peroxidation products. This damage has been implicated in the cause of certain diseases and has an impact on aging processes.

► Defective Protein Folding Disorders
 ► Free Radicals
 ► Proteomics in Ageing

2-Oxoglutarate-Dependent Hydroxylase

Definition

2-oxoglutarate-dependent hydroxylase is a member of the family of hydroxylases that employ Fe(II) as

non-heme, and require oxygen in the form of dioxygen with one oxygen atom being transferred to the prime substrate, and the other to 2-oxoglutarate yielding succinate and CO₂.

► Hypoxia Inducible Factors

Oxygen Sensing

Definition

Oxygen sensing refers to an oxygen-dependent enzymatic activity that is directly linked to the regulation of hypoxia-induced responses, and is inherently dependent on ambient oxygen pressures.

► Hypoxia Inducible Factors

Oxytocin

Definition

Oxytocin is a peptide hormone produced in the hypothalamus and stored in the posterior pituitary gland. Its major physiological functions are to stimulate uterine smooth muscle contraction at birth and stimulate milk secretion.

► Hypothalamic and Pituitary Diseases Genetics

p21

- Ras

P300/CBP

Definition

Related proteins p300 and CBP (cAMP-response-element-binding protein (CREB)- binding protein)) are transcriptional co-activators that act with other factors to regulate gene expression, and play roles in many pathways involved in cell-differentiation and signal transduction.

- Catalytic RNA

p53

Definition

P53 is a phosphoprotein and transcription factor encoded by the *p53* gene, which acts as a powerful tumor suppressor gene, and is mutated in many cancer cells or tumors. It plays a critical role as a sensor for a variety of stress signals, triggering protective responses such as cell-cycle arrest, apoptosis, and DNA repair. Mutations or chromosomal deletions of the p53 locus are associated with a poor prognosis in chronic B-cell lymphocytic leukaemia (B-CLL).

- Jun/Fos
- Leukemia
- RNA Polymerase III
- Senescence
- Tumor Suppressor Genes

PABP

- Poly(A)-Binding Protein

PABPN1

- Nuclear Poly(A) Binding Protein 1

PAC

Definition

Phage artificial chromosomes are large, extrachromosomal DNA constructs (see also ►BAC, ►YAC). Artificial chromosomes have become important tools allowing the establishment of comprehensive genomic libraries, and also has potential as gene delivery vectors for gene therapy.

- Hereditary Neuropathies, Motor and/or Sensor

PAC Transgene

- YAC/PAC/BAC Transgene

Pacemaker

Definition

The mammalian CNS contains a biological clock that drives remarkably precise circadian rhythms. An

ensemble of coupled neurons in the suprachiasmatic nucleus (SCN) of the hypothalamus comprises of the principal circadian pacemaker, which is capable of both sustaining its own oscillation (daily light/dark cycle) and of synchronizing signals to local circadian oscillators in peripheral tissues.

► [Circadian Clocks](#)

Pachyonychia Congenita

Definition

Pachyonychia congenita is an autosomal dominant inherited disorder caused by mutations in keratin genes, contributing to structural stability of cells of the outer root sheath of the hair follicle. Hallmark of the disease is thickened nails. However, additional variable involvement of hair, sebaceous glands and oral mucosa can occur, depending on which keratin gene is mutated (e.g. palmoplantar keratoderma and oral leukoplakia in K6a, K16 mutants, versus pilosebaceous cysts, twisted hairs and natal teeth in K6b or K17 mutants).

► [Heritable Skin Disorders](#)
 ► [Intermediate Filaments](#)

Paclitaxel

Definition

Paclitaxel and its analogue docetaxel bind to β -tubulins of microtubules. Heterodimers can no longer be released from ► [microtubules](#). Dynamic instability, necessary for the capturing and segregating of chromosomes during mitosis, is no longer possible; cell division is stopped. Both taxols are used for cancer treatment.

► [Cytoskeleton](#)

PAGE

► [Polyacrylamide Gel Electrophoresis](#)

2D-PAGE Database

Definition

These databases present 2-dimensional electrophoresis patterns in the www. Clicking on spots gives the user the identity of the protein species and further information such as Mr (relative molecular mass), isoelectric point, and links to genome, protein sequence, pathway and other databases. In more developed 2D-PAGE databases, mass spectrometry data and functional data allows the mining of proteomics data.

► [Two-Dimensional Gel Electrophoresis](#)

Paget's Disease

Definition

Paget's disease is a generalized skeletal chronic-progressive disease – often familial – in which bone turnover is increased, leading to thickening and softening of the skull and bending of weight-bearing bones.

► [Hyper- and Hypoparathyroidism](#)
 ► [Osteitis](#)

PAK

Definition

PAK stands for p21-Activated (serine/threonine) kinase.

► [Signal Transduction: Integrin-Mediated Pathways](#)
 ► [Receptor Serine/Threonine Kinase](#)

Pal-CoA

Definition

Pal-CoA stands for Palmitoyl-Coenzyme A. Palmitic acid as well as other fatty acids must be activated before they can be transferred to proteins or lipids. Enzymes designated acyl-CoA synthetase link the carboxyl-group of the fatty acid to the – SH group of

coenzyme A. Synthesis of acyl-CoA requires ATP, and the energy of the ATP-hydrolysis is stored in the labile thioester-bond between the carbon chain and CoA. Pal-CoA is an amphipatic molecule with a hydrophilic head group, the CoA part, and a hydrophobic moiety, the acyl chain. The total concentration of long chain acyl-CoAs inside cells is in the range between 5 μ M and 150 μ M. However, the concentration of free Pal-CoA is very low (less than 200 nM), with most of it being bound to acyl-CoA binding proteins (ACBP). This helps to prevent possible damaging effects caused by the detergent-like properties of acyl-CoAs.

► [Fatty Acid Acylation of Proteins](#)

Palindrome

Definition

Palindrome designates a DNA sequence that reads the same on each strand of DNA, when the strand is read in the 5' to 3' direction. It consists of adjacent inverted repeats.

► [Nuclear Hormone Receptors: Regulators of Gene Transcription and Cellular Signaling](#)

Palmitoylation

Definition

Palmitoylation refers to post-translational modification of proteins, and describes the addition of palmitate, a fatty acid with 16 carbons, to a cysteine residue.

► [Fatty Acid Acylation of Proteins](#)

► [Ras Signalling](#)

Palmoplantar Keratoderma

Definition

Palmoplantar keratoderma is an autosomal dominantly inherited skin disorder characterized by thickening of the epidermis of the palms and soles.

► [Heritable Skin Disorders](#)

PAMP

Definition

PAMP stands for pathogen-associated molecular pattern.

► [Inflammatory Response](#)

Pancreatic Insufficiency

Definition

Pancreatic insufficiency describes a reduced production or release of digestive enzymes by the exocrine pancreas. Pancreatic insufficiency usually leads to malabsorption and lack of growth.

► [Epistasis in Cystic Fibrosis](#)

Pancreatitis

Definition

Pancreatitis is the inflammation of the pancreas.

► [Hyper- and Hypoparathyroidism](#)

Panhypopituitarism

► [Hypothalamic and Pituitary Diseases, Genetics](#)

Pannexins

Definition

Pannexins are intercellular channel forming proteins found in the brain that share structural features with gap junctions in vertebrates and invertebrates.

► [Gap Junctions](#)

Pannus

Definition

Pannus can have different meanings: it describes a hypertrophic cell mass in the joints due to chronic synovitis. Synovitis is adherent to cartilage and locally invasive at the cartilage bone interface, and finally can destroy the affected joint. Pannus also describes a pathological state of the eye, characterized by proliferation of fibrous or vascular tissue under the epithelium in the anterior cornea, due to chronic inflammation in the eye.

► [Rheumatoid Arthritis](#)

Papilloma Viruses

Definition

Papilloma viruses constitute a large virus family characterized by small icosahedral, non-enveloped capsids, circular doubled-stranded DNA genomes and a strict tropism for epithelia for their replication. The steps for viral replication are closely linked to the differentiation of the keratinocytes. Depending on the HPV (human papillomavirus) type, the productive infection results in different types of benign tumours of skin (wart, papilloma) and genital mucosa (condyloma). Depending on the presence of high risk – HPV types and chromosomal integration of viral genomes, these lesions can convert to cancer. High risk type genital papillomaviruses like HPV16, HPV18, 45, and 56 are the cause of cervical carcinoma, one of the most frequent tumours in females.

► [Viral Oncogenesis](#)

PAR

Definition

PAR describes a kinase that has been shown to phosphorylate Dishevelled (Dvl) and induce Wnt/ β -catenin signaling

► [Wnt/Beta-Catenin Signaling Pathway](#)

Paracrine

Definition

Paracrine refers to the action of cytokines, growth factors or hormones that act locally on cells other than the producing cell.

► [Growth Factors](#)

► [Rheumatoid Arthritis](#)

► [Wound Healing](#)

Paralog/Paralogous Genes/ Paralogous Proteins

Definition

Paralogs are homologous genes present in the genome of the same species that arise by duplication events, and code for proteins with similar but not identical functions. Paralogous proteins are proteins created by a duplication event within one species. In contrast, ► [orthologs](#) are genes that are related by vertical descent from a common ancestor, and encode proteins with the same function in different species.

► [Double-Strand Break Repair](#)

► [Gene Duplications](#)

► [Homeodomain Transcription Factors](#)

► [Protein Domains](#)

► [Sequence Annotation in Evolution](#)

Paralysis Agitans

► [Parkinson's Disease: Insights from Genetic Causes](#)

Parathyroid Hormone

Definition

Parathyroid hormone (PTH) is an 84-amino-acid polypeptide secreted by the two pairs of parathyroid glands adjacent to the back of the thyroid glands in the neck.

► [Hyper- and Hypoparathyroidism](#)

Parathyroid Hormone-Related Protein

Definition

PTHrP is secreted by many types of malignant tissues, and produces hypercalcemia by activating PTH/PTHrP receptors in bone and kidney. There are three isoforms of this polypeptide consisting of 139, 141 and 171 amino acids in the human.

►Hyper- and Hypoparathyroidism

Paratope

Definition

Paratope refers to local surface sites on antibodies that react with antigen determinant sites on antigens. They are formed from parts of the variable regions of the Fab fragment of the immunoglobulin.

►Peptide Chips

Paraxial Mesoderm

Definition

Paraxial mesoderm refers to that part of the mesoderm that lies bilaterally, immediately adjacent to the notochord and neural tube of the embryo prior to the formation of ►somites.

►Muscle Development

►Somitogenesis

Paresthesia

Definition

Paresthesia is an abnormal sensation of the skin, for example, such as burning, prickling, tickling or tingling.

►Hyper- and Hypoparathyroidism

PARG

►Poly(ADP-Ribose) Glycohydrolase

Parkinson's Disease: Insights from Genetic Causes

MELISA J BAPTISTA, MARK R COOKSON,
KATRINA GWINN-HARDY

National Institute on Aging and National Institute of Neurological Diseases and Stroke, National Institutes of Health, Bethesda, MD, USA

baptista@mail.nih.gov

Synonyms

Paralysis Agitans, Dopamine Responsive Parkinsonism

Definition

Parkinson's disease (PD) belongs to a group of conditions known as movement disorders. Clinically the disease is characterized by at least two of the cardinal features of bradykinesia (slowness of movement and a decrease in spontaneous movement), gait difficulty, postural instability, rigidity, resting tremor and asymmetry, as well as by a good response in the early stages to dopamine replacement therapy.

Characteristics

Clinical

Parkinsonism is a term used to apply to all disorders where at least two of the four cardinal features named above occur. Parkinson's disease is the most common cause of parkinsonism. Symmetric onset, lack of response to a reasonable trial of ►dopamine replacement therapy, prominent early dementia, eye movement difficulties, or severe, early ►dysautonomia (such as orthostatic hypotension), all argue in favor of an alternative cause of parkinsonism, such as progressive supranuclear palsy (PSP) or multiple system atrophy (MSA). James Parkinson first described Parkinson's disease in his essay on Shaking Palsy in 1817. Although classically a movement disorder as described above, the disease is systemic, and can cause problems with the skin (ruddiness and oiliness due to seborrheic keratosis), cognition (especially problems in executive function), mood (such as depression), and dysautonomia (cardiac sympathetic dysfunction, constipation,

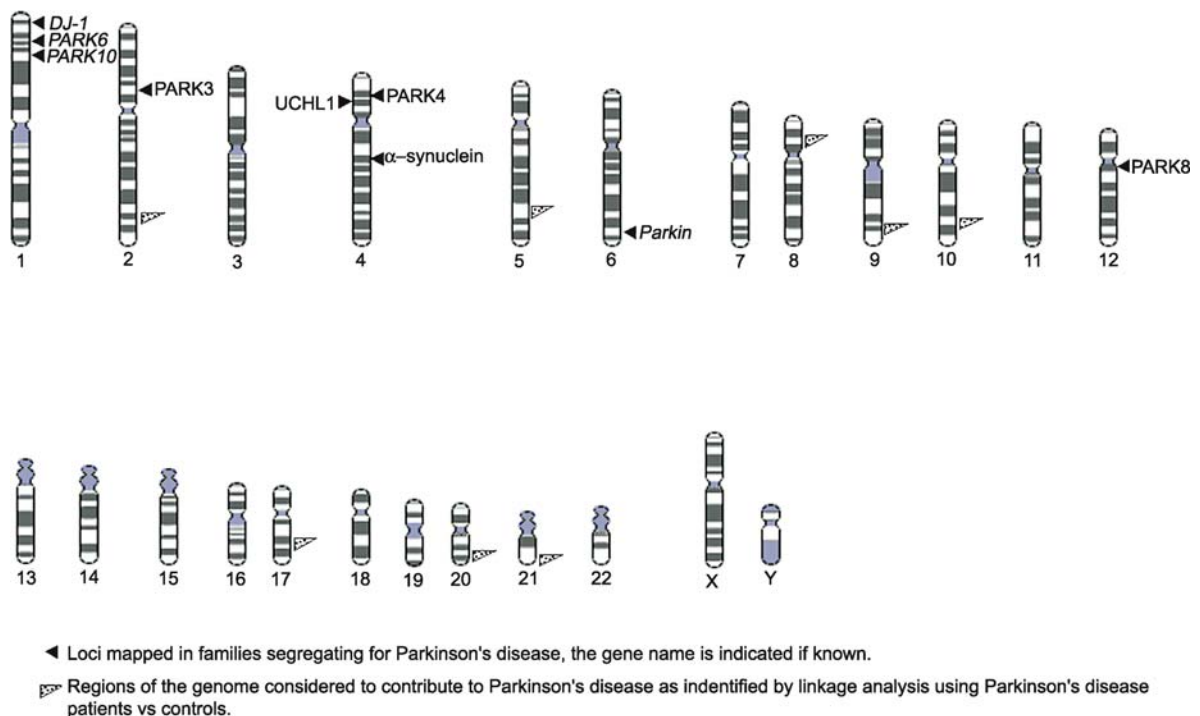
and urinary difficulties). The pathological hallmarks of Parkinson's disease are the presence of Lewy bodies and a marked loss of dopamine-producing neurons in the ►**substantia nigra** (1). The neurotransmitter ►**dopamine** plays a crucial role in the initiation and maintenance of movement and thus its loss leads to a number of the above symptoms and signs. Hence dopamine replacement therapy is quite helpful in the early stages of this illness. However, not all symptoms are improved by dopamine replacement strategies, and there is currently no treatment known to slow the progression of the disease. Surgical treatment, including deep brain stimulation (DBS) in which a stimulating electrode is placed in deep brain structures (often the subthalamic nucleus), has been shown to be beneficial in some patients, but like dopamine replacement medication, is not curative. No therapies are known to slow the progression of this illness although neuroprotection trials are underway.

Genetic Causes of Parkinson's Disease

To date, several mutations have been identified as Mendelian causes of parkinsonism, and linkage has been made to several other regions, showing that in rare cases Parkinson's disease may be inherited as a ►**monogenic** trait (Fig. 1). (For more details see Hardy et al. 2003 (2), and Gwinn-Hardy 2002 (3)). Although

PD most often affects individuals over the age of 60 years, this illness may also affect younger people. Often, individuals with a younger age of onset of symptoms and a family history have a causative genetic mutation. Epidemiological surveys support the contention that the contribution of genetic compared to environmental factors is greater in younger patients. The role of susceptibility genes in PD is also supported by epidemiological studies that show a higher frequency of disease in patient relatives as compared to controls. It is often supposed that genetic variability plays a role in the development of later onset "typical" PD. Susceptibility genes may work in concert with environmental factors to influence disease incidence in the general population (see below).

The identification of multiple genes that produce a similar, if not identical, phenotype suggests that the gene products involved may have common cellular and molecular effects. As there are three genes unambiguously associated with PD (*α-synuclein*, *parkin* and *DJ-1*) plus a fourth (*UCHL1*) that is possibly causative in one family, there is a unique opportunity to delineate the pathway that leads to PD by understanding the molecular relationships between the gene products. The section below will deal with each of the gene products and the hypothesized cellular events that are thought to drive pathogenesis.



Parkinson's Disease: Insights from Genetic Causes. Figure 1 Regions of the human genome, encoding *PARK* loci, and regions with significant linkage to Parkinson's disease. Adapted from Riess et al., 2002.

Cellular and Molecular Regulation

α -Synuclein (PARK 1)

Alpha-synuclein was the first PD gene to be discovered in 1997 by Polymeropoulos and colleagues in an Italian family called the Contursi kindred. Affected individuals typically develop fulminant parkinsonism in the fourth or fifth decade, although the phenotype is variable. To date, in this and other families, two α -synuclein mutations have been identified (at position 30 alanine is substituted for proline [A30P] and at position 53 alanine is substituted for threonine [A53T]), both altering the amino acid composition of the protein. One of the oddities in this is that alanine at position 53 is poorly conserved throughout evolution. For example, rats and mice both have a threonine at the same position; the reason this is normal in rodents but causes disease in humans remains unclear. Alpha-synuclein has been shown to be a major component of Lewy bodies, including in familial and sporadic cases without any mutation in that gene, supporting the importance of this protein in general towards our understanding of Parkinson's disease.

The important function(s) of α -synuclein are still under investigation, as is the way that causal mutations affect these. Wild type α -synuclein appears to play an important role in synaptic function. It is associated with presynaptic vesicles; the protein binds to certain lipids in intracellular membranes. Knocking out α -synuclein in mice produces subtle changes in synaptic transmission, probably by reducing the number of vesicles that are normally present, which in turn may affect presynaptic protein amounts.

In Lewy bodies, α -synuclein is present in an insoluble, fibrillar form that can be recreated *in vitro* using purified protein. Events that promote fibrillation may include some oxidative modifications, high protein concentrations or the presence of some kinds of lipids. However, the role of mutations in this process is apparently complex. Several groups have shown that A30P does not promote the formation of mature fibrils, but instead creates a relatively soluble intermediate species referred to as oligomers or protofibrils.

Mutant forms of α -synuclein have a number of effects at the level of the whole cell. They can increase sensitivity of cells to various damaging events, including oxidative stress, dopamine toxicity, mitochondrial damage and proteasome inhibition. The precise mechanisms involved are likely to be multifactorial. Evidence suggests that the mutant forms have more extreme effects than wild type α -synuclein, consistent with their dominant cellular actions. However, at high levels of expression, wild type α -synuclein also has damaging effects. Several studies have shown that there is variability in the susceptibility of neurons to the damaging effects of mutant α -synuclein,

although it is notable that mouse models often show damage to spinal cord motor neurons rather than midbrain cells.

Parkin (PARK2)

Numerous mutations associated with parkinsonism have been identified in *parkin*. Originally described in Japanese teenagers and young adults as causal for autosomal recessive juvenile parkinsonism (ARJP), *parkin* has also been found to be important in other forms of parkinsonism. The original description was a disease characterized not only by an early age of onset as noted above, but also, by sleep benefit and foot dystonia. Subsequent reports in the non-Japanese literature, particularly in Caucasians, describe a later age of onset, no foot dystonia, and symptoms identical to "typical" PD, possibly with a slower than typical rate of progression. Patients with parkin mutations are not thought to generally have Lewy bodies although *post mortem* reports are limited to date. Parkin is one of several proteins that "tag" target intracellular proteins for degradation by adding chains of the small protein ubiquitin. The loss of functional parkin is hypothesized to compromise the ability of neurons to degrade potentially damaging proteins (for example, those that have an abnormal conformation) or proteins that may be toxic if allowed to accumulate in the cell. There is some evidence that one of the proteins that parkin acts upon is α -synuclein itself. Parkin has many other targets and it may be that these are also important in the pathogenesis of PD.

DJ-1 (PARK7)

Mutations in the gene encoding DJ-1 have been shown to cause PD in two European families and one from Venezuela. Mutations are inherited in an autosomal recessive fashion, and are postulated to result in disease due to a loss of function of the resulting protein. The precise function of the protein, and its relation to either α -synuclein or parkin is unclear at this time. No pathological evaluation of any affected members has been reported to date.

Ubiquitin C-Hydrolase-1 (UCHL1, PARK5)

A single mutation in this gene has been identified altering the amino acid at position 93, and leading to an abnormal enzyme, in a single sibling pair. Because it has only been described in that sibling pair to date, its importance in understanding Parkinson's disease remains to be confirmed. Nonetheless, further evidence for the involvement of this enzyme with PD comes from studies of a common genetic variant which appears to protect against the onset of PD. The function of UCHL1 is to recycle the ubiquitin "tag" added by enzymes such as parkin (see above), also supporting its role conceptually in the pathway which leads to

Parkinson's disease. The mutation identified reduces this activity. The way in which the mutation in *UCHL1* affects enzyme function is complex, and in fact there appear to be two distinct normal enzyme activities, one of which may affect α -synuclein processing, further supporting a potential importance of this enzyme.

Other Loci

The genes highlighted above were identified in families with apparent Mendelian parkinsonism. In the same way, five additional loci have been reported to date, for which the causative gene and mutation remain to be elucidated (Fig. 1).

A Comment on Possible Environmental Causes of Parkinson's Disease

There have been many proposed environmental insults that may lead to the onset and/or progression of Parkinson's disease, including pesticides and/or drugs. Exposure to the neurotoxin 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (►MPTP), a contaminant of the abused drug heroin, can reproduce many of the symptoms of PD; however Lewy bodies are not present in affected individuals. MPTP has become an important tool in the study of disease in animal models. A second model is the chemical used by exterminators, rotenone, which is an inhibitor of mitochondrial complex I activity (4).

Both of these toxins cause the loss of dopaminergic neurons. Outside of this observation, their relation to PD is unclear. However, accumulation of α -synuclein is seen in some models developed with both of these substances. Neither are known to cause Lewy bodies. Accumulation of ubiquitinated proteins has been reported for rotenone treatment but is not generally a feature of the MPTP model. It is not clear if the absence of Lewy bodies in these models represents a fundamental difference between the model systems and the genetic causes of PD, or, alternatively, if there is a peculiarity of PD in humans that promotes Lewy body formation not seen in animals. It is worth noting, however, that there are causes of parkinsonism in humans (such as PSP described above) which do not cause Lewy bodies. Parkinsonism has many causes, some of which lead to Lewy bodies, and those causes associated with Lewy bodies probably lead to the most common form of Parkinson's disease.

Clinical Relevance

The monogenic causes of Parkinson's disease described above are rare. In all likelihood, the risk of developing Parkinson's disease for individuals without a positive family history of disease will be due to a combination of genetic predisposition and environmental triggers. However, the study of the genetic

causes of Parkinson's disease is important for a number of reasons. Current treatment for Parkinson's disease is merely symptomatic; no interventions exist which interrupt the inexorable disease progression or which protect the nervous system from damage. Neuroprotective strategies in Parkinson's disease (PD) are being developed. Pre-symptomatic diagnosis is of extreme importance should such interventions be identified; it is thought that by the time a patient presents with motor manifestations, approximately 70% of the neurons in the substantia nigra have been lost and therefore the maximal desired impact of therapeutics is not possible. The families in which Parkinson's disease occurs in a Mendelian fashion will be key for the study and determination of such early diagnostic features. Therefore, the study of the genetic causes of Parkinson's disease remains of value, as it allows us to unravel the underlying biological mechanisms leading to Parkinson's disease, which in turn will allow development of better treatments and diagnostic methods.

References

1. Hughes AJ, Daniel SE, Kilford L et al (1992) Accuracy of clinical diagnosis of idiopathic Parkinson's disease. A clinico-pathological study of 100 cases. *J Neurol Neurosurg Psychiatry* 55:181–184
2. Hardy J, Cookson MR, Singleton A (2003) Genes and Parkinsonism. *Lancet Neurol* 2:219–226
3. Gwinn-Hardy K (2002) Genetics of parkinsonism. *Mov Disord* 17(4):645–656
4. Betarbet R, Sherer TB, Mac Kenzie G et al (2000) Chronic systemic pesticide exposure reproduces features of Parkinson's disease. *Nat Neurosci* 3:1301–1306
5. Riess O, Kruger R, Schulz J (2002) Spectrum of phenotypes and genotypes in Parkinson's disease. *J Neurol* 249 (Suppl 3):15–20

PARN

►Poly(A) Ribonuclease

PARP1

►Poly(ADP-Ribose) Polymerase (PARP)–1

Parsimony

Definition

Parsimony, like Neighbour-Joining, is a frequently used algorithm for construction of phylogenetic trees. In principle, no assumptions about distances are made. The algorithm finds the tree that explains the resulting proteins with a minimal number of substitutions.

► [Sequence Annotation in Evolution](#)

Partial Specific Volume

Definition

Partial specific volume denotes the molecular volume per unit weight. In ultracentrifugation, it is historically measured in units of ml/g. Typical values are 0.73 ml/g for proteins.

► [Analytical Ultracentrifugation](#)

Particle-Coated DNA

Definition

For use in gene gun vaccination, plasmid DNA is coated onto gold particles.

► [DNA-based Vaccination](#)

PAS Domain

Definition

PAS domains are involved in many signalling proteins where they are used as a signal sensor domain. The PAS domain was named after three proteins that it occurs in: Per- (period circadian protein), Arnt (Ah receptor nuclear translocator protein), and Sim (single-minded protein). PAS domains appear in prokaryotes and eukaryotes. Several PAS-domain proteins are known to detect their signal by way of an associated co-factor. Haeme, flavin, and a 4-hydroxycinnamyl chromophore are used in different proteins. LOV domains and ► [PyP](#) are subsets of the PAS domain.

► [Photoreceptors](#)

Patch Clamping

CLAUDIA EDER

Institute of Physiology, Medical Faculty Charité,
Berlin, Germany

claudia.eder@charite.de

Definition

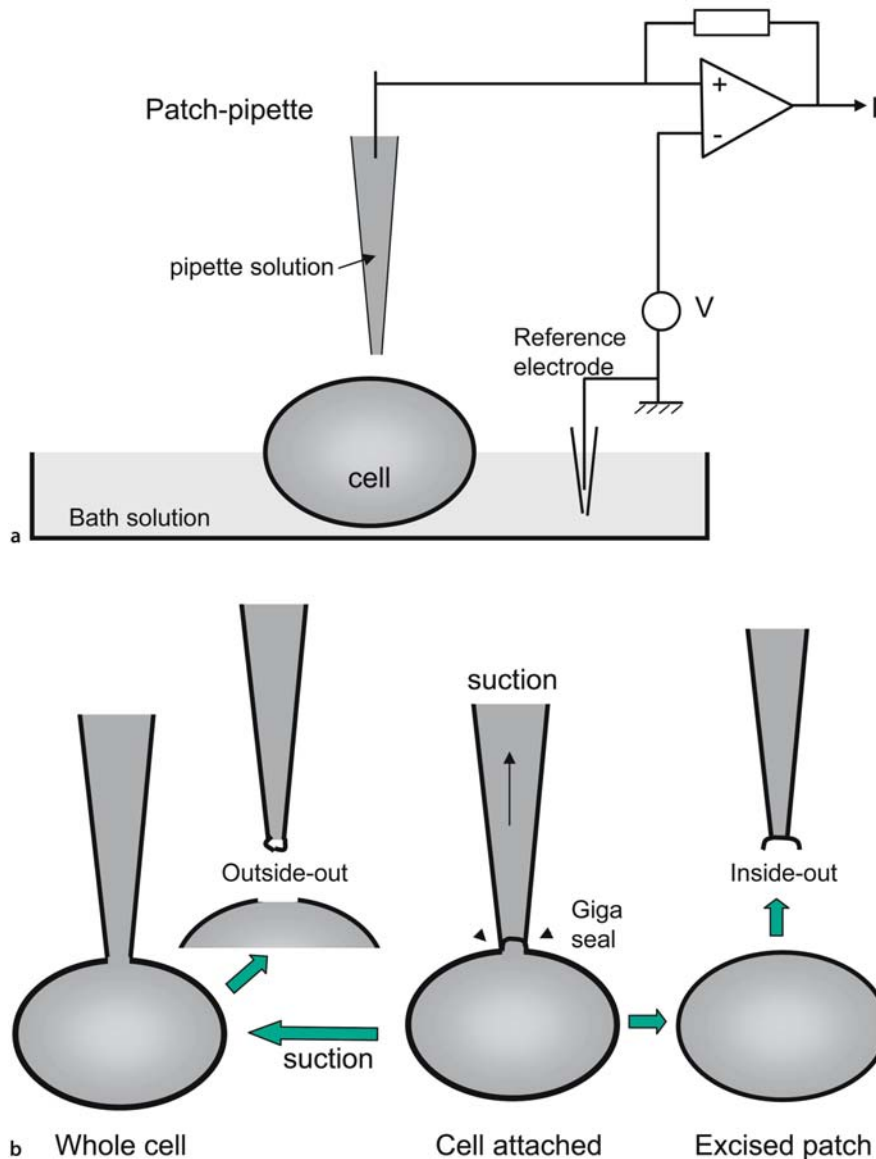
Patch clamping is a powerful method for studying ion currents in ► [biological membranes](#). In patch clamp experiments, charge movement caused by ion flux through ion channels or transporters is measured in a closed electrical circuit (1, 2). Patch clamping allows recordings of either single channel currents or of currents from entire cells or organelles. Single channel currents can be measured in patches of the membrane in intact cells or in excised patches. The sum of single channel currents produces the whole cell current of a cell. The patch clamp technique has made it possible to characterize the biophysical and pharmacological properties of ion channels and electrogenic ion transporters. Patch clamp recordings can be performed under *in vitro*, *in situ* or *in vivo* conditions.

Ion channels and ion transporters are membrane proteins that control the flow of ions across cellular membranes and therefore influence the behavior of all cells (3). By regulating intracellular ion concentrations and membrane potential, ion channels and transporters modulate cell volume, excitability, proliferation, migration, secretion and other cellular processes.

Description

Patch Clamp Set-up

The patch clamp set-up consists of at least a vibration-isolated table, a microscope, a micromanipulator, the patch clamp amplifier and a personal computer. The patch pipette filled with electrolyte solution is held in a pipette holder, which is mounted on a micromanipulator so that the patch pipette can be moved towards a cell. A silver-silver chloride wire connects the patch pipette solution and the patch clamp pre-amplifier (which is called the headstage). Another silver-silver chloride wire in the bath chamber is connected to the headstage and serves as the ground electrode. The headstage contains a current-to-voltage converter and is connected to the amplifier (Fig. 1a). Voltage- and current-clamp protocols are generated and controlled by a personal computer, which is also used for data storage. In some cases, additional micromanipulators are used to position stimulation pipettes, additional patch pipettes or superfusion systems.



Patch Clamping. Figure 1 (a) schematic representation of a patch-clamp setup. (b) Different patch-clamp configurations. The patch-pipette is brought into contact with the cell surface membrane, slight suction is applied to achieve a dense seal (giga-seal); cell-attached configuration. Whole cell configuration is reached when further suction is applied to break the membrane inside the tip. Excised patch configuration (inside-out and outside-out) is reached when the piece of membrane attached to the tip is excised from the cell. For further details see text.

Patch Clamp Configurations

In conventional patch clamp recordings, glass or fused quartz pipettes having a tip diameter of about 1 μm are used. Pipette tips are usually fire-polished immediately before using, because a very clean surface is required for these recordings. To carry out the patch clamp recording, the fire-polished pipette filled with electrolyte solution is brought into close contact with the surface of a cell (Fig. 1). Application of gentle suction to the inside of the pipette increases the resistance

between the glass pipette and the cell membrane reaching values of up to 100 gigaohms. The tight contact between the pipette tip and the membrane, which is called a **giga-seal**, prevents ion flux between pipette solution and bath solution. Thus, the small patch of membrane within the tip is electrically isolated, so that measurements of currents that are carried by ions flowing into or out of that cell patch through ion channels are possible. This configuration is called the cell-attached configuration. Depending on

the exact size of the membrane patch under the pipette and the density of ion channels in the membrane, activity of one or more ion channels can be detected. In the cell-attached configuration it is possible to alter the composition of the extracellular medium in the pipette, but there is no direct access to the inside of the patch, i.e. the intracellular milieu remains intact (Fig. 1).

Once the cell-attached configuration is obtained, other recording configurations can be produced (1, 2). Due to the mechanical strength of the giga-seal, the membrane patch at the pipette tip can be pulled away from the surrounding cell. Withdrawing the pipette will excise a patch with the internal membrane surface facing the bath solution and the external surface facing the pipette solution. This patch clamp configuration is called inside-out configuration. It is mainly used for the testing of various intracellular channel modulators, such as calcium, ATP, kinases and phosphatases, because these agents can easily be exposed to the intracellular side of the membrane by bath perfusion.

Alternatively, after reaching the cell-attached configuration, additional suction can be used to break the membrane within the tip, thus creating the whole-cell configuration. Whole-cell recordings permit diffusional exchange of the pipette solution with cytoplasmic constituents. After the establishment of the whole-cell configuration, pulling the pipette away from the cell will excise a patch with the extracellular side of the membrane facing the bath solution and the intracellular side facing the pipette filling solution. This patch clamp configuration is called outside-out configuration.

The disadvantage of outside-out and whole-cell recordings is that normal second messenger systems can be dialyzed out, which may result in current run-down. Therefore, the perforated patch-clamp method has been applied to achieve electrical access to the cell while minimizing cytoplasmic dialysis (4). This method involves briefly dipping the pipette tip in normal pipette solution and then backfilling with solution to which nystatin, amphotericin or gramicidin have been added. After giga-seal formation, these molecules form small pores in the membrane patch within the pipette. The pores lower the access resistance of the patch enough to allow voltage- or current-clamp of the whole cell membrane. These pores are exclusively selective for monovalent ions. They do not allow divalent ions and larger molecules such as proteins to pass. In ▶perforated-patch recordings normal intracellular Ca^{2+} buffering and second messenger mechanisms within the cell remain intact.

Patch Clamp Protocols

In general, the membrane currents of a cell are measured in the ▶voltage clamp mode of the patch clamp technique where the membrane potential is

controlled and the current activated at a certain potential is monitored (3). The holding potential is usually clamped near the resting membrane potential of the cell. Ion currents are studied using either voltage pulse or voltage ramp protocols. Voltage pulses are usually applied from a negative holding potential to a certain potential for a duration of a few milliseconds or seconds. This approach allows investigations of current kinetics, such as voltage- and time-dependent activation and inactivation behavior. Application of voltage ramps means that the membrane voltage is steadily increased from a negative to a positive potential (or in the opposite direction) within a few milliseconds or seconds. The advantage of voltage ramp protocols is that the activation threshold and reversal potential of currents can be easily estimated and changes in current amplitude can be visualized over the whole voltage range. The disadvantage is that time-dependent gating (opening and closing kinetics) will distort and complicate the data.

▶Current clamp recordings are used to monitor changes in the resting membrane potential of a cell (3). Furthermore, by application of various current pulses, electrical behavior, such as the pattern of action potentials, can be observed in excitable cells.

Combination of the Patch Clamp Technique and Single Cell PCR

The combination of the patch clamp technique and a molecular biological technique, namely single cell reverse transcription polymerase chain reaction (▶RT-PCR), permits correlation of the functional properties of individual cells with their specific gene expression profile. The patch clamp technique enables the harvesting of messenger RNAs (mRNAs) from a single living cell. During whole cell recording of ion channel activity, the cell interior is dialyzed with the internal solution contained in the patch pipette. After a sufficiently long recording, the electrode solution contains RNA. To obtain the remaining cytoplasmic RNA, suction is applied to the interior of the pipette before termination of the recording. The pipette contents are then expelled into a sterile reaction tube and used for subsequent RT-PCR experiments.

Clinical Applications

Ion channel dysfunction can lead to the appearance of clinical signs and symptoms. Such diseases are called ▶channelopathies. Three major types of channelopathy have been described so far: i) genetic, in which ion channels function abnormally or fail to function as a result of mutations, ii) autoimmune, in which antibodies perturb channel function and iii) transcriptional, which result from changes in the expression of non-mutated channel genes. Several efforts are currently

being made to develop new drugs that are capable of modulating ion channel activity, i.e. acting specifically either to restore ion channel functions or to inhibit ion channel activity. Although patch clamping remains the sole technique with the sufficiently high time resolution and sensitivity required for precise and direct characterization of ion channel properties, the conventional patch clamp technique is not practical for high throughput drug screening. The conventional patch clamp technique requires a skilled operator to manipulate the glass pipette onto the cell manually. It is a slow, labor intensive and thus expensive technique. This fact underlies current efforts directed towards improving the throughput of the conventional patch clamp technique in order to facilitate drug screening and to assay the function of genes that encode ion channel and transporter proteins directly. As an approach for high throughput screening for ion channel modulators, automated patch clamp systems have been developed (5, 6). These systems are often based on an array of planar chips. They allow multidrug and multiconcentration applications on a single cell and recordings can be performed in a great number of cells simultaneously.

Therapeutic Consequences

A number of common diseases, including epilepsy, cardiac arrhythmia, hypertension and type II diabetes, are primarily treated by drugs that modulate ion channels. Current studies aim at the development of more specifically acting drugs in order to minimize unwanted side effects. High throughput automated patch clamp systems will facilitate drug screening. The subsequent development of new drugs will lead to a better understanding and finally to a more efficient treatment of various channelopathies.

► [Ion Channels/Excitable Membranes](#)

References

1. Hamill OP, Marty A, Neher E et al (1981) Improved patch-clamp techniques for high-resolution current recording from cells and cell-free membrane patches. *Pflügers Arch* 391:85–100
2. Sakmann B, Neher E (1995) Single channel recording. Plenum Press, New York
3. Hille B (2001) Ion channels of excitable membranes. Sinauer, Sunderland, MA
4. Horn R, Marty A (1988) Muscarinic activation of ionic currents measured by a new whole-cell recording method. *J Gen Physiol* 92:145–159
5. Fertig N, Blick RH, Behrends JC (2002) Whole cell patch clamp recording performed on a planar glass chip. *Biophys J* 82:3056–3062
6. Klemic KG, Klemic JF, Reed MA et al (2002) Micro-molded PDMS planar electrode allows patch clamp electrical recordings from cells. *Biosens Bioelectron* 17:597–604

Pathomechanism

Definition

Pathomechanism denotes the biochemical mechanism leading to the disease phenotype.

► [Mutagenesis Approaches in Yeast](#)

Pathway Particle

Definition

Pathway particle refers to a collection or network of proteins that assemble in a specific way within cells, and contain all of the catalytic activities required for a particular metabolic pathway, for example, the purine pathway.

► [Nucleotide Biosynthesis](#)

Pattern Recognition Receptors

Definition

Pattern recognition receptors (PRR) are receptors of the innate immune system (► [innate immunity](#)) that recognize common molecular patterns on pathogen surfaces. They bind pathogen-associated molecular patterns (PAMP).

► [Inflammatory Response](#)

Patterson Function/Patterson Map

Definition

In X-Ray structure analysis, the Patterson function is defined as the convolution of the electron density with its own inverse, or equivalently, the Fourier transform of the diffraction intensities. The resulting map (Patterson map) calculated using this equation is essentially a map of all interatomic vectors of the crystal contents. The Patterson map can be used to locate anomalous scatterers (for instance heavy atoms) in the crystal.

► [MAD Phasing](#)

► [Structure Determination by X-Ray](#)

PCR

Definition

A method for rapidly amplifying a small amount of DNA using a heat-stable polymerase and two oligonucleotide primers, one complementary to the (+) strand at one end of the sequence to be amplified, and one complementary to the (−) strand at the other end. In a series (typically 30) of temperature cycles, the DNA is repeatedly denatured, annealed to the primers and a daughter strand extended from the primers. As the daughter strands themselves act as templates for subsequent cycles, DNA fragments matching both primers are amplified exponentially. Thus, the original DNA need be neither pure nor abundant. PCR has become widely used not only in research, but in clinical diagnostics and forensic science.

- ▶ Large-Scale Homologous Recombination Approaches in Mice
- ▶ Morpholino Oligonucleotides, Functional Genomics by Gene ‘Knock-down’
- ▶ Recombinant Protein Expression in Bacteria
- ▶ Single-Cell Gene Expression Profiling: Cell-level Biology by Multiplexed Expression Fluorescence in Situ Hybridization
- ▶ SNP Detection and Mass Spectrometry
- ▶ Thermodynamic Properties of DNA

most commonly found protein domains in metazoan genomes.

- ▶ Adherens Junctions
- ▶ Adhesion Molecules
- ▶ Epithelial Cells
- ▶ Hereditary Neuropathies, Motor and/or Sensory
- ▶ Tangier Disease
- ▶ Tight Junctions

PE

- ▶ Position Effect

Pectus Carinatum

Definition

Pectus carinatum designates a protrusion of the chest wall.

- ▶ Marfan Syndrome

PDGF

Definition

PDGF stands for Platelet-Derived Growth Factor. It is a widely expressed growth factor that stimulates proliferation of a large number of cell types. It is found in platelets and secreted during blood clotting where it may stimulate tissue healing.

- ▶ Growth Factors

Pectus Excavatum

Definition

Pectus excavatum refers to a depression in the chest wall, which is one symptom of Marfan syndrome.

- ▶ Marfan Syndrome

PDZ Domains

Definition

PDZ domains are modular protein interaction domains that function in the polarized sorting of proteins, and in the assembly of supramolecular signaling complexes. These domains of approximately 90 amino acids are named for the first three proteins they were detected in: PSD-95, Discs large and ZO-1. They are among the

Definition

Pedigree describes a family tree diagram that shows how a particular genetic trait or disease has been inherited.

- ▶ Mendelian Forms of Human Hypertension and Mechanisms of Disease

P-Element

Definition

P-element denotes a class of transposable elements commonly used in *Drosophila* for gene transfer and mutagenesis.

► *Drosophila* as a Model Organism for Functional Genomics

Pemphigus

Definition

Pemphigus comprises of a group of rare autoimmune blistering diseases that can affect the skin and/or mucous membranes.

► Desmosomes

PEN-2

Definition

PEN-2 (Presenilin enhancer protein 2) functions as an activating component of the γ -secretase complex.

► Alzheimer's Disease

Penetrance

Definition

Penetrance defines the proportion to which individuals that carry a pre-disposing mutation develop the associated disease under a defined set of environmental conditions. Penetrance is more or less high in Mendelian diseases and tends to be low for complex genetic disorders.

► Atopy Genetics
 ► Familial Dilated Cardiomyopathy
 ► Genetic Epidemiology
 ► High-HDL Syndrome
 ► Huntington's Disease
 ► Microarrays in Pancreatic Cancer
 ► Morpholino Oligonucleotides, Functional Genomics by Gene 'Knock-down'
 ► Repeat Expansion Diseases
 ► Splicing

Peptamers

► Peptide Aptamers

Peptide

Definition

Peptides are molecules formed by the linking of various amino acids in a defined order. The individual amino acid residues are covalently linked by an amide bond (sometimes referred to as a peptide bond) between the carboxyl and amino groups.

► PNA Chips

► Proteases and Inhibitors

Peptide Aptamers

PIERRE COLAS

Aptanomics, Lyon, France

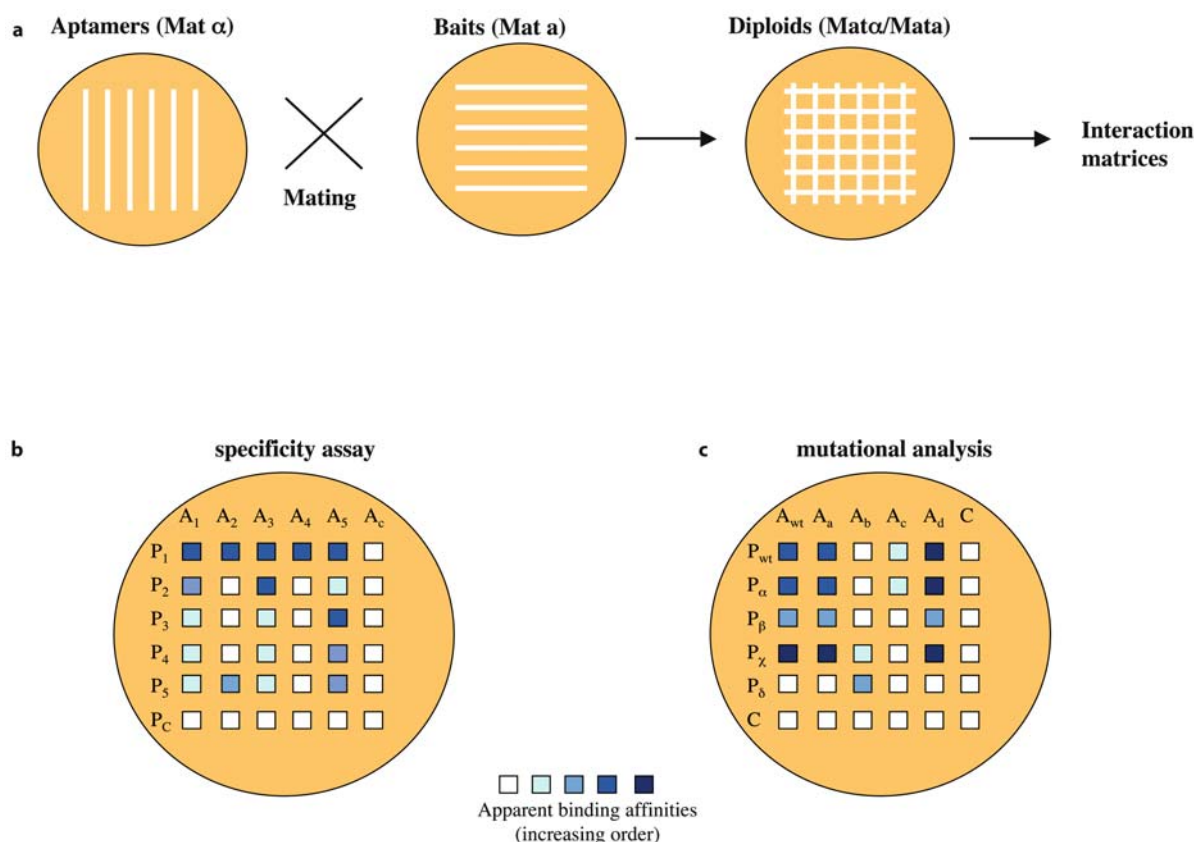
pierre.colas@aptanomics.com

Synonyms

Perturbagens, Peptamers

Definition

Peptide aptamers (from the Latin aptus: to fit) are an emerging class of recognition molecules that typically consist of a constant protein scaffold displaying a variable peptidic sequence inserted within the scaffold (1, 2). Two key features distinguish peptide aptamers from other combinatorial proteins (3). First, the variable region is inserted within the coding sequence of the protein scaffold in such a way that it is fused and thus constrained at both its amino- and carboxy-termini. This double constraint protects the aptamers from proteolysis and promotes increased binding affinity by reducing their conformational freedom. Second, peptide aptamers are generally isolated from combinatorial expression libraries through the use of intracellular selection methods. This contrasts with most combinatorial biology methods where ligands are selected *in vitro* and are thus often not suitable for intracellular targets. The ► yeast two-hybrid system is



Peptide Aptamers. Figure 1 Two-hybrid mating assays to determine peptide aptamer binding specificity and to map target binding site and aptamer pharmacophore. (a) Plasmids directing the expression of aptamers fused to an activation domain on the one hand, and of proteins fused to a DNA binding domain (baits) on the other hand, are transformed into yeast strains of opposite mating type. The transformants are streaked linearly on their respective selection media, as depicted. The streaks are then crossed on rich medium to allow yeast to mate at each intersection. The resulting diploid yeasts are then grown on selective medium. Different two-hybrid reporter genes can be used. They are typically carried by a plasmid that is co-transformed with the bait plasmids. Here, the *lacZ* gene is employed, conferring a blue phenotype on a X-gal-containing medium. The intensity of the blue color reflects the apparent binding affinity. Other reporter genes such as GFP or luciferases can be used to perform more quantitative binding assays. (b) specificity assay of a collection of peptide aptamers selected for their ability to bind a protein P_1 . P_{2-5} are proteins that belong to the same family as P_1 . P_c and A_c are an unrelated bait and aptamer, respectively. A typical result is schematized, where aptamer A_4 is highly specific for P_1 , aptamer A_2 exhibits a dual specificity for P_1 and P_5 , and the other aptamers show various specificity profiles against the protein family. (c) mutational analysis of a target-aptamer complex. P_{wt} and A_{wt} are the non-mutated protein and aptamer, respectively. $P_{\alpha-\delta}$ and A_{a-d} are different mutated variants (typically single aminoacid substitutions). A_a and P_a carry silent mutations. A_c and P_β carry loss-of-function mutations, whereas A_d and P_γ carry gain-of-function mutations (here, binding is assimilated to function). A_b and P_δ both carry loss-of-function mutations that compensate for each other and restore binding. Such mutational analyses greatly benefit from the use of reporter genes supporting a more accurate quantification of the binding affinities.

widely employed to select peptide aptamers for their ability to bind a given target protein. Peptide aptamers are also isolated for their ability to confer a selectable phenotype on a cell population, through a **transdominant genetic effect**. Fewer aptamers have been selected through extracellular display systems or have been designed from naturally occurring protein-binding surfaces. Peptide aptamers specifically bind target

proteins and modulate their function, often by inhibiting their interaction with other proteins. Hence, peptide aptamers are powerful tools for unraveling molecular regulatory networks and for identifying and fully validating new therapeutic targets. Moreover, peptide aptamers hold great promise in guiding the discovery of new therapeutic molecules that inhibit intracellular protein interactions.

Characteristics

Molecular Design

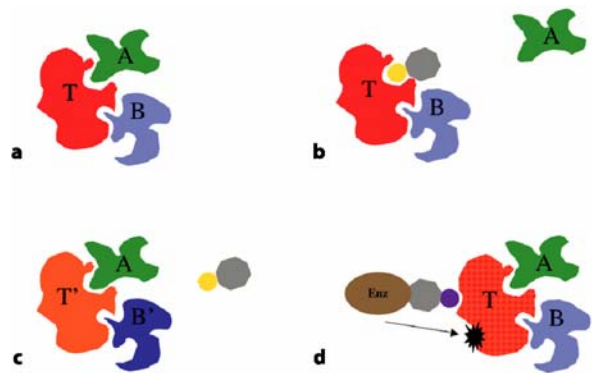
So far, the most frequently used peptide aptamer scaffold has been thioredoxin, a small ubiquitous disulfide reductase that maintains intracellular proteins in their reduced state. To construct combinatorial libraries, oligonucleotides of random sequence are cloned into the cDNA of thioredoxin, within its active site (2). Other scaffolds such as green fluorescent protein (GFP) or a catalytically inactive mutant of staphylococcal nuclease have also been used. Most libraries reported so far have featured variable peptidic loops of 20 amino acids, but this length can be adjusted as desired.

Specificity and Affinity

Peptide aptamers exhibit precise specificities for their target proteins. Typically, they can discriminate between different members of a protein family (e.g. the cyclin-dependent kinases) or even between different allelic variants of a given protein (e.g. Ras). The determination of their binding profile towards a protein family or a panel of mutants can be advantageously performed through two-hybrid interaction mating assays (Fig. 1a,b). The binding affinity of peptide aptamers can lie within the nanomolar range and can be further improved by random mutagenesis of the peptidic loop, followed by a more stringent selection (2). Semi-quantitative two-hybrid assays can accurately determine relative binding affinities between a target protein and a collection of peptide aptamers.

Inhibition of Protein Function

Most peptide aptamer collections selected for their ability to bind a target protein contain aptamers that inhibit the function of this protein and that produce compelling phenotypes when expressed in cellular models (1). Examples include aptamers targeting cell-cycle regulators (Cdk2, E2F, cyclin J), viral proteins (HPVE6 and E7, HBV core protein), a small G-protein (Ras), a small G-protein activator (Rho-GEF), a receptor tyrosine kinase (EGF-R) and a transcription factor (Stat3). Some of these aptamers have been shown to produce phenotypes when expressed in *Drosophila*. In all cases, it has been shown or strongly suggested that peptide aptamers interfered with their target protein by preventing it from interacting with one or more of its natural partners (Fig. 2). When aptamers are selected for their ability to produce a selectable phenotype, the challenge is to identify their target protein. This can be done by screening two-hybrid cDNA libraries against the selected aptamer, which is used as a bait protein. A complementary method consists of producing the selected aptamer in *E.coli*, coupling it to a solid



Peptide Aptamers. Figure 2 Peptide aptamers to manipulate regulatory networks. (a) a protein complex is depicted, where a target protein T interacts simultaneously with two partners A and B. (b) a peptide aptamer selected against T disrupts the T-A interaction without affecting the T-B interaction. (c) a similar protein complex formed between T' (closely related to T), A and B' (closely related to B) is not affected by the aptamer because it discriminates between T and T'. (d) a peptide aptamer derivative is built by fusing the catalytic domain of an enzyme to an aptamer selected against T, which is normally not a substrate of this enzyme. The aptamer derivative directs the specific modification of T by the enzyme (for example ubiquitination or phosphorylation).

phase and capturing the target protein from a cellular lysate. The captured proteins are identified by mass spectrometry.

Manipulations on Molecular Regulatory Networks

Peptide aptamer derivatives have been designed to enhance the range of manipulations that can be exerted on regulatory networks, beyond the disruption of protein interactions (1). Such derivatives consist of fusion proteins comprising a targeting domain (a peptide aptamer) and an effector domain (e.g. an enzyme catalytic domain or an addressing sequence). Peptide aptamer derivatives have been shown to direct the specific ubiquitination of their target protein in living cells or to modify its intracellular localization (Fig. 2d).

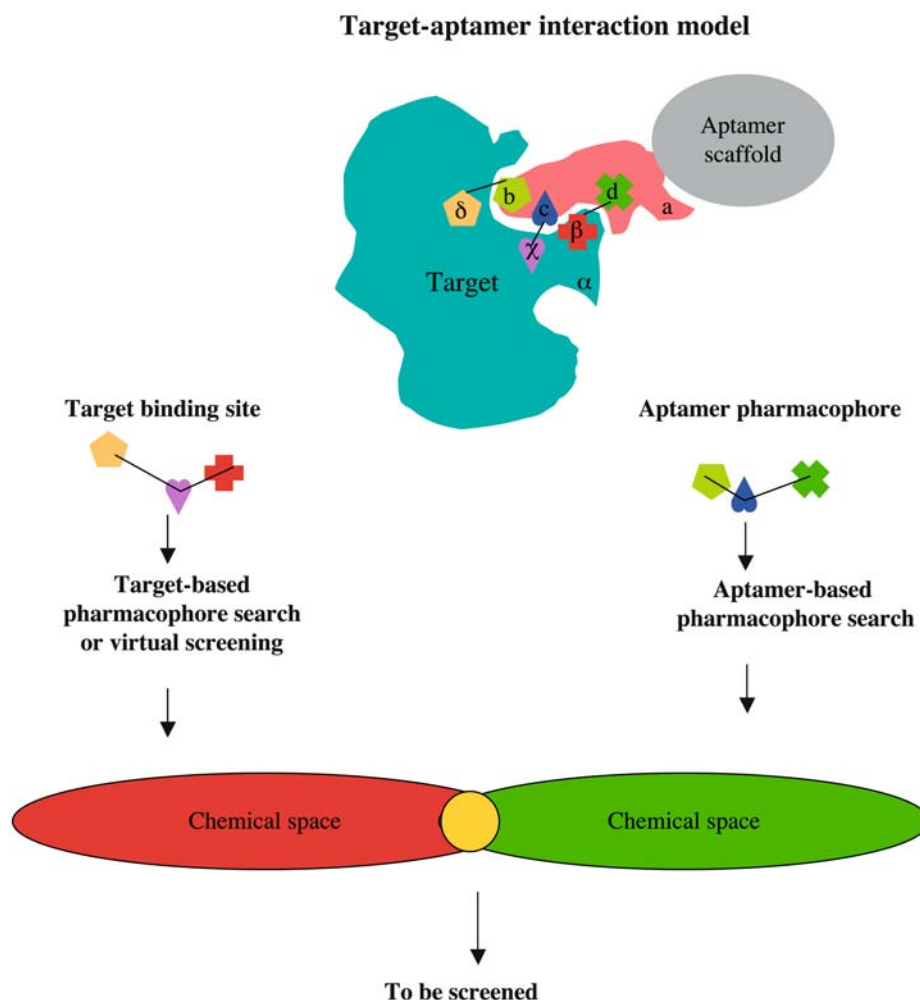
Clinical Relevance

Identification and Validation of Therapeutic Targets

The problem of validating therapeutic targets remains highly complex because it relates to fundamental biological questions and it suffers from the inadequate nature of many of the proposed methods. Biological processes are finely controlled by highly intricate protein interaction networks. Methods that inhibit or abolish gene expression (RNAi, antisense technology, gene knockout) often introduce severe molecular perturbations within such networks. In some cases,

no clear phenotype is obtained because compensatory or redundant mechanisms within protein families maintain homeostasis. In other cases, a strong phenotype is observed because of extensive secondary

deregulation within the network. Therefore, putative therapeutic targets may be either wrongly ruled out or wrongly validated if evaluated solely by such disruptive methods. Peptide aptamers introduce finer



Peptide Aptamers. Figure 3 Peptide-aptamer based discovery of small molecule drugs. A model of the interaction (not to scale) between a target protein and its peptide aptamer is obtained from structural studies and the mutational analysis depicted in Figure 1c. A simple case is presented, with no contribution of the aptamer scaffold (other than the double constraint imposed on the peptidic loop) to target binding. The model defines a binding site on the target protein, and a pharmacophore on the peptide loop of the aptamer. The colored symbols represent various physico-chemical descriptors that mediate binding (such as hydrogen bond donors and acceptors, hydrophobic and charged patches, defined spatial constraints, etc.). Computational chemistry software is then used to define small focused libraries enriched in small organic molecules that fit the target binding site and/or that conform to the aptamer pharmacophore. For a target-based or an aptamer-based pharmacophore search, the first step consists of defining the spatial distribution of a small set of descriptors from the structure of the target binding site (δ - χ - β) or from the structure of the binding determinant on the aptamer (b-c-d), respectively. The software then analyzes a virtual chemical library assembled from commercially available molecules and optionally from molecules that can be easily synthesized. Those molecules that do not comply with the defined sets are excluded. This low-resolution method is performed rather quickly with limited computational power. Virtual screening is a more accurate method that can be used whenever the target binding site is amenable to reliable molecular docking (typically deep pockets). Here, an exhaustive atomic representation of the pocket rather than a limited set of descriptors is defined. All molecules from the virtual library are screened against this pocket and are assigned a score that reflects the predicted fit. This method is much slower than pharmacophore search (one molecule per minute) and requires a higher computational power. The small focused libraries defined by these *in silico* approaches are then screened.

molecular perturbations that, like those exerted by small molecule drugs, are a direct consequence of the targeting of a specific binding site on the target. Hence, peptide aptamers validate therapeutic targets with higher confidence levels than other methods and they also provide a first indication of the druggability of the targets. Moreover, phenotypic selections of peptide aptamers allow the identification of new therapeutic targets, which are *de facto* validated prior to their identification.

Aptamers as Biotherapeutics

In view of the increasing therapeutic successes of ►monoclonal antibodies, it is conceivable to use peptide aptamers as biotherapeutic molecules. However, the usual issues of stability, toxicity and immunogenicity will probably require the exploration of various scaffold proteins. An additional difficulty will arise when targeting intracellular proteins, although it has been shown that peptide aptamers can penetrate living cells when fused to ►protein transduction domains (PTDs).

Aptamers as Guides to the Discovery of Therapeutic Molecules

Peptide aptamers are powerful guides to the discovery of small molecule drugs. Selected from combinatorial unbiased libraries, peptide aptamers “interrogate” protein surfaces to find new potential druggable sites. Various screening assays (such as yeast two-hybrid or fluorescence polarization) can be set up to discover small molecules that disrupt the interaction between a peptide aptamer and its target protein. Those hits that displace this interaction by targeting the protein are candidate mimics of the aptamer. To enhance the hit rates, a structure-based approach can be undertaken to circumscribe the chemical spaces to be screened. The structure of the aptamer bound to its target can be solved by NMR or the complex can be co-crystallized. The outcome of this approach is facilitated when the structures of the target protein and the aptamer scaffold are already known (as for thioredoxin). Besides these structural studies, the modeling of the interaction is greatly facilitated by performing quantitative two-hybrid assays with mutated variants of both target protein and peptide aptamer (Fig. 1a,c). A model of interaction defines a binding site on the target and a ►pharmacophore on the aptamer, which modern computational chemistry tools can convert into small molecule candidates to be screened (Fig. 3).

Peptide Aptamers to Exploit Protein Interactions for New Therapies

There is now mounting evidence that protein-protein interactions constitute a huge but still largely untapped pool of therapeutic targets. Although typical binding

interfaces are rather large, they contain “hot spots” that concentrate binding energy and can be targeted by small molecules. Moreover, many protein interactions can be inhibited through allosteric mechanisms, without the need for the aptamer or the therapeutic molecule to directly interfere with the binding interface.

Considerable resources have been invested collectively in establishing highly complex protein interaction maps in order to decipher molecular regulatory networks and to identify putative targets. The first challenge is now to identify which interactions are worth pursuing as therapeutic targets. The second major challenge is to discover small molecule drugs that disrupt these interactions. Peptide aptamers appear to be ideal tools to help achieve these two exciting endeavors.

References

1. Hoppe-Seyler F, Crnkovic-Mertens I, Tomai E et al (2004) Peptide aptamers: specific inhibitors of protein function. *Current Molecular Medicine* 4:529–538
2. Geyer CR, Brent R (2000) Selection of genetic agents from random peptide aptamer expression libraries. *Methods in Enzymology* 328:171–208
3. Colas P (2000) Combinatorial protein reagents to manipulate protein function. *Current Opinion in Chemical Biology* 4:54–59

Peptide Chips

RUDOLF VOLKMER-ENGERT

Molecular Libraries, Institute for Medical Immunology, Charité – University Clinical Center Berlin, Berlin, Germany
rve@charite.de

Definition

Peptide chips are miniaturized peptide arrays that are used as assay systems for large-scale and high-throughput biological analysis. In a peptide microarray format, peptides as capture molecules are immobilized in a small area on planar solid surfaces and probed for various biological or biochemical activities. Peptide arrays refer to a format of flat surfaces on which peptide sequences are organized in rows and columns which can be independently arranged. The array format uses a precise, spatially ordered arrangement of compounds that allows them to be examined side-by-side. This is in contrast to the library approach in which a huge set of related compounds are tested in batch followed by an additional selection step to isolate the candidate molecules with desired biological properties.

Peptide arrays can be generated on different planar solid surfaces by several methods, including deposition of pre-synthesized peptides with the aim of robots or by parallel peptide synthesis methodologies directly on a planar solid support. The density of the sequences on a peptide array ranges from 20 peptides per cm^2 (peptide macroarray) up to over a few thousands in a square cm area (arrays of minute size so called microchips). The typical peptide microarray is produced with densities of approximately 200 sequences per cm^2 on a glass slide and is called peptide chip.

Description Introduction

The examinations of protein activities and function have been carried out historically on the analysis of single molecules. However, the rapid pace of discovery of new gene products by genomic and proteomic initiatives have necessitated the design of alternative strategies for analyzing protein function. The possibility to determine thousands of different parameters in a single experiment perfectly fits the needs of so called “omic” approaches in biology. As demonstrated by DNA microarray experiments (“▶DNA chip”) the array format is a robust, reliable and established method for the global analysis of gene expression and allows a global view into the ▶transcriptome of a living cell or organism on the basis of a single experiment (1). But, it became obvious that the information obtained from DNA chips is not sufficient to understand life in general. The expression level of genes or mRNA does not close correlate with the abundance of functional protein molecules that are the principal components in organizing and mediating the physiological functions of any living cell. A proteome comprises not only sequence information about all the proteins in a cell, but also their quantity, localization, activity, modifications, interaction networks and regulation. In general, protein and peptide arrays are able to support proteomic research. Whereas protein arrays (reviewed in 2) are suited to investigate which protein interact with each other, peptide arrays enable one to precisely characterize molecular recognition events at the amino acid level (3). Peptides are chemically quite resistant compounds and, in contrast to proteins, usually have no folding requirements to achieve their active conformation. Beside their privileged role as biologically active molecules, peptides are accessible in a high degree of structural diversity through modern peptide synthesis approaches and molecular biology. Their chemistry is well established and routine. Chemical synthesis allows incorporation of non-natural building blocks, preparation of branched and cyclic structures and labeling with chromophores. Thus, peptides are perfect probe molecules for a robust screening assay.

Peptide Array production

Normally, two main principles are used for the preparation of peptide arrays. Firstly, the *in situ* peptide synthesis performed directly on the array surface. Secondly, the robotic supported immobilization of pre-synthesized peptide derivatives on modified array surfaces.

The direct peptide synthesis on a surface is seen as effective means of miniaturization of traditional chemical reactions. Peptides were prepared in parallel on the surface by a step by step solid phase peptide synthesis approaches. Two methods are reported in literature: (i) The light-directed, spatially addressable parallel chemical synthesis (reviewed in 4) and (ii) the SPOT synthesis concept (reviewed in 3 and 5).

The first one is a technology that permits extreme miniaturization of the peptide array formats. Addressable peptide arrays have been generated by this approach, using photomasks and photolability-protected amino acids. The synthesis involves multi-cycles with each cycle using light irradiation through a photomask followed by adding one type of amino acid to specific locations where the photo-labile protecting group on the terminal of the growing chain has been removed upon light irradiation. The amino acids used in the synthesis are not conventional but protected at the amino position with a photo-labile protecting group. Although promising for high-density microchips, the process has its limitations because production of photomasks is time consuming, expensive and requires a high-grade clean-room environment. Furthermore, most of the amino acids containing photolabile protecting groups are not commercial available and the photochemistry used is less efficient than conventional

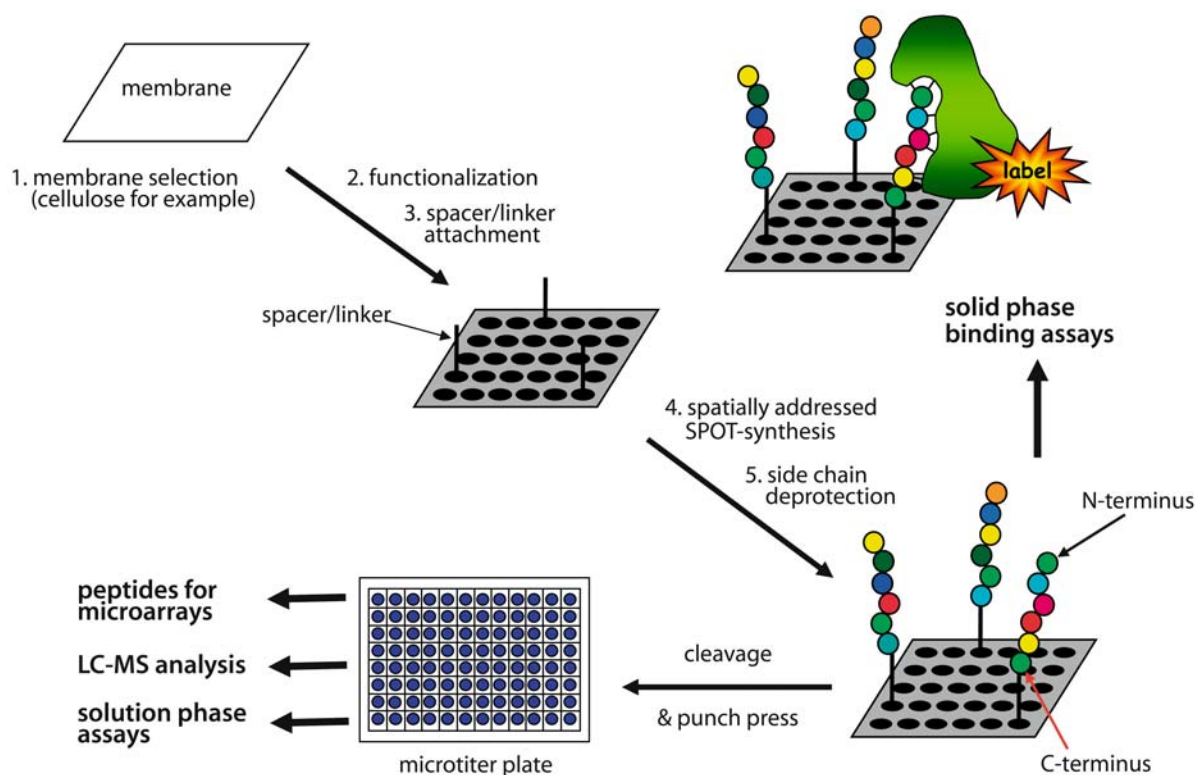
▶Boc chemistry or ▶Fmoc chemistry. That results in a poor stepwise yield and reduces the quality of the surface-bound peptides. As a consequence higher rates of false positive and false negative results are observed in the screening assay. An interesting alternative is the use of a novel photogenerated reagent chemistry combined with digital ▶photolithography and standard Boc chemistry (6). It based on an addressable photo-generated acid-formation using digital photolithography. Subsequently, the generated acid deprotects the t-Boc group of the growing peptide chain and releases the amino-terminus which is then amidated by an activated amino acid. Up to 4000 features in a 1.5 cm^2 area were synthesized on a microchip format (4).

The SPOT technique developed by Ronald Frank (7) consists of the stepwise synthesis of peptides on planar supports, such as amino-functionalized cellulose membranes or aminated polypropylene, applying standard Fmoc chemistry. The basic principle involved the positionally addressed delivery of small volumes of activated amino acid solutions directly onto a coherent

membrane sheet. The areas wet by the resulting droplets can be considered as micro-reactors provided that a non-volatile solvent system is used. The functional groups fixed on the membrane surface react with the pipetted reagents and conventional solid phase syntheses occur. The volumes dispensed as well as the physical properties of the membrane surface and the solvent system define the size of the resulting spots and together with the minimum distance between the spots the number of peptides which can be synthesized per membrane area. The SPOT technique is technically very simple and flexible and does not require any expensive laboratory automation or synthesis hardware. The degree of miniaturization using the SPOT synthesis concept is significant lower than the microchip format described above. In general up to 20 peptides can be prepared in a square cm area. Peptide arrays prepared by the SPOT technique became popular tools for studying numerous aspects of molecular recognition. Over the past decade ►B-cell epitope and ►paratope mapping, characterisation of numerous protein-protein interactions, enzyme-substrate recognition, inhibitory activity, peptide-metal ion interaction, peptide nucleic acid binding, ►T-cell

epitope mapping and the direct action of peptides *in vivo* have all been explored using synthetic peptide arrays on cellulose membranes prepared by the SPOT technique. Moreover, the SPOT technology has become a highly parallel and high throughput peptide synthesis tool. It enables the parallel synthesis of up to 2000 peptides in a very short time (one week) and at amount of approximately 50–100 nmol. Hence, SPOT synthesis is the prerequisite for the preparation of peptide microarrays (peptide chips). Figure 1 schematically shows the basics and applications of the SPOT technique.

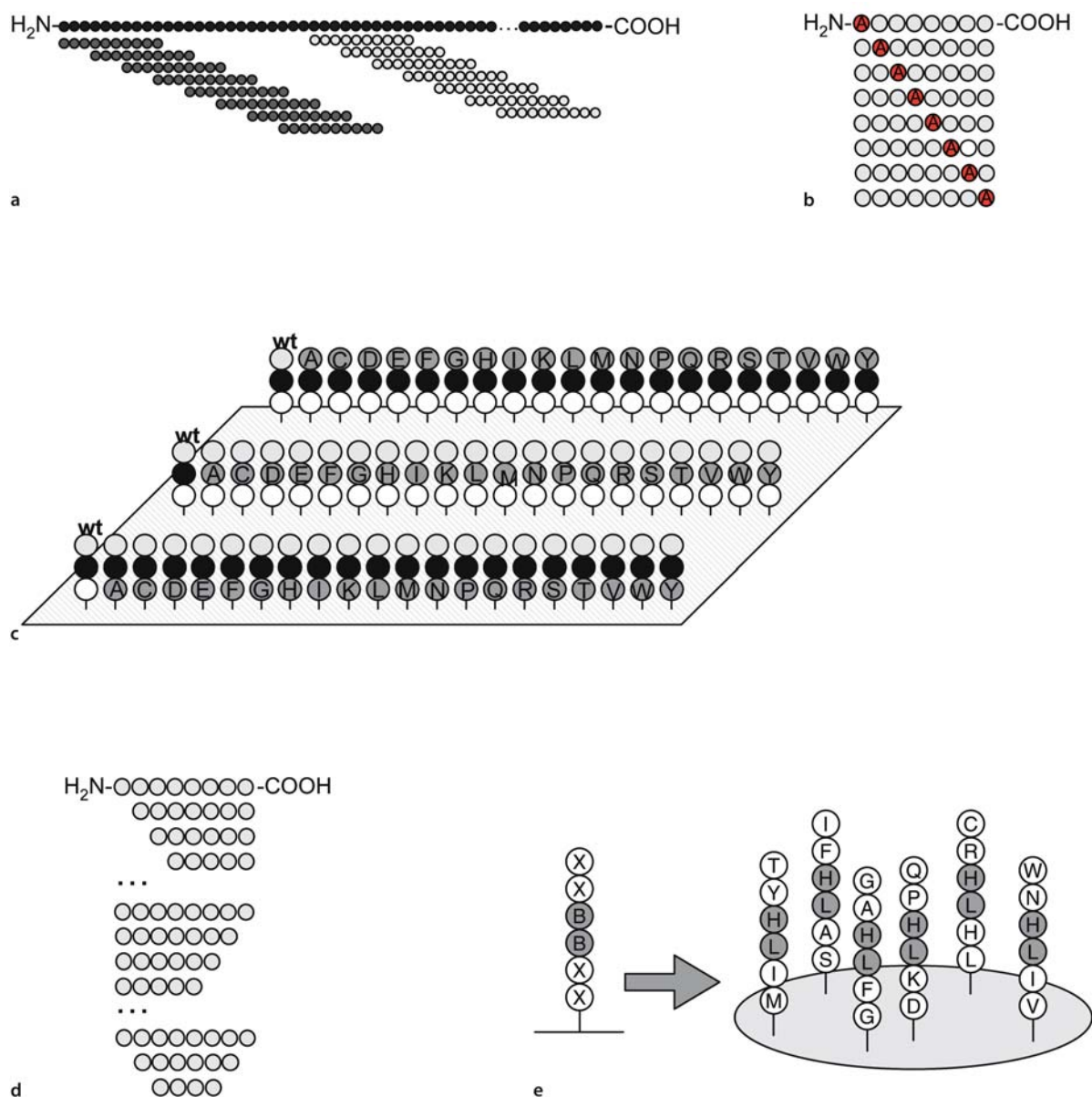
When large numbers of peptide arrays with the same sequences are required, immobilization of pre-synthesized peptides is more economical than *in situ* synthesis. In general long peptide sequences have to be purified to obtain high-quality products. Therefore immobilization is also the method of choice for that kind of peptides. Typically immobilization takes place on surface-functionalized glass slides, but self-assembled monolayers on gold surface have also been employed as a surface-modification tool. Chemoselective immobilization reactions are of particular interest in preparation of peptide arrays because they allow



Peptide Chips. Figure 1 Schematic view of the SPOT synthesis technique. Peptide arrays prepared by SPOT synthesis can be applied to solid phase binding assays as well as solution phase binding assays. With the possibility to cleave the compounds from the planar support peptide could be analyzed and applied to the fabrication of peptide chips.

control over both the orientation and the density of the attached peptide. Different chemoselective reactions were used for peptide microarray preparation. Only a few examples are given. An aldehyde function at the surface of a glass slide in combination with aminoxy-acetyl moieties in the peptide or cysteinyl residues was used for the preparation of peptide microarrays on glass slides. Also the native chemical ligation method is well suited for effective attachment of N-terminally

labelled peptides to thioester-modified glass slides. A more sophisticated reaction for an oriented immobilization of peptide derivatives is the Diels-Alder reaction between bezoquinone groups on self-assembled monolayers and cyclopentadiene-peptide conjugates. In the most cases of chemoselective peptide immobilization a spacer molecule is inserted between the peptide and the surface. This is done in respect to the accessibility of the surface-bound peptides to the proteins used in the



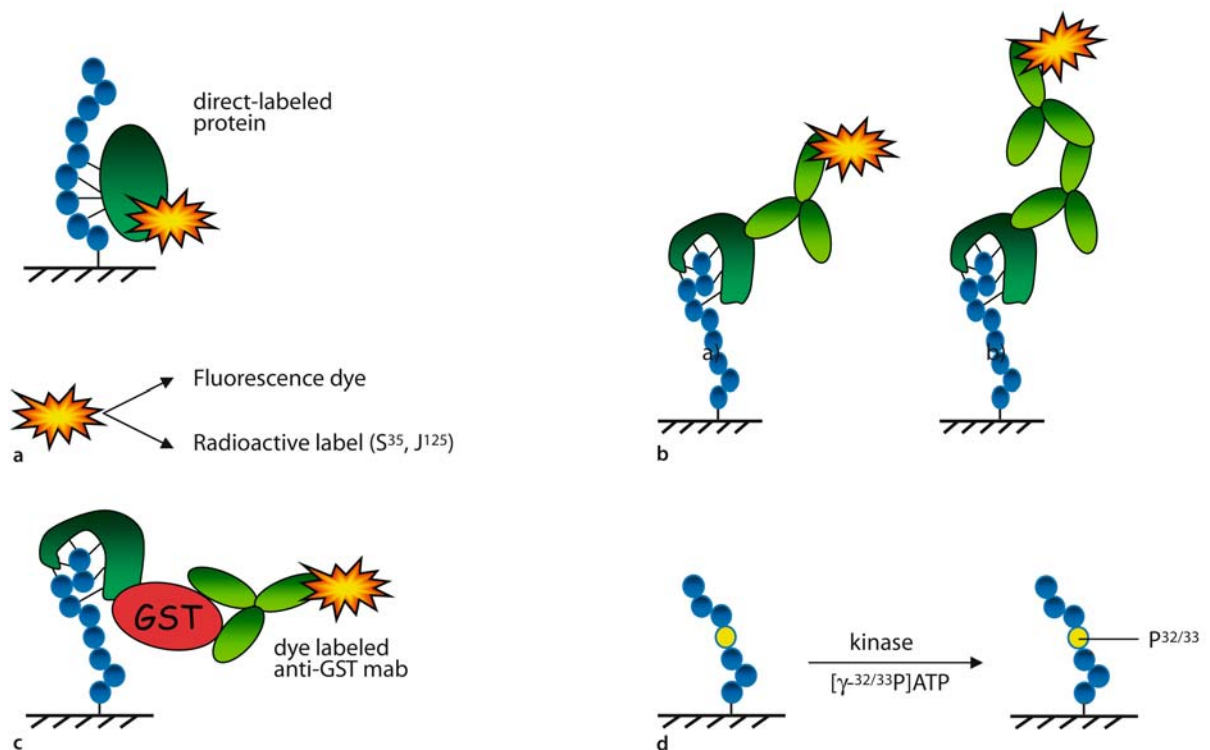
Peptide Chips. Figure 2 (a) The primary sequence of the protein to be examined is used to generate short linear overlapping 10mer peptides shifted by three amino acids (peptide scan). (b) Amino acid substitution scan (alanine scan) of an octamer peptide. (c) Complete substitutional analysis of a trimer peptide. (d) Truncation library with N-terminal, C-terminal, and bidirectional stepwise truncation. (e) Peptide mixtures with defined position B and randomized position X generated on one spot.

screening. Insertion of hydrophilic dextran structures between the surface and the presented peptide or incorporation of 1-amino-4,7,10-trioxa-13-tridecanamine succinimic acid as a spacer moiety have been described. Microarrays prepared by immobilization of pre-synthesized peptides on glass slides are predominantly used up to now for the identification and optimization of ►kinase substrates. They are produced typically with a peptide density of approximately 200 sequences per cm².

Peptide Library Types

A considerable number of different library types have been used for the investigation molecular recognition events. Two general types can be defined: knowledge-base libraries, comprising peptides with sequences that are derived from naturally occurring proteins, and libraries that are designed *de novo*, that is, either consisting of randomly generated single peptides or peptide mixtures based on combinatorial principles. The first knowledge based library is of particular interest for the mapping of protein-protein binding

sites, identification of the actual phosphor-acceptor residue of a known protein as a kinase's target and B-cell epitope or T-cell epitope mapping. Identification of the binding site of interests is achieved by a scan of overlapping peptides derived from the sequence of one or both interaction partners (Fig. 2a). The entire sequence of the protein whose epitope to another protein has to be determined is synthesized as a group of overlapping peptides usually between 8 and 15 amino acids in length which are shifted between one and five positions. Incubation of such a peptide scan with the binding partner and detection of peptide-bound protein by immunological methods, enzymatic reactions with chemoluminescent or chromogenic substrates, fluorescence or autoradiography reveals one (linear epitope) or more (discontinuous epitope) stretches of the protein which are part of the binding site. For closer characterization of the key residues of such binding regions i. e. amino acids with side chains which are effectively in contact with the receptor protein, substitution analyses are carried out by exchanging every position – but only one at a



Peptide Chips. Figure 3 Assay types and detection methods applied to peptide microarrays. (a,b,c) show the detection of protein binding to an immobilized peptide. (a) Detection of binding is mediated by the fluorescence or radioactive label directly incorporated at the protein. (b) In the case a fluorescently labelled non neutralization antibody is applied. In case b) binding is mediated by a secondary labelled antibody. (c) Protein-peptide interaction is mediated via a protein tag (here a GST tag) in combination with a fluorescently labelled antibody. (d) Detection of peptide phosphorylation on peptide microarrays. The array is incubated with the kinase of interest in the presence of [γ-³² or ³³P]ATP, and autoradiography is used for detection.

time – by either alanine (Fig. 2b) or by all other L-amino acids (Fig. 2c) resulting in all possible single site substitution analogs. Key residues may not be exchanged by any other or only by physicochemically similar amino acids without loss of binding. Truncation libraries (Fig. 2d) are used to determine the minimal length of the peptide sequence that is still recognized by the protein of interest. For *de novo* detection of B cell epitopes and kinase substrates both combinatorial approaches and randomly generated libraries of single peptides have proven to be useful. Combinatorial libraries have one or more defined amino position and a number of randomized or degenerated positions. Only one amino acid is introduced at the defined positions, while a mixture of amino acids is introduced at the randomized position. This results in a sublibrary of different sequences in each single spot (Fig. 2e).

Assays and detection

Detection methods in analyzing any array are required to offer high throughput, high signal-to-noise ratio, relatively low instrumentation costs, good resolution, and reproducible results. Many applications have used fluorescently labeled detection methods, because they are simple and stable to manipulate, they provide highly sensitive and resolved results and they are compatible with the standard array scanners developed for DNA microchips. Although radioactivity is also suitable for analyzing arrays, especially for enzymatic phosphorylation, due to sensitivity and specificity as well as the possibility of fluorescence detection, the use of isotope-labeled molecules raises safety concerns. Chemiluminescence is also highly sensitive however at present gives relatively lower resolution and relatively limited dynamic range. This detection method is mostly applied to the field of SPOT synthesis. Figure 3 shows assays as well as detection methods that are typically applied to peptide microarrays.

References

1. Fehrenbach E, Zieker D, Niess AM et al (2003) Microarray technology – the future analyses tool in exercises physiology. *Exerc Immunol Rev* 9:58–69
2. Zhu H, Snyder M (2001) Protein arrays and microarrays. *Curr Opin Chem Biol* 5:40–45
3. Reineke U, Volkmer-Engert R, Schneider-Mergener J (2001) Applications of peptide arrays prepared by the SPOT-technology. *Curr Opin Biotechnol* 12:59–64
4. Gao X, Pellois JP, Na Y et al (2004) High density peptide microarrays. *In situ* synthesis and applications. *Mol Divers* 8:177–187
5. Frank R (2002) The SPOT-synthesis technique. Synthetic peptide arrays on membrane supports – principles and applications. *J Immunol Methods* 267:13–26
6. Pellois JP, Zhou X, Srivannavit O et al (2002) Individually addressable parallel peptide synthesis on microchips. *Nat Biotechnol* 20:922–926
7. Frank R (1992) Spot synthesis: An easy technique for the positionally addressable, parallel chemical synthesis on a membrane support. *Tetrahedron* 48:9217–9232

Peptide Mass Fingerprinting

Definition

Peptide Mass Fingerprinting (PMF) is a mass spectrometry based method for protein identification. The protein is cleaved by an enzyme with high specificity (trypsin, Lys-C, Arg-C, Asp-N, etc.) or chemically (CNBr). The generated peptide mixture is analyzed by ▶[Matrix-Assisted Laser Desorption/Ionisation \(MALDI\)](#) or electrospray ionisation (▶[ESI](#)) mass spectrometry. The determined set of masses (mass fingerprint) is characteristic for the protein present and is used to search peptide masses generated by theoretical fragmentation of protein sequences of databases.

▶[Mass Spectrometry: MALDI](#)

Peptide Mass Map

Definition

A protein can be digested by a protease (usually trypsin). The mass spectrum of the resulting peptides provides a peptide mass map. Because trypsin cleaves after Lys and Arg, the mass map is dependent on the sequence of the protein. Therefore, the peptide mass map can be used to identify the protein after database searching.

▶[Proteomics in Cancer](#)

Peptide Release Factor 2

Definition

Peptide release factor 2 is the enzyme that is responsible for recognizing the codons UAA and UGA in prokaryotes, and for causing termination of protein synthesis.

▶[Translational Frameshifting, Non-standard Reading of the Genetic Code](#)

Peptidomimetic

Definition

Peptidomimetic molecules are molecules that mimic structural features of peptides, but lack some or all of

the amide bonds that link the amino acids in peptides together.

►Protein Prenylation

Peptidyl Prolyl *cis/trans* Isomerases

CORDELIA SCHIENE-FISCHER

Max Planck Research Unit for Enzymology of Protein Folding, Halle, Germany

schiene@enzyme-halle.mpg.de

Synonyms

A short term for peptidyl prolyl *cis/trans* isomerization is prolyl isomerization. According to the recommendations of the Nomenclature Committee of the IU/BMB, enzymes catalyzing prolyl isomerization were named peptidyl prolyl *cis/trans* isomerases in 1984. The colloquial terms prolyl isomerase and proline isomerase are frequently in use.

The alternative name rotamase, launched later on, does not fit the rules of IU/BMB, communicating an imprecise picture of the enzyme function. Also in use is the term immunophilin. Peptidyl prolyl *cis/trans* isomerase cannot be considered synonymous with immunophilins. An immunophilin is molecularly described in terms of high affinity for the immunosuppressant drugs cyclosporin A, FK506 or rapamycin characterizing some members of the cyclophilin and FKBP family. Regarding immunosuppressant affinity, many proteins not related to peptidyl prolyl *cis/trans* isomerases are immunophilic (ubiquitin, G-actin, interleukin 8, and many others), whereas many peptidyl prolyl *cis/trans* isomerases do not exhibit immunophilic properties. Alternative names for FK506 and rapamycin are tacrolimus and sirolimus, respectively.

Definitions

Peptidyl prolyl *cis/trans* isomerases (PPIases, E.C. 5.1.2.8) are enzymes evolved to accelerate the *cis/trans* isomerization of peptide bonds preceding prolyl residues in peptides and proteins. Due to the imidic structure, prolyl bonds have particular electron distribution and steric constraints resulting in both slow conformational interconversions and comparable equilibrium populations of the *cis* (dihedral angle $\omega \approx 0^\circ$) and the *trans* (dihedral angle $\omega \approx 180^\circ$) conformer in digopeptides and unfolded proteins (1). (Fig. 1)

At present, there are three families known, where the individual members can exhibit PPIase activity. They are termed ►cyclophilins (Cyp), ►FK506-binding proteins (►FK506 (Tacrolimus)) (FKBP) and parvulins, which are unrelated in their amino acid sequences,

have distinct substrate specificities and are sensitive to different types of inhibitors. The immunosuppressive drug ►cyclosporin A (CsA) and ►sanglifehrin avidly bind to and inhibit the enzyme activity of cyclophilins. FKBP are able to bind the immunosuppressive drugs ►FK506 or ►rapamycin. These compounds represent reversible tight-binding inhibitors for the majority of FKBP. Until now, a nonpeptidic reversible inhibitor of comparable affinity is not known for parvulins.

All three families comprise prototypic enzyme species consisting of a catalytic domain of 92 (parvulins), 109 (FKBP) or 163 (cyclophilins) amino acids. Many PPIases complement their catalytic domain with N- or C-terminal polypeptide segments of different chain lengths and functionality.

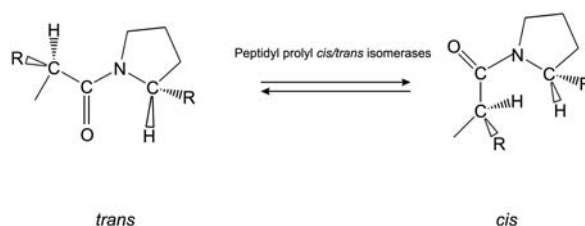
Regarding the substrate specificity, the PPIase families make use of a distinct set of amino acids flanking the proline residues. For cyclophilins, catalytic properties are relatively unaffected by the amino acid preceding proline. FKBP show a preference for substrates with a hydrophobic amino acid in the same position. The Ess1/Pin1 subfamily of the parvulins requires (PO₃H₂)-Ser/(PO₃H₂)-Thr-Pro moieties for high catalytic efficacy.

PPIases are abundant, ubiquitously expressed proteins. All genomes analyzed so far encode at least one PPIase but often far more. Among the 470 predicted coding regions in the bacterium *Mycoplasma genitalium*, which has the smallest known genome of any independently living organism, the FKBP-like trigger factor is the only PPIase present. In *Escherichia coli* there are 2 cyclophilins, 5 FKBP and 3 parvulins, in *Saccharomyces cerevisiae* there are 9 cyclophilins, 3 FKBP and a single parvulin. Presently, the SwissProt database (www.us.expasy.org/sprot/) annotates 16 cyclophilins, 18 FKBP and 2 parvulins for the human genome (2).

Characteristics

Assaying PPIase Activity

PPIases can be identified by their ability to accelerate the *cis* to *trans* isomerization of tetrapeptide substrates of the type Succinyl-Yaa-Xaa-Pro-Zaa-4-nitroanilide (Xaa, Yaa, Zaa are any natural amino acid) in an assay using the conformational specificity of proteases like chymotrypsin or trypsin as a kinetic probe. Practically,



Peptidyl Prolyl *cis/trans* Isomerases. Figure 1

the isomerization reaction is detected by means of the uv/visible absorbance of the released 4-nitroaniline. At high concentrations of the isomer-specific protease, the *trans* isomer of the substrate (~80–95%, dependent on Xaa) rapidly cleaves the anilide bond releasing the chromogenic 4-nitroaniline. Because the protease is not able to catalyze the cleavage of the remaining *cis* isomer (~5–20%) the *cis* to *trans* isomerization of inert conformer is the prerequisite for complete proteolytic cleavage, and thus forms the rate-limiting step for the remaining fraction of the reaction products produced. In the presence of PPIases the relaxation time of the uncatalyzed first-order isomerization reaction of about 125 s at 10°C can be decreased to the millisecond level. By dissolving substrates in dry LiCl/trifluoroethanol the content of *cis* isomer is occasionally increased allowing a higher signal to noise ratio of the assay. Other PPIase assays are based on the rapid disturbance of the *cis/trans* equilibrium of a proline-containing peptide followed by monitoring reequilibration kinetics in the presence of PPIases. As detection methods fluorescence, uv/visible and CD spectroscopy have been utilized. In favorable cases, dynamic NMR spectroscopy allows investigation of the kinetics of prolyl isomerization in peptides and proteins in an unperturbed equilibrium by magnetization transfer experiments. Prototypic PPIases are perfectly evolved enzymes in that they achieve relaxation times of the catalyzed reaction in the range of milliseconds concomitant with low substrate affinity. For oligopeptide substrates, both the Michaelis constant K_M and the turnover number k_{cat} are large, yielding k_{cat}/K_M values close to the diffusion-controlled limits for enzyme reactions ($k_{cat}/K_M > 10^7 \text{ M}^{-1} \text{ s}^{-1}$).

PPIases in ► Protein Folding

Prolyl isomerization often forms a late step in protein refolding because rapidly formed folding intermediates retain the prolyl bond in nonnative conformation and thus retard folding processes. The acceleration of slow folding events by PPIases may serve to avoid accumulation of aggregation-prone intermediates. Therefore, PPIases belong to the class of folding helper enzymes. In support of an *in vivo* role in protein folding, a PPIase of the cyclophilin family was shown to be involved in triple helix formation in type I procollagen in chick embryo fibroblasts. The trigger factors, a prokaryotic subfamily of FKBP, are localized to the large subunit of ribosomes, and can bind to nascent chains suggesting a crucial role in *de novo* protein folding in bacteria.

Several PPIases are able to accelerate the rate of prolyl bond limited kinetic phases in the *in vitro* refolding of ribonuclease T1 (RNase T1). This enzyme has two prolyl bonds (Tyr38-Pro39; Ser54-Pro55) in *cis* conformation in the native state. The site-directed

mutagenized, reduced and carboxymethylated variant Ser54Gly/Pro55Asn is used as standard protein substrate for PPIases. With $k_{cat}/K_M = 1.1 \times 10^6 \text{ M}^{-1} \text{ s}^{-1}$ the *E. coli* trigger factor catalyzed refolding of denatured RNase T1 variant Ser54Gly/Pro55Asn exemplifies a particularly efficient PPIase effect on protein folding *in vitro* (3). The efficacy of trigger factors results from presenting extended secondary binding sites for the unfolded protein substrate. The link between acquisition of biological activity and molecular details of the folding process was shown for the CH3 domain of the heavy chains of IgG. The formation of the essential *cis* isomer of the Phe34-Pro35 bond precedes protein dimerization constituting the rate-limiting step of this reaction. PPIase catalysis shifts the rate-limiting step from prolyl isomerization to fast chain collapse.

The identification of PPIases as crucial proteins in a variety of biological processes shows that these enzymes may play a biological role that is more general than just affecting *de novo* protein folding.

PPIases in Cell Signaling

Beside unfolded polypeptides, PPIases are able to make use of native protein substrates. This property allows them to influence those proteinaceous ligands that express biological activity only in one of the isomeric forms of a particular prolyl bond (2). Proline-directed protein kinases and phosphatases exhibit such a kind of isomer specificity in phosphate transfer reactions on oligopeptides and proteins. PPIases are able to affect the kinetics of phosphate transfer reactions thereby influencing the dynamics of subsequent phosphorylation-dependent chain rearrangements in the targeted proteins.

Definite identification of natural PPIase substrates has been reported as an exception. A large number of binding proteins, which may include the putative cellular substrates, were identified by chemical cross-linking, affinity chromatography, the yeast two-hybrid system or co-purification experiments for many PPIases. Because the detection of a catalyzed prolyl isomerization *in situ* proved to be very difficult, the role of PPIase catalysis *in vivo* has to be analyzed by the generation of PPIase variants with reduced enzymatic activity or by the use of specific cell-permeable PPIase inhibitors.

For example Cyp40 inhibits the DNA binding activity of the transcription factor c-Myb. This effect could be abrogated by the inhibitor CsA. The parvulin-like PPIase Pin1 with its preference to catalyze $(\text{PO}_3\text{H}_2)\text{-Ser}/(\text{PO}_3\text{H}_2)\text{-Thr-Pro}$ moieties interacts with a series of mitotic phosphoproteins, including Polo-like kinase-1, Cdc25C and Cdc27, and its catalytic activity is involved in cell division control in eukaryotes. This human parvulin also interacts with the C-terminal domain of RNA polymerase II; thereby altering the

interaction with proteins required for transcription of essential cell cycle genes. Inactivation of Pin1 induces mitotic arrest and ▶apoptosis.

A mitochondrial complex comprising the adenine nucleotide translocase of the inner membrane and matrix Cyp22 forms the permeability transition pore, which can open transiently allowing free permeation of low molecular weight molecules playing a central role in apoptosis. Immunosuppressive and nonimmunosuppressive cyclophilin inhibitors antagonize mitochondrial-mediated apoptosis by inhibition of the opening of the ▶mitochondrial permeability transition pore.

PPIases are found to form heterooligomeric complexes with different receptors. The intracellular ▶calcium release channels on the endoplasmic or sarcoplasmic reticulum, the ryanodine receptor and the inositol 1,4,5 triphosphate receptor were found to be associated with either cytosolic FKBP12 or FKBP12.6 acting as modulatory proteins to stabilize the channel gating functions. FKBP and cyclophilins with additional ▶tetratricopeptide repeat (TPR) motifs are found to be associated with ▶hsp90 as components of nuclear receptor heterocomplexes like unactivated steroid receptor complexes or the unliganded aryl hydrocarbon receptor. Prototypic PPIases contribute to the control of receptor and non-receptor protein kinases. For example, the phosphorylation of TGF β family type I receptor by the type II receptor chain is inhibited by FKBP12. Similarly, the interaction of Cyp18 with the SH2 domain of the interleukin-2 tyrosine kinase as a substrate controls the catalytic activity of the neighboring protein tyrosine kinase domain.

The enzymatic activity of PPIases is probably involved in many processes of pathophysiological relevance.

▶Familial amyotrophic lateral sclerosis associated mutant Cu/Zn superoxide dismutase-1 (SOD) induces apoptosis of neuronal cells associated with high levels of reactive oxygen species. Over-expression of wild type Cyp18, but not of Cyp18 variant Arg55Ala with a reduced enzymatic activity of about 0.1% of wild type Cyp18, protected cells from death after SOD Val148Gly expression, underlining the importance of PPIase activity in mutant SOD-mediated apoptosis. The FKBP-type PPIase macrophage infectivity potentiator (Mip) protein of *Legionella pneumophila*, the causative agent of Legionnaires' disease, represents a PPIase activity dependent virulence factor necessary for optimal intracellular survival of the pathogen.

Clinical Relevance

The prototypic PPIases Cyp18 and FKBP12 are thought to be important as presenter proteins for the immunosuppressive drugs CsA, FK506 and rapamycin in human (4, 5). A composite surface created by the drug and the surrounding amino acid residues of the respective PPIase enables PPIase/drug complexes to

recruit other cellular proteins involved in signal transduction. Both the Cyp18/CsA and the FKBP12/FK506 complexes interact with the protein phosphatase 2B (calcineurin) in an inhibitory manner, thereby inactivating the nuclear factor of activated T cells (NFAT) followed by the blockage of interleukin-2 transcription and suppression of T cell expansion. However lymphocytes of patients treated with CsA do not show dramatic effects on calcineurin activity. For the FKBP12/rapamycin complex, the mammalian target of rapamycin (mTOR), a phosphatidylinositol kinase-related protein kinase was identified as an interaction partner, resulting in the inhibition of the interleukin-2 stimulated transition from G1 to S phase in T cell replication. By the down-regulation of the T cell activation program, the drugs can prevent allograft rejection, and are prophylactic for the prevention of graft-versus-host disease following bone marrow transplantation.

Whereas rapamycin has a potential as an antitumor agent, the drugs CsA and FK506 have been shown to induce increased incidence of malignancy. CsA induces phenotypic changes, including invasiveness of non-transformed cells by a cell-autonomous mechanism. Apparently, alterations in the expression pattern of many PPIases coincide with malignant transformation. Pin1 is highly over-expressed in breast cancer. Deletions in the region of the *Cyp40* gene might be a late event in breast tumor progression. The prototypic cyclophilin COAS2, which has 83.6% sequence identity to the constitutively expressed Cyp18 was found to be over-expressed almost exclusively in aggressive metastatic or chemotherapy resistant tumors.

Incorporation of Cyp18 from the host cells into nascent human immunodeficiency virus type 1 (HIV-1) virions *via* direct interaction with the HIV-1 Gag polyprotein p55gag is required for virion infectivity. HIV-1-associated Cyp18 mediates HIV-1 entry into host cells *via* interaction with the surface receptor ▶CD147 of the targeted cell. (Me-Ile-4)cyclosporin (SDZ NIM 811) is a 4-substituted cyclosporin derivative that is devoid of immunosuppressive activity but retains full capacity for binding to Cyp18 and exhibits potent anti-HIV-1 activity.

▶Protein Folding

References

1. Fischer G (2000) Chemical aspects of peptide bond isomerization. Chem. Soc. Rev. 29:119–127 and references therein
2. Fischer G, Aumüller T (2003) Regulation of peptide bond *cis/trans* isomerization by enzyme catalysis and its implication in physiological processes. Rev Physiol Biochem Pharmacol. 148:105–50 and references therein

3. Schmid FX et al (1996) Catalyzed and assisted protein folding of Ribonuclease T1. *Biological Chemistry* 377:417–424 and references therein
4. Fischer G et al (1989) Cyclophilin and peptidyl-prolyl cis-trans isomerase are probably identical proteins. *Nature* 337:476–478
5. Harding MW et al (1989) A receptor for the immunosuppressant FK506 is a cis-trans peptidyl-prolyl isomerase. *Nature* 341:758–760

Peptidylglycine Amidation

Definition

Peptidylglycine amidation designates the oxidative process by which peptide hormones are amidated at their N-terminus. Almost half the known peptide hormones require an N-terminal amide to exert their biological function.

► [Enzyme Catalyzed Post-translational Hydroxylation of Proteins](#)

Peptidyl-Transferase Center

Definition

Peptidyl-transferase center is the site of peptide bond synthesis on the ribosome.

► [Genetic Code](#)

Percentage Chimaerism

Definition

Percentage chimaerism comprises of the amount of the mouse that was contributed by descendants of the ES cells (based on coat colour contribution).

► [Mouse Genomics](#)

Perforated Patch Recording

Definition

Perforated patch recording is a special configuration in patch clamp recording that allows to measure whole-cell currents or membrane potential changes, while the cell interior remains intact. Pore-forming molecules

allow electrical access to the cell, but prevent wash-out of intracellular constituents.

► [Patch Clamping](#)

Perichondrium

Definition

Perichondrium consists of flattened mesenchymal cells surrounding a skeletal element.

► [Bone and Cartilage](#)

Perinatal

Definition

The term perinatal refers to a time span around birth/delivery.

► [Bone and Cartilage](#)

Periosteum

Definition

Periosteum refers to the inner layer of mesenchymal cells surrounding a skeletal element, adjacent to the region of prehypertrophic and hypertrophic chondrocytes. At later stages, lining of the cortical bone surface facing the musculature contains osteoprogenitor cells.

► [Bone and Cartilage](#)

Peripheral Membrane Proteins

Definition

In contrast to integrale membrane proteins, peripheral membrane proteins do not contain a hydrophobic sequence of amino acids long enough to span a membrane. They are peripherally attached to the outer or inner leaflet of a lipid bilayer by hydrophobic modifications such as glypidation, palmitoylation or isoprenylation, or by protein-protein interactions with integral membrane proteins.

► [Biological Membranes](#)

► [Fatty Acid Acylation of Proteins](#)

Peripheral Nervous System

The peripheral nervous system (PNS) is a part of the nervous system and consists of the nerves (spinal nerves, cranial nerves) and neurons that reside or extend outside the central nervous system (CNS). The PNS can be divided into two major parts: the somatic nervous system and the autonomic nervous system. The somatic nervous system consists of sensory neurons that send sensory information running from stimulus receptors via the peripheral nerve fibers to the central nervous system, and of motor neurons running from the CNS to the muscles and glands - called effectors - that take action.

The autonomic nervous system is divided into three parts: the sympathetic nervous system, the parasympathetic nervous system and the enteric nervous system. The autonomic nervous system controls smooth muscle of the viscera (internal organs) and glands. Diseases of the PNS can affect the nerve roots, ganglia, plexi, autonomic nerves, sensory nerves, and motor nerves.

► [Hereditary Neuropathies, Motor and/or Sensor](#)

Periplasm

Definition

Periplasm refers to the space between the inner and outer membrane in Gram negative bacteria.

► [Camel as a Model for Functional Genomics](#)

Permeability Barrier

► [Skin \(Permeability\) Barrier](#)

Peroxin

Definition

Peroxin is a protein that is required for the formation, maintenance or proliferation of peroxisomes.

► [Peroxisomal Disorders](#)

Peroxiredoxin

Definition

Peroxiredoxins are peroxidases that reduce various hydroperoxides by thiols. Typically, the reducing substrates are proteins containing a CXXC motif. The Prx family is characterized by a catalytic triad composed of a cysteine residue, an arginine, and a threonine or serine.

► [Free Radicals](#)

Peroxisomal Disorders

RONALD J.A. WANDERS

Pediatrics and Clinical Chemistry, University of Amsterdam, Academic Medical Centre, Amsterdam, The Netherlands
r.j.wanders@amc.uva.nl

Definition

Peroxisomal disorders are relative newcomers in the arena of genetic diseases in man in which there is an impairment in one or more peroxisomal functions, either caused by a defect in peroxisome biogenesis or a single peroxisomal enzyme deficiency. They include a range of different diseases with the ► [cerebro-hepato-renal](#) syndrome, usually called ► [Zellweger syndrome](#), as prototype. Zellweger patients show a great number of abnormalities and the clinical picture is dominated by the typical craniofacial features and by neurological aberrations. Facial features include a large anterior fontanel with widely spaced sutures, a broad/full forehead, micrognathia, external ear deformity, low/broad nasal bridge, shallow orbital ridges and redundant skin folds in the neck. The neurological picture is dominated by the profound hypotonia, resulting in poor sucking requiring gavage feeding, depressed neonatal and deep tendon reflexes and a flat occiput. Other neurological abnormalities include an abnormal Moro response, hypo/areflexia and epileptic seizures (1).

Studies in Zellweger syndrome patients have been instrumental in the identification of additional peroxisomal disorders, which now comprise some twenty different diseases. All peroxisomal disorders identified today are autosomal recessive disorders with one exception, which is ► [X-linked adrenoleukodystrophy](#).

Characteristics

Functions and Biogenesis of Peroxisomes

Functions of Peroxisomes

►**Peroxisomes** are subcellular organelles just like mitochondria (►**Mitochondria – Biogenesis and Structural Organization**) and ►**lysosomes**, and are now known to catalyse a variety of different functions in the eukaryotic cell. These functions include (1):

1. Fatty acid beta-oxidation

Peroxisomes are able to degrade fatty acids just as mitochondria do. The mitochondrial and peroxisomal ►**beta-oxidation** systems are different however, one major important difference being that mitochondria and peroxisomes oxidize different fatty acids. Indeed, whereas mitochondria oxidize the bulk of the long chain fatty acids derived from dietary sources and lipolysis, peroxisomes oxidize only a few fatty acids, notably 1) ►**very long chain fatty acids** with chain lengths of 24 carbon atoms or more, 2) ►**pristanic acid** and 3) ►**di- and trihydroxycholestanic acids**. The latter are intermediates in the synthesis of the primary bile acids, ►**cholic acid** and ►**chenodeoxycholic acid** from cholesterol in the liver.

2. Ether phospholipid biosynthesis

Peroxisomes play an indispensable role in the synthesis of a specific set of phospholipids called ►**ether phospholipids**; plasmalogens are the main type of ether phospholipid. Ether phospholipids resemble the normal di-acyl phospholipids in many respects, the only difference being the presence of an ether bond rather than an ester bond at the *sn*-1 position of the glycerol backbone.

3. Fatty acid alpha-oxidation

Some fatty acids cannot undergo beta-oxidation but can only be broken down by an alternative mechanism, called ►**alpha-oxidation**, in which the terminal carboxyl group is released as CO₂. ►**Phytanic acid** is the prototype fatty acid undergoing alpha-oxidation, which occurs in peroxisomes only.

4. Glyoxylate detoxification

Peroxisomes catalyse the detoxification of ►**glyoxylate** to ►**glycine**. This reaction is catalysed by the peroxisomal enzyme alanine glyoxylate aminotransferase (AGT).

Peroxisome Biogenesis

In addition to the improved knowledge about the metabolic functions of peroxisomes, much has been learned recently about ►**peroxisome biogenesis**. First of all, it is now firmly established that peroxisomal matrix and membrane proteins are synthesized on free polyribosomes in the cytosol. Matrix proteins contain one of two different targeting signals called ►**PTS1** and ►**PTS2** that allow specific recognition by the

peroxisome. Two soluble receptors, named ►**PTS1** and ►**PTS2** receptor encoded by the ►**PEX5** (►**PEX Gene**) and ►**PEX7 genes**, respectively, recognise these signals and carry the proteins involved to the peroxisome, followed by transfer of the proteins across the peroxisomal membrane and recycling of the receptors. Many of the proteins involved in this complicated process have been identified in recent years. These proteins are called ►**peroxins** and are encoded by their respective, so-called, ►**PEX genes** (2).

Peroxisomal Disorders

The peroxisomal disorders are usually subdivided into two groups 1) the peroxisome biogenesis disorders (PBDs) and 2) the single peroxisomal enzyme deficiencies.

Peroxisomal Biogenesis Disorders

Apart from Zellweger syndrome, a number of different peroxisomal disorders have been identified as belonging to this category. These include ►**neonatal adrenoleukodystrophy** (NALD), ►**infantile Refsum disease** (IRD) and ►**rhizomelic chondrodysplasia punctata** (RCDP). ZS, NALD and IRD show overlapping signs and symptoms, ZS being the most severe, NALD intermediate and IRD the least severe. Together, ZS, NALD and IRD, are referred to as the “Zellweger spectrum”. In ZS, NALD and IRD, peroxisome biogenesis is completely deficient, resulting in the lack of morphologically identifiable peroxisomes and as a consequence, the complete loss of peroxisomal functions. This explains why ZS, NALD and IRD patients have elevated very long chain fatty acid levels, deficient plasmalogens and elevated phytanic acid levels.

►**Complementation studies** have shown that there is profound genetic heterogeneity within the ZS/NALD/IRD group, with eleven different complementation groups identified so far. The different genes mutated in each of these groups have all been identified in recent years (Table 1).

Rhizomelic chondrodysplasia punctata (RCDP) is different from ZS, NALD and IRD, not only in clinical terms but also in its biological and molecular basis. Patients who are suffering from RCDP show a disproportionally short stature primarily affecting the proximal parts of the extremities, a typical facial appearance, including a broad nasal bridge, epicanthus, high arched palate, dysplastic external ears and micrognathia, congenital contractures, characteristic ocular involvement, dwarfism and severe mental retardation with spasticity. Roentgenological studies have shown a series of abnormalities including symmetrical shortening of the femur and humerus with irregular and broad metaphyses, calcific stippling mainly of the epiphyses, absent femoral head nucleus,

Peroxisomal Disorders. Table 1 Peroxisomal disorders and the genes involved

Peroxisomal Disorders	Gene Name	Exons	Chromosome	Protein (AAs)
Peroxisome biogenesis disorders				
PBD Type 1: Zellweger spectrum	PEX1	24	7q21-q22	1283
ZS/NALD/IRD	(PXMP) PEX2	4	8q21.1	626
	PEX3	12	6q23-q24	373
	(PXR1) PEX5	15	12p13.3	602&639
	PEX6	17	6p21.1	980
	PEX10	6	1p36.32	326
	PEX12	3	17q21.1	359
	PEX13	4	2p15	403
	PEX16	11	11p12-p11.2	346
	(PXF) PEX19	8	1q22	299
	PEX26	5	22q11.2	305
PBD Type 2: Rhizomelic Chondrodysplasia Punctata (RCDP) 1	PEX7	10	6q21-q22.2	323
Single peroxisomal enzyme deficiencies				
X-linked adrenoleukodystrophy	ABCD1	10	Xq28	745
AcylCoA oxidase deficiency	ACOX1	14	17q25.1	660
D-Bifunctional protein deficiency	HSD17B4	24	5q2	736
Alpha-MethylacylCoA racemase deficiency	AMACR	5	5p13.3-p12	382
RCDP2 (DHAPAT deficiency)	GNPAT	16	1q42.11-q42.3	680
RCDP3 (alkylDHAP synthase deficiency)	AGPS	20	2q33	658
Refsum disease (phytanoylCoA hydroxylase deficiency)	PHYH	9	10p15-p14	338
Hyperoxaluria Type 1 (alanine glyoxylate aminotransferase deficiency)	AGXT	11	2q37.3	392
Acatalsasaemia	CAT	13	11p13	526

coronal clefts of vertebrae, increased intravertebral disc spaces, cupping of dorsal ribs and a barrel-shaped thorax. In RCDP, peroxisomal biogenesis is only partially deficient as a consequence of mutations in the ►*PEX7*-gene coding for the PTS2 receptor. The normal routing of PTS1 proteins explains why peroxisomal beta-oxidation is completely normal in RCDP whereas plasmalogen biosynthesis and phytanic acid alpha-oxidation are deficient.

Single Peroxisomal Enzyme Deficiencies

The different single peroxisomal enzyme deficiencies can be classified according to the pathway involved. These are:

a) Peroxisomal fatty acid beta-oxidation disorders

X-linked adrenoleukodystrophy (XALD) is the prototypical peroxisomal fatty acid beta-oxidation disorder with a variety of different phenotypes. ►**Childhood cerebral ALD (CCALD)** and ►**adrenomyeloneuropathy (AMN)** are the most frequent phenotypes, together accounting for > 80% of all cases. CCALD presents in young boys in the middle of the first decade, starting with behavioural abnormalities, decline in school performance and deterioration of vision and hearing. CCALD is relentlessly progressive with death occurring within two years. AMN presents much later in life with neurological deficits. Patients gradually develop

spastic paraparesis often in combination with disturbed vibration sense in the legs and sphincter dysfunction. The sensorimotor neuropathy in AMN is predominantly axonal. In all XALD phenotypes, very long chain fatty acids accumulate as a result of their defective beta-oxidation in the peroxisome. The gene involved in XALD codes for a peroxisomal half-ABC transporter, which is involved in the transmembrane transport of very long chain fatty acids from the cytosol into the peroxisome (3).

D-bifunctional protein deficiency is another peroxisomal beta-oxidation disorder and ranks among the more frequent peroxisomal disorders. Interestingly, the clinical presentation of affected patients is very different from XALD, being very severe and resembling Zellweger syndrome in many respects, including the craniofacial dysmorphism and neurological abnormalities.

Acyl-CoA oxidase deficiency is much less frequent and has been described in some fifteen patients in the literature only. All patients show severe neurological abnormalities, resembling those observed in the peroxisome biogenesis disorder NALD. The last beta-oxidation disorder is 2-methyl-acylCoA racemase (AMACR) deficiency, described in only a few patients. Patients with this deficiency show a slowly progressive polyneuropathy resembling that observed in classical Refsum disease.

b) Disorders of ether phospholipid biosynthesis

The two single enzyme deficiencies in this category are dihydroxyacetonephosphate acyl transferase (DHAPAT) deficiency, and alkyl-dihydroxyacetonephosphate (alkyl-DHAP) synthase deficiency. The clinical signs and symptoms of these two disorders closely mimic those described in rhizomelic chondrodysplasia punctata with the characteristic rhizomelic shortening, especially of the upper extremities.

c) Peroxisomal fatty acid alpha-oxidation disorders

Refsum disease is the only peroxisomal disorder in which peroxisomal fatty acid oxidation is deficient. Cardinal manifestations include retinitis pigmentosa, cerebellar ataxia, chronic polyneuropathy and an elevated protein level in CSF with normal cell count. Less constant features include sensory neural hearing loss, anosmia, ichthyosis, skeletal malformations and cardiac abnormalities. The clinical picture is often that of a slowly developing, progressive peripheral neuropathy manifested by severe motor weakness and muscular wasting especially of the lower extremities. Patients with Refsum disease are unable to alpha-oxidize phytanic acid as a consequence of the deficiency in the peroxisomal enzyme ►phytanoyl-CoA hydroxylase.

d) Disorders of Glyoxylate Detoxification

Primary hyperoxaluria Type 1 (PH1) is the only peroxisomal disorder belonging to this group. Patients show recurrent calcium oxalate nephrolithiasis and nephrocalcinosis frequently leading to progressive renal insufficiency and death before the age of twenty. In primary hyperoxaluria Type 1, alanine glyoxylate aminotransferase, a peroxisomal enzyme, is deficient, leading to the production of excessive amounts of oxalic, glyoxylic and glycolic acids which are excreted in urine.

Cellular and Molecular Regulation

As described above, the peroxisomal disorders are clinically and genetically heterogeneous with the involvement of a large variety of different genes which either code for proteins involved in peroxisome biogenesis or for proteins involved in different metabolic functions, be it an enzyme or a transporter protein. Peroxisomes are generally regarded as highly flexible organelles playing different roles in different organs. Furthermore, peroxisomes are highly inducible, at least in rodents, by 1) hypolipidaemic agents such as clofibrate, 2) plasticizers but also 3) high-fat diets. Apart from the increase in the number of peroxisomes, these agents also produce a profound induction of the fatty acid oxidation enzymes in peroxisomes. The nuclear hormone ►peroxisome-proliferator-activator-receptor alpha (PPAR-alpha) plays a crucial role in this induction process. Activated PPAR-alpha forms a heterodimer with the retinoic X-receptor (RXR) and binds to response elements present in a variety of different genes, including the ones coding for some of the peroxisomal beta-oxidation enzymes, thus causing enhanced transcription (4).

Clinical Relevance

The group of peroxisomal disorders is diverse with some twenty different diseases. In most of these there is neurological involvement, which is usually quite severe. Current estimates suggest that the group of peroxisomal disorders as a whole has a combined incidence of about one in 5000. Thanks to the improvement in clinical recognition of patients and the development of adequate laboratory methods, patients can be classified and the underlying molecular defect can be established. Furthermore, prenatal diagnosis of all peroxisomal disorders can now be done using either biochemical or molecular methods.

References

1. Wanders RJA (2004) Metabolic and molecular basis of peroxisomal disorders: A review. *Am J Med Genet* 126A:355–375

2. Wanders RJA, Waterham HR (2005) Peroxisomal disorders I: biochemistry and genetics of peroxisome biogenesis disorders. *Clin Genet* 67:107–133
3. Moser HW, Smith KD, Watkins PA et al (2001) X-Linked Adrenoleukodystrophy. In: Scriver CR, Beaudet AL, Sly WS, Valle D (eds) *The Metabolic & Molecular Bases of Inherited Disease*. Mc Graw-Hill, New York, pp. 3257–3301
4. Kersten S, Desvergne B, Wahli W (2000) Roles of PPARs in health and disease. *Nature* 405:421–424

Peroxisome

Definition

Peroxisomes are subcellular organelles present in all body cells, except in the mature erythrocyte catalysing a range of important metabolic functions. One of the most important metabolic processes of the peroxisome is the β -oxidation of long and very long chain fatty acids. The peroxisome is also involved in bile acid synthesis, cholesterol synthesis, amino acid and purine metabolism. The importance of the peroxisome and these processes is underscored by the existence of numerous genetic disorders associated with defects in the peroxisome.

► [Peroxisomal Disorders](#)

Peroxisome Biogenesis

Definition

Formation of peroxisomes.

► [Peroxisomal Disorders](#)

Peroxisome Proliferator-Activated Receptor

Definition

PPAR is a member of the ► [nuclear hormone receptor](#) subfamily of transcription factors. PPAR regulates transcription of many genes and is believed to be involved in adipocyte differentiation.

► [Diabetes Mellitus, Genetics](#)

Peroxisome Targeting Signal

► [PTS](#)

Perturbagens

► [Peptide Aptamers](#)

Pertussis Toxin

Definition

Pertussis toxin (PTX) is produced by *Bordetella pertussis*. It ADP-ribosylates a cysteine residue in most inhibitory G-protein α (G_{ia}) subunits, thereby uncoupling the modified G proteins from the receptor.

► [G-Proteins and G-Protein Mutations in Human Disease](#)

Pes Planus

Definition

Flat foot.

► [Marfan Syndrome](#)

PET

► [Positron Emission Tomography](#)

Petite Mutant

Definition

Petite mutant denotes a yeast mutant that leads to small (“petite”) colonies caused by respiratory deficiency.

► [Mutagenesis Approaches in Yeast](#)

Peutz-Jeghers Syndrome

DIETER E. JENNE

Max Planck Institute of Neurobiology, Martinsried,
Germany
djenne@neuro.mpg.de

Synonyms

Polyps-and-spots syndrome; pigment spot polyposis

Definition

Peutz-Jeghers Syndrome first described in 1921 by a Dutch physician named Peutz (1) and further studied by Jeghers in 1949 (2) is characterized by the occurrence of gastrointestinal hamartomatous polyps and abnormal melanocytic pigmentation of the lips, buccal mucosa and hands in an individual, who may have inherited this syndrome in an autosomal dominant fashion. Inactivating mutations in the *STK11* (*LKB1*) gene (3, 4), a serine threonine ►kinase of the nucleus and cytoplasm, have been identified in the majority of patients. Carriers of a dysfunctional *STK11* allele often suffer from

multiple benign tumors and polyp-induced gastrointestinal complications, but also bear an increased lifetime risk of developing various cancers within and outside the gastrointestinal tract (5, 6). Surveillance and screening protocols for PJS patients and their relatives have been developed.

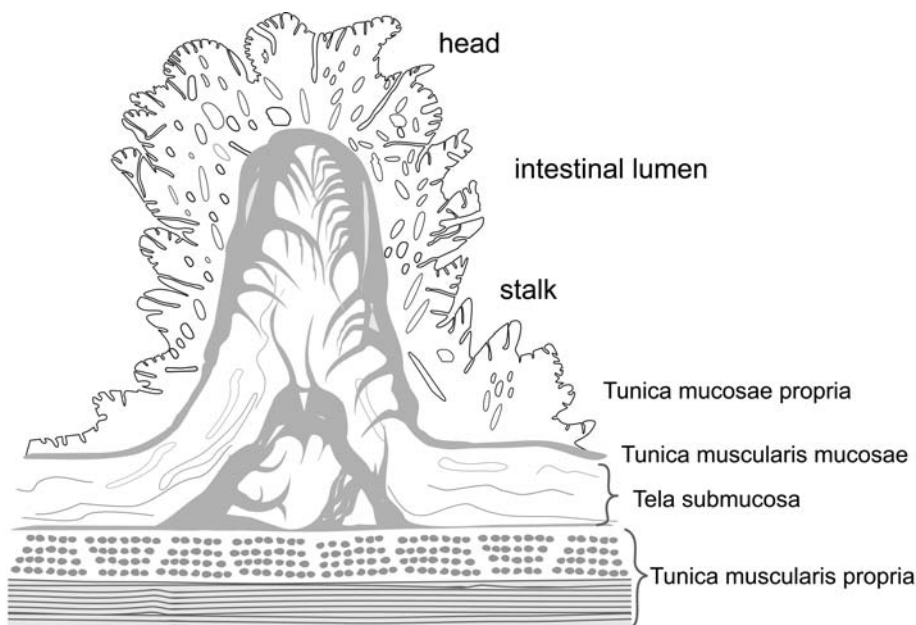
Characteristics

Epidemiology

The disease incidence is uncertain. Estimates fall between 1:8,300 and 1:120,000 per live birth. All races and both genders are equally affected. An incidence of 1:30,000 is calculated for the UK on the basis of 2,000 known individuals with PJS. *De novo* mutations in the *STK11* gene on 19p13.3 account for about 10–40% of all PJS cases. The syndrome shows high penetrance during life, although onset, clinical severity, relapse of symptoms and clinical course are highly variable.

Clinical Features

The predominant symptoms are caused by the formation and growth of multiple polyps in the gastrointestinal tract (Fig. 1). Rectal bleeding, recurrent colicky abdominal pains, in particular after meals and



Peutz-Jeghers Syndrome. Figure 1 Schematic drawing (cross section) of an early-stage Peutz-Jeghers polyp. Smooth muscle cells (dark-grey) from the tunica muscularis mucosae and from the inner circular layer of the tunica muscularis propria proliferate and expand into the tela submucosa. Once fused they form a continuous smooth muscle bundle, which connects the head of the polyp with the tunica muscularis propria. Thin ramifications (arborization) originating from the muscularis mucosae divide the head of the polyp into several sectors (not shown). The tree-like branching of smooth muscle cells connects the inner circular layer of the tunica muscularis propria with the tunica muscularis mucosae and occupies the connective tissue space (tela submucosa) (courtesy of Robert Schorner, Photolab, Max-Planck-Institute of Neurobiology).

in the morning, and increased peristalsis of the small intestine are the most common symptoms which patients report. Pain attacks are caused by invaginations of small polyps (intussusception), whereas bowel obstructions are rare complications of larger polyps, which may manifest as acute abdomen. Invaginations of polyps that develop in the sigma or rectum can lead to anal prolapse. Chronic blood loss as a result of erosions and ulcerations on the surface of polyps can cause the general symptoms of anemia (achrocythemia), like fatigue, adynamia, syncopes and growth retardation in children. In two-thirds of all cases this gastrointestinal bleeding is not visible. One third of patients will experience the first gastrointestinal symptoms before the age of 10, another third during the second decade of life. After the age of 30 the growth of polyps decreases and new polyps are rare. Extraintestinal polyps of the gallbladder and nasopharynx have been reported in a few cases.

The second cardinal feature is abnormal mucocutaneous pigmentation (1–6 mm macules) around the mouth and the nasolabial folds, on the lips (Fig. 2) crossing the vermilion border and affecting the buccal mucosa. This clinical sign is very helpful in the demarcation of PJS from other inherited hamartomatous polyposis syndromes. Uncommon sites of pigmentation are the dorsal and volar aspects of the hands and feet, the perianal and genital region and the rectal mucosa. With advancing age, in particular after puberty, pigmentation often fades and sometimes disappears completely. Buccal pigmentation tends to persist in PJS and is rarely mimicked by ordinary freckles.

Additional clinical signs and symptoms can originate from hormone producing benign gonadal tumors resulting in precocious puberty and menstrual irregularities in females. Ovarian sex cord tumors with annular

tubules (►SCTAT) sometimes produce excessive estrogens. Sertoli cell tumors can trigger gynecomastia and growth acceleration in male patients.

Individuals with PJS bear a significantly higher relative and absolute risk for all types of cancer at all stages of life. The lifelong cumulative risk of dying from intestinal or extraintestinal malignancies is 48% by the age of 57 and 93% by the age of 65. The site with the highest relative risk for cancer is the small intestine, followed by the stomach, pancreas, colon, esophagus, ovary, lung, uterus and breast in decreasing order. Counting all cancers in PJS patients between the age of 15 and 64, cancer most frequently develops in the breast (54%), colon, (39%), pancreas (36%), stomach (29%), ovary (21%), lung (15%) and small intestine (13%).

Histopathology

Peutz-Jeghers polyps have a non-neoplastic hamartomatous appearance (Fig. 1). They consist of a branching framework of connective tissue and smooth muscle cells lined by normal intestinal epithelium with numerous goblet cells. Epithelial and goblet cells are sometimes surrounded by smooth muscle cells (pseudoinvasion), which may be mistaken for an invasion of the tela muscularis mucosae and submucosa and the beginning of an adenocarcinoma. The muscular and fibrous framework of the stalk and head adds high mechanical stability to Peutz-Jeghers polyps. This explains the relatively frequent complications induced by these hamartomatous polyps.

In contrast to adenomatous polyps, dysplasia, specifically of the epithelial cell layer, and aneuploidy are absent. A direct malignant transformation of a benign hamartomatous polyp into cancer is unlikely. Secondary erosions and inflammatory infiltrates, however, occur frequently in larger polyps and may contribute to the occasional formation of adenomatous as well as carcinomatous foci. The most common location of Peutz-Jeghers polyps is the small intestine (jejunum, ileum, duodenum in that order), followed by the large intestine and stomach. Population differences in the distribution of polyps along the gastrointestinal tract may exist.

Pigment spots of the facial skin and buccal mucosa develop in most individuals with PJS before the fifth year of life, but are rarely present at birth. The margins of macules are irregular and fuzzy. Histologically, hyperplasia of melanocytes is found at the epidermal-dermal junction. Transport to and deposition of melanosomes in the basal layer of keratinocytes is disturbed. The number of melanosomes is increased in the dendrites of melanocytes and thus abnormal extracellular deposition of melanin and melanin-containing macrophages are seen in the basal layer of the epidermis. Pigment spots from PJS patients



Peutz-Jeghers Syndrome. Figure 2 Typical pigment spots on the lips and periorally in a Peutz-Jeghers patient (courtesy of Dr. W. Back and Dr. S. Löff, Mannheim Clinics, University of Heidelberg).

display the typical histopathological and electron microscopical features of lentigo simplex.

Cellular and Molecular Regulation

Structure of the *STK11* Gene

To date, 75 different germ line mutations have been identified in *STK11* (LKB1). *STK11* encodes a highly conserved serine threonine kinase (STK), which is present in vertebrates, *Drosophila* and *Caenorhabditis elegans* (Par4). The human gene consists of 10 exons which span about 23 kb on 19p13.3 and is transcribed in telomere-centromere direction in many tissues and cell types. Exons 1 and 2 are separated by an extremely rare intron type with unusual splice junction sequences. Intron 2 is removed in an U12 snRNA-dependent manner. The open reading frame codes for a polypeptide chain of 433 amino acid residues, which can be divided into three segments, a kinase domain (residues 46 to 314) and two regulatory domains at the N- and C-termini.

Functional Features of the *STK11* Protein

The kinase domain is most closely related to the yeast SNF1/cyclic AMP-activated protein kinase subfamily. Its catalytic activity is regulated by a number of cellular proteins, which determine its subcellular localization, catalytic activity and phosphorylation state. Endogenous *STK11* is mainly localized in the nucleus, but a significant fraction is retained in the cytoplasm. Mutation of the nuclear localization signal at position 38–43 prevents its accumulation in the nucleus. A cofactor protein called STRAD (ste20 related adapter) binds to *STK11*, activates it and re-directs *STK11* to the cytoplasm. As a consequence both *STK11* and STRAD become phosphorylated. An in-frame deletion of the four residues 303–306 (Ile-Arg-Gln-His) found in a PJS family abolishes STRAD interactions and cytoplasmic retention. A number of threonine and serine residues of *STK11* can be modified by phosphorylation, but only Thr336 and Ser431 in over-expressed *STK11* were shown to be required for suppression of cell growth in certain *STK11* deficient cell lines. Thr336 is phosphorylated in cells exposed to ionizing radiation. The C-terminus, furthermore, displays a typical consensus site (Arg-Arg-Leu-Ser) for phosphorylation by the cyclic AMP-activated protein kinase A (PKA) and a **►prenylation** site (Cys-Lys-Gln-Gln). Ser428 (equivalent to Ser431 in the mouse sequence) is indeed phosphorylated by PKA and the p90 ribosomal S6 protein kinase (RSK) in endogenously expressed *STK11*. Cys430 is farnesylated in cultured cells. These two sites are of functional significance since truncation of the terminal 20 residues by a *STK11* mutation has triggered the full-blown PJS syndrome. Both functional sites are conserved and essential to properly regulate cell polarity in

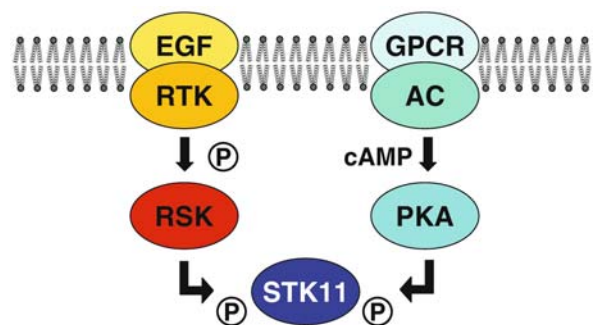
Drosophila. As observed in *C. elegans* and *Drosophila*, *STK11* functions in establishing tissue polarity and ventral-dorsal tissue patterning. Reduction of its activity may, therefore, facilitate the local overgrowth and disorganization of differentiated cells.

Cellular Functions

Defects in **►apoptosis** regulation have been inferred to explain benign **►hamartoma** formation in accordance with the reduced number of apoptotic epithelial cells in PJS polyps. Apoptotic epithelial cells on the tip of intestinal villi and at the base of crypts contain higher cytoplasmic levels of *STK11* than those in PJS polyps. Besides this *in situ* correlation between elevated cytoplasmic *STK11* levels and apoptosis in epithelial cells, *STK11* associates with the tumor suppressor p53 and translocates to mitochondrial membranes during apoptosis, a process that appears to amplify the apoptotic cascade.

Mouse embryonic fibroblasts can be immortalized by *STK11* inactivation. These cells then become resistant to culture stress-induced senescence. On the other hand these cells cannot be transformed by activated RAS and SV40 large T antigen. The complete inactivation of *STK11* prevents the hamartoma-to-adenocarcinoma transition, but unblocks cellular proliferation of local stromal and epithelial cells.

Multiple rationales have been proposed for the growth suppressive effect of *STK11*. Its ability to mediate a G1 cell-cycle arrest only in certain *STK11* deficient cell types on one hand and its presence in many normal



Peutz-Jeghers Syndrome. Figure 3 The activity of *STK11* is regulated by phosphorylation (indicated by the P) at multiple serine and threonine sites through the actions of multiple upstream kinases, which have been poorly characterized. Ser428 e.g. can be phosphorylated via different signaling pathways, either by the cAMP dependent kinase A (PKA) or the p90 ribosomal S6 protein kinase (RSK) which in turn is activated by the epidermal growth factor (EGF) receptor, a receptor tyrosine kinase (RTK). cAMP is generated by adenyl cyclase (AC) whose activity is regulated by G-proteins coupled to various seven transmembrane domain receptors (GPCRs).

proliferating cell types on the other hand, however, illustrates the complex mechanisms that control the activity and multiple functions of STK11 with regard to cellular proliferation. The complex crosstalk between the stroma and the epithelial cell compartment in the small intestine probably involves membrane receptors and secretory signaling molecules from different cell types. Several observations support this suggestion. Phosphorylation of STK11 at Ser428 by PKA is triggered by elevated cAMP levels as well as by epidermal growth factor (EGF) *via* p90 ribosomal S6 protein kinase (RSK). Moreover, signal processing *via* certain G-proteins may also be altered by STK11 depending on its phosphorylation state. Recently, STK11 was shown to interact with activator of G-protein signaling 3 (AGS3). The more STK11 functions and interacting proteins are disclosed, the slighter the hope that either STK11 or any interacting partner will be a suitable target for drugs with predictable effects in tumor patients.

Mouse Models for Peutz-Jeghers Syndrome

An essential role of STK11 in early embryonic development was demonstrated by targeted disruption of both gene copies in mice. Development was normal until day 8 (E8.0) and thereafter heavily disturbed, so

that all embryos died by day 11 (E11.0). Obvious problems were the defects in tissue vascularization, neural tube closure and placenta development. Heterozygous mice with one inactivated and one normal copy of the gene were viable at birth and grew normally until they developed numerous polyps, predominantly in the glandular stomach between 20 and 45 weeks of age. Unfortunately the mechanism of hamartoma formation in heterozygous animals has so far remained unclear. At present, all theoretical possibilities, allelic loss (loss of heterozygosity) by mitotic recombination or interstitial chromosomal deletions, acquired second hit of the normal allele, epigenetic inactivation by promoter methylation and ▶**haploinsufficiency** are still being considered.

Clinical Relevance

Diagnosis and Genetic Testing

Clinical and histopathological findings are still crucial for the diagnosis of Peutz-Jeghers syndrome although molecular genetic testing of the *STK11* gene is widely available. In patients without a known family history, either two histologically confirmed PJS polyps from the gastrointestinal tract or any PJS polyp in combination with characteristic mucocutaneous pigmentation are sufficient diagnostic criteria. For patients with a

Peutz-Jeghers Syndrome. Table 1 Differential diagnosis of familial hamartomatous polyposis syndromes

Syndrome	Polyp characteristics	Inheritance & etiology
Peutz-Jeghers	hamartomatous polyps with smooth muscle core mainly in small intestine	STK11 (LKB1), 19p13.3 autosomal dominant
Juvenile polyposis	juvenile polyps, primarily in colon with cystic dilatations	SMAD4, 10q23.31 BMPR1A, 10q23.31 autosomal dominant
Cowden's disease	hamartomas in various tissues including gut (colon & stomach)	PTEN, 10q22-23 autosomal dominant

Peutz-Jeghers Syndrome. Table 2 Differential diagnosis of familial facial lentiginosis syndromes

Syndrome	Major Findings	Inheritance (autosomal dominant)
Peutz-Jeghers	facial pigment spots, polyps	STK11 (LKB1), 19p13.3
Carney complex	multiple myxoma, endocrine tumors	PRKAR1A, 17q22-24 second locus at 2p16?
Bannayan-Riley-Ruvalcaba	perigenital pigment spots juvenile polyps	PTEN, 10q22-23
Laugier-Hunziker	oral and cutaneous spots subungual pigmentation	unknown
LEOPARD	lentiginosis, electrocardiographic abnorm., ocular hypertelorism, pulmonary stenosis, abnormalities of genitalia, retardation of growth, deafness	PTPN11, 12q24.1

first-degree relative with Peutz-Jeghers syndrome, either typical pigment spots or any number of PJS polyps are the minimum criteria for diagnosis. If these clinical criteria are met or a *STK11* mutation is already known among a first-degree relative, genetic testing can be considered.

Sequence analysis of the nine coding exons of *STK11* reveals disease-causing mutations in more than 70% of all familial cases. *STK11* mutations in sporadic cases are found with lower frequencies. ►Nonsense mutations that cause the truncation of the polypeptide chain are of conclusive significance for the diagnosis, but ►missense mutations are more difficult to assess. Natural amino acid variations for *STK11* have not been found in the healthy human population. Any substitution of residues that are shared by the various mammals and higher vertebrates are highly suspicious. In dubious cases bioinformatic tools can predict the possible functional impact of an amino acid substitution on the structure and function of *STK11*, using 3D-structural and comparative models.

Other genes that cause Peutz-Jeghers syndrome have so far not been identified. Linkage to the *STK11* locus on 19p13.3 was excluded in some PJS families. In one large Indian family the gene defect appears to be linked to 19q13.3. Failure to identify *STK11* mutations can have many reasons. The widely used exon scanning strategies are not suited to discovering promoter mutations, splicing enhancer defects, polyadenylation anomalies, intronic mutations generating new splice sites, intragenic insertions or exon deletions, intragenic inversions and other long range rearrangements of the *STK11* locus. Mutation screening at the transcript level using blood cell mRNA is recommended when coding exons are found to be normal. Since *de novo* mutations account for a significant number of PJS cases, somatic mosaicism and segmental forms of *STK11* mutation must be considered as well.

Management of PJS Families

The clinical diagnosis and final proof of *STK11* mutations in a PJS patient has serious consequences for the patient and his relatives. Clinical evaluation, ►genetic counseling and genetic testing should be offered to the parents and at-risk relatives. The type of mutation, however, is not useful in predicting onset, severity, clinical course and various complications of the disease in asymptomatic individuals.

An extensive medical surveillance programme is recommended irrespective of *STK11* mutations to all patients who meet the clinical criteria for PJS and to asymptomatic, at-risk relatives in familial PJS. In cases of sporadic PJS without *STK11* mutations, it is not possible to assess the risk to relatives. In sporadic cases with *de novo* mutations in *STK11*, the parents should be carefully examined for subtle signs of the disorder

and gonadosomal mosaicism of one parent should be considered. *STK11* testing can be performed prenatally and before embryo implantation if an *STK11* mutation is known in one parent. The ethical problems and legal situation associated with such requests should be clarified with a genetics professional.

Adequate follow-up care and cancer surveillance must be offered to affected individuals and asymptomatic at-risk family members in early life (at the age of 10). Physical examination of breasts, abdomen, pelvis and testes and a complete blood count should be performed every year and complemented by ultrasound investigations of the pelvis (women), testes (men), abdomen and pancreas. Every other year, esophagogastro-duodenoscopy and colonoscopy and in the case of abdominal complaints a small bowel follow-through (contrast enema) are recommended. All polyps larger than 10 mm should be removed by electrocautery snare either during endoscopy or surgically after laparotomy. The need for repeated bowel resections and emergency interventions should be minimized. Mammography should be started later at the age of 25 and repeated initially at intervals of 5 years, after the age of 38 every second year and after the age of 50 annually. The beneficial effects of special medical care and surveillance on the morbidity and mortality of PJS patients have not yet been assessed in prospective studies.

►Colorectal Cancer

References

1. Peutz JLA (1921) Very remarkable case of familial polyposis of mucous membrane of intestinal tract and nasopharynx accompanied by peculiar pigmentations of skin and mucous membrane. *Nederl. Maandschr. Geneesk* 10:134–146
2. Jeghers H, McKusick VA, Katz KH (1949) Generalized intestinal polyposis and melanin spots of the oral mucosa, lips and digits; a syndrome of diagnostic significance. *N Engl J Med* 241:993–1005
3. Hemminki A, Markie D, Tomlinson I et al (1998) A serine/threonine kinase gene defective in Peutz-Jeghers syndrome. *Nature* 391:184–187
4. Jenne DE, Reimann H, Nezu J et al (1998) Peutz-Jeghers syndrome is caused by mutations in a novel serine threonine kinase. *Nat Genet* 18:38–43
5. Giardiello FM, Brensinger JD, Tersmette AC et al (2000) Very high risk of cancer in familial Peutz-Jeghers syndrome. *Gastroenterology* 119:1447–1453
6. McGarrity TJ, Kulin HE, Zaino RJ (2000) Peutz-Jeghers syndrome. *Am J Gastroenterol* 95:596–604

PEX Gene

Definition

PEX gene defines the name of any gene coding for a ▶[peroxin](#) that is required for the formation, maintenance and proliferation of peroxisomes.

▶[Peroxisomal Disorders](#)

PEX7 Gene

Definition

PEX7 is the gene coding for the ▶[PTS2](#) (Peroxisome targeting signal type 2) receptor, which is involved in directing PTS2 proteins to peroxisomes and mutated in rhizomelic chondrodysplasia punctata (RCDP) Type 1.

▶[Peroxisomal Disorders](#)

PFAM Database

Definition

PFAM database refers to Protein Families Database (<http://www.sanger.ac.uk/Software/Pfam/>), which is a large collection of multiple sequence alignments and hidden Markov models covering many common protein domains and families. For each protein family in PFAM you can look at multiple alignments, view protein domain architectures, and examine species distribution etc.

▶[Functional Assays](#)

▶[Protein Databases](#)

PFGE

Definition

Pulsed field gel electrophoresis.

▶[Hereditary Neuropathies, Motor and/or Sensory](#)

PGC

▶[Primordial Germinal Cells](#)

PGD

▶[Preimplantation Genetic Diagnosis](#)

PHA

▶[Pseudohypoaldosteronisms](#)

PHA 1

▶[Type 1 Pseudohypoaldosteronism](#)

PHA 2

▶[Type 2 Pseudohypoaldosteronism](#)

Phage Artificial Chromosome

▶[PAC](#)

Phage Display

Definition

The phage display is a technology that displays protein or peptide libraries on the surface of a multitude of

phages which also contain the encoding gene (each phage contains the gene for its encoded protein or peptide). Proteins or peptides reactive with a specific molecule can therefore be isolated by a binding reaction (which eliminates the non-binders) and multiplied because of the association of the binding protein and the corresponding encoding gene in the same particle.

► [Immunological Methods, the Use of Antibodies to Discover the Function of Novel Gene Products](#)

► [Monoclonal Antibodies](#)

Phage Surface Display

► [Protein Interaction Analysis: Phage Display](#)

Phagemid

Definition

A phagemid is a vector combining elements of a plasmid and the M13 phage genome. These vectors offer several advantages such as easy preparation and high yield of dsDNA (double stranded DNA) for cloning, easy maintenance due to antibiotic resistance, and they allow modulation of display valency. They contain a bacterial and a phage origin of replication, the phage Packaging Signal (PS), an antibiotic resistance gene allowing for selection of phagemid-containing *E. coli* and the gene for one coat protein allowing fusion. The phagemids lack all the other phage genes, and for phage propagation require the aid of a super-infecting

► [Helper Phage](#)

► [Protein Interaction-Phage Display](#)

Phagocytosis

Definition

Phagocytosis is a phenomenon of ingestion and digestion. It is performed by certain specialised cells like macrophages, polymorphonuclear leukocytes and amoebae. This process involves chemotaxis, adherence, phagosome and phagolysosome formation.

► [Actin Cytoskeleton](#)

► [Cytokine Receptors](#)

► [Rho, Rac, Cdc42](#)

Pharmaco

Definition

Pharmaco refers to a field of research in which the individual drug response of human beings is studied by tools of genome research.

► [SNP Detection and Mass Spectrometry](#)

Pharmacogenetics

► [Pharmacogenomics](#)

Pharmacogenomics

MICHEL EICHELBAUM, MATTHIAS SCHWAB
Dr Margarete Fischer-Bosch Institute of Clinical
Pharmacology, Stuttgart, Germany
michel.eichelbaum@ikp-stuttgart.de

Synonyms

► [Pharmacogenomics/Pharmacogenetics](#)

Definition

Pharmacogenetics, a term coined by Vogel in 1959 (1), deals with the interindividual differences in response to drugs caused by mutations of proteins leading to alteration of pharmacokinetics and pharmacodynamics of drugs. It uses genetic information to predict an individual's drug response. The term pharmacogenomics has been introduced in recent years to describe the progressive transition from genetics to genomics, acknowledging that the genome is more than the sum of its genes. It is a genome-wide approach to identify genes that contribute to a specific disease, to identify new drug targets, design new drugs and target drugs to specific patient populations.

Characteristics

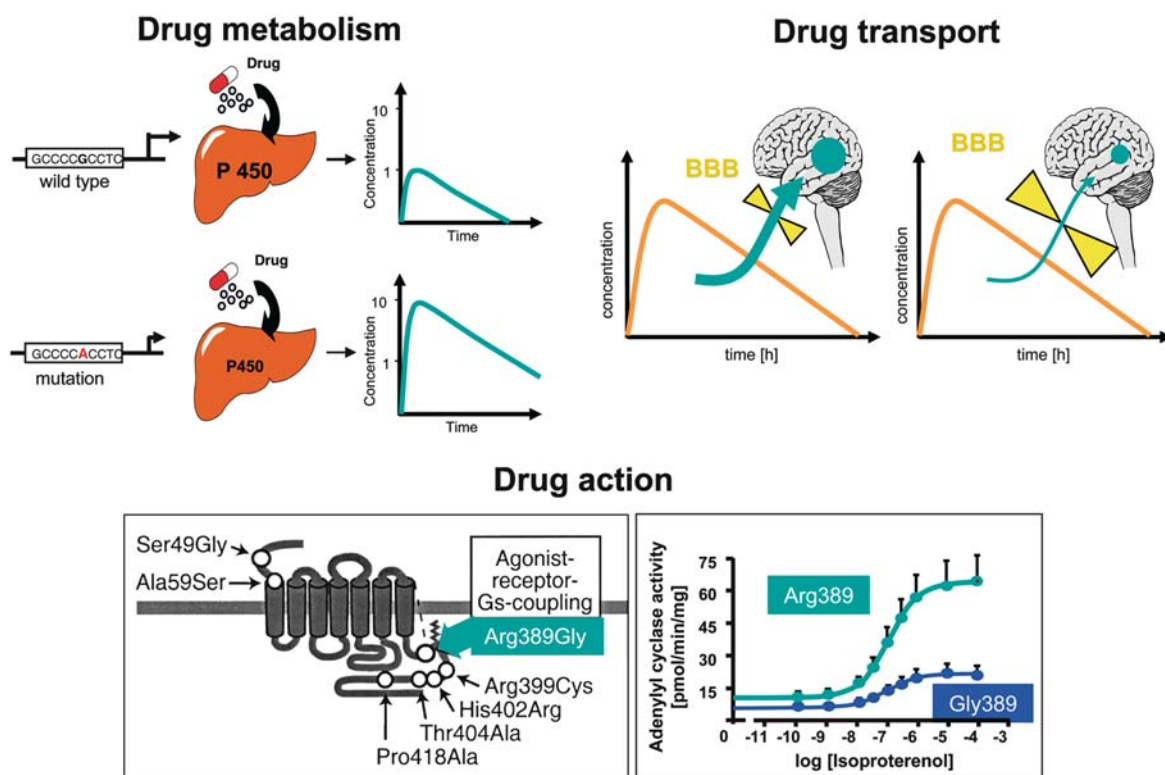
There are in principle three mechanisms by which
► [genetic polymorphisms](#) or ► [rare mutations](#) in proteins that are involved in the metabolism, transport

and action of drugs lead to interindividual differences in response to and toxicity of drugs (Fig. 1).

The rate at which drugs are eliminated from the body is a major determinant of both the intensity and duration of drug action and toxicity. Drug metabolizing enzymes, in particular ►cytochrome P450 enzymes, play a pivotal role in the elimination of most drugs. Variability of drug metabolism is responsible for the pronounced interindividual differences in plasma concentrations when patients receive the same dose of a drug. As a consequence, variability in drug action and side effects/toxicity ensues. During the last 30 years, mutations have been identified for many phase 1 and phase 2 enzymes catalyzing the biotransformation of drugs. These mutations affect either the expression or catalytic properties of the enzymes. In the case of mutations leading to a loss of function, administration

of a standard dose of a drug will lead to very high plasma concentrations resulting in an exaggerated response, side effects or toxicity. At the other end of the spectrum, gene amplifications of enzymes resulting in ultra rapid metabolism of drugs have been identified as a mechanism for poor response. In the case of prodrugs, which require bioactivation for therapeutic efficacy, loss of enzyme function due to polymorphism of the enzyme is associated with a loss in efficacy (Table 1).

But even if the drug dose is individualized by therapeutic drug monitoring in order to achieve the same plasma concentrations substantially, variability in therapeutic response and side effects will still be observed because concentrations at the site of action vary substantially. It is increasingly recognized that transfer of drugs in and out of cells is not a passive



Pharmacogenomics. Figure 1 Different pharmacogenetic mechanisms responsible for interindividual drug response. Drug metabolism: Mutations in cytochrome P450 genes lead to decreased or deficient enzyme expression in human liver. Patients carrying these mutations given the same dose of a drug as wild type individuals will accumulate drug concentrations in their plasma that lead to a higher risk of developing adverse drug reactions. Drug transport: Absorption and elimination of drugs as well as transfer of a drug from the blood into different tissues depend on active transport processes via membrane transporters. P-glycoprotein, the product of the MDR1 gene, is located in the blood brain barrier (BBB) and mutations in this gene associated with decreased expression of P-glycoprotein will lead to an enhanced transfer of drugs and/or metabolites into the brain. Thus, despite the same plasma concentrations, different concentrations may be achieved at the site of action. Drug action: Generally, drugs act via their interaction with specific receptor proteins such as β -adrenoceptors. Despite the same target concentration at the site of action, mutations in the receptor gene may cause alterations in receptor activity, leading to differences in drug response.

Pharmacogenomics. Table 1 Examples of genetic polymorphisms influencing drug response in humans

Gene	Drug
Drug-metabolizing enzymes	
CYP2C9	losartan, NSAID, phenytoin, tolbutamide, S-warfarin
CYP2C19	diazepam, mephenytoin, proguanil, proton pump inhibitors
CYP2D6	antiarrhythmics, antidepressants, β -blocker, codeine, 5-HT ₃ antagonists, neuroleptics
Dihydropyrimidine dehydrogenase	5-fluorouracil
Glutathione transferases	various cytotoxic agents
Thiopurine S-methyltransferase	azathioprine, 6-mercaptopurine
UGT1A1	irinotecan
Drug transporters	
ABCB1 (MDR1)	antiepileptics, cyclosporine, digoxin, HIV protease inhibitor
OATP1B1 (OATP-C)	methotrexate, statins
Serotonin transporter (5-HTT)	antidepressants
Drug targets	
β_2 adrenoceptor	β_2 agonists
Gs protein α	β -blocker
ALOX5	leukotriene receptor antagonists
APO E	statins

process depending on physicochemical properties, lipophilicity and protein binding but also involves active transfer by transport proteins.

Transporters are membrane proteins that maintain cellular homeostasis through import and export of endogenous compounds. Because of their localization in intestinal, hepatic and renal epithelial cells, these transport proteins are important for the absorption, bioavailability and elimination of drugs. Moreover, they play an important role in targeting drugs to organs because they are localized in blood-organ barriers such as the blood-brain and blood-placenta barrier. Thus, mutations either affecting the expression or changing the affinity of the transporter can alter the absorption and elimination of drugs. Due to their localization in blood-organ barriers, polymorphism in drug transporters can lead to different concentrations at the site of action although blood concentrations are similar. Among the many transporters identified MDR1, MRPs, OATPs, OCTs, OATs and nucleoside transporters are of particular interest since they transport not only endogenous compounds but also exogenous substrates including drugs. For all these transporters mutations have been identified which modify either their expression or function.

Finally, the same concentration of a drug at the site of action does not necessarily mean an identical response, because mutations at drug targets (receptors, neurotransmitter transporters, signaling pathways) can profoundly alter response. In the case of β_2 adrenoceptor polymorphisms, marked differences in the bronchodilatory response to β_2 agonists such as albuterol have been described. Moreover, the effectiveness of generally ineffective drugs such as gefitinib in patients with lung cancer can be explained by sensitizing mutations of the epidermal growth factor receptor in the tumor.

With the complete sequence of the human genome available, it is hoped that individualized medicine will soon become a reality. The expectations are that with the use of genomic information we can determine an individual's drug response and select the appropriate dose of the drug. This would allow achievement of the optimal therapeutic response, avoid therapeutic failure and minimize side effects and toxicity. Although many genes responsible for inherited differences in the metabolism, transport and action of drugs have been identified, this new knowledge has not been translated into clinical practice. With the exception of a very few examples of drug metabolizing enzymes, the

contribution of genetic polymorphisms to individual differences in drug effects and toxicity are not well understood. Moreover, most of these studies have focused on the consequences of a single gene polymorphism for altered drug response. This approach, however, neglects the fact that the drug response phenotype like most disease phenotypes is a complex polygenic trait with nongenetic factors contributing to the manifestation of the phenotype. The extent to which genetic factors contribute to the drug response/toxicity phenotype will depend on whether the candidate gene is a gene of major, moderate or minor effect. There are also misconceptions with respect to the information provided by a pharmacogenetic test. Even in the case of a gene with maximum effect, the presence or absence of a mutation will not provide a straight forward yes or no answer, but rather the likelihood that in a subject with a given mutation an event will occur or not. The highest positive predictive value of a genetic test will be observed for genes with a major effect. In the case of drug metabolizing enzymes, mutations leading to a loss of function will result in higher drug concentrations. If these higher drug concentrations are associated with toxicity, the likelihood that a patient with this genotype will develop toxicity is very high if he is exposed to the same dose as the patients who carry the wild type of the gene. However, the negative predictive value (likelihood that a patient without the mutation will not have toxicity) can be rather poor if non-genetic factors leading to high drug concentrations, which are associated with drug toxicity, are neglected. If a patient who carries a wild type gene is concomitantly treated with a drug that inhibits the enzyme, the patient will develop the phenotypic high concentration that is usually associated with the presence of two mutant alleles, a phenomenon called phenocopying.

Clinical Relevance

For all major classes of drugs, a substantial proportion of patients will not respond or respond only partially when standard doses of a particular drug are administered. Possible mechanisms underlying the marked interindividual variability in drug response and toxicity include heterogeneity of the disease and such clinical variables as age, gender, diet, co-administration of drugs and renal and hepatic functions. In addition it has become apparent during the last three decades that genetic factors can be responsible for significant heterogeneity in the efficacy and toxicity of therapeutic agents.

It has been suggested that, depending on the drug, genetic factors account for 20–95% of the variability in drug disposition and effects (2, 3). In a recent systematic review by Philipps and co-workers (4), among 27 drugs frequently cited in adverse drug reaction (ADR) studies, 59% were metabolized by at

least one enzyme with a variant allele known to cause poor metabolism. Conversely, only 7–20% of randomly selected drugs were metabolized by enzymes known to be polymorphically expressed. This analysis suggests that genetic variability in drug metabolizing enzymes is a contributor to the incidence of ADR. Altogether, the effects of most drugs are determined by many proteins and it will be the composite of genetic polymorphisms in multiple genes that determine drug response, along with non-genetic factors.

To date the best-characterized genetic polymorphisms influencing variability of drug response are mutations in drug metabolizing enzymes. They have a profound effect on the pharmacokinetics and pharmacodynamics of drugs.

Cytochrome P450 2C9

Cytochrome P450 2C9 (CYP2C9) is a major enzyme involved the metabolism of the more potent S-enantiomer of the oral anticoagulant warfarin. Two mutations (*CYP2C9**2 and *3), which have only 20% and 5% respectively of the wild type enzyme activity, have been identified as major risk factors for severe bleeding complications on the initiation of anticoagulant therapy, if patients receive the standard dosage.

Cytochrome P450 2C19

Cytochrome P450 2C19 (CYP2C19) exhibits a genetic polymorphism that produces three distinct phenotypes in the population, the poor metabolizer (PM), the heterozygous (hetEM) and the extensive metabolizer (EM). Following administration of the same dose of proton pump inhibitors, which are mainly metabolized by CYP2C19, PM have a 3–12× higher drug exposure than EM. In view of the close relationship between drug concentration and inhibition of acid secretion, with up to 12 fold difference among patients it is becoming apparent that, with the current practice of administering the same dose to all patients, Hp (*Helicobacter Pylori*) eradication and ulcer healing will be affected by the CYP2C19 genotype. Apart from antibiotic resistance, multivariate analysis showed that CYP2C19 polymorphism was the most important factor influencing Hp eradication success.

Cytochrome P450 2D6

Approximately 25% of all drugs in clinical use are metabolized by cytochrome P450 2D6 (CYP2D6). Case reports and case control and observational studies clearly indicate that PM are at a much higher risk of developing concentration related ADR. On the other hand, ultra rapid metabolism, which is partly explained by *CYP2D6* gene amplification, has been identified as a factor responsible for lack of drug response.

Moreover, in the case of drugs where efficacy depends solely on the formation of an active metabolite, such as the analgesic effect of codeine mediated by the formation of morphine, loss of analgesia has been demonstrated in PMs (5).

Thiopurine S-Methyltransferase

Thiopurine S-methyltransferase (TPMT) catalyzes the S-methylation of thiopurines (6-mercaptopurine and azathioprine). TPMT enzyme activity is controlled by a genetic polymorphism. Approximately 1 in 200 individuals has a very low TPMT activity, 10% have intermediate activity and 89% show normal/high activity. Numerous clinical studies in childhood ALL, rheumatic diseases and IBD have demonstrated that TPMT-deficient patients are at high risk of severe and sometimes fatal hematotoxicity due to the accumulation of cytotoxic metabolites after treatment with standard doses of thiopurines unless a dose-adjusted administration of AZA is used. Prospective determination of erythrocyte TPMT activity or *TPMT* genotype is advocated as a routine procedure prior to therapy in order to avoid drug toxicity.

Dihydropyrimidine Dehydrogenase

5-Fluorouracil (5FU) is widely prescribed for the treatment of solid tumors (e.g. colorectal cancer) and is metabolized to a major extent by dihydropyrimidine dehydrogenase (DPD) to inactive dihydrofluorouracil. DPD exhibits genetic polymorphism in humans. G > A transition at the exon 14 5'-splice consensus sequence characterizes about 50% of non-functional DPD alleles, leading to skipping of exon 14. Total DPD deficiency causes an inherited deficiency in pyrimidine metabolism and is usually associated with thymidine-uraciluria and potentially neurological disorders, whereas heterozygotes appear less likely to exhibit a phenotype in the absence of 5FU treatment. There is currently no strong consensus as to the optimal panel of DPD genotypes that should be determined to provide adequate sensitivity and specificity of DPD genotyping for guiding 5FU therapy.

UDP-Glucuronosyl Transferase

Reduced expression of the UDP-glucuronosyl transferase 1A1 (UGT1A1), which has been associated with increased blood concentrations of unconjugated bilirubin, has been linked to polymorphisms in the *UGT1A* coding region (Crigler-Najjar syndromes) or differences in the number of TA repeats in the *UGT1A1* promoter region (Gilbert's syndrome). Irinotecan is used to treat several human cancers and dose-limiting toxicities of irinotecan (e.g. diarrhea, leucopenia) have been associated with higher levels of the SN-38 metabolite, which is conjugated by UGT1A1. A 7-fold higher likelihood of diarrhea and/or leukopenia was

found in patients with UGT1A1 polymorphisms when compared to patients inheriting the wild type genotype.

References

1. Vogel F (1959) Moderne Probleme der Humangenetik. *Ergeb Inn Med Kinderheilkd* 12:52–125
2. Weinshilboum R (2003) Inheritance and drug response. *N Engl J Med* 348:529–537
3. Evans WE, Relling MV (2004) Moving towards individualized medicine with pharmacogenomics. *Nature* 429:464–468
4. Phillips KA, Veenstra DL, Oren E et al (2001) Potential role of pharmacogenomics in reducing adverse drug reactions: a systematic review. *JAMA* 286:2270–2279
5. Zanger UM, Raimundo S, Eichelbaum M (2004) Cytochrome P450 2D6: overview and update on pharmacology, genetics, biochemistry. *Naunyn Schmiedeberg Arch Pharmacol* 369:23–37

Pharmacophore

Definition

Pharmacophore refers to properties or functional groups that are essential for affinity of a ligand to its biological target.

- ▶ [Molecular Docking](#)
- ▶ [Peptide Aptamers](#)
- ▶ [QSAR](#)
- ▶ [Structure-based Drug Design](#)
- ▶ [Virtual Screening](#)

Phase

Definition

In the context of circadian rhythms, the term 'phase' describes the momentary state of an oscillation within a period. Phase relation refers to the distance between the phases of defined variables of two oscillatory processes.

- ▶ [Circadian Clocks](#)

Phase Problem

Definition

The central problem in crystallography arises from the fact that the electron density distribution can be

calculated from the amplitudes and phases of diffracted X-rays, but only the amplitudes can be measured. Solving the phase problem means estimating phases well enough to produce a useful electron density function.

► [3D Structure Determination by X-Ray](#)

Phases of the Cell Cycle

Definition

Cell growth and division is divided into 4 phases. In G1-phase (gap 1 phase) the cell grows and prepares for duplication of its DNA in S-phase (synthesis phase). After DNA synthesis is complete, the cell enters G2-phase (gap 2 phase), during which further growth occurs and the DNA is checked for integrity and if necessary repaired. In M-Phase (mitosis), the DNA is condensed into the characteristic mitotic chromosomes and the two copies (chromatids) are separated into two daughter cells. The cell cycle can last from between 10 to 30 hours. Cells that are no longer dividing leave the cell cycle and go into the resting G0-phase (gap zero phase).

► [Cell Cycle - Overview](#)

► [Chromosomal Instability Syndromes](#)

Phenocopies

Definition

Phenocopies are clinically indistinguishable disorders that have specifically different genetic causes and disease mechanisms.

► [Atopy Genetics](#)

► [Common \(Multifactorial\) Diseases](#)

► [Familial Dilated Cardiomyopathy](#)

► [Genetic Predisposition to Multiple Sclerosis](#)

Phenomics

ROBERT GERLAI

Psychology Department, University of Toronto at Mississauga, Mississauga, Canada
robert_gerlai@yahoo.com

Synonyms

Phenotyping; phenotypical characterization; phenotypic screening; phenotypical profiling

Definition

Phenomics is an emerging scientific field focused on phenotypical characterization of ► [mutation](#)-induced changes. It will integrate multidisciplinary biological and genetics research as well as technological advancements and will help organize and mine complex phenotypical data, thereby facilitating multifactorial understanding of complex biological phenomena.

The ease with which genetic mutations can be induced or introduced with targeted mutagenesis (e.g. ► [gene targeting](#) with the use of ► [homologous recombination](#) in ► [embryonic stem cells](#)) or random mutagenesis (e.g. with the use of ► [chemical mutagens](#) such as ENU (ethyl nitroso urea) in mammalian organisms, such as the mouse (► [Large-Scale ENU Mutagenesis Approach in Mice](#)) has created a significant need for phenotypical analysis. Developments in computer technology, instrumentation and ► [bioinformatics](#), as well as in numerous biological disciplines including neuroscience will help to meet the demand set by the molecular revolution. A new field, phenomics, is being born that will integrate multidisciplinary research with the goal of understanding the complex phenotypical consequences of genetic mutations at the organismal level. Although this essay will use examples of behavioral phenotyping to illuminate the role of phenomics in neuroscience, the general conclusions drawn should be valid for all fields of biological investigation.

Characteristics

Genetics has become a major driving force in biology. With the development of ever more powerful recombinant DNA technologies has come the need for sophisticated phenotypical screening tools. Understanding the function of genes even in such a complicated organ as the mammalian brain or studying brain diseases of our own species is now becoming possible.

For most biomedical research the preferred model organism has become the house mouse (*Mus musculus domesticus*) partly because the genome of this species has high sequence homology with that of the human. The recognition that phenotypical characterization of this species is crucial has led to the Mouse Phenome Project (3). Grant funding agencies, e.g. the National Institutes of Health (NIH, USA), have also issued requests for applications to stimulate phenotyping of genetically engineered mice (► http://www.nih.gov/science/models/mouse/genomics/priority_setting_genomics.pdf). It is forecast that the demand for sophisticated and fast phenotyping tools will broaden

the perspective of these analyses both conceptually and technologically.

An Example of Phenotyping Tools: New Ways to Probe Brain Function with Behavioral Analysis

The ultimate output of the brain is behavior and behavioral analysis has the potential to reveal functional alterations of any circuit or any neurobiological process of the brain. However, there have been serious practical issues that limited the utility of behavioral phenotyping. Behavioral tests were time consuming and thus investigators could perform only a limited number of them. The danger of too narrow a focus was that the consequences of the mutation could not be revealed properly. By now it has been widely accepted that broad-spectrum test batteries should be conducted. The questions of how many tests are needed and what should be included in a test battery are debated. It is probable, however, that novel mutations induce unexpected functional changes that may not be captured by a limited number of tests. Furthermore, the ►phenotype, including behavioral traits, is influenced not only by the effects of a large number of genes that interact with each other and modify each other's effects (►epistasis) but also by an almost certainly even larger number of environmental factors that also interact with the genetic effects and lead to the phenomenon termed genotype environment interaction. In summary, the number of "phenes" (or traits) and the number of tests needed to properly quantify them is likely to be enormous.

In the behavioral phenotyping of rodents including the mouse, some test batteries have already gained recognition [for review see (1)]. For example, the SHIRPA protocol is a conglomeration of previously characterized and individually developed tests. CANTAB (►<http://www.camcog.com>) is a system that has some ingenious computerized behavioral paradigms that allow the user to compare behavioral functions across multiple species, from rodent to man. The question of how to organize and design test batteries has been debated. The issue of whether one needs standard or custom designed tests is also not trivial (4, 5, 6). One argument for standardization is clearly the ability to compare results from lab to lab (4). However, the issue of rigidity in standardization leading to the inability to properly evaluate potentially unique genetic effects has been pitted against the advantages of cross-lab comparability (6). The emerging consensus appears to be a compromise; one needs to have a set of standard tests, a sort of reference point, but the need for creative thinking and custom made phenotypical test applications is also recognized (5). Numerous other questions have also been considered. For example, investigations as to how the order of tests in a battery may influence the outcome of the study and whether one test may

interfere with another have been started. The problems associated with lab-specific environmental factors and their interaction with the genotype of the studied subjects, i.e. genotype environment interaction, has been brought to attention and false positive and negative findings due to inappropriate control of the environment and/or the ►genetic background (2) have also been pointed out [for review see (1)]. Finally, the importance of better understanding of the ecology of the mouse and utilization of biologically/ethologically relevant paradigms has been extensively discussed [for review see (1)].

Often, the rationale behind test batteries is that they are organized hierarchically; the investigator starts from broader, less specific, tests that are sensitive to numerous factors and subsequently employs more and more specialized tests that finally tease out the details of the functional alterations of the brain. The question, however, of what the organizing principle should be has not been explicitly addressed. Pharmaceutical and biotechnology research companies may benefit from test batteries that are primarily organized according to disease target. For example, a test battery aimed at ►Alzheimer's disease research may need to cover major domains of cognition, e.g. attention, short- and long-term memory and executive function, in a manner similar to the way these behavioral phenomena are examined in the human clinic. Others may prefer organizing the test battery in a way that would allow one to tap into different neurobiological mechanisms or the function of neuroanatomical areas of the brain. For example, in the analysis of memory, one may be interested in whether acquisition, consolidation, retention or recall is affected by the mutation. One may also be interested in what brain area is affected, e.g. whether procedural learning (perhaps cerebellar function), relational learning (perhaps the hippocampus), elemental learning (perhaps the cortex) or emotional learning (perhaps the amygdala) is altered. While test batteries based upon these different organizing principles are clearly not orthogonal or mutually exclusive, they may yield different answers when employed. However, irrespective of the organization of the test battery, it has been generally appreciated that multiple tests tapping into the same phenotype (behavioral trait in this case) while employing idiosyncratic procedures and representing unique performance demands must be conducted to avoid false findings.

Speed Versus Quality

From the above it is clear that a test battery represents a significant challenge. Optimizing a battery is not trivial. The primary problem is practical; phenotyping is space and time intensive. The danger is that speed may be increased at the expense of quality. Steps have been taken to address this problem. One solution is

scalability, i.e. the increase of the number of pieces of apparatus (and thus experimental subjects) one runs in parallel. The second solution is to increase the information density of the test, i.e. the number of phenotypical measures that one can obtain from a single test. The third is to increase the flexibility of the test apparatus to enable it to tap into a broader spectrum of domains of gene function. Increased processing speed and memory capacity now allow computers to control several pieces of apparatus in which, e.g. animal behavior is monitored. Computerization also makes it possible to record numerous phenotypical measures at the same time. For example, detection of motor patterns is now possible using force transducer technology or 3D computer vision (1). The latter pioneered by PsychoGenics Inc. (Tarrytown, NY, USA) and Noldus Info. Tech. B.V. (Wageningen, The Netherlands) is particularly sophisticated and will allow quantification of a large number of behavior elements, motor and posture patterns in an automated manner. The “SmartCube” of PsychoGenics can also be equipped with numerous instruments including food receptacles, shock grids and conditioned stimulus delivery devices, thus facilitating both complex behavioral quantification and stimulus presentation. Thus, one can design and conduct countless types of behavioral paradigms in a precisely controlled manner and without the interference of a human experimenter. Another device based on the principles of scalability, increased information density and flexibility is the Intellicage developed by NewBehavior Inc. (Zurich, Switzerland, <http://www.newbehavior.com>). This system employs transponder-based technology to monitor the whereabouts of several mice in the same (intelli)cage. As the device can differentiate the individual mice by means of implanted commercially available microchips, the computer based monitoring system can tell which mouse is working to obtain reward from a receptacle, which one is running in the middle of the cage and which one is sitting in the corner frustrated by all this technological sophistication. The Intellicage is not only sophisticated but also mimics the natural habitat, a mouse community. It has been developed on the basis of information gathered in field studies and the semi-natural enriched environment it provides may facilitate the high-throughput analysis of several behavioral phenomena ranging from learning or anxiety to numerous aspects of social interaction.

Bioinformatics to the Rescue

The amount of phenotypical data gathered using such devices can be staggering. Bioinformatics tools and multivariate statistical methods may be required to properly and concisely extract information from phenotypical tests. Furthermore, phenotyping may

include a large number of different disciplines. For example, in addition to behavioral tests, *in vivo* multielectrode recording from individual neurons can be performed and the behavioral and electrophysiological quantification of brain function may be conducted *in vivo* in the same, freely moving, mouse, further increasing the demand for complex mathematical procedures in data analysis and data mining. Emergent patterns based on phenotypical “mosaic pictures” may be discovered that reflect certain neurobiological mechanisms or disease states better than individual measures. In summary, investigation of complex phenotypes will require new analytical procedures. Comprehensive databases will help the investigator. The seeds of such databases already exist. For example, the Internet accessible public domain ‘Induced Mutant Resource’ (or IMR) database (<http://www.jax.org/resources/documents/imr/>) together with another database called TBASE (<http://tbase.jax.org/>) from The Jackson Laboratory (Bar Harbor, ME, USA) provides comprehensive listing of mutant mice along with their phenotypical profiles. Open discussion forums where scientists can share data and ideas beyond the details usually provided in peer-reviewed journals have also been suggested. Bioinformatics tools that were developed to cope with the large amount of genetic information coming out from sequencing of genomes or gene expression analyses will be also crucial for organizing and interpreting phenotypical data.

Clinical Relevance

Phenomics will be defined not only by the sophistication of the experimental paradigms but also by technological complexity. Hardware and software engineers as well as biologists, bioinformaticians and statisticians will co-develop test paradigms and equipment that will enable the investigator to characterize the phenotype of genetically altered organisms in a detailed manner. Phenomics will be a crucial approach in both academic as well as industrial research and may lead to a significant paradigm shift in the genetic modeling and analysis of human diseases. These phenotyping tools will also be useful for the analysis of the effects of pharmaceutical agents, the so called small molecules, and thus the development of drug therapies.

References

1. Gerlai R (2002) Phenomics: Fiction or the Future? *Trends in Neurosciences* 25:506–509
2. Gerlai R (1996) Gene targeting studies of mammalian behavior: Is it the mutation or the background genotype? *Trends in Neurosciences* 19:177–181
3. Paigen K, Eppig JT (2000) A mouse phenome project. *Mammalian Genome* 11:715–717
4. Staay FJ van der, Steckler T (2002) The fallacy of behavioral phenotyping without standardization. *Genes, Brain and Behavior* 1:9–13

5. Wahlsten D (2001) Standardizing tests of mouse behavior: Reasons, recommendations, and reality. *Physiology and Behavior* 73:695–704
6. Würbel H (2002) Behavioral phenotyping enhanced – beyond (environmental) standardization. *Genes, Brain and Behavior* 1:3–8

Phenotype

Definition

Phenotype is the physical appearance of an organism, a tissue or a cell, determined by the genetic make-up (“genotype”) together with the environment. The term is also used to define the consequences of a particular mutation, or to describe certain cellular or molecular characteristics. Phenotypic traits are not necessarily genetic.

- Cancer Invasion and Metastasis: Cellular, Molecular and Clinical Aspects
- COPD and Asthma Genetics
- Dominant Alleles
- Epistasis in Cystic Fibrosis
- Familial Dilated Cardiomyopathy
- Heritable Skin Disorders
- Huntington’s Disease
- Large-Scale ENU Mutagenesis in Mice
- Large-Scale Homologous Recombination Approaches in Mice
- Mutagenesis Approaches in the Zebrafish
- Neurofibromatosis Type 1 (NF1), Genetics
- Protein Interaction Analysis: Suppressor Hunting
- Recessive Alleles
- Schizophrenia Genetics
- SRY – Sex Reversal

Phenotypical Characterization

- Phenomics

Phenotypical Profiling

- Phenomics

Phenotypical Screening

- Phenomics

Phenotyping

- Phenomics

Phenylalanine Hydroxylase Deficiency

- Phenylketonuria

Phenylketonuria

ULRICH LANGENBECK

Institute of Human Genetics, University Hospital,
Frankfurt/Main, Germany
u.langenbeck@em.uni-frankfurt.de

Synonyms

PKU; phenylalanine hydroxylase deficiency;
Fölling disease

Definition

PKU (► OMIM *261600, ► <http://www.ncbi.nlm.nih.gov/Omim/>) is an inborn error of metabolism with a deficient or absent activity of the enzyme phenylalanine hydroxylase (PAH, phenylalanine 4-monooxygenase, ► EC 1.14.16.1, ► <http://www.chem.qmw.ac.uk/iubmb/>). The enzyme defect is due to mutations in the PAH gene on human chromosome 12q24.1 and is inherited as an autosomal recessive Mendelian trait. Almost 500 different mutations are now known. The most severe PAH defects render tyrosine, the product of the enzymic reaction, an essential amino acid. With such defects, the PAH substrate phenylalanine (phe) can only be cleared through protein synthesis and, less efficiently, through minor alternate metabolic routes and urinary excretion. The clinical chemical results are elevated blood levels of phe (i.e. hyperphenylalaninemia) and of acidic phe metabolites. Without strictly

managed dietary treatment this will lead to an irreversible state of mental retardation.

Excluded from this definition are the various types of atypical severe phenylketonuria with genetic defects of PAH cofactor (tetrahydrobiopterin) recycling or synthesis, e.g. dihydropteridine reductase deficiency (OMIM *261630) and GTP cyclohydrolase deficiency (OMIM #233910, *600225). They make up only about 1 per cent of all cases of hyperphenylalaninemia.

Characteristics

PKU deserves interest from a multitude of points of view. It has become the textbook example of a classic inborn error of metabolism which covers almost all aspects of modern biochemical genetics of man, including biochemical pathogenesis, diet therapy, population-wide newborn screening and maternal disease. PKU also contributes to the study of population history because the distribution of different mutations may serve as a tracer of ancient population movements. And, research in PKU contributed one of the first disease models in mice (1, 2)

Clinical Characteristics

The clinical phenotype of untreated PKU is characterized by severe mental retardation, a seizure disorder, microcephaly, fair complexion of skin and hair, eczema and, sometimes, a peculiar 'mousy' odor. None of these signs would allow a specific diagnosis. It was Föllings discovery in 1936 of a highly specific sign (the urine of some mentally retarded patients turns green on addition of an acid solution of ferric chloride) that made possible the delineation of PKU as a clinical entity. The next step, also done by Fölling, was the identification of phenylpyruvic acid (PPA) as the agent involved in the color reaction.

PPA, the transamination product of phe, is chemically unstable and its excretion may be low. Furthermore, the ferric chloride reaction is rather insensitive. It therefore was Fölling's discovery of increased blood levels of phe that finally established a diagnostic criterion with 100% sensitivity. Testing for hyperphenylalaninemia enabled the diagnosis of PKU in all suspected cases, e.g. among the residents of institutions for the mentally retarded where about 1% of the elderly are afflicted with severe types of PKU.

In the young population, the unfavorable and irreversible course of the disease has almost disappeared in the industrial nations due to newborn screening and subsequent initiation of diet therapy. Nevertheless, reversible cognitive features may be observed when, at times, the rules of diet therapy are not obeyed. Such signs of temporarily elevated blood levels of phe are headache and the inability to work with concentration. If the diet was discontinued early, e.g. at the age of 6, adults may manifestly reduced intelligence scores.

Those with normal intelligence seem to possess, as a protective mechanism, a reduced capacity for phe transport from blood to brain.

With regular monitoring of blood phe and with firm adherence to the low-phe diet therapy well into adulthood, the regular outcome is a PKU person with a normal mental phenotype and successful social adaptation.

As a phe-dependent embryo-fetopathy, maternal PKU has emerged as a second clinical manifestation of PKU, this time in the heterozygous offspring of homozygous mothers who never knew of their disease or who forgot about their metabolic disease after treatment was discontinued at an earlier age. If maternal blood levels exceed 20 mg per dl, offspring will be mentally retarded in over 90%, will be microcephalic in about 75% and will have heart malformations in about 12% of cases. Birth weight will be reduced.

Therapeutic Management

The dietary treatment of PKU commenced in 1953 when Bickel, Gerrard and Hickmans documented improved mental development of PKU children on a low-phe diet. This early success made any controlled study unethical. It therefore remained unknown in the beginning that 2–5% of people with classic PKU will more or less escape mental retardation without dietary treatment and the mild non-PKU hyperphenylalaninemia (MHPA) has only recently been established as an entity that does not require a PKU diet. Diet therapy must start within the first weeks of life in order to avoid the irreversible cognitive sequels of hyperphenylalaninemia.

The rules of dietary treatment concern (i) the level of phe that should be attained and (ii) the question of whether or when the treatment could cease. Presently, treatment centers aim at 300 micromoles phe per liter (i.e. 5 mg/dl) and life-long dietary surveillance is now uniformly recommended.

To prevent the maternal PKU embryo-fetopathy, pregnant PKU women must be treated as strictly as PKU children and optimal treatment will start pre-conceptionally.

Additional therapeutic options are (i) a supplementation with the cofactor tetrahydrobiopterin which stabilizes a number of different enzyme mutants, thus acting as a 'pharmacological chaperone' and (ii) supplementation with large neutral amino acids which inhibit the transport of phe from blood to brain (see above). (iii) A still experimental option is dialysis with phe ammonia-lyase (PAL, EC 4.3.1.5) containing cartridges. PAL converts phe into ammonia and *trans*-cinnamic acid. The latter product becomes metabolized to hippuric acid, which is finally excreted into urine. (iv) Gene therapy is presently being investigated with genetically engineered keratinocytes as one of the possible future methods.

Newborn Screening

During pregnancy, the metabolic defect of the fetus with PKU is corrected through transplacental outflow of phe into the maternal circulation. Thus it is only after birth that hyperphenylalaninemia ensues. In classic cases, then, blood levels may exceed 40 mg/dl, the normal level being about 1 mg/dl.

The semi-quantitative bacterial inhibition test invented by Guthrie and Susi measures the outgrowth in agar plates of *Bacillus subtilis* spores in the presence of the phe analog β -thienylalanine. Quantitative data are obtained by enzymatic screening with phe dehydrogenase, E.C. 1.4.1.20. Tandem mass-spectrometry has emerged as a general and highly specific method for screening for a larger number of inborn errors of metabolism, including PKU.

PKU screening has become paradigmatic in a number of ways. To be socially and economically acceptable, any screening procedure must fulfill all of the following three criteria. (i) It must be robust, sensitive, and specific. (ii) The disease to be screened for must have an appreciable incidence at birth and (iii) the disease must be treatable. It therefore would make no sense to screen for PKU in Finland where the disease is almost absent and it is highly controversial to screen for ►Duchenne muscular dystrophy as long as its lethal course cannot be averted by therapeutic measures.

Population Genetics of PKU Mutations

Population history (bottlenecks and migration), population structure (degree of inbreeding, religious and geographic isolates) and chance (so-called genetic drift) determine the frequency of given mutant PKU alleles and of PKU incidence in general. Countries with a high PKU incidence are Turkey (1/2,600), Ireland (1/4,500), and Scotland (1/5,300). The incidence in Poland and Germany is about 1 in 7 to 8,000, whereas in UK, France and Switzerland it is comparatively low (about 1/15,000). In countries with high PKU incidence, pediatricians must recall the possibility of maternal PKU when caring for a child with unexplained microcephaly and congenital heart disease.

Of special interest is the geographical distribution of mutations and the regional changes (clines) of their frequencies. It shows, for example, that the mutation R408W (arginine in position 408 changed to tryptophan) originated once in Ireland and once in Eastern Europe. The distribution of other mutations allows estimates of their age and may give hints as to the peopling of countries.

Mouse Models of PKU

Artificially created mouse mutants have become indispensable objects in the study of the pathogenesis and therapy of human inherited diseases.

In the 1970s, experimental PKU models were created by treating mice with high doses of phe together with the PAH inhibitor *para*-chlor-phenylalanine. Since 1990, real mouse PAH mutants are available. They are created by male germline mutagenesis with N-ethyl-N-nitrosourea (ENU). The second of these mutants, the *Pahenu2* mouse, is due to a phe → serine substitution in exon 7 of the PAH gene and has a severe metabolic phenotype.

Cellular and Molecular Regulation

Research in PKU also set the stage for genotype-phenotype correlations in inborn errors of metabolism. In the beginning, the mutant enzymes were studied in patient liver biopsies. With the advent of molecular techniques, it became possible to analyze enzymes with defined mutations in transfected cell lines and relate these data to the clinical outcome.

Now, that the 3D-structure of the PAH enzyme has been elucidated, it has become possible to deduce the mechanism of action of a given mutant, as it was possible to do with the hemoglobinopathies when the 3D-structure of the adult hemoglobin tetramer had been defined in the 1960s. Despite the power of the modern approaches, it has become increasingly evident that predictions may fail. Although the PAH enzyme defect obeys the rules of Mendelian inheritance, PKU turns out to be a multifactorial disorder with environmental and (still unknown) genetic factors contributing to clinical variability, even within families (3).

Types of Mutations

The PAH gene was cloned in 1983 and the first PAH mutation, a deletion, was identified in 1985. At present, almost 500 mutations are known, and most of them were discovered between 1990 and 1995. With 63%, missense mutations are the most frequent type, followed by 13% small deletions, 11% splice mutations and 5% nonsense mutations. There are 7% putative silent mutations, e.g. V399V, where a DNA base change A → T in codon 399 changes the valine codon CAA to the valine codon CAT. Some of the putatively silent mutations and also some missense mutations turned out to be splice mutations, because they had created a new splice site.

With increasing age and wider geographical distribution of a mutation, there is a growing probability that a second mutation will be acquired 'in *cis*' by one of the mutant alleles, either by a mutational event or by homologous crossing-over in a compound heterozygote. The beta hemoglobin sickle cell (HbS; HBB glu 6 val) mutations Hb C Harlem / Georgetown (HBB glu 6 val + asp 73 asn), and Hb C Ziguinchor (HBB glu 6 val + pro 58 arg), provided the first examples of *cis*-mutations, six being known by now for Hb S alone. There are 12 known *cis* mutant alleles of the

PAH gene. The ancient R408W mutation, for example, has acquired a *cis* mutation in each of two different alleles, E183Q in one and D315Y in another. Undetected *cis*-mutations may mislead the interpretation of disease alleles and of innocuous DNA polymorphisms. In the cell, the PAH enzyme exists in an equilibrium state of homo-dimers and homo-tetramers. Each subunit is composed of 452 amino acids (aa) and structured into three functional domains, the N-terminal regulatory domain (aa 1-142), the middle catalytic domain (aa 143-410) and the C-terminal tetramerization domain (aa 411-452). With regard to this structural basis, PKU mutations may be classified as active site mutations, dimer interface mutations and domain structure mutations (4). Somewhat unexpectedly, not all active site mutations lead to classic PKU, e.g. Y277D that concerns an evolutionarily highly conserved position. Domain structure mutations cause a broad spectrum of PKU phenotypes. They are pathogenic mostly by interfering with the overall protein stability, e.g., when a large hydrophobic aa is replaced by a smaller one (leucine by valine) or *vice versa* (valine replaced by methionine) or when a hydrophobic aa is replaced by a polar one (phe by serine).

Biochemical Parameters of Mutant Enzymes

Enzymatic parameters that can be altered or nullified by a mutation are substrate and cofactor affinities, maximum catalytic velocity, protein stability at different temperatures and amount of immunoreactive material. If an enzyme is easily available, like erythrocyte glucose-6-phosphate dehydrogenase, all these parameters can be estimated for characterizing its mutants. However, because PAH is primarily a liver enzyme, only a few early studies were made on native liver biopsy samples. They established a clear relationship between residual PAH activity and severity of PKU phenotype. MHPA patients were found to have $\geq 5\%$ of normal PAH activity. Modern *in vitro* expression analysis (IVE) uses transfected cell lines (COS, human kidney), yeast and *E. coli* to analyze enzyme activity, immunoreactive protein, specific activity and amount of mRNA, all expressed as % of wild type. Most of these data do more or less correlate with the clinical phenotype. An interesting exception is the catalytic domain mutant A309V with high activity *in vitro* and a severe phenotype *in vivo*. Although the mutation density is 2–3 \times higher in the catalytic domain sequence than in the two other domains, IVE data indicate that most PAH mutations cause PKU by interfering with normal folding and by destabilizing the enzyme protein with consequent formation of inactive protein aggregates. PKU has therefore been called a 'protein folding disorder'.

Mild mutations and most recently also more severe ones have been shown to be tetrahydrobiopterin-responsive. Their different mutant sites are found all over the protein, pointing to a general stabilizing effect. This effect is now therapeutically exploited.

Classification of PKU and the Prediction of Metabolic Phenotype

A reliable prediction of the metabolic phenotype early in life is important for two reasons; management of diet therapy can then be individualized and informing parents about the disease will be based on validated prognostic features. In the pre-genomic era, Blaskovitz had devised a three days protein-loading test for evaluating the phe tolerance of six months old babies. In the severe types of PKU, very high phe blood concentrations could ensue during this test. Alternative measures to define disease severity were the initial phe levels in newborns, loadings with tritiated phe or unlabelled phe and, in older patients (5 years), the daily amount of ingested phe that was tolerated while keeping phe blood levels at 5 mg/dl. For classical, moderate, mild PKU and MHPA, these amounts were set at ≤ 20 , 20–25, 25–50 and ≥ 120 ('normal diet') mg phe per kg body weight per day.

A classification system for adults is based on phe blood levels during 'normal diet', classical PKU 20, moderate PKU 10 and MHPA 4 mg/dl, respectively. A metabolic model has been published (5) which enables translation of such values into kinetic parameters of phe disposal. These clinical classifications of metabolic phenotype can now be related to the different PKU mutations, so that predictions have become possible in newborns (6). The system is based on the observation that in compound heterozygotes the metabolic phenotype is determined in an additive fashion. Homozygotes for a null mutation and compound heterozygotes for two different null mutations will have classic PKU, e.g. R408W/R408W and R408W/R158Q. Compound heterozygotes of a null and a mild mutation ('functional hemizygotes') display a mild metabolic phenotype; two moderate or mild mutations act additively. Classic, mild, moderate and MHPA mutation are scored in a geometric series as 1, 2, 4, and 8 respectively and the sum of the two scores predicts the metabolic phenotype: 2 \rightarrow classic; 3 \rightarrow moderate; 4 \rightarrow moderate/mild; 5 and 6 \rightarrow mild; 8 \rightarrow mild/MHPA; 9, 10, 12, and 16 \rightarrow MHPA.

Pathogenesis of Metabolic and Cognitive Phenotype

The above predictive system (6) was evaluated in 184 patients, and discrepancies were found in about 20% of cases. This indicates that individuals may differ with regard to the 'handling' of their mutations. This

postulate is corroborated by observations of siblings with identical genotype but different phenotypes and it receives further support from the finding that the mutations I65T and R261Q are overrepresented in patients with discordant phenotypes. It can, at present, only be surmised which factors modulate the metabolic effect of mutations. They may act at the level of the PAH enzyme (e.g. capacity of chaperones and proteases), at the level of intermediary metabolism (capacity of secondary metabolite pathways) or at the level of renal physiology (urinary excretion).

A great number of hypotheses have been formulated to explain the finding of varying cognitive capacities in untreated or 6 years only treated PKU patients. Beside the surmised influence of neurotransmitters, the only validated relation is the blood-brain barrier phe transport and the corresponding intra-cerebral concentration of phe. The latter, as measured with magnetic resonance spectroscopy, correlates inversely with the cognitive capabilities and is in the range seen in heterozygotes when intelligence in the patients has remained normal. This points to phe itself as the major pathogenic factor in the cognitive outcome of PKU.

► **Predictive Genetic Testing**

References

1. Scriver CR, Kaufman S (2001) Hyperphenylalaninemia: Phenylalanine hydroxylase deficiency. In: Beaudet AL, Sly WS et al (eds) *The metabolic & molecular bases of inherited disease*, 8th ed., vol. II, pp 1667–1724. McGraw-Hill, New York and chapters on PKU in the former editions since 1960
2. Scriver CR, Hurtubise M, Konecki D et al (2003) *PAHdb2003*: What a locus-specific knowledgebase can do. *Hum Mutat* 21:333–344. Data from additional articles in the same issue are mentioned in the paragraphs above
3. Scriver CR, Waters PJ (1999) Monogenic traits are not simple: lessons from phenylketonuria. *Trends Genet* 15:267–272
4. Jennings IG, Cotton RGH, Kobe B (2000) Structural interpretation of mutations in phenylalanine hydroxylase protein aids in identifying genotype-phenotype correlations in phenylketonuria. *Eur J Hum Genet* 8: 683–696
5. Langenbeck U, Zschocke J, Wendel U et al (2001) Modelling the phenylalanine blood level response during treatment of phenylketonuria. *J Inher Metab Dis* 24:805–814
6. Guldborg P, Rey F, Zschocke J et al (1998) A European multi-center study of phenylalanine hydroxylase deficiency: classification of 104 mutations and a general system for genotype-based prediction of metabolic phenotype. *Amer J Hum Genet* 63:71–79

Pheochromocytomas

Definition

Ph(a)eochromocytomas are mostly benign catecholamine producing tumors of chromaffine tissues. More than 80% of these tumors are found in the adrenal medulla.

► **G-Proteins**

Phlebotomy

Definition

Phlebotomy or *venea sectio* describes the collection of blood from a vein section or usually a puncture.

► **Hemochromatosis**

Phosphatidyl Serine

Definition

Phosphatidyl serine (PS) is a membrane phospholipid that is located in the inner cytoplasmic leaflet of the cell membrane. In apoptotic cells, PS is translocated within the cell membrane so that it appears on both the inner and the outer surface. PS represents an important surface molecule that helps to recognize the cell as an apoptotic one for phagocytes.

► **Apoptosis**

► **Apoptosis, Regulation and Clinical Implications**

► **Biological Membranes**

Phosphatidylinositides

► **Phosphoinositides**

Phosphatidylinositol 3-Kinase

- PI3K

Dephosphorylated PLB inhibits binding of SERCA2 to Ca^{2+} , whereas phosphorylation facilitates Ca^{2+} binding and transport into the SR lumen, enhancing cardiac function.

- Cardiac Signaling: Cellular, Molecular and Clinical Aspects

3'-Phosphoadenylyl-Sulfate

- Tyrosine Sulfation of Proteins

Phospholipase A2

- PLA2

Phosphodiester

Definition

Phosphodiester refers to any molecule, such as nucleic acids, in which two parts are joined through a phosphate group.

- Splicing

Phosphor-Imaging

Definition

Phosphor-imaging plates present an alternative to X-ray film. The radiation image is stored in proportion to its intensity distribution on a storage phosphor screen. It can be visualized with the appropriate technical equipment.

- Protein Microarrays as Tool for Protein Profiling and Identification of Novel Protein-Protein Interactions

Phosphoinositides

Definition

Phosphoinositides (inositol phospholipids; synonym: ► **Phosphatidylinositides**) are phospholipids with an inositol ring in their headgroup, which can be phosphorylated at one or several positions, and are called phosphoinositides. They play important roles in signal transduction and in the recruitment of peripheral membrane proteins in eukaryotic cells.

- G-Proteins
- Protein and Membrane Transport in Eukaryotic Cells
- Rho, Rac, Cdc42

Phosphorylation

Definition

Phosphorylation refers to a chemical reaction by which a phosphate group is added to a molecule. The activity of many proteins, and subsequently many cellular processes, are regulated by phosphorylation and dephosphorylation of hydroxyl-containing amino acids (serine and threonine or tyrosine) by protein kinases and phosphatases, respectively.

- Cancer Invasion and Metastasis: Cellular, Molecular and Clinical Aspects
- Kinases
- Nuclear Hormone Receptors: Regulators of Gene Transcription and Cellular Signaling
- Protein Microarrays as Tool for Protein Profiling and Identification of Novel Protein-Protein Interactions
- Splicing

Phospholamban

Definition

Phospholamban (PLB) is a 52 amino acid, homopentameric protein that regulates the cardiac Sarco-plasmic Reticulum Ca^{2+} -ATPase (SERCA2).

Photobleaching

Definition

Photobleaching denotes the light induced irreversible destruction of a fluorophore's fluorescence properties. This can occur during the course of exposing a fluorophore to a light source, or it can be regulated by controlled exposure of a discrete ROI (►[region of interest](#)) to intense laser light. After ca. 10^5 excitation-emission cycles, most common fluorophores tend to undergo photodeconstruction, i.e. an irreversible transition to a non-fluorescent state.

- FCS
- FRAP
- FRET

Photoconversion

- Uncaging and Photoconversion/Activation

Photolithographic Synthesis

Definition

Photolithographic synthesis refers to a manufacturing technology of biomolecular patterning, where UV light is used to activate or deactivate chemical species.

- Uncaging and Photoconversion/Activation

Photolithography

Definition

Photolithography is a process used in semiconductor device fabrication to transfer a pattern from a photomask to the surface of a wafer or substrate.

- PNA Chips

Photomorphogenesis

Definition

Photomorphogenesis refers to light dependent changes in morphology and/or development.

- Photoreceptors

Photoreceptor Cells

Definition

Photoreceptors are photosensitive cells of the retina that are specialized in converting light into neuronal signals. In the human retina there are two types of photoreceptors, rods and cones.

- Cell Polarity

Photoreceptors

MARTIN ENGELHARD¹, KLAUS PETER HOFMANN²

¹Max Planck Institute of Molecular Physiology, Dortmund, Germany

²Institute of Medical Physics and Biophysics, Charité, and Center of Biophysics, Humboldt University, Berlin, Germany

martin.engelhard@mpi-dortmund.mpg.de

kph@charite.de

Definition

Living cells react to external stimuli, in order to perform their functions in organisms or to adapt to their environment. Stimuli are realized in physical or chemical signals which are often detected by specialized membrane receptor proteins. For the perception of light Nature has developed exceedingly diverse solutions, which are in many – if not all – cases driven to the physical limit of single photon detection. Beyond every day's experience, light controls much more than 'sees the eye'; almost every organism monitors and reacts towards illumination. In physical terms, photoreceptors serve to capture the short lived excited state caused by photon absorption, in order to channel the energy of the excited state into a conformationally or chemically altered ►[chromophore-protein interaction](#).

A high percentage of the photonic energy is thereby stored in the chromoprotein. Examples include *cis/trans* isomerization in ►**rhodopsins** or phytochromes and cysteinyl adduct formation in phototropins. By extending the local alterations near the chromophore to functionally defined domains, the receptor protein becomes competent to form a long lived signaling state. This is recognized by specialized coupling components of a downstream regulatory network. Biological outputs include neuronal signaling, behavioral responses, circadian adjustment, DNA gene regulation or protein expression.

Characteristics

Absorption of light serves many biological functions. The free energy of visible light can be fed into energy transduction and photosynthesis in plants and microorganisms or provide a signal for information flow as in vision or be used for the entrainment of a circadian clock. Typically, each family of ►**chromoproteins** performs more than one of these purposes by one or several of its members (1). The archetype in signal transduction is the visual pigment rhodopsin, the “visual purple” in the retinal rods and cones (2). It has numerous cousins in the still expanding family of ►**retinal proteins**. Vertebrate proteins include those serving circadian entrainment, a role most likely played by the melanopsins located in the photosensory ganglions of the retina and also by pinopsin which is expressed in the avian pineal (3). Photoisomerases (►**RGR**) located in retinal pigment epithelium use photonic energy for the backisomerisation of photolysed all-*trans*-retinal into the 11-*cis* form needed for visual pigment regeneration. Bacterial rhodopsins which display some structural similarities with vertebrate rhodopsins are found throughout all three kingdoms of life and serve sensory as well as energy needs of the cell. A new class of retinylidene proteins, channelopsins, are light activated proton and/or cation channels. Another family of photoreceptors, ►**cryptochromes** encompass flavoproteins which are sensitive to blue light. These receptors, which share sequence homology with light dependent DNA repair photolyases, have been detected in plants and as circadian clock components in animals (4). Finally, phytochromes with phytochromobilin as chromophore, absorbing predominantly the red and far-red region of the spectrum are involved in plant growth and development.

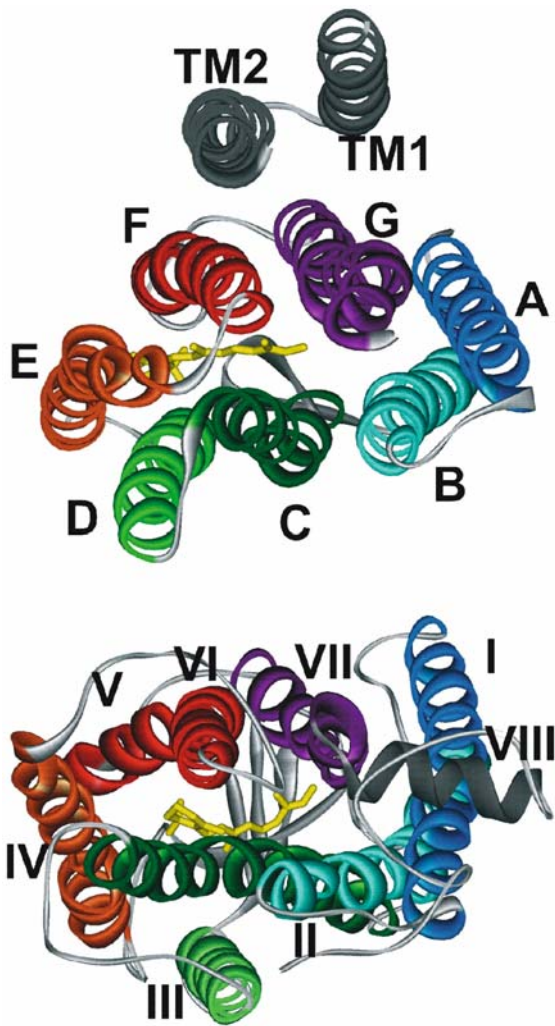
Photoreceptors are found not only in cell membranes but also as soluble proteins or as a part of larger multidomain proteins in the interior of the cell. Intracellular photoreceptors often appear as domains of larger polypeptides. In this way, ►**LOV domains** harboring the flavin chromophore can be directly connected to histidine kinase effectors. Such an

arrangement may guarantee the most efficient transmission of the light signal from receptor to effector. Receptors that bind diffusible extracellular ligands have to connect to both sides of the plasma membrane to carry the signal to the cell interior. Although there is no such requirement for photoreceptors, the membrane bound proteins are in their general topology strikingly similar to “normal” receptors not carrying a chromophore. This may reflect a common evolutionary origin and/or the same signal processing principle. Examples include visual signal transduction, where rhodopsin shares the seven transmembrane structure of other G-protein coupled receptors (►**GPCRs**). Another well-characterized representative is bacterial ►**phototaxis**. In this latter example the receptor comprises a seven transmembrane protein (SR I and II) and a tightly bound transducer (►**Htr** I and II), which in turn displays homology to the cytoplasmic domain of eubacterial ►**chemotaxis** receptors. Interaction partners are G-proteins like transducin (Gt) or histidine kinases like CheA, activating the visual cascade in rods and cones or the two component system in eubacterial photo- and chemotaxis, respectively. These two receptor systems are currently the best understood and reveal instructive similarities and differences (2, 5, 6). Their elucidation may help to understand common principles in photoreceptor signaling and even signal transfer for stimuli other than light.

Molecular Interactions

Topology

The extracellular (intradiscal) and intracellular regions of seven helix receptors (Fig. 1) each consist of three interhelical loops (E(xtracellular) and C(ytoplasmic), respectively), and terminal (COOH- or NH₂-) tail regions. The cytoplasmic domain of the transducer protein in bacterial phototaxis displays short connecting loops of transmembrane helices and is highly homologous to the corresponding section of chemotaxis receptors. It has been shown that chimeric proteins convey the ability for phototaxis to *E. coli*, demonstrating the functional identity of chemo- and photo-taxis signaling pathways. The complex between sensory rhodopsin and the transducer forms in a 2:2 stoichiometry. In the rhodopsin of vertebrate rods, a fourth cytoplasmic loop is introduced by anchoring the C-terminal tail to the membrane *via* two Cys residues which carry palmitates in the native structure. It also contains the short helix H-VIII, adjacent to H-VII, which runs parallel to the cytoplasmic surface. Rhodopsin is contained in the flat disc vesicles, which fill the outer segment of the rod cell like a stack of coins. Rod cells share with all sensory cells a bipartite outer/inner segment structure with a thin connecting cilium. Transducin (Gt) is bound to the cytoplasmic surface of the discs and, by interaction



Photoreceptors. Figure 1 Crystal structure of rhodopsin (lower panel) and the sensory rhodopsin II-transducer complex (upper panel) as viewed from the cytoplasm. Corresponding helices display the same color. Helix VIII (rhodopsin) and TM1, TM2 are depicted in grey.

with light-activated rhodopsin, catalytically converts into the active GTP-bound form which in turn activates a cyclic GMP phosphodiesterase. This effector hydrolyses the second messenger cyclic GMP. Its low concentration results in the closure of cyclic GMP dependent ion channels.

Receptor Activation and Regeneration

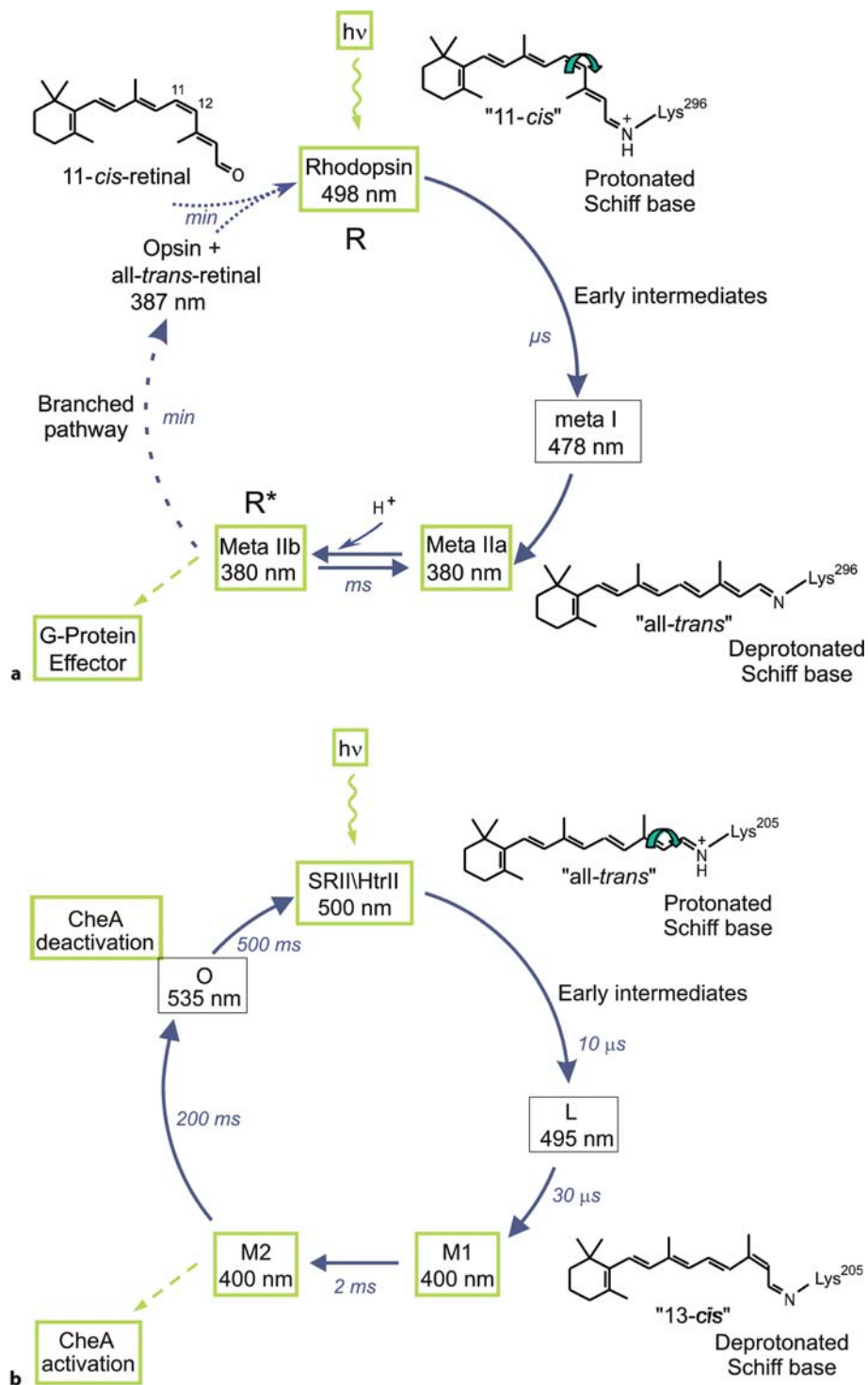
The absorption of a photon causes isomerization of the retinal chromophore between *cis* and *trans* configurations, followed by thermal relaxations which ultimately lead into the signaling state. The covalent retinylidene linkage between chromophore and apoprotein is

provided by a ►Schiff base. In the dark ground state, the nitrogen of the Schiff base is protonated and stabilized by a counter ion, provided by the negatively charged carboxylic group on a Glu or Asp residue in the third helix. After light excitation the retinal chromophore undergoes fast initial motions on the excited-state potential surface. Eventually, retinal reaches a region in the configurational space, where the transition to the product state can take place. The subsequent reactions leading to the signaling state were in both systems initially characterized by colored intermediates (such as K, L, M in the case of sensory rhodopsin II or Batho, Lumi, Meta I, Meta II in the case of rhodopsin, denoted according to the respective UV/Vis absorption spectra). The conversion into the activating conformation is marked by a UV absorbing “bleached” product (M or Meta II), in which the Schiff base is still intact but deprotonated (Fig. 2).

The decisive difference between vertebrate and bacterial rhodopsins concerns the reversibility of the light induced pathway (Fig. 2). In bacteria, thermal reactions restore the original all-*trans* ground state. Contrarily, after light excitation of the 11-*cis* retinal-opsin ground state complex, vertebrate rhodopsins follow a unidirectional mechanism which finally leads to the hydrolysis of the retinylidene Schiff-base and the export of all-*trans* retinal. This is enzymatically reisolomerized into 11-*cis*-retinal in neighboring pigment epithelium cells. In addition, the back transport of 11-*cis* retinal to opsin requires several enzymatic and transport steps. This retinoid cycle takes minutes to be completed. The question arises as to why nature has chosen such an elaborate cycle to regenerate the light sensitive receptor. Not even the absorption of blue light allows the restoration of the original 11-*cis*-retinal-opsin complex. This is likely to be the price to be paid for the exceedingly stable ground state provided by this form of retinal (6 and citations therein). Apparently, this specific 11-*cis* retinal-protein configuration is necessary for the complete inhibition of thermal receptor activation in the dark on the one hand and single quantum excitation on the other hand.

Properties of the Signaling State

Properties of the ►signaling state are best understood when considering the dark ground state first. A major stabilizing element is the salt bridge from the Schiff base nitrogen to its counter ion (Asp⁷⁵ in SR/Htr, Glu¹¹³ in rhodopsin). The X-ray structure of the SR II/Htr II complex revealed a tight interaction between helix F_{SRII} and TM2_{HtrII}. Vertebrate rhodopsin displays special interactions between helices, including the tripeptide Glu¹³⁴-Arg¹³⁵-Tyr¹³⁶, highly conserved throughout GPCRs, which contributes to stabilizing the hydrogen bond network between helices III and VI. Met²⁵⁷ in H-VI is surrounded by a Leu cluster



Photoreceptors. Figure 2 The photoreaction cycle of (a) rhodopsin and (b) sensory rhodopsin II. Only the intermediates involved in the direct formation and sustenance of the signaling state are depicted.

extending to H-II and H-III. The NPxxY(x)_{5,6}F motif extends over H-VII and H-VIII causing a rectangular kink by hydrophobic interaction between Tyr³⁰⁶ and Phe³¹³. Certain interactions must persist to reach the

signaling state. They comprise the hydrophobic interaction between Tyr³⁰⁶ and Phe³¹³ in rhodopsin and may include a strong hydrogen bridge between Tyr¹⁹⁹ in SR with Asn⁷⁴ in Htr.

The transition into the signaling state is enabled in both systems by a breakage of the Schiff base stabilizing salt bridge. It is brought about by proton transfer from the Schiff base to the counter ion and thus defines the M or Meta II photointermediate. A second step reveals itself as a spectrally silent conversion, termed M1 \rightarrow M2 or MetaIIa \rightarrow MetaIIb. It is found either by kinetic analysis of the photocycle in SR or by proton uptake to Glu¹³⁴, which breaks interactions around Glu¹³⁴-Arg¹³⁵-Tyr¹³⁶. Light-induced helix motion, which has been demonstrated for both receptors, is in SR/Htr definitely linked to M1 \rightarrow M2. Helix F/ VI tilts out of the helix bundle, with the hinge probably located at Pro 175 or a Gly-Gly tandem in SR and H-VI, respectively. Helix F motion in SR would tangentially collide with TM2, thus inducing its rotation (Fig. 1).

Regulatory Mechanisms

Once the signaling state of the receptor is reached, the signal has to be processed by the downstream components of signaling pathways. Although mechanistic details of the signaling machineries in living cells are quite diverse they do exhibit a number of operating principles in common. To meet the cell's need, the incoming signal has to be amplified and must be switched off on demand. In bacterial signal transduction, the signaling cascades usually belong to regulatory systems which integrate over different extracellular inputs and which may additionally incorporate information from the physiological state of the cell. The cells also have to control the effector activity over several orders of magnitude. Unlike bacterial signal transduction, the rod cell is highly specialized serving like the olfactory cells only one purpose. In vision the system has to cope with light intensities ranging from single quanta to brilliant sun light as encountered in snow, i.e. 11 orders of magnitude. Depending on the species, a substantial part of the adaptation utilizes the transduction machinery in the sensory cell itself.

Amplification of the Light Signal

The ►**amplification** of the incoming signal is different for the rhodopsin signal transduction chain from that of the bacterial photoreceptors. The activation of a G-protein, as a catalytic process, has the capacity for "particle amplification" in the first step of signal relay. Amplification is thereby driven by the GTPase cycle of the G-protein. Signal transfer from SR/Htr to the His-kinase CheA however follows a stoichiometric signal transfer. This receptor family and its circuit elements show high homology to the chemotaxis system. For the latter it has been shown that the receptors form patches

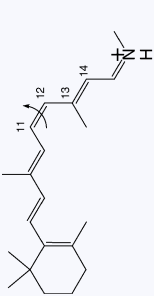
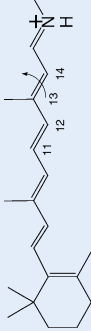
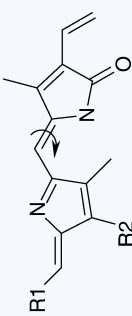
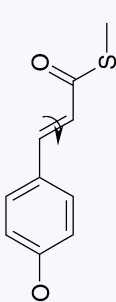
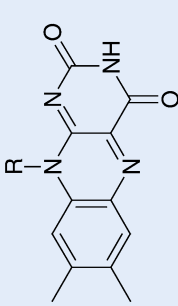
and higher order receptor assemblies. Amplification based on the idea that the activation of one receptor may affect numerous neighboring receptors would be fundamentally different from that involving G-proteins, because there is no energy fed into the system except the photo- or chemo-activation.

For vertebrate rhodopsin, a supramolecular arrangement (rows of dimers) was also reported. Although a constitutive dimerisation of rhodopsin has been challenged, based on the diffusional freedom of rhodopsin in dark-kept intact rod cells, a "functional" dimerisation of GPCRs may well play a role in different stages of their activation/ deactivation. The actual amplification at the G-protein stage, more than hundred in the rod cell, is much lower in the microvillar photoreceptors of insects. Their transduction chain *via* Gq/phospholipase C, feeds into a heteromultimeric signaling complex ("transducisome") of ►**trp channels**. Subcellular translocation of such channels has been shown to be involved in long term adaptation. Less clear is the role of a rod transducisome, which forms through an unusual set of glutamic-acid-rich proteins of unknown function. While a translocation of components of this complex has never been observed, Gt and arrestin undergo translocation between the inner and outer segment and through the thin connecting cilium. Extended illumination depletes the outer segment of Gt but enriches it with arrestin.

Deactivating Interactions and Adaptation

Light activation must be reset on a subsecond time scale, to allow repeated activation of the transduction system. In the case of the bacterial photo- and chemoreceptors, the deactivation and adaptation of sensor kinase activity is an intricate interplay of two proteins which methylate and demethylate specific residues located in the cytoplasmic domain of the transducer (or chemoreceptor). In this way the phosphate flow is regulated by integral feedback control. In the rod, the dark level of cyclic GMP is restored by a feedback loop activating a guanylate cyclase *via* calcium-binding protein(s). A complete reset also implies the deactivation of the cascade at all its stages, namely, receptor, G-protein and effector. The Gt-phosphodiesterase complex is deactivated by the intrinsic GTPase activity of the G-protein, up-regulated by RGS-9, a member of the "regulators of G-protein signaling" family. Rhodopsin is deactivated by interaction with rhodopsin kinase, phosphorylation and subsequent binding of arrestin. Ultimate deactivation in the dark requires regeneration of rhodopsin with fresh, metabolically supplied 11-*cis*-retinal. This closes the rhodopsin cycle and makes the receptor available for renewed light activation.

Photoreceptors. Table 1 Brief compilation of various photoreceptor classes

Chromophore	Primary reaction	Photosensor	Cellular location	Signaling cascade (components)	Physiological response
	11-cis → all-trans	Rhodopsin	Specialized photoreceptor membrane Plasma membrane	G-protein Second messenger	Neural response Circadian synchronization
	all-trans → 13-cis	Bacterial Rhodopsin	Bacterial plasma membrane	Two component system His kinase	Phototactic response
	cis → trans	Phytochrome	Cytoplasm Nucleus	Two component system His kinase (prokaryotes) Ser/Thr kinase Transcription factors (plants)	Photomorphogenesis Phototactic response Circadian synchronization
	trans → cis	Xanthopsin	Cytoplasm	Not known	Phototactic response
	Electron transfer? Covalent adduct formation	Cryptochrome Phototropin LOV domain	Cytoplasm Cytoplasm	Kinase activation DNA-binding In the dark: ubiquitination Two component system His kinase (prokaryotes) Light regulated Ser/Thr kinase activity (plants)	Photomorphogenesis Circadian synchronization Photomorphogenesis

References

1. Van Der Horst MA, Hellingwerf KJ (2004) Photoreceptor proteins, "Star Actors of Modern Times": a review of the functional dynamics in the structure of representative members of six different photoreceptor families. *Acc Chem Res* 37:13–20
2. Okada T, Ernst OP, Palczewski K et al (2001) Activation of rhodopsin: New insights from structural and biochemical studies. *Trends Biochem Sci* 26:318–324
3. Hattar S, Lucas RJ, Mrosovsky N et al (2003) Melanopsin and rod-cone photoreceptive systems account for all major accessory visual functions in mice. *Nature* 424:75–81
4. Brudler R, Hitomi K, Daiyasu H et al (2003) Identification of a new cryptochrome class. Structure, function, and evolution. *Mol Cell* 11:59–67
5. Klare JP, Engelhard M (2004) The archaeal sensory rhodopsin II/transducer complex: a model for transmembrane signal transfer. *FEBS Lett* 564:219–224
6. Filipek S, Stenkamp RE, Teller DC et al (2003) G protein-coupled receptor rhodopsin: A Prospectus. *Annu Rev Physiol* 65:851–879

Photorelease

Definition

Photorelease describes a technique of the release of a molecule from a precursor compound using light illumination.

► [Uncaging and Photoconversion/Activation](#)

Photoremovable Protecting Group

Definition

Photoremovable protecting group is a photolabile chemical group that temporarily blocks an otherwise reactive functional group in a molecule. Photoirradiation removes the protecting group and restores the reactive functionality.

► [Uncaging and Photoconversion/Activation](#)

Phototaxis

Definition

Phototaxis is the ability of bacteria to respond to light. The signal transduction is analogous to that of the chemotaxis system.

► [Photoreceptors](#)

Phototrigger (of Biomolecules)

Definition

Phototrigger is a photo-labile precursor that may release biologically active molecules under irradiation of light.

► [Uncaging and Photoconversion/Activation](#)

PHP

► [Pseudohypoparathyroidism](#)

Phylogeny

Definition

Phylogeny means the evolutionary history including the entity of all relationships between species. The natural way to present a phylogeny graphically is a tree.

► [Nuclear Hormone Receptors: Regulators of Gene Transcription and Cellular Signaling](#)
 ► [Sequence Annotation in Evolution](#)

Physical Maps

Definition

Physical Maps refers to genome maps defined by the distance in base pairs.

► [Chromosome 21 Disorders](#)

Phytanic Acid

Definition

Phytanic acid is a branched-chain fatty acid (3,7,11,15-tetramethylhexadecanoic acid) present in dietary components, notably meat, and elevated in certain peroxisomal disorders including ► [Refsum disease](#).

► [Peroxisomal Disorders](#)

Phytanoyl-CoA Hydroxylase

Definition

Phytanoyl-CoA hydroxylase is a peroxisomal enzyme catalysing the first step in fatty acid α -oxidation.

► [Peroxisomal Disorders](#)

progression of cell cycle, diminished apoptosis and enhanced cell survival, but also in cancer.

► [Breast Cancer](#)

► [Gut Epithelium](#)

PI3K

Definition

Phosphatidylinositol 3-kinase (PI3K) is a heterodimeric enzyme, comprising of a 110 kDa catalytic subunit (p110) and an 85 kDa regulatory subunit (p85). In addition to its p110 binding site, the p85 subunit contains SH2 and SH3 domains, which mediate its interactions with other cellular proteins. Phosphoinositide 3-kinase (PI3K) family enzymes phosphorylate the phosphoinositide class of lipids at the 3-OH position of the inositol ring. Upon activation, PI3Ks phosphorylate phosphatidylinositol (4,5)-bisphosphate (PtdIns(4,5)P₂), a constitutive membrane component, to create the lipid second messenger phosphatidylinositol (3,4,5)-trisphosphate (PtdIns(3,4,5)P₃) (and its immediate metabolic product PtdIns(3,4)P₂). PtdIns(3,4,5)P₃ induces the membrane translocation and/or conformational changes of target proteins, such as a variety of enzyme classes, e.g. protein kinases, phospholipases and guanine nucleotide exchange factors.

► [Cytokines](#)

► [Kinases](#)

► [Signal Transduction: Integrin-Mediated Pathways](#)

PIC

► [Preinitiation Complex](#)

Piezo-Electricity

Definition

Piezo-electricity describes the electric polarization resulting from the application of mechanical stress, especially in certain crystals. Crystals with these properties are used as piezo-elements in electronic devices, to translate electricity into mechanical movement.

► [Protein Microarrays as Tool for Protein Profiling and Identification of Novel Protein-Protein Interactions](#)

Pigment Spot Polyposis

► [Peutz-Jeghers Syndrome](#)

PI3K Pathway

Definition

PI3K is a ubiquitous lipid kinase involved in receptor signaling transduction by tyrosine kinase receptors, which may be initiated through activation of the plasma membrane receptor ► [HER2](#). This pathway involves activation of phosphatidylinositol 3-kinase (► [PI-3K](#)) and protein kinase B (PKB) = ► [Akt](#), resulting in regulation of various proteins, and ultimately in

Pilomatricomas

Definition

Pilomatricomas are benign, hair matrix derived tumors which are characterized by an exterior zone of densely packed, small basophilic cells, a transitional zone of cells displaying a gradual loss of nuclei and an inner zone of 'shadow' cells consisting of enucleated, eosinophilic cellular ghosts.

► [Skin and Hair](#)

PIP2

Definition

PIP2 stands for phosphatidyl inositol 4,5-bisphosphate. It is a membrane phospholipid that is cleaved by the enzyme phospholipase C to inositol 1,4,5 trisphosphate and diacylglycerol. PIP2, or its break-down products, directly or indirectly regulate the activity of proteins.

- ▶ Adhesion Molecules
- ▶ Focal Complexes/Focal Contacts
- ▶ Microvilli
- ▶ Signal Transduction: Integrin-Mediated Pathways

Pituitary Diseases

- ▶ Hypothalamic and Pituitary Diseases, Genetics

PJS

- ▶▶ Peutz-Jeghers Syndrome

PKC

Definition

PKC stands for Protein kinase C, which comprises of a family of serine/threonine kinases.

- ▶ Receptor Serine/Threonine Kinase
- ▶ Signal Transduction: Integrin-Mediated Pathways

PKD Domain

Definition

PKD domains were first identified in polycystin-1 protein and contain immunoglobulin-like folds. They are probably involved in protein-protein interactions.

- ▶ Polycystic Kidney Disease, Autosomal Dominant

PKD1 and PKD2

Definition

PKD1 and PKD2 refer to two genes which code for polycystin-1 and polycystin-2. Mutations in these genes have been identified as the cause of ▶ Autosomal Dominant Polycystic Kidney Disease.

- ▶ Polycystic Kidney Disease, Autosomal Dominant

PKR (Double-Stranded RNA-Dependent Protein Kinase)

Definition

PKR is an interferon- and tumor necrosis factor-inducible stress signal kinase, which depends strictly on double stranded (dsRNA), a hallmark of viral infection, for its activation. PKR is the major mediator of antiviral and inflammatory responses. PKR activation inhibits protein synthesis by phosphorylating the α -subunit of eukaryotic translation initiation factor 2 (eIF-2 α).

- ▶ RNA Interference in Mammalian Cells

PKU

- ▶ Phenylketonuria

PLA2

Definition

Phospholipase A2 is a member of the class of heat-stable, calcium-dependent enzymes catalyzing the hydrolysis of the 2-acyl bond of 3-n-phosphoglycerides to release free fatty acids. The enzyme has a molecular weight of 30,000 daltons.

Plakins

Definition

Plakins comprise of a family of large multifunctional proteins, also called cytolinker proteins, which have the capacity to bind multiple components of the cytoskeleton. The family includes plectin, BP230, envoplakin, periplakin, desmoplakin amongst others.

- ▶ Cytoskeleton
- ▶ Hemidesmosomes
- ▶ Intermediate Filaments

Plant Functional Genomics

- ▶ Plant Genomics

Plant Genomics

THOMAS ALTMANN
Institute of Biochemistry and Biology, University
Potsdam, Potsdam-Golm, Germany
taltmann@rz.uni-potsdam.de

Synonyms

Genome analysis in plants; plant functional genomics

Definition

Plant genomics refers to the analysis of plant genomes, which includes physical mapping of the genome, genome sequencing and annotation (prediction and identification of genes), the elucidation of the (biochemical) functions of the identified genes and gene products and their roles in determining the characteristics of the organisms.

Characteristics

The major/ultimate goal of plant genomics is to provide detailed and comprehensive knowledge of the molecular processes underlying plant life. As for any other organism, the central aims of plant genome analysis are the identification of all the genes of a plant, the determination of the biochemical function of the corresponding gene products and the elucidation of the roles these genes play in plant metabolism, growth and development and in interaction with the environment. Thus, the major principles of plant genomics are

nearly identical to those of human, animal or microbial genomics, but several features distinguish the characteristics of plant genome analysis. These include the broad range of different important plant species to be studied, the enormous size and complexity of many plant genomes, the opportunities to establish (large) experimental populations including immortalized segregating populations and the ease of genetic transformation (at least for a number of model plant species).

Justification of Plant Genomics

Plants are the dominant form of life on our planet, representing more than 99% of the total terrestrial biomass. They grow in all regions of the earth, occur with an enormous diversity of properties including their (biochemical) composition and constitute the basis and essential component of any ecosystem. Plants represent the essential prerequisite for all human and animal life and welfare. As naturally growing renewable resources with an adjusted ecological balance, plants are not only absolutely indispensable for the production of bulk nutrients for human food and animal feed but they also provide us with essential vitamins and medicine, with fibers for clothing and many raw materials for industrial use and for building and construction. As a result of these various different uses of plants and plant products, a wide range of plant species are in the focus of plant genome analysis. These range from model organisms used to rapidly advance our understanding of the principles of plant life to major crops, trees and medicinal plants. Thus, plant genome analysis reaches far beyond understanding the molecular mechanisms which permit plants to function and interact with the environment. The results of plant genomics are expected to support very strongly plant improvement for human and animal needs and plant use in a better, sustainable and more efficient way. These considerations and expectations triggered initiation and conductance of significant plant genome analysis programs worldwide. In addition to the importance of plant genome analysis for human society, plants exhibit a number of properties that make them ideally suited to serve as model systems and to advance our basic understanding of life. Regardless of the particular aim of the work, either basic scientific investigations or studies on applied aspects, the advances in plant genomics make plant research one of the most exciting research branches of all.

Front-Runners in Plant Genomics

A special feature of plant genomics is the variety of the species examined, ranging from model species serving as front-runners to crops with highly complex genomes. Due to its favorable properties, the most widely used plant model species is *Arabidopsis thaliana*, a member of the *Brassicaceae* family related to oilseed

rape and cabbage (8). Its small size, numerous seeds produced by a single individual *via* self-pollination and the ease of chemical mutagenesis made it an ideal object for genetic studies in plants. The very small size of *ca.* 135 million (Mbp) base pairs distributed among the five chromosomes and the low complexity of its genome made it the first choice for plant genome analysis. The achievements of the international *Arabidopsis* Genome Initiative (AGI) include the first complete BAC-based physical map established for any organism and the first complete plant genome sequence at the end of the year 2000 (9). As the second multi-cellular organism (after the nematode *Caenorhabditis elegans*) with its genome fully sequenced, *Arabidopsis* represents one of the central biological model systems for basic research. With the genome sequenced and a comprehensive inventory of the genes of this model species at hand, the *Arabidopsis* (genome) research community has now taken the challenge to decipher the function of all the genes. This includes for every gene the determination of the site, condition and level of activity (expression), the (sub-cellular) localization and the biochemical function of the gene product, its position in regulatory or metabolic networks and its interaction with other factors and its role in plant growth and development, metabolism and interaction with the environment (functional genome analysis) (1). The worldwide *Arabidopsis* functional genome analysis projects are coordinated by the multinational *Arabidopsis* steering committee (MASC; ►http://www.Arabidopsis.org/info/2010_projects/MASC_Info.jsp), which has been founded according to the excellent experience of the predecessor committee that coordinated the international *Arabidopsis* genome-sequencing project. This work is strongly supported *via* a wide range of important resources and novel techniques, which include very comprehensive sequence-indexed gene k.o. populations (see: ►<http://signal.salk.edu/cgi-bin/tdnaexpress>, ►<http://www.mpiz-koeln.mpg.de/GABI-Kat>, ►<http://flagdb-genoplante-info.infobiogen.fr/projects/fst>, ►<http://www.Arabidopsis.org/abrc>), full-length cDNA collections (see: ►<http://pfgweb.gsc.riken.go.jp/projects/raflcdna.html>, ►http://www.Arabidopsis.org/abrc/ssp_clone.jsp), gene-chips for genome-wide expression analyses (see: ►<http://www.affymetrix.com/products/arrays/specific/arab.affx>, ►<http://www.catma.org>) and other novel techniques such as proteome analysis and metabolic profiling, the latter having been originally developed in *Arabidopsis* and now adopted for many other organisms (3). Supported by a comprehensive database, which serves as the major information resource (TAIR, ►<http://www.Arabidopsis.org>), *Arabidopsis* research continues to accumulate the most extensive information on plant gene/genome function.

As mentioned earlier, *Arabidopsis* is “only” the front-runner of plant genomics and several crop species’ genomes are under very extensive analysis. The example of *Arabidopsis thaliana* genome sequencing has shown how well the tremendous workload can be distributed and the high financial expenses can be shared among various national genome programs. This led to the formation of new international sequencing consortia, concentrating on sequencing of the rice genome as a model for grasses (►<http://rgp.dna.affrc.go.jp/IRGSP/index.html>) and of the poplar genome, as a model for trees (►<http://www.ornl.gov/sci/ipgc>). Like the situation in the dicots, where *Arabidopsis* was selected for sequencing, rice was chosen for sequencing according to its relatively compact genome as the first monocot plant species and as a representative of the grasses, which include as other important crops wheat, maize and sorghum. While the rice genome has a size of *ca.* 430 Mbp distributed among 12 chromosomes, the evolutionarily related genomes of the other cereal crops are much larger: sorghum (750 Mbp), maize (2,400 Mbp), barley (5,000 Mbp) and wheat (16,000 Mbp). In contrast to *Arabidopsis*, however, rice does not only serve as a model for monocots / grasses but is a very important crop itself. It feeds half of the world’s population with a production area of more than 150 million hectares and more than 580 million metric tonnes in yield (FAO statistics). Rice is the predominant staple for many countries in Asia and the Pacific, in Latin America and the Caribbean and in Africa. In developing countries, rice accounts for 715 kcal/capita/day, 27% of dietary energy supply, 20% of dietary protein and 3% of dietary fat (6). In spring 2002, a draft version of the almost complete genome sequence of “Japonica-variety” rice was assembled by Researchers of Syngenta (4). At the same time a draft sequence of an “Indica-variety” was generated by researchers from China (10). The completely finished high quality genomic sequence of the rice genome, established by the International Rice Genome Initiative was available at the end of 2004. This will boost rice research as much as the availability of the genome sequence in *Arabidopsis* did and similar resources for functional analysis (including sequence indexed k.o. populations are currently being established by the International Functional Genomics Consortium (►<http://www.iris.irri.org/IRFGC>). The third plant genome sequence soon to be finished will be that of the poplar tree. The *Populus* genus includes poplars, cottonwoods, and aspens – fast-growing trees widely used in forestry research. These fast growing trees produce seeds prolifically, and can be genetically modified in the laboratory. The tree to be sequenced is a female *Populus balsamifera* (black cottonwood or balsam poplar). With a size of 550 Mbp organized in 19 chromosomes, the genome of this model tree is only

4 times as big as that of *Arabidopsis thaliana*. The complete sequencing and annotation data are available on the consortium web site (►<http://genome.jgi-psf.org/poplar0/poplar0.info.html>).

The moss *Physcomitrella patens* is being developed as a model system for plant biological studies and will support plant functional genomics. Bryophytes are the oldest living branch of land plant evolution. They are separated by approx. 450 million years of evolution from seed bearing plants. To fully understand and employ land plant evolution and plant diversity, mosses are to be added to the current list of model plants. The dominance of the haploid gametophyte in the life cycle of this moss facilitates genetic analysis. To date, this moss is the only known terrestrial plant with an efficient system for homologous recombination in its nuclear DNA, making gene-targeting strategies as easy as in yeast. In contrast to seed bearing plants, the dominating generation in the moss life cycle is the haploid gametophyte. Therefore *Physcomitrella* is an ideal model for gene-function analyses. Loss-of-function mutants created by a targeted gene knockout approach (homologous recombination) can be analyzed immediately without complex back-crosses (5). The *Physcomitrella* genome is about 3.5× as big (~480 Mbp) as the *Arabidopsis* genome and is distributed among 27 chromosomes. Today more than 100,000 EST entries exist in databases and represent approximately 25,000 genes. Preliminary analysis of genes and ESTs strongly indicate that these sequences are highly similar to those of the corresponding genes of seed bearing plants.

With the availability of the genome sequences, the multi-parallel analysis techniques (“x-omics” such as transcriptomics, proteomics and metabolomics) can be applied as efficiently in these plant species as in the model plant *Arabidopsis* (2). The functional genome analysis thus performed will further enhance the massively increasing amount of available data to be stored, analyzed and interpreted. As a consequence, bioinformatics developed in parallel to the analytical techniques will play the same essential role in plant genomics as in human, animal or microbial genomics. As does the research in these other kingdoms, plant biology becomes more and more an information and computable science research and has moved into plant systems biology.

Beside this shift of paradigms and the broad increase in basic knowledge, plant genomics influences the way in which our staple crops are produced and will be produced in future.

Genomics of Other Plant Species

As mentioned above, humans depend on a wide range of crop plants that provide bulk nutrients such as carbohydrates (cereals), starch (potato, cassava or

yams) and proteins (legumes). Plants provide fat, oil, wax and fiber (cotton, soy or rapeseed). Furthermore, we cultivate plants as fruits or vegetables and as fodder plants for our livestock. But plants also contain natural stimulants (cacao, coffee) and provide wood, cork, rubber, dyes and tanning agents (indigo, henna, eucalyptus). They supply insecticides (neem tree, Dalmatian pyrethrum or tobacco). Many plants have a positive influence on our health; many substances from medicinal plants are either directly used or served as lead compounds for the development of new drugs. Ten of the world’s 25 top-selling drugs in the late nineties derived from natural sources. The number of plant species with very large genomes that have to be subjected to genomes analysis is therefore far greater than that of animals (including models such as mouse, rat and primates and our farm animals). This challenge is being met in several ways.

General and basic gene functions will mainly be clarified in the principal plant model systems such as *Arabidopsis* and rice. Many of the corresponding genes in crops and other important plant species will carry out the same or very similar functions and the knowledge gained through the functional genomics/systems biology approaches in the principal models will be highly useful to direct targeted analyses in the other plants species. This “bridging concept” requires a good sequence information basis for the crops that is being provided by comprehensive expressed sequence tag (EST) libraries. They represent snapshots of the genes expressed in a given tissue and/or at a given developmental stage and provide a good representation of the expressed regions of the genome if a sufficient number of different libraries are analyzed from a range of tissues, developmental stages and after various environmental challenges. As only genic sequences are analyzed independent of the genome size, this approach is highly cost efficient, especially for species with very large genomes. Up to the present, more than 20 million ESTs (from all kind of species) are in public databases. Platforms for comparative plant genomics are being developed at different places in the world including the “reconstructomics” platform using the SPUTNIK program (►<http://mips.gsf.de/proj/sputnik/>). SPUTNIK is a tool to analyze “genomeless genomes”, i.e. to estimate the gene content represented by the ESTs and to establish unigene sets for plant species for which no genome sequence information is available. A further highly important tool for the verification of the proposed gene function in crops is TILLING (targeted identification of local lesions in genomes) (7). This refers to a procedure for the targeted identification of mutant alleles that was first developed in *Arabidopsis* and is now applied to many other species including animals and other plants. Thus, genes selected for detailed analysis on the basis of the

information provided by the model system will be tested for their functions in the crop plant through the identification of corresponding mutants and their characterization.

However, the gene functions specific to certain plant families will remain largely undiscovered and the basis of model species will be broadened stepwise to representatives from other families. These will include tomato for the *Solanaceae* and *Lotus japonicus* or *Medicago truncatula* for legumes. The latter play an important role in providing protein and are able to fertilize themselves by fixation of atmospheric nitrogen through symbiosis with soil bacteria. Genome sequencing programs on tomato, soybean, *Medicago*, maize etc. have been initiated or will soon be initiated. However, they are not currently directed towards the same level of completion as the *Arabidopsis* or rice genome sequences. Instead, so-called draft sequence versions will be established (low coverage) or sequencing will be concentrated on gene-rich regions. Methods developed for these purposes, such as the enrichment of weakly methylated DNA or the separation from repetitive sequences based on re-association kinetics, lead to cost savings and first insights into crop species genomes. Both methods have just proven their functionality in the sequencing of gene-rich regions in the maize genome. The sequences thus obtained for a wider range of plant species will be anchored to the genomes of the principal models, *Arabidopsis* and rice and great emphasis will be put on comparative genomics in plants in the future.

Application of Plant Genomics to Crop Improvement

One aspect of plant genomics that is radically different from genome analysis in humans and model animals and which is closely related to genome analysis in farm animals is its direct connection to practical application. Crop genome analysis provides breeders with a wealth of information and novel tools to enhance and speed up the breeding process. Molecular markers such as microsatellites (simple sequence length polymorphisms, SSLPs) or single nucleotide polymorphisms (SNPs) provide the means to identify favorable genome segments as quantitative trait loci (QTL) in experimental populations or breeding populations. Marker-assisted selection (MAS) is increasingly applied, e.g. in backcrossing programs used to introduce favorable genes/alleles from exotic germplasm.

These procedures provide the means to increase the efficiency of the (classical) breeding process by addition of the new knowledge gained through plant genome research *via* genetic engineering to achieve the necessary boost in high quality food production for the global human population. All this is happening against the background of a world population still growing and a reduction of available resources like water, arable

land and energy *per capita*. Plant genome research is a basic element to support the development of a sustainable and ecologically balanced agriculture. Moreover, it also supports the production of new and improved products. The diversity of the uses of plants mentioned above shows the potential of the uses of plant genomics to benefit humans and the environment. Through the possible use of exotic breeding material and the efficient use of such germplasm *via* MAS assisted backcrossing, it also becomes clear that plant genomics research provides the means to enhance the use of natural genetic diversity and to support its preservation. These natural resources offer great potential for further crop improvement and the information gained through plant genome analysis will strongly support their protection.

Considering this potential and the need to enhance crop improvement procedures, several plant genome programs have been initiated that have a major focus on crops. In the year 1998 the “National Plant Genome Initiative” was founded in the USA (► http://www.nsf.gov/bio/dbi/dbi_pgr.htm). In January 2003 this initiative presented the strategy plan for the second program stage (2003–2008). In contrast to the first period, which mainly focused on building up resources and establishing technologies, these tools will now be used to address a wide range of important biological questions in plant science. The second phase will involve a continued elucidation of the genome structure and organization of crops like rice, maize, tomato and *Medicago* as key plant species. Furthermore, the transfer of knowledge about gene functions from key plants to other crops will help to understand complex processes like tuber or bulb development, wood formation, fruit development and nutrient uptake and use in plants. Plant genomics will expand into the analysis of biodiversity, studies of ecology and ecosystems and the development of sustainable production systems, renewable resources and novel biomaterials. It will greatly benefit from advances in bioinformatics, the integration and interaction of plant genome databases, the development of standard operating principles and the development of new algorithms to analyze plant genomics data. Furthermore, education and training of students and young researchers will be important aspects of plant genomics, as will the outreach to broaden the recognition in our society of the importance of plant genomics. National plant genomics research programs also exist in Canada, Brazil, China, Japan and Australia, with major foci on rice, maize, wheat, soybean, rapeseed and sugarcane. In Europe, several larger national plant genomics programs with major emphasis on model and crop plants, including *Arabidopsis*, rice, wheat, barley, maize, rapeseed, grapes, tomato, potato, *Medicago*, sugar beet etc. have been started in the UK (GARNET,

2000, ►<http://www.york.ac.uk/res/garnet/garnet.htm>), France (Génoplante, 1999, ►<http://www.genoplante.org>), Germany (GABI, 1999, www.gabi.de), the Netherlands (BioSystems Genomics, 2002, ►<http://www.biosystemsgenomics.nl>), Sweden (SCTFG, 1999, ►<http://www.upsc.nu>) and Spain, 2003, (<http://www.mcyt.es>). In addition, efforts are being made to concentrate national activities in Denmark, Finland, Hungary and Italy and others will follow. To enhance the output of these various initiatives, further increase in co-operation across programs and setting cross-links between programs will be necessary to make developed resources, technologies and information freely available and thus enhance the efficiency of plant genome research and increase the output achieved with the limited available financial resources. The European Research Area Net Plant Genomics (►<http://www.erapg.org>) is a European wide activity directed towards these goals.

A Vision of Plant Genomics

Plant genomics represents the bridge between classical molecular biology and the holistic view into complex systems. The term “plant system biology” thus emerged during the last few years as it did for several other groups of species. As in the general concepts of genomics, many commonalities exist at this level also among the studies of various species. The use of systematic genomics, transcriptomics, proteomics and metabolomic technologies and the increasing use of bioinformatics to construct models of complex biological systems are shared between the different kingdoms. However, plant systems biology may broaden much faster from the analysis of the properties and mechanisms of the organisms themselves to the incorporation of the interaction of the organisms with their environment (both biotic and abiotic influences). Plant performance, that of our field crops in particular, is very strongly affected and limited by environmental conditions. Tolerance or resistance to adverse conditions thus offers the greatest potential to increase and secure yields in sustainable production systems. Thus, while plant systems biology as a logical expansion from plant functional genomics will probably develop similarly to that of other systems (from analysis of simple cellular processes to studies at the organ and whole plant level), the long term vision may be an extension into the study of plants as interacting partners within ecological systems.

References

1. Chory J, Ecker JR, Briggs S et al (2000) National Science Foundation-Sponsored Workshop report: “The 2010 Project” – Functional genomics and the virtual plant. A blueprint for understanding how plants are built and how to improve them. *Plant Physiol* 123:423–425
2. Fiehn O, Kloska S, Altmann T (2001) Integrated studies on plant biology using multiparallel techniques. *Curr Opin Biotech* 12:82–86
3. Fiehn O (2002) Metabolomics – the link between genotypes and phenotypes. *Plant Mol Biol* 48:155–171
4. Goff SA, Ricke D, Lan TH et al (2002) A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. japonica). *Science* 296:92–100
5. Hohe A, Egener T, Lucht JM et al (2004) An improved and highly standardised transformation procedure allows efficient production of single and multiple targeted gene-knockouts in a moss, *Physcomitrella patens*. *Curr Genet* 44:339–347
6. Kennedy G, Burlingame B, Nguyen VN (2002) Nutritional contribution of rice: impact of biotechnology and biodiversity in rice-consuming countries. The International Rice Commission (Twentieth Session); Bangkok, Thailand
7. McCallum CM, Comai L, Greene EA et al (2000) Targeted screening for induced mutations. *Nat Biotech* 18:455–457
8. Meinke DW, Cherry JM, Dean C et al (1998) *Arabidopsis thaliana*: a model plant for genome analysis. *Science* 282:662–682
9. The Arabidopsis Genome Initiative (2000) *Nature* 408:796–815
10. Yu J, Hu S, Wang J et al (2002) A Draft Sequence of the Rice Genome (*Oryza sativa* L. ssp. indica). *Science* 296:79–92

Plasma Membrane

Definition

Plasma membrane designates the partially permeable membrane forming the outer limit of a cell, separating the chemically distinct intracellular milieu (the cytoplasm) from the extracellular environment. The plasma membrane regulates the flow of materials in and out of the cell.

► [Biological Membranes](#)

► [Microvilli](#)

► [Vesicular Traffic](#)

Plasmid/Plasmid Vector

Definition

Plasmid is a small circular DNA molecule that replicates independently of the genome, and is used extensively as a vector for DNA cloning. Plasmid vectors for cloning generally contain a replication origin, an antibiotic-resistance gene, and other varying sites. Expression vectors, for instance, also contain a ribosomal binding site, a promoter and often a

transcription terminator. In the context of clinical gene transfer, an expression construct is usually inserted in the plasmid, and the bacterium *E. coli* is used for plasmid propagation.

- ▶ Clinical Gene Transfer
- ▶ Recombinant Protein Expression in Bacteria
- ▶ Two-Hybrid System

Pleiotropic/Pleiotropic Gene Effect

Definition

A gene is referred to as pleiotropic when it leads to many different phenotypic expressions. Affected individuals can have one or more of a range of signs and symptoms. For example, the Marfan gene is pleiotropic with widespread effects and can cause long fingers and toes, dislocation of the ▶ lens of the eye, and dissecting aneurysm of the ▶ aorta.

- ▶ Manic Depression
- ▶ Marfan Syndrome

Plexiform Neurofibromas

Definition

Plexiform neurofibromas refers to benign tumors originating from cells of the Schwann cells or endoneural mesenchym of peripheral nerve fibers. Plexiform characterises local infiltrating growth causing pain and neurological failures. These are benign tumours associated with major nerve trunks and are usually found in young children.

- ▶ Neurofibromatosis Type 1 (NF1), Genetics

Pluri-/Multipotent

Definition

The terms define the property of a stem cell that enables it to differentiate into all (pluripotent) or several (multipotent) cell types of the ca. 200 different tissues in the body. Pluripotentiality (pluripotency) is a property of embryonic stem cells. However, pluri/multipotent cells are not able to form a complete organism in a morphological manner, which is only possible for totipotent stem cells.

▶ ES Cell Differentiation as a Model System for Functional Genomics

- ▶ Neural Stem Cells
- ▶ Stem Cells: an Overview

PMF

- ▶ Peptide Mass Fingerprinting

PML Nuclear Bodies

Definition

Promyelocytic leukemia protein (PML) is a tumor suppressor protein, whose function is disrupted by the translocation t(15;17) in acute promyelocytic leukaemia (APL). PML has a wide range of functions *in vivo*, and is involved in suppression of cell growth, multiple pathways of apoptosis, transcriptional regulation, and genome stability. PML proteins contain zinc finger motifs, and are typically concentrated within discrete speckled multiprotein subnuclear domains, designated as PML nuclear bodies. PML nuclear bodies are also targeted by many DNA viruses in early infection, and are the site of viral transcription initiation. Subsequently, the viral early proteins localize to, and eventually disrupt, PML nuclear bodies.

- ▶ Bloom Syndrome

PMS

Definition

Postmeiotic segregation refers to the separation of alleles that occurs at the first ▶ mitosis after meiosis (postmeiotic segregation) leading to new gene combinations (▶ recombination).

- ▶ Hereditary Nonpolyposis Colorectal Cancer

PNA

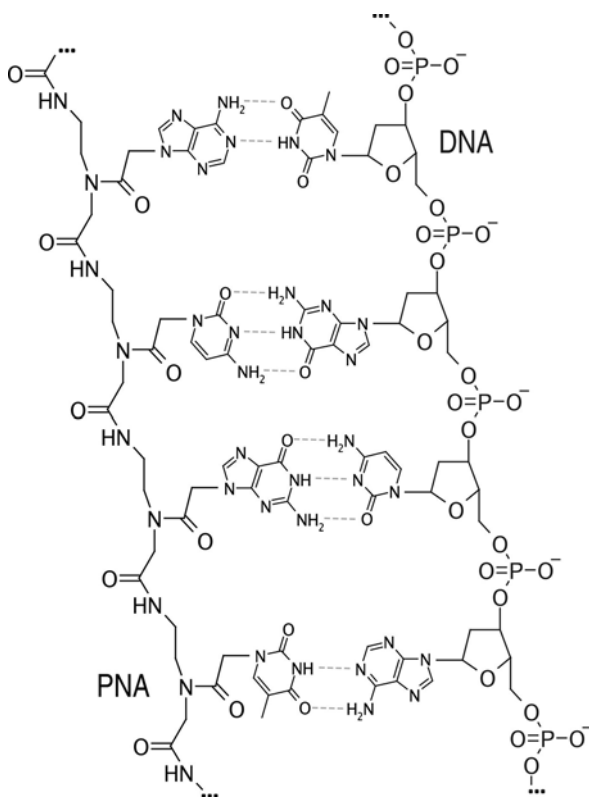
- ▶ PNA Chips

PNA Chips

ANETTE JACOB, OLE BRANDT, JÖRG D. HOHEISEL
Functional Genome Analysis, Deutsches
Krebsforschungszentrum, Heidelberg, Germany
j.hoheisel@dkfz.de

Definition

Peptide nucleic acid (PNA) is an artificial mimic of nucleic acid, which was invented more than a decade ago (7). The molecule combines many features that are similar to those of natural nucleic acids with characteristics that are rather different or even unique to PNA. First and foremost, N-(2-aminoethyl)-glycine units linked by ►amide bonds substitute the (deoxy)ribose-phosphate backbone and the nucleobases are attached via a methylene carbonyl link to the glycine moiety. Nevertheless, the distances and angles within PNA oligomers much resemble the conformation of nucleic acids (Fig. 1). One format of application is the microarrays that are made for a highly parallel analysis of nucleic acids. To this end, very many different PNA molecules are attached onto a small planar surface,



PNA Chips. Figure 1 Structures of PNA and DNA.

e.g. a microscope slide. Upon addition of a DNA or RNA sample, hybridisation occurs on the chip at PNA-oligomers that are of complementary sequence to molecules in the sample. Recorded by detecting a label that is attached to the DNA or RNA, various forms of information can be deduced from such studies, such as for example, sequence variations or changes in the amount of an individual RNA molecule within a complex mixture.

Description

PNA binds to complementary single-stranded DNA or RNA in a sequence-specific manner. Because of the PNA's uncharged nature, the duplexes are substantially more stable than the equivalent double-strand molecules formed by DNA and RNA alone. The neutral amide backbone also allows PNA to hybridise to nucleic acids in the absence of salt, since no positive ions are required for counteracting the interstrand repulsion that occurs between two negatively charged nucleic acid molecules. As a result, the target DNA has fewer secondary structures and is therefore more accessible to a PNA probe than it would be to a DNA oligonucleotide. Being an artificial molecule, PNA is not degraded by nucleases or proteases. This makes it a versatile tool in assays that are subjected to environmental constraints.

Applications of PNA in biotechnology divide into reactions that occur in homogenous solution and analyses that take place on a solid support. Important examples of the former assay format are the use of PNA as an antisense reagent, as a DNA-opener and as a tool for PCR-clamping or the detection of single nucleotide polymorphisms. A comprehensive and up-to-date overview on the various applications of PNA is provided by Nielsen et al. (12). For all its advantages, however, PNA has not had quite the impact that was initially predicted. The limited solubility of PNA molecules is especially an obstacle that hampers many applications in solution. For the analysis of nucleic acids on microarrays, however, this limitation is not so restricting, since the PNA concentration is low and the molecules are constrained by their physical attachment to the solid support. ►DNA-microarrays, also named DNA chips, are an assay format that is widely used in biotechnology and biomedical research, since they permit molecular analyses of complex samples, such as genome-wide ►transcriptional profiling, typing of ►single nucleotide polymorphisms or epigenetic studies. Moreover, microarrays will become even more important in future, once the various types of analysis are established to a quality standard that will, for example, allow their use in routine medical diagnostics or food-processing control. Substituting PNA for DNA oligomers as the chip-bound capture molecules could

contribute substantially to achieving higher sensitivity and improved assay robustness. Also, taking advantage of the chemical differences between the PNA probe and the nucleic acid analyte, their interaction can be detected directly, with neither molecule carrying a label. In consequence, the processing is much simplified, facilitating automation. Simultaneously, the degree of bias caused by introduction of a label is avoided and the effort in sample preparation is reduced significantly.

Synthesis

As with DNA microarrays, the probes for PNA chips can either be pre-fabricated and attached to the chip surface by means of robotic devices or synthesised directly *in situ*. With respect to synthesis chemistry, PNA does not differ from ►peptides. Therefore, most procedures developed for the synthesis of peptides can be used for the production of PNA-oligomers and *vice versa*.

Parallel synthesis in 384-well filter-bottom microtiter plates has been established. This releases adequate numbers of PNA-molecules that are of good quality and in sufficient yield for the production of microarrays. In this process, the consumption of reagents per individual oligomer is reduced significantly compared to standard column-based synthesis protocols. Furthermore, procedures for the attachment to the chip surface exist, which select for the binding of full-length molecules (1). In the analysis of biological samples, this leads not only to better reproducibility but also to a higher dynamic range of detection and thus an improved accuracy.

Alternatively, *in situ* synthesis on the array by the SPOT method (8), in which the individual monomers are placed sequentially at the relevant positions of the chip in small volumes of liquid, has been described. Unfortunately, following this protocol, throughput is limited with current liquid handling systems. Also, only a very limited number of array copies can be produced simultaneously. However, drawing on expertise from ink jet technology, for example, handling of small volumes is improving and should soon allow the quick synthesis of many different molecules in many copies. In addition, the control of synthesis by ►photolithography was described for peptide arrays even before the technique was used for the production of DNA chips (3). In combination with PNA monomers that can be deprotected by light, procedures could be established that are equal in throughput to current production processes for DNA microarrays. All *in situ* synthesis formats, however, have the disadvantage that not only full-length molecules but also shorter derivatives that result from the incomplete yield during each synthesis step will be on the arrays.

Basic Hybridisation Characteristics

Overall, hybridisation to PNA chips is performed similarly to hybridisation to DNA microarrays. The main difference is the stability of the duplex formed. Comparable to results on PNA-DNA interaction in solution, the stability on a solid support of a duplex formed by a DNA molecule and a surface-attached PNA oligomer of mixed base composition is in general 1–1.5°C higher per base pair than that of the corresponding DNA-DNA hybrid. Therefore, both higher hybridisation and washing temperatures are used on PNA chips or the length of the PNA oligomers is reduced. The dissociation temperature of PNAs can be predicted from their sequence with reasonable accuracy and the sequence specificity of binding is high. In addition, PNAs that contain a ‘chiral box’ at their centre exhibited even higher mismatch recognition than normal molecules. ►Chirality can be introduced into PNA-monomers by substituting D-lysine for the usual N-aminoethylglycine backbone (6). ►Hybridisation to complementary molecules in a parallel rather than anti-parallel fashion, potentially possible because of the PNA’s achiral nature, is strongly disfavoured on PNA chips.

Label-Free Diagnostics

Especially for routine applications in a clinical setting, it would be advantageous to avoid labelling of the probes and the target molecules altogether. Besides speeding up the analysis process and reducing cost, all labelling steps introduce an additional degree of variation. In addition, such procedures should assist in the integration of the assays into automated analysis systems, which are eventually required for the analysis of patient samples. From its hybrid nature – being chemically rather different from nucleic acids while nevertheless behaving very similarly – PNA offers some unique opportunities to achieve such ends.

►Surface Plasmon Resonance

One means for the direct detection of DNA-binding to microarrays is surface plasmon resonance, although it is not unique to PNA arrays. The microarray surface is illuminated with polarised light. Reflection of the light occurs at a conductive film at the interface between the chip and the fluid containing the sample. Upon hybridisation of DNA molecules to the chip-bound PNA oligomers, the refractive index changes, resulting in a resonance signal that can be detected. The refractive index is directly proportional to the mass change at the particular position. This allows a real-time monitoring of the hybridisation process. The variation in signal intensities for a single C to T conversion was found to be as high as 300-fold (2). To date, however, the sensitivity of the procedure is still

inadequate for the analysis of complex mixtures of analyte molecules.

Electrochemical and Direct Electronic Detection

Beside the obvious benefit of avoiding extra labelling steps, a direct electrical detection of binding events would have the additional advantage that the connection to the electronic data analysis could be simplified even further. In addition, the read-out could be much faster than with other systems and real-time measurement during the hybridisation process would be possible. Conductance measurement (4) or **▶ impedance-based detection** (5) have been used for signal recording. As above, there are still limits in the sensitivity of the assay. In order to circumvent this dilemma, various electrochemical detection methods based on electrochemically active substances such as methylene blue or Hoechst 33258 are being investigated. Micro-fabricated PNA arrays were equipped with electrodes made of gold. Different PNA molecules that had a cysteine at their N-termini were bound to the electrodes *via* the thiol groups. The advantage of using PNAs is the smaller adhesion of the positively charged electrochemical compound to the neutral backbone, thereby reducing the background signal. Subsequent to the hybridisation of a DNA-fragment and appropriate washing steps for the removal of unspecific binding, the array is incubated in a solution that contains the electrochemical substance. Hoechst 33258, for example, binds to the minor groove of PNA-DNA duplexes. This produces an anodic peak current that is measured by linear sweep voltammetry. In all systems based on electrical measurements, the complexity of the array is still limited at present, thereby turning the advantage of continuous and parallel measurement at all array features into a handicap as far as arrays with very many spots are concerned. However, the number of array features required for many routine assays in the microarray format will anyway be relatively small.

Secondary Ion Mass Spectrometry

A third process seems sensitive enough to circumvent the above-mentioned limitation on sample complexity. In addition, the hardware required for the detection process could be adapted to deal with large feature numbers on the chip, although detection is performed sequentially rather than in parallel as with electrical measurement. A combination of PNA microarrays and detection of bound nucleic acids by time-of-flight secondary ion **▶ mass spectrometry** (TOF-SIMS) could permit a direct analysis of genomic DNA (1). SIMS is very sensitive in detecting fragments (PO_2^- and PO_3^-) of the phosphate ions, which are an integral part of nucleic acids but missing entirely in PNA. In addition, a DNA fragment contains very many phosphates. Since both the detectable marker and signal amplification are

therefore intrinsic to the target molecule, analyses could be performed without any prior amplification by means such as PCR and without labelling. In this respect, it is also of advantage that relatively long fragments can be hybridised to PNA-oligomers under low salt or no salt conditions and that the duplex stability is high.

Subsequent to a hybridisation of nucleic acids to a PNA microarray and the removal of non-specifically bound molecules by appropriate washing steps, the binding of nucleic acids is identified by mass spectrometry. A primary ion beam is directed at the microarray surface, causing the fragmentation and release of molecule fragments present at this position. However, just a monolayer of molecules is actually affected. Concurrently with fragmentation, the released molecules are ionised by the primary ion beam. Negatively charged ions are accelerated into the flight tube of the mass spectrometer, in which they are separated according to their mass-to-charge ratios. Phosphates and therefore the respective signals should only be present when nucleic acid has hybridised to the complementary PNA molecule at a given position.

From the difference in backbone structure of PNA and nucleic acids, other fragments, such as parts of the sugar molecules, could also be used as an indicator for the binding of nucleic acids. Since the sugar molecules of RNA and DNA are different at the 2'-position, the procedure could even distinguish between the two molecular classes. This process could even allow discrimination of living cells or organisms, producing a signal from both DNA and RNA, from dead ones, which should exhibit a signal that only represents the relatively long-lived DNA.

Acknowledgements

The work of the authors was funded by the German Federal Ministry of Education and Research (BMBF) as part of the "Leitprojekt Medizin", Nanotechnology and DHGP programmes.

▶ Chip Technologies, Basic Principles

References

1. Brandt O, Feldner J, Stephan A et al (2003) PNA microarrays for hybridisation of unlabelled DNA samples. *Nucleic Acids Res* 31:e119
2. Burgener M, Sanger M, Candrian U (2000) Synthesis of a stable and specific surface plasmon resonance biosensor surface employing covalently immobilized peptide nucleic acids. *Bioconjug Chem* 11:749–754
3. Fodor SPA, Read JL, Pirrung MC et al (1991) Light-directed spatially addressable parallel chemical synthesis. *Science* 251:767–773
4. Hahm J and Lieber CM (2003) Direct ultrasensitive electrical detection of DNA and DNA sequence variations using nanowire nanosensors. *Nano Letters* 4:51–54

5. Macanovic A, Marquette C, Polychronakos C et al (2004) Impedance-based detection of DNA sequences using a silicon transducer with PNA as the probe layer. *Nucleic Acids Res* 32:e20
6. Menchise V, De Simone G, Tedeschi T et al (2003) Insights into peptide nucleic acid (PNA) structural features: the crystal structure of a D-lysine-based chiral PNA-DNA duplex. *Proc Natl Acad Sci USA* 100: 12021–12026
7. Nielsen PE (ed) (2004) *Peptide Nucleic Acids: Protocols and Applications*, 2nd edn. Horizon Bioscience
8. Weiler J, Gausepohl H, Hauser N et al (1997) Hybridisation based DNA screening on peptide nucleic acid (PNA) oligomer arrays. *Nucleic Acids Res* 25:2792–2799

PNET

Definition

PNET stands for Primitive Neuroectodermal Tumours.

► [Medulloblastoma](#)

Pneumothorax

Definition

A pneumothorax is an accumulation of air in the pleural cavity.

► [Marfan Syndrome](#)

PNS

► [Peripheral Nervous System](#)

Podophyllotoxins

Definition

Podophyllotoxins are natural anticancer products, acting as chemotherapeutics by inhibition of DNA synthesis in a cell cycle dependent manner. They form a complex with topoisomerase II and DNA, thus

inducing double stranded DNA breaks, and preventing repair by topoisomerase II binding and the entry phase of cell division. Commonly used podophyllotoxins include etoposide and teniposide.

► [Multi-Drug Resistance](#)

Point Mutation

Definition

Point Mutation defines a change or deletion/insertion of a single base in a gene, which changes the nucleotide composition of the resulting RNA. In protein encoding genes, it can lead to a single amino acid replacement in the protein with or without loss or change of function.

► [Mouse Genomics](#)

► [tRNA](#)

Point Spread Function

Definition

An optical system produces a complex three-dimensional light distribution of an ideal point light source in the image space in and near the image plane. This light distribution represents the point spread function of the optical system. For an aberration-free, diffraction-limited optical system, the slice of the diffraction pattern in the image plane is designated as Airy function, which is the mathematical description of the diffraction pattern of a circular aperture. Above and below the image plane, the diffraction pattern changes periodically along the optical axis, so that bright and dark Airy-disk-like patterns appear alternately with diameters increasing proportional to the distance from the image plane.

► [Fluorescence Microscopy: Single Particle Tracking](#)

Poisson-Boltzmann Electrostatics

Definition

Poisson-Boltzmann electrostatics describes classical treatment of electrostatic interactions in solution using

a continuum model for the solvent and accounting for salt effects and a non-uniform dielectric constant in the system. When no mobile ions are present and the dielectric constant is uniform throughout space, the Poisson-Boltzmann equation reduces to Coulomb's law.

► [Molecular Dynamics Simulation in Drug Design](#)

Pol

Definition

Pol is the retrovirus gene that encodes reverse transcriptase (RT) and integrase (IN).

► [Retroviruses](#)

Pol I

► [RNA Polymerase I Transcription](#)

Polaris

Definition

Polaris is a protein component of the intraflagellar transport system, which is required for ciliogenesis. Mice with a deletional mutation of the polaris gene exhibit defects in cilia formation, randomization of left-right axis determination, and cystic renal disease.

► [Autosomal Dominant \(Inherited Disorder\)](#)

► [Polycystic Kidney Disease, Autosomal Dominant](#)

Polarizability

Definition

Polarizability defines the degree to which the electron distribution of an atom or a molecule can be deformed by an electric field.

► [Molecular Dynamics Simulation in Drug Design](#)

POLG

Definition

POLG stands for DNA polymerase γ , the polymerase that is involved in replication of mitochondrial DNA. It is encoded by the nucleus and transported into the mitochondria.

► [Mitochondrial Myopathies](#)

Poly(A) Polymerase

Definition

Poly (A) Polymerase is the enzyme adding the ► [poly \(A\) tail](#) to eukaryotic mRNA precursors during nuclear processing, using ATP as activated building blocks. Related enzymes exist in the cytoplasm and are probably involved in regulating mRNA stability, and/or translational efficiency by elongating poly (A) tails of specific mRNAs.

► [Polyadenylation](#)

Poly(A) Ribonuclease

Definition

Poly (A) ribonuclease (PARN) is an enzyme that removes the poly (A) tail from an mRNA molecule.

► [RNA Stability](#)

Poly(A) Tail

Definition

Poly (A) tail is a homopolymeric sequence of AMP residues at the 3' end of almost all eukaryotic mRNAs. It is added during maturation of eukaryotic mRNA in a posttranscriptional polyadenylation reaction at the 3' end. In mammalian cells, a newly synthesized poly (A) tail is about 250 nucleotides long. In the cytoplasm, gradual shortening of the poly (A) tail results in a heterogeneous size distribution.

► [Polyadenylation](#) and ► [RNA Stability](#)

Poly(A)-Binding Protein

Definition

Poly(Adenin)-binding protein (PABP) is a multifunctional binding protein that plays an important role in mRNA stability and protein translation. PABP is found complexed to the 3-prime poly(A) tail of eukaryotic mRNA, and is required for poly(A) shortening and translation.

- Fragile X Syndrome
- Polyadenylation
- RNA Stability
- Translational Control in Eukaryotes

Poly(ADP-Ribose) Glycohydrolase

Definition

Poly(ADP-ribose) glycohydrolase (PARG) catalyses the degradation of poly(ADP-ribose) into fragments and ultimately into free monomeric ADP-ribose.

- Poly(ADP-Ribosyl)ation, Pathophysiology

Poly(ADP-Ribose) Polymerase (PARP)–1

Definition

Poly(ADP-ribose) polymerase (PARP)–1 is the PARP isoform catalysing the bulk of poly(ADP-ribosyl)ation under conditions of DNA strand breakage. Other, less frequently used designations for the same enzyme are “NAD⁺ ADP-ribosyltransferase (polymerising)” [pADPRT]; “ADP-ribosyltransferase” [ADPRT], or “poly(ADP-ribose) synthetase” [PARS] (EC 2.4.2.30).

- Poly(ADP-Ribosyl)ation, Pathophysiology

Poly(ADP-Ribosyl)ation

Definition

Poly(ADP-ribosyl)ation is a posttranslational protein modification in eukaryotic cells carried out by

members of the family of poly(ADPribose) polymerases (PARPs), which catalyse the covalent transfer of ADP-ribosyl moieties from NAD⁺ to “acceptor” proteins (initiation reaction), as well as repeated cycles of ADP-ribosyl transfer to previously transferred ADP-ribosyl moieties (“elongation”), and also branching of the growing chains, thus creating highly complex polymers of protein-linked ADP-ribose.

- Poly(ADP-Ribosyl)ation, Pathophysiology

Poly(ADP-Ribosyl)ation, Pathophysiology

ALEXANDER BÜRKLE

Molecular Toxicology, Department of Biology,
University Konstanz, Konstanz, Germany
alexander.buerkle@uni-konstanz.de

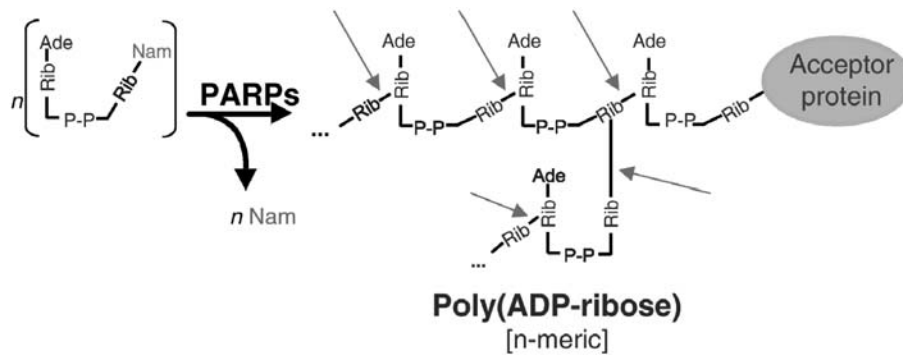
Definition

► Poly(ADP-ribosyl)ation is a posttranslational protein modification in eukaryotic cells (1-3). This modification is carried out by members of the family of poly(ADP-ribose) polymerases (PARPs), which catalyse the covalent transfer of ADP-ribosyl moieties from NAD⁺ to “acceptor” proteins (initiation reaction) as well as repeated cycles of ADP-ribosyl transfer to previously transferred ADP-ribosyl moieties (“elongation”) and also branching of the growing chains, thus creating highly complex polymers of protein-linked ADP-ribose (Fig. 1). Poly(ADP-ribosyl)ation was originally discovered as a biochemical reaction dependent on DNA strand breaks, but among the presently known PARPs, activation by DNA strand breaks could only be detected for PARP-1 and PARP-2, which both localise to the cell nucleus. In the context of living cells PARP-1 is the major catalyst of poly(ADP-ribose) formation under conditions of DNA damage induced by ionising radiation, alkylating agents and oxidants. The enzyme catalysing the catabolism of poly(ADP-ribose) by hydrolysis of the ribose-ribose linkages in the polymer is termed poly(ADP-ribose) glycohydrolase (PARG) (Fig. 1). The enzymes involved in poly(ADP-ribose) metabolism and their genes are listed in Table 1.

Characteristics

Catalytic Activity of PARP-1 and Life Cycle of Poly(ADP-ribose)

PARP-1 is the best-characterised member of the PARP family. While its gene (termed *ADPRT* in man and *Adprt1* in mouse according to official nomenclature)



Poly(ADP-Ribosyl)ation, Pathophysiology. Figure 1 Schematic representation of the structure of poly(ADP-ribose). Ade, Adenine; Nam, Nicotinamide; PARG, poly(ADP-ribose) glycohydrolase; PARPs, poly(ADP-ribose) polymerases; Rib, Ribose; grey arrows, sites of poly(ADP-ribose) cleavage by PARG.

is constitutively and abundantly expressed in proliferative tissues, its catalytic activity is stimulated dramatically by double-stranded DNA with interruptions in the sugar-phosphate backbone (i.e., single-strand or double-strand breaks), thus making poly(ADP-ribose) a cellular immediate early response to DNA strand breakage. PARP-1 has been detected in all eukaryotic species studied except yeasts and displays a highly conserved domain structure. Its N-terminal DNA-binding domain binds to DNA strand breaks *via* two zinc fingers, which causes an immediate and dramatic activation of the catalytic centre within the C-terminal NAD⁺-binding domain. In intact cells PARP-1 itself is the major acceptor protein to be modified with poly(ADP-ribose), this automodification occurring in a specific centrally located domain. However, additional acceptor proteins have been found in living cells, e.g. histones and topoisomerases. The existence of poly(ADP-ribose) in living cells is transitory and depends on the presence of DNA strand breaks, due to rapid turnover of the ADP-ribose polymer.

Cellular Functions of PARP-1

Cytoprotection and Maintenance of Genomic Stability

Various strategies have been employed to abrogate poly(ADP-ribose)ation, including competitive low-molecular-weight PARP inhibitors, expression of dominant negative PARP-1 versions, PARP-1 antisense RNA, or *Adprt1* knockout mice. The results have demonstrated that poly(ADP-ribose)ation contributes to the recovery of proliferating cells from low-level DNA damage as induced, e.g. by alkylating agents, oxidants or ionising radiation. This effect has been linked mechanistically with an involvement of PARP-1 in DNA base-excision repair. Furthermore, poly(ADP-ribose)ation counteracts the induction of genomic instability by DNA damage (2).

Cell Death by Energy Depletion

In contrast to the above cytoprotective function, PARP-1 overactivation may lead to cell suicide, due to severe depletion of its substrate, NAD⁺, and consequently of ATP. This cytotoxicity mechanism has most prominently been identified in various non-proliferative cell types *in vivo* and *in vitro*, by comparing *Adprt1*-knockout with wild-type mice and derived cells, or by using PARP inhibitors (see below, "Clinical Relevance").

Transcriptional Regulation

Over the last few years numerous physical and functional interactions of PARP-1 with transcription factors have been described, including AP-2, B-MYB, DF1-4, E47, NF-κB, p53, PC1, Oct-1, RXR, HTLV-1 tax, TBP, TEF-1 and YY1. Their interaction with PARP-1 has either positive or negative effects on the specific transactivation process, depending on the interaction partner. An involvement of PARP-1 was also reported in the control of the E2F-1 promoter. Consistent with the above, primary *Adprt1*-knockout fibroblasts displayed down-regulation of genes involved in regulation of cell-cycle progression, mitosis, DNA replication or chromosomal processing or assembly, while some genes encoding extracellular matrix or cytoskeletal proteins, implicated in cancer or ageing, were up-regulated.

Of great functional relevance is the finding that PARP-1 acts as a cofactor for ►NF-κB transactivation, a signalling pathway playing a key role in immune and inflammatory responses. This was shown by the lack of the lipopolysaccharide (LPS)-induced increases in tumour necrosis factor-α (TNF-α) production, inducible nitric oxide synthase (iNOS) expression and nitric oxide (NO) production in *Adprt1*-knockout mice and by the dramatic protection of such mice from endotoxic shock. The co-transactivator function of PARP-1 apparently depends on the presence of PARP-1 protein,

Poly(ADP-Ribosyl)ation, Pathophysiology. Table 1 Synopsis of enzymes involved in poly(ADP-ribose) formation or degradation and their genes

Enzyme	Gene*	Alternative symbols	Locus ID	Chromosomal location
NAD ⁺ ADP-ribosyl transferase (polymerising); <i>N.B. Common alternative designations are the following: poly(ADP-ribose) polymerase-1 [PARP-1]; poly(ADP-ribose) synthetase [PARS]</i>	<i>ADPRT</i>	<i>PARP, PARP-1, PARS, ADPRT1, pADPRT-1, PPOL</i>	142	1q41-q42
NAD ⁺ poly(ADP-ribose) polymerase-like 1	<i>ADPRTL1</i>	<i>KIAA0177, PH5P, PARPL, VPARP, VAULT3</i>	143	13q11
NAD ⁺ poly(ADP-ribose) polymerase-like 2	<i>ADPRTL2</i>	<i>ADPRTL3, PARP-2, ADPRT2, pADPRT-2</i>	10038	14q11.2-q12
NAD ⁺ poly(ADP-ribose) polymerase-like 3	<i>ADPRTL3</i>	PARP-3	10039	3p22.2-p21.1
NAD ⁺ poly(ADP-ribose) polymerase-like 4**	<i>ADPRTL4</i>	<i>PARP-4</i>	23783	22q11.1
Tankyrase ; TRF1-interacting ankyrin-related ADP-ribose polymerase	<i>TNKS</i>	<i>PARPL</i>	8658	8q
Poly(ADP-ribose) glycohydrolase [PARG]	<i>PARG</i>		8505	10q11.23

*Approved Human Genome Organization (HUGO) Gene Nomenclature Committee symbol; **Protein predicted

with NF- κ B and PARP-1 forming a stable nuclear complex independent of DNA binding. Primary cells from *Adprt1*-knockout mice are also impaired in NF- κ B transactivation induced by various stimuli involved in inflammatory and genotoxic stress signalling. PARP inhibitors, however, did not block the transcriptional activation of a NF- κ B-dependent reporter gene in wild type cells. Neither the enzymatic activity nor the DNA binding activity of PARP-1 was required for NF- κ B-dependent transcriptional activation in *Adprt1*-knockout cells complemented with different PARP-1 mutants. PARP-1 interacted *in vitro* directly with both NF- κ B subunits (p50 and p65), each subunit binding to a different PARP-1 region. In cells, PARP-1 was shown to activate the promoters of the iNOS and P-selectin genes in a NF- κ B-dependent manner upon stimulation of the cells with inflammatory stimuli or co-transfection of p65 (4).

Recently, a loss of several stress-activated transcription factors as well as decreased expression of genes for cytokines and cellular adhesion molecules was observed in glial cells from *Adprt1* knockout mice. In this system it was shown that augmented expression of some but not all of these genes is dependent of PARP-1 catalytic activity (5). This might explain why PARP inhibitors can lead to functional improvement in

in vivo models of inflammation or ischemia-reperfusion damage (see below).

Role in Apoptosis

Apoptosis is associated with dramatic changes in the poly(ADP-ribosyl)ation system. There is proteolytic cleavage of PARP-1 into two fragments by caspase-3 during the execution phase of apoptosis, which serves as a biochemical marker for apoptosis. This cleavage is thought to abrogate the responsiveness of PARP-1 to DNA strand breaks. On the other hand, despite PARP-1 cleavage, there is evidence from several different cell systems for massive accumulation of poly(ADP-ribose) (i.e., the product of PARP activity) in apoptotic cells. Furthermore, PARP-1 activation is required for translocation of apoptosis-inducing factor (AIF) from the mitochondria to the nucleus and AIF is necessary for PARP-1-dependent but caspase-independent cell death (6). This may represent an additional mechanism by which PARP-1 can kill cells, perhaps independent of NAD⁺ depletion and transcriptional regulation of inflammation-related genes.

Clinical Relevance

In contrast to the physiological, “cytoprotective” functions discussed above, PARP-1 over-activation

can lead to cell death and dysfunction or degeneration of tissues. Many experiments have been reported in animal models of a wide variety of common, debilitating or even lethal human disorders over the last few years that all provide evidence for a pathogenic role of PARP(-1) activity. Typical experimental strategies were (i) to demonstrate PARP activation in the target tissue under conditions of induction of pathology, and (ii) to analyse the phenotype of *Adprt1* knockout mice or normal animals treated with PARP inhibitors. The various paradigms are discussed below, but it must be noted that very little, if anything, is known as yet about the role of PARP activity in human patients. Likewise no clinical trials of PARP-inhibitory compounds have been reported to date. Therefore it is impossible to predict exactly what clinical relevance PARP inhibitors are going to have.

Ischemia-Reperfusion Damage

Ischemia of tissues followed by reperfusion is known to induce massive release of ROS and NO in the affected area. The resulting DNA damage and PARP(-1) over-activation appears to be crucial for the ensuing acute cell death, which at the macroscopic level is typically manifested as an infarct. Apart from acting as a drain for cellular NAD⁺ in the target cells, PARP-1 may also be involved in this context as a co-transactivator for NF- κ B (see above) and perhaps of other stress-induced transcription factors in macrophages or microglia, thus mediating the inflammatory component of tissue damage. The critical role of PARP(-1) in ischemia-reperfusion damage was demonstrated in a variety of organs.

Adprt1 gene knockout or PARP inhibitors can reduce the volume of brain infarct and attenuate the associated neurological dysfunction. PARP inhibitors seem to be effective even when administered after the onset of ischemia. An important pathogenic component is the excessive release of glutamate and the resulting excitotoxicity. It is interesting to note that a selective role of PARP-1 in excitotoxicity induced by NMDA but not AMPA (two glutamate agonists interacting with distinct receptors) has been observed in *Adprt1*-knockout mouse brain. Protective effects by PARP inhibition have also been reported in rat brain cortex under global cerebral ischemia without reperfusion. Protection was also afforded to the inner retinal elements under conditions of retinal ischemia-reperfusion damage.

In *Adprt1*-knockout mice cognitive and motor deficits after traumatic brain injury are attenuated.

The size of myocardial infarcts after ischemia-reperfusion can be significantly reduced by abrogating PARP(-1) activity. Myocardial contractility and relaxation, coronary blood flow, and endothelial function were all significantly improved and myocardial

high energy phosphate content was preserved in the PARP inhibitor-treated animals, whereas activation of P-selectin and the adhesion molecule ICAM-1 was suppressed.

Ischemia-reperfusion damage in liver or in the kidney is attenuated by PARP inhibitors.

Increased mucosal PARP activity has been seen after occlusion-reperfusion of the superior mesenteric artery of rat and mouse, followed by mucosal damage and hyper-permeability of the intestine. There was attenuation of such alterations and reduction of neutrophil infiltration of the ileum by PARP inhibitors or in *Adprt1*-knockout mice subjected to the same ischemia-reperfusion protocol.

Toxin-Induced Parkinson Syndrome

A common model for Parkinson's disease is the administration of 1-methyl-4-phenyl-1,2,3,6-tetrahydropyridine (MPTP) to animals. Its metabolite is known to provoke formation of ROS in the dopaminergic neurones of the substantia nigra, thus leading to selective neuronal death and Parkinson's syndrome. Dopaminergic neurones from *Adprt1* knockout mice were protected against MPTP-induced cell death, mirrored by increased resistance to clinical disease.

Chronic Heart Failure

Chronic heart failure induced in rats by chronic coronary artery ligation led to increased nitrotyrosine formation and PARP activation in the myocardium and intramural vasculature, depressed left ventricular performance and impaired vascular relaxation of aortic rings. The PARP inhibitor N-(6-oxo-5,6-dihydro-phenanthridin-2-yl)-N,N-dimethylacetamide (PJ34) improved both cardiac dysfunction and vascular relaxation (7).

Septic or Haemorrhagic Shock

Adprt1-knockout mice are resistant to lethal septic shock induced by injection of LPS or by caecal ligation and puncture. This seems to be mediated by a deficiency of NF- κ B transactivation, thus reducing the typical LPS-induced increases in serum levels of TNF- α , IL-6 and IL-10, and of iNOS expression and NO production. *Adprt1*-knockout mice further exhibited a lower degree of inflammation in the gut and lung. The contribution of PARP to bacterial LPS-induced acute lung injury and vascular failure was assessed in pigs. LPS infusion evoked a decline in systemic arterial pressure and a biphasic elevation of pulmonary arterial pressure, while lung compliance declined and respiratory resistance increased. Pre-treatment with the PARP inhibitor 3-aminobenzamide (3-AB) eliminated the rise in pulmonary and total respiratory resistance but did not alter the systemic and pulmonary hemodynamic responses to LPS infusion, suggesting a role for PARP

activation in the changes of lung mechanics after LPS-induced acute lung injury.

Adprt1-knockout mice are also resistant to haemorrhagic shock and the rapid decrease in blood pressure after resuscitation, displaying an increased survival time and reduced lung neutrophil sequestration. In rats undergoing haemorrhage and resuscitation, the PARP inhibitor 5-aminoisoquinolinone (5-AIQ) reduced multi-organ injury and dysfunction but did not prevent circulatory failure.

Diabetes Mellitus and Diabetes-Related Endothelial Dysfunction

Type-1 (juvenile) diabetes results from selective destruction of insulin-producing pancreatic β -cells during islet inflammation. Cytokines, ROS and NO, which are all released during this process, contribute to β -cell death. *Adprt1*-knockout mice are completely resistant to diabetes induction by streptozotocin. Furthermore, knockout mouse derived β -cells in culture or wild type cells treated with PARP inhibitors are also protected from cell death induced by exposure to ROS, NO or streptozotocin.

Many diabetic patients suffer from retinopathy, nephropathy, neuropathy and accelerated atherosclerosis, all of which are preceded by dysfunction of endothelial cells. PARP activation seems to play a role in the pathogenesis of endothelial dysfunction in diabetes. Treatment of diabetic mice with a PARP inhibitor allowed maintenance of normal vascular responsiveness despite persisting severe hyperglycaemia.

Other Acute and Chronic Inflammatory Disorders

P selectin and circulating xanthine oxidase are involved in neutrophil infiltration of the lung associated with acute pancreatitis. Up-regulation of P-selectin, generation of platelet-activating factor and neutrophil recruitment were all prevented by PARP inhibition or by scavenging of H_2O_2 .

Crohn's disease is a chronic inflammatory bowel disease characterised by oxidant-induced tissue injury and increased intestinal permeability. *I110*-knockout mice are used as a model for Crohn's disease, as they display increased intestinal permeability, pro-inflammatory cytokine release and nitrosative stress. Here the PARP inhibitor 3-AB normalised colonic permeability, attenuated inflammation, and reduced secretion of TNF- α and IFN- γ , expression of iNOS, and nitrotyrosine levels. Recently a chemically modified antisense oligonucleotide (ISIS 110251) blocking expression of PARP-2 was tested for efficacy in *I110*-knockout mice. This oligonucleotide reduced PARP-2 mRNA levels in liver by up to 80%. In *I110*-knockout mice with established colitis, treatment with ISIS 110251 normalised the colonic epithelial barrier and transport function, reduced pro-inflammatory cytokine secretion

and iNOS activity and attenuated inflammation. These data indicate a significant role for PARP-2 in mediating colonic inflammatory disease.

The PARP inhibitors 6(5H)-phenanthridinone and benzamide attenuated experimental allergic encephalomyelitis (EAE) in rats, reducing clinical score, neuroimmune infiltration and expression of inflammatory mediators such as iNOS, IL-1 β and IL-2, cyclooxygenase-2, TNF- α and INF- γ in the spinal cord of myelin-immunised rats. This effect seemed to be due to a reduction of NF- κ B and AP-1-mediated production of pro-inflammatory cytokines.

Excitotoxic brain lesions initially result in the primary destruction of brain parenchyma, followed by microglial cell migration towards the sites of injury and release of large quantities of ROS, which causes secondary damage that accounts for most of the loss of brain function. Microglial migration was found to be controlled by expression of the integrin CD11a, which in turn is regulated by PARP-1/NF- κ B. Antisense-mediated down-regulation of PARP-1 or CD11a abrogated microglial migration almost completely and protected neurones from secondary damage.

Intra-thoracic administration of zymosan-activated plasma leads to an increase in lung infiltration by neutrophils and serves as a model for acute inflammation. The PARP inhibitor 5-AIQ reduced the degree of lung injury and attenuated the expression of P selectin and ICAM-1 as well as the recruitment of neutrophils into the injured lung. The up-regulation of P selectin and ICAM-1 in human endothelial cells exposed to oxidative stress or to TNF α was also attenuated by 5-AIQ.

Adprt1-knockout mice were protected against bacterial meningitis-associated central nervous system complications including blood-brain barrier breaching and increase in intracranial pressure. This was accompanied by a reduction in meningeal inflammation and improved clinical status. A similar effect was achieved by PARP inhibition using 3-AB.

Outlook: ADP-Ribosylation Inhibitors as Therapeutic Drugs for Humans

It is obvious that PARP-inhibitory substances with appropriate pharmacokinetic profiles possess a significant potential as drugs for a "cell-rescue therapy", in view of the above prominent role of PARP-1 overactivation in a wide range of highly prevalent disabling and even lethal conditions (7). Improved PARP inhibitors, which ideally should target specific cell types, could turn out to be very useful tools for therapy and/or prevention.

With regard to inhibitor specificity, it should be noted that, by definition, all existing PARP inhibitors inhibit PARP-1, but the first-generation inhibitors at least, interfered also with other ADP-ribosyl transfer

reactions, such as mono-ADP-ribosylation of proteins, cyclic ADP-ribose formation or NAD⁺ glycohydrolysis, albeit at different IC₅₀ levels. In addition, the novel PARPs are likely to be inhibited as well, probably at similar IC₅₀ levels as PARP-1. It will be desirable to further increase inhibitor selectivity, so as to minimise possible side effects arising from collateral inhibition of enzymes other than PARP-1.

References

1. Bürkle A (2004) Poly(ADP-Ribosyl)ation. Landes Bioscience, Georgetown
2. Bürkle A (2001) Physiology and pathophysiology of poly(ADP-ribosyl)ation. *BioEssays* 23:795–806
3. Chiarugi A (2002) Poly(ADP-ribose) polymerase: killer or conspirator? The 'suicide hypothesis' revisited. *Trends Pharmacol Sci* 23:122–129
4. Hassa PO, Hottiger MO (2002) The functional role of poly(ADP-ribose)polymerase 1 as novel coactivator of NF-κB in inflammatory disorders. *Cell Mol Life Sci* 59: 1534–1553
5. Ha HC, Hester LD, Snyder SH (2002) Poly(ADP-ribose) polymerase-1 dependence of stress-induced transcription factors and associated gene expression in glia. *Proc Natl Acad Sci USA* 99:3270–3275
6. Yu SW, Wang H, Poitras MF et al (2002) Mediation of poly(ADP-ribose) polymerase-1-dependent cell death by apoptosis-inducing factor. *Science* 297:259–263
7. Virag L, Szabo C (2002) The therapeutic potential of poly(ADP-ribose) polymerase inhibitors. *Pharmacol Rev* 54:375–429

Polyacrylamide Gel Electrophoresis

Definition

Polyacrylamide gel electrophoresis denotes the separation of proteins according to their molecular weight using a polyacrylamide gel to which an electric current is applied.

► [Recombinant Protein Expression in Bacteria](#)

► [Two-Dimensional Polyacrylamide Gel Electrophoresis](#)

Polyadenylation

ELMAR WAHLE¹, NIELS H. GEHRING²

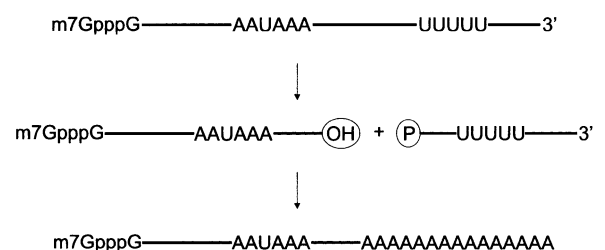
¹Institute for Biochemistry, Martin-Luther-University Halle-Wittenberg, Halle, Germany

²University Clinic Heidelberg, Children's Hospital, Heidelberg, Germany
niels.gehring@med.uni-heidelberg.de

Definition

The 3' ends of all eukaryotic mRNAs are generated by post-transcriptional processing of a precursor RNA generated by ► [RNA polymerase II](#). As part of these processing reactions, which take place in the cell nucleus, the 5' end receives a 7-methyl guanosine ► [cap](#), whereas the 3' end receives a ► [poly\(A\) tail](#) – a string of adenylate residues not encoded in the genome. These two end modifications are characteristic for all mRNAs and only for mRNAs, and thus define an mRNA molecule as such. By means of specific binding proteins, cap and poly(A) tail function synergistically in translation, and during ► [mRNA decay](#) they are the first targets to be attacked by hydrolytic enzymes before the coding sequence is degraded. 3' end processing leading to polyadenylated RNA consists of two partial reactions. The RNA is first cleaved endonucleolytically, then the upstream cleavage fragment receives the poly(A) tail by polymerization from ATP, whereas the downstream fragment is degraded (Fig. 1). The efficiency of 3' processing can be regulated so that the output of mature mRNA and therefore of the encoded protein is adjusted to physiological needs. Genes frequently have two or more potential cleavage/polyadenylation sites and, in some mRNA precursors, the choice among these sites can also be regulated and determine the kind of protein made.

The only mRNAs that do not receive a poly(A) tail are those encoding the histone mRNAs synthesized during the S phase of the cell-cycle. Their 3' ends are generated by a different type of cleavage reaction that leaves a conserved stem-loop structure at the end of the



Polyadenylation. Figure 1 An outline of the 3' end processing reaction. The primary transcript containing the cap and the polyadenylation signal sequences is first cleaved between the AAUAAA polyadenylation signal and the downstream element. Cleavage generates a 3' hydroxyl and a 5' phosphate. The downstream fragment is degraded, whereas the upstream fragment, containing the coding sequence, is polyadenylated.

message; no poly(A) is added. The stem-loop structure, associated with a specific protein, plays the same role for the histone message as the poly(A) tails for all other messages.

Characteristics

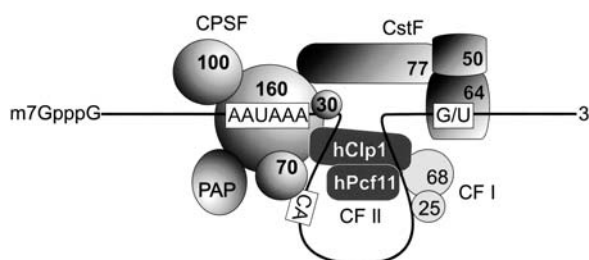
RNA Sequences Directing Polyadenylation

In mammals, two sequence elements are sufficient to direct 3' end processing and polyadenylation of a messenger RNA. The highly conserved AAUAAA hexanucleotide sequence (poly(A) signal) is located between 10 and 30 nucleotides upstream of the cleavage site. Therefore, this signal remains in the upstream cleavage fragment, destined to become the mature mRNA and is required not only for cleavage but also for subsequent polyadenylation. In contrast, a less conserved GU- or U-rich element (downstream sequence element) is found downstream of the cleavage site; it is thus removed from the mRNA by cleavage. The phosphodiester bond to be cleaved is defined by the spacing between the AAUAAA sequence and the downstream sequence element. The cleavage reaction preferentially occurs after an adenosine residue, usually at a CA dinucleotide sequence. Sequences upstream of AAUAAA have been shown to increase the efficiency of some poly(A) signals and are commonly found in the poly(A) sites of viral genes. Nonetheless, these upstream sequence elements do not appear to be essential for general 3' end processing and have been identified in only a small number of cellular mRNAs.

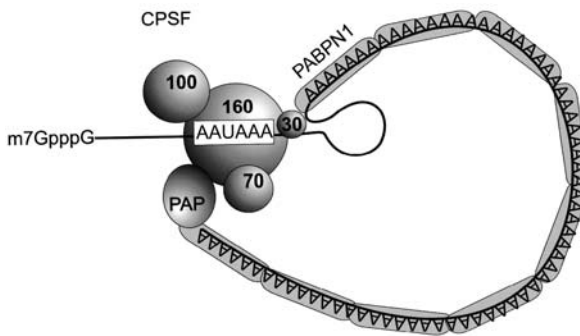
Factors and Mechanisms

3' processing can be carried out in a cell-free system derived from HeLa cell nuclear extract. Proteins involved in the reaction have been identified by the biochemical fractionation of such extracts and reconstitution of cleavage and polyadenylation. ▶ **Cleavage and polyadenylation specificity factor** (CPSF) contains at least four subunits and is required for both partial reactions. This complex specifically binds the AAUAAA polyadenylation signal. The downstream sequence element is recognized by ▶ **cleavage stimulation factor** (CstF), a heterotrimer. RNA binding of the two proteins is cooperative. ▶ **Cleavage factor I** (CF I), probably a heterodimer, is also believed to be involved in substrate RNA binding at an early stage of the reaction, but no details are known. ▶ **Cleavage factor II** (CF II) contains at least two subunits. One of the CF II subunits is likely to bind ATP, as predicted from its sequence, and the cleavage reaction is ATP-dependent. Beyond this, the role of CF II is unknown. ▶ **Poly(A) polymerase**, the enzyme carrying out poly(A) tail synthesis in the second step, is also required or at least stimulatory for the cleavage reaction. Most if not all of these proteins assemble in a large complex on the substrate RNA before any reaction takes place. It is

unknown which of the many polypeptides catalyzes hydrolysis of the phosphodiester bond. Fig. 2 shows a largely hypothetical view of the cleavage complex. Based on a comparison to the 3' processing system of *S. cerevisiae*, one can expect that additional polypeptides involved in the reaction remain to be discovered. The second step of 3' processing, polyadenylation, is tightly coupled to cleavage *in vivo* and under normal reaction conditions *in vitro*, but can be uncoupled easily *in vitro* by the use of 'pre-cleaved' RNA substrates ending at or close to the normal cleavage site. CstF, CF I and CF II are only required for the cleavage reaction and are dispensable for polyadenylation uncoupled from cleavage. Under these conditions, poly(A) synthesis requires both CPSF and poly(A) polymerase and, in addition, the ▶ **nuclear poly(A) binding protein 1** (PABPN1). Poly(A) polymerase can synthesize poly(A) from ATP by elongation of a primer RNA, but has only marginal catalytic activity on its own and no sequence specificity. The enzyme is stimulated synergistically by CPSF and PABPN1. During the reaction, CPSF remains bound to the AAUAAA sequence and PABPN1 covers the growing poly(A) tail. Cooperative recruitment of poly(A) polymerase by these two proteins thus depends on the presence of appropriate binding sites in the RNA and results in processive poly(A) tail synthesis, which stops at a length of approximately 250 nucleotides. A cartoon of the polyadenylation complex is presented in Fig. 3. According to sequence database information and a limited number of experimental studies, there are multiple versions (splice variants and/or different genes) of many of the polypeptides mentioned above.



Polyadenylation. Figure 2 A hypothetical view of the cleavage complex. CPSF is shown as a heterotetramer binding the AAUAAA sequence, and CstF is shown as a heterotrimer recognizing the downstream sequence. The two complexes interact as indicated. The arrangement of the other proteins is hypothetical. The 70 kD subunit of CPSF is shown touching the CA cleavage dinucleotide based on a suggestion that it might be the endonuclease of the complex. Experimentally verified RNA binding polypeptides are CPSF 30K and 160K, CstF 64K and both subunits of CF I. PAP, poly(A) polymerase.



Polyadenylation. Figure 3 A hypothetical view of the processive polyadenylation complex. CPSF remains bound to the AAUAAA sequence and contacts PAP through its 160 kD subunit. PABPN1 covers the growing tail with monomers touching each other; RNA binding is slightly cooperative. PABPN1 also touches poly(A) polymerase, increasing the enzyme's affinity for RNA. An additional contact shown between PABPN1 and CPSF 30K has been experimentally proven for the isolated polypeptides, but its significance in the context of the polyadenylation reaction is unknown.

The biological relevance of this diversity remains largely unexplored. In addition to the poly(A) polymerase involved in nuclear mRNA polyadenylation, a second type of such enzymes exists in the cytoplasm. These are only remotely related in sequence to the nuclear enzymes and appear to function in post-transcriptional regulation by regulated poly(A) tail extension on certain mRNAs.

Polyadenylation in the Context of mRNA Production

A number of observations suggest that the 3' processing reaction as outlined above does not take place as an isolated event *in vivo*. Instead, 3' processing is coupled to other aspects of mRNA synthesis. The substrate on which the cleavage complex assembles is the nascent RNA. In living cells, cleavage and polyadenylation are efficient only when the precursor RNA is made by RNA polymerase II – transcripts of RNA polymerase I or III are processed poorly. Similarly, processing *in vitro* is more rapid with precursor RNAs made by RNA polymerase II in the extract as compared to synthetic substrates. While some of the dependence on RNA polymerase II *in vivo* may be explained by the fact that the cleavage reaction is stimulated by the 5' cap, which in turn is only added to RNA polymerase II transcripts, it is believed that there is also a more direct involvement of the polymerase. Some of the 3' processing factors bind the C-terminal domain of the largest subunit of the enzyme. Consequently, these proteins are thought to be delivered to the nascent RNA through their transient association with the polymerase rather than from

solution. Transcription termination of RNA polymerase II also depends on the assembly of the processing complex on the polyadenylation signal in the growing RNA chain, although probably not on the processing reaction itself. It is currently unknown how the signal of polyadenylation site recognition is transmitted to the elongating RNA polymerase.

In addition, cleavage/polyadenylation is known to be stimulated by the presence of an upstream intron and, in turn, splicing of the 3' terminal exon of an mRNA is stimulated by cleavage and polyadenylation.

Clinical Relevance

Mutations in polyadenylation signals can cause a variety of diseases by reducing the expression of the affected gene.

The **thalassemias** are characterized by a defective synthesis of either the α - or the β -globin chain. A large number of mutations in the polyadenylation signals of both the α - and the β -globin gene that usually lead to a mild clinical manifestation of thalassemia have been described and are commonly detected in compound with more severe thalassemia mutations. This indicates that the poly(A) site mutations by themselves lead only to a partial inactivation of the affected gene and thus do not cause a strong phenotype. This is probably due to the presence of extended transcripts arising from failure to process the mutated polyadenylation signal and processing at a cryptic downstream signal. The resulting mRNA species can be translated into normal polypeptides and thus contribute to the formation of functional $\alpha_2\beta_2$ hemoglobin. Furthermore, at least some normally polyadenylated mRNAs are produced, although their expression is markedly reduced.

Clinical studies have revealed an association between the risk of venous **thrombosis** and the presence of a G \rightarrow A mutation at position 20210 of the **prothrombin** gene (coagulation factor II). This mutation affects approximately 1–2% of the Northwest European population and represents one of the most common genetic risk factors for the occurrence of thromboembolism. Elevated prothrombin plasma concentrations in the carriers of the mutation are thought to play the key role in the pathogenesis of thrombosis. The 20210 G \rightarrow A mutation affects the 3' terminal nucleotide of the prothrombin mRNA 3' untranslated region and changes the CG dinucleotide at the cleavage site into a CA dinucleotide. The CG dinucleotide presumably constitutes an unfavorable cleavage site, which is responsible for a relatively inefficient 3' end processing of the prothrombin pre-mRNA. Mutation to a CA causes an increase of cleavage site recognition and an elevation of correctly 3' end processed mRNA. This leads to increased protein synthesis and thus links the prothrombin 3' end processing efficiency to the hemostasis of the blood-clotting cascade.

An inborn deficiency of ►**arylsulfatase A** (ASA) causes the lysosomal accumulation of cerebroside sulfate within nervous tissue and the characteristic symptoms of ►**metachromatic leukodystrophy** (MLD). However, an ASA deficiency is also commonly found in apparently healthy individuals and has been termed arylsulfatase A pseudodeficiency. The ASA pseudodeficiency allele may be clinically relevant in compound heterozygotes with an MLD allele. Pseudodeficiency is caused by a mutation changing the major polyadenylation signal from AAUAAC to AGUAAC, which severely diminishes ASA-mRNA levels.

Connections have also been made between 3' processing of mRNA and cell-cycle progression and cancer. Nuclear extracts from cells treated with DNA damaging agents have a defect in 3' processing. This may be caused by an inhibitory interaction between cleavage stimulation factor (CstF, see above) and the protein BARD1, which, in turn, associates with the tumor suppressor BRCA1. BRCA1 and BARD1 are thought to be involved in DNA repair. It has been suggested that this mechanism may serve to inhibit inappropriate cleavage and polyadenylation of nascent transcripts associated with RNA polymerase stalled at sites of DNA damage. Many tumors show enhanced activity of poly(A) polymerase and increased levels of transcription of two different poly(A) polymerase genes. On the other hand, modest over-expression of poly(A) polymerase is deleterious for growth of tissue culture cells. The enzyme is down-regulated by phosphorylation of its serine- and threonine-rich C-terminal domain during mitosis. This may contribute to the general shutdown of gene expression at this phase of the cell-cycle.

The inherited disease ►**oculopharyngeal muscular dystrophy** (OPMD) is caused by moderate and genetically stable trinucleotide expansions in the PABPN1 gene, which lead to corresponding expansions of a naturally occurring tract of ten consecutive alanine residues at the very N-terminus of the protein. Addition of a single alanine results in a recessive phenotype, whereas additions of up to seven alanines are dominant. The mutations cause the aggregation of PABPN1 in the form of regular fibrils in the nuclei of muscle cells. The mechanism leading to cell damage is unknown but is unlikely to involve a defect in mRNA polyadenylation.

►**Recombinant Protein Production in Mammalian Cell Culture**

References

1. Minvielle-Sebastia L, Keller W (1999) mRNA polyadenylation and its coupling to other RNA processing reactions and to transcription. *Curr Op Cell Biol* 11:352–357

2. Proudfoot NJ, Furger A, Dye MJ (2002) Integrating mRNA processing with transcription. *Cell* 108:501–512
3. Wahle E, Rügsegger U (1999) 3'-End processing of pre-mRNA in eukaryotes. *FEMS Microbiol Rev* 23:277–295
4. Zhao J, Hyman L, Moore C (1999) Formation of mRNA 3' ends in eukaryotes: Mechanism, regulation and interrelationship with other steps in mRNA synthesis. *Microbiol Mol Biol Rev* 63:405–445

Polyarticular

Definition

Polyarticular is defined by the involvement of more than five joints.

►**Rheumatoid Arthritis**

Polycistronic Constructs

Definition

Polycistronic constructs are expression plasmids that encode two or more antigens. They are produced to increase the amount of immunogenic information delivered by a DNA-biased vaccine. Constructs are made by incorporating ►**internal ribosomal entry sequences** (IRES) between the antigen encoding genes. These polycistronic expression units allow coexpression of several antigens under control of a single promoter.

►**DNA-based Vaccination**

Polycystic Kidney Disease, Autosomal Dominant

MICHAEL SUTTERS

Division of Renal Medicine, Johns Hopkins University
School of Medicine, Baltimore, MD, USA
msutters@jhmi.edu

Synonyms

Adult onset polycystic kidney disease

Definition

Disease Characteristics

Autosomal dominant polycystic kidney disease (►ADPKD) is the commonest inherited disease of humans, has a prevalence of 1:1000 and shows no racial or geographic predilections. In patients with ADPKD, multiple renal cysts develop throughout adult life, causing the loss of renal function that is the principal source of morbidity. Although cysts are also observed in the liver and pancreas, patients with ADPKD rarely develop dysfunction of these organs. In the majority of affected individuals, progressive renal failure typically leads to the requirement for renal replacement therapy by mid to late adulthood, but disease progression is very variable, even within families, raising the possibility that environmental factors and modifying genes have an important influence on the ADPKD phenotype. In addition to cyst formation, a range of other features has been associated with this disease, including hypertension, cerebral aneurysms, heart valve abnormalities, colonic diverticuli and hernia formation.

Molecular Genetics

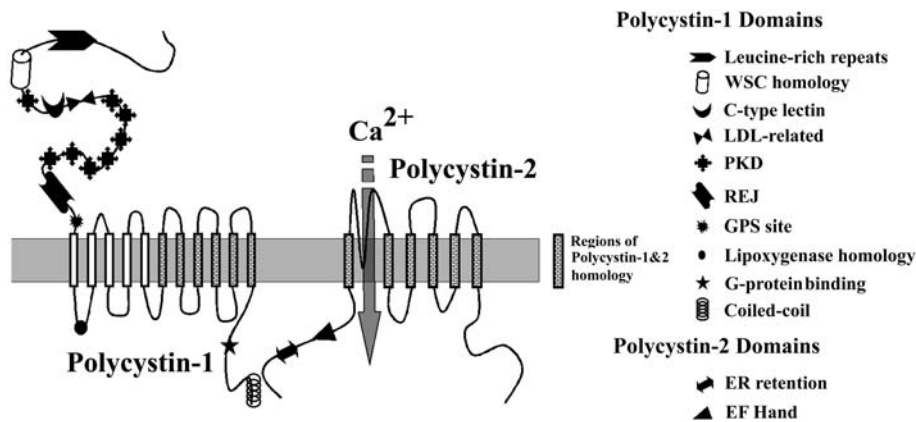
ADPKD arises as a consequence of mutations in one or other of two genes, ►PKD1 and ►PKD2, which were identified in 1994 and 1996 (1, 2). Disease characteristics in patients with mutations in the PKD1 gene (ADPKD type 1) are identical to those of patients with mutations in the PKD2 gene (ADPKD type 2) with the exception that disease progression is generally less rapid in the latter. It is therefore likely that the PKD1 and PKD2 gene products, named ►polycystin-1 and ►polycystin-2, function in the same pathway. Many of the disease-specific mutations of these two genes encode truncated or nonsense polypeptides, suggesting that ADPKD arises as a consequence of loss of function of polycystin-1 or polycystin-2. The loss of function model has gained support from studies in transgenic mice. Renal cyst formation was observed in homozygous PKD1 or PKD2 knockout embryos (neither transgenic model survived beyond 16 to 18 days post-conception) but not in heterozygous animals, which lived into adulthood and displayed delayed and rather rare cyst formation. Hence, the pattern of cyst formation characteristic of human disease appeared to require loss of both PKD alleles. More recently, evidence that cyst formation requires inactivation of both PKD alleles has emerged from a human study, which also provided an explanation for the sporadic and focal nature of cyst formation in ADPKD. In this study, human cyst epithelia were found to be ►monoclonally derived, and exhibited ►loss of heterozygosity for the PKD1 gene (3). Hence, in ADPKD, each cyst probably arises from a single renal tubule cell, not because of the ubiquitous germline mutation, but as a

consequence of a somatic mutation or “second hit” to the sole remaining normal PKD allele in that cell. Subsequent clonal proliferation of the affected cell would then give rise to the cyst epithelium. This model accounted for the observation that, despite the universal presence of the germline mutation in ADPKD, the kidneys appear not to exhibit a generalized abnormality. The second hit hypothesis for pathogenesis in ADPKD was subsequently confirmed in the currently most convincing animal model for the disease. This model arose from the fortuitous development of a transgenic mouse with an unstable PKD2 allele (4). In this mouse, an early germline inactivation of one PKD2 allele gave rise to a heterozygote background upon which subsequent sporadic inactivation of the remaining intact PKD2 allele created a phenotype closely resembling human disease.

Characteristics

Polycystin-1

Polycystin-1 is a 4302 amino acid protein with a large extracellular domain, 11 transmembrane passes and a smaller cytoplasmic carboxyl terminus. The extracellular domain displays a range of motifs indicating the capacity for interactions with extracellular proteins and carbohydrates. In particular, a series of so-called ►PKD domains suggest the potential for ligand binding, raising the possibility that polycystin-1 acts as a transmembrane receptor. This view gains further support from the presence of a ►heterotrimeric G protein binding and activation domain in the cytoplasmic carboxyl terminus. A ►G protein-coupled receptor proteolytic site (GPS) lies close to the cell membrane in the extracellular amino terminus of polycystin-1. Polycystin-1 is physiologically cleaved at this position and this cleavage event appears to be functionally important. However, the way in which cleavage modulates the function of polycystin-1 remains unknown. A lipoxxygenase homology (LH) domain lies within the first intracytoplasmic loop and probably mediates interactions with lipid or lipoprotein. The combination of a GPS and an ►LH domain appears to be the hallmark of all polycystin-1-like proteins, which are forming an expanding family with diverse tissue distribution and, probably, diverse functions. A motif resembling a sea urchin protein named “►receptor for egg jelly” (the REJ protein is found in the cell membrane of sea urchin sperm) lies in a region of the extracellular domain of polycystin-1, proximate to the cell membrane. In sea urchin fertilization, binding of the REJ protein to a component of the egg jelly activates cation channels in the sperm cell membrane. The resultant cation influx triggers the acrosome reaction. The presence of the REJ motif raises the possibility that polycystin-1 might similarly modulate membrane cation flux in response to extracellular



Polycystic Kidney Disease, Autosomal Dominant. Figure 1 Polycystins 1 and 2 are transmembrane proteins. The coiled coil domain in the cytoplasmic carboxyl terminus of polycystin-1 mediates binding with the cytoplasmic carboxyl terminus of polycystin-2. The extracellular amino terminus of polycystin-1 probably interacts with other proteins (via leucine rich repeats, PKD domains and the LDL-A region) and/or with carbohydrate (via the WSC domain and the C-type lectin region). Conformational changes induced by such interactions might be transduced to modulate intracellular signal transduction pathways (via the carboxyl cytoplasmic terminus G protein binding and activation site or coiled coil domain) or to influence the channel properties or membrane localization of polycystin-2.

signals. In fact, this seems rather likely since polycystin-1 is known to interact with polycystin-2 (via a [coiled-coil domain](#) in the polycystin-1 cytoplasmic carboxyl terminus) and polycystin-2 is a cation channel.

Polycystin-2

Polycystin-2 is a six transmembrane domain protein with overall homology to voltage activated calcium channels. This protein has been characterized as a calcium permeable non-selective cation channel (5) and has been identified as the founder member of a novel sub-class of so-called transient receptor potential proteins ([TRP Channel](#)). Members of the TRP family are thought to function as the pathway for [store-operated calcium entry](#) following a calcium release stimulus. The polycystin-2 carboxyl terminus exhibits an endoplasmic reticulum (ER) retention signal, a coiled coil domain and an [EF hand](#) motif typical of voltage gated calcium channels. The EF hand is a calcium-binding domain that might account for the sensitivity of polycystin-2 to inhibition by calcium. The polycystin-2 coiled-coil domain, located in the carboxyl terminus, is required for homotypic interactions between polycystin-2 molecules; interactions with polycystin-1 appear to require a domain in the carboxyl terminus of polycystin-2 that lies outside this coiled coil region.

Cellular and Molecular Regulation

Polycystins 1 and 2 Function in the Same Pathway

The precise roles of polycystins 1 and 2 remain uncertain. Polycystin-1 is probably resident at the plasma membrane with a basolateral distribution and

within junctional complexes, but polycystin-1 immunoreactivity has also been identified in the central apical cilia of renal epithelial cells. Debate continues over whether polycystin-2 is confined to the ER or if this protein might also function in the plasma membrane. More recent studies indicate that polycystin-2 co-localizes with polycystin-1 in the apical [cilia](#). As discussed above, observations of human disease and knockout-mouse phenotypes indicate that the two proteins almost certainly act in the same pathway with respect to the generation of cysts. Direct genetic evidence that polycystins 1 and 2 are indeed functional in one pathway arose from studies in [C. elegans](#), where the polycystin-1 homologue [Lov-1](#) colocalized with the polycystin-2 homologue in the cilia and cell bodies of male sensory neurons (6). Mutations in either protein lead to loss of the male ability to locate the hermaphrodite vulva; double mutations (inactivating Lov-1 and polycystin-2) produced no additive effect. Subsequent studies have confirmed that polycystins 1 and 2 are functionally inter-dependent in a number of other pathways as listed below.

1. The effect of exogenously over-expressed polycystin-1 to induce G0/G1 cell cycle arrest in MDCK cells is mediated by activation of the [JAK-STAT](#) pathway and is dependent upon polycystin-2.
2. Polycystin-2 appeared to antagonize polycystin-1 activation of Gi/o-type G-proteins.
3. Channel activity attributable to polycystin-2, exogenously over-expressed in Chinese hamster ovary cells, was dependent upon co-expression of a

polycystin-1 cDNA with an intact coiled-coil domain.

4. In a reconstituted membrane system, the non-selective cation channel activity of polycystin-2 was subject to spontaneous run-down that was reversed by addition of the polycystin-1 cytoplasmic carboxyl terminus.

Are All Polycystin Functions Relevant to the ADPKD Phenotype?

The polycystins appear to be involved in numerous processes, but no predominant function has yet emerged. In view of this apparent diversity, it is possible that some of the functions attributed to the polycystins are not related to cystogenesis, and perhaps not to any of the features of the ADPKD phenotype. A number of observations give support to this possibility. It is apparent from transgenic mouse studies that the polycystins play developmental roles, influencing cardiac (polycystins 1 and 2) and blood vessel development (polycystin-1), skeletal configuration (polycystin-1) and even left-right axis determination (polycystin-2). However, in the light of the double hit hypothesis, which suggests that *de novo* cyst formation could take place throughout adult life, it seems reasonable to propose that it is the polycystin functions expressed in renal epithelial cells in the terminally differentiated stage of development that are important in the pathogenesis of ADPKD. It is more difficult to envisage how the ADPKD phenotype could emerge as a delayed consequence (manifesting in adulthood) of some developmental dysregulation. These questions will be resolved when a conditional polycystin knock-out mouse is developed. Polycystin functions might be arranged along a developmental, time-based axis, but there is also evidence for a spatial separation suggested by the observation that polycystin-2 does not always co-localize with polycystin-1. Since it is very likely that ADPKD arises from disruption of a single pathway involving both proteins, the involvement of polycystins 1 and 2 in divergent pathways could indicate that such pathways are not relevant to ADPKD.

One way of organizing the large amount of information that is accumulating is to focus on those functions of the polycystins that are plausibly related to cyst formation. These functions could be considered the clinically relevant functions of the polycystins and might be expected to operate in the single (putative) pathway whose disruption in terminally differentiated renal epithelial cells leads to the cyst forming phenotype.

Clinical Relevance Polycystins and Cyst Formation

The initiator of cyst formation is a sporadic “second hit” to the single normal PKD allele in an individual

renal tubule cell (the first hit being the germ line mutation affecting all cells). According to a plausible model for cystogenesis, the single affected cell proliferates and gives rise to an epithelium deficient in the contextual responses that maintain normal tubular architecture. As a consequence, small evaginations develop from the tubule lumen, which separate from the tubule to form a small cyst rather than a tubule-like structure. Cyst expansion proceeds as a consequence of continued cyst epithelial cell proliferation accompanied by fluid accumulation within the cyst lumen. Hence, three basic processes probably underlie cystogenesis, 1) the transition from stable epithelial population to proliferation, 2) loss of the morphological constraints determining normal tubule architecture and 3) a shift in the balance of transepithelial transport from reabsorption to secretion. Each of these could arise as a consequence of the failure of affected cells to maintain a terminally differentiated phenotype. The evidence that polycystin pathways are involved in these three processes is discussed below.

Proliferation

As described above, polycystin-1 is known to inhibit cell proliferation in concert with polycystin-2 through activation of the JAK-STAT pathway. Therefore, loss of polycystin-1 or 2 would be predicted to result in increased cell proliferation.

Loss of the Morphological Constraints Determining Normal Tubule Architecture

Little is known of the processes maintaining tubule architecture in mature renal epithelial cells. Even when cell division occurs, after acute renal injury for example, the regenerating epithelium in normal individuals reconstitutes within the tubular basement membrane rather than budding out to form cysts. Perhaps the co-localization of the polycystins to the apical cilia in renal tubule epithelial cells provides a clue to cystogenesis. It has been proposed that the cilia function as a mechano- or chemo-sensory organ. If, in ADPKD, renal epithelial cells should lose information from the cilia, they may fail to detect environmental cues required for the maintenance of the tubular configuration of the epithelial sheet; for instance chemical gradients or direction of flow of luminal fluid. An increasing number of murine models for cystic kidney disease have been demonstrated to arise from mutations in proteins involved in cilia formation and function including ►polaris, ►cystine and ►inversin. Furthermore, there is evidence that the polycystins are required for one normal function of the cilia, namely the mediation of flow-dependent cell calcium responses in which ciliary deflection triggers alterations in membrane calcium channel activity. The dependence of this ciliary function upon polycystin-1

was demonstrated in renal epithelial cell lines derived from homozygous polycystin-1 knockout mice (isolated at E15.5) that exhibited loss of the normal cell calcium response to flow (7).

A Shift in the Balance of Transepithelial Transport from Reabsorption to Secretion

The proximal and distal renal tubular epithelia, from which cyst lining epithelia are derived, are predominantly reabsorptive, transporting ions and water from the apical (luminal) to the basolateral (blood) compartments. In cysts as in tubules, the apical membrane of the epithelial cells faces the lumen. These cyst epithelial cells become progressively flattened and de-differentiated in appearance as disease progresses, and lose the capacity for avid sodium reabsorption. In addition to diminished reabsorption, aberrant secretion across the cyst epithelium is probably required for fluid accumulation within the cyst lumen. In classical secretory tissues, such as colon and lung, transepithelial chloride secretion is driven by the opening of apical chloride channels. Two chloride conductance pathways that mediate secretion have been identified, the ►cAMP activated cystic fibrosis transmembrane conductance regulator (►CFTR) and the calcium activated chloride channel ►Cl_{Ca}. If sodium reabsorption is down-regulated in the cyst epithelium, then constitutive basal activity of either chloride channel might play a permissive role in the gradual accumulation of cyst fluid. However, since interruption of polycystin pathways might promote dysregulation of heterotrimeric G protein signal transduction and/or intracellular calcium homeostasis, it is conceivable that these chloride channels might be pathologically activated by up-regulation of ligand-induced stimulation of cAMP or intracellular calcium concentration in ADPKD.

References

1. The European Polycystic Kidney Disease Consortium (1994) The Polycystic Kidney Disease 1 Gene Encodes a 14 kb Transcript and Lies Within a Duplicated Region on Chromosome 16. *Cell* 77:881–94
2. Mochizuki T, Wu G, Hayashi T et al (1996) PKD2, A Gene For Polycystic Kidney Disease That Encodes An Integral Membrane Protein. *Science* 272:1339–1342
3. Qian F, Watnick T, Onuchic LF et al (1996) The Molecular Basis of Focal Cyst Formation in Human Autosomal Dominant Polycystic Kidney Disease Type I. *Cell* 87:979–987
4. Wu G, Drummond I, Cai Y et al (1998) Somatic Inactivation of PKD2 Results in Polycystic Kidney Disease. *Cell* 93:177–188

5. Gonzalez-Perrett S, Kim K, Ibarra C et al (2001) Polycystin-2, the protein mutated in autosomal dominant polycystic kidney disease (ADPKD), is a Ca²⁺-permeable nonselective cation channel. *PNAS* 98:1182–1187
6. Barr MM, De Modena J, Braun D et al (2001) The *Caenorhabditis elegans* autosomal dominant polycystic kidney disease gene homologs *lov-1* and *pkd-2* act in the same pathway. *Current Biology* 11:1341–1346
7. Nauli S, Alenghat FYL, Williams E et al (2003) Polycystins 1 and 2 mediate mechanosensation in the primary cilium of kidney cells. *Nat Genet* 33:129–137

Polycystin

Definition

Polycystin-1 is the protein product of the PKD1 gene. Mutations in the PKD1 gene cause ADPKD.

►Polycystic Kidney Disease, Autosomal Dominant

Polygenic/Polygenetic

Definition

Polygenic/polygenetic designates a mode of disease transmission or phenotypic trait determined by multiple genes, each with a comparatively small but additive effect.

- Crohn Disease
 ►Drosophila Model of Cardiac Disease
 ►Genetic Screening in Populations
 ►Large-Scale ENU Mutagenesis in Mice
 ►Schizophrenia Genetics

Polyglutamine Disease, the Emerging Role of Transcription Interference

J. PAUL TAYLOR, TIM W. RATZLAFF

Department of Neurology, University of Pennsylvania
 School of Medicine, Philadelphia, PA, USA
 jpt@mail.med.upenn.edu

Synonyms

Huntington's disease; Kennedy's disease; dentatorubral pallidoluysian atrophy; spinocerebellar ataxia; inherited neurodegenerative disease; trinucleotide repeat expansion disease; Repeat extension disease

Definition

Polyglutamine Disease

The polyglutamine diseases are a group of nine hereditary neurological disorders caused by pathological expansion of a CAG trinucleotide repeat in the disease gene. These include Huntington's disease (►HD), Kennedy's disease (►KD), dentatorubral pallidoluysian atrophy (►DRPLA), and six forms of dominantly inherited spinocerebellar ataxia (►SCA1–3, 6, 7 and 17) (1). They are all late-onset neurodegenerative disorders for which there is currently no treatment. The expanded CAG repeat in the disease genes results in elongated stretches of polyglutamine in the disease proteins, making them toxic to neurons. Thus, the polyglutamine diseases appear to share a common molecular basis determined by the toxic gain of function associated with expanded polyglutamine. The phenotypes of the different polyglutamine diseases vary considerably. HD is characterized by degeneration of cortical and subcortical grey matter and the clinical picture is dominated by dementia, psychiatric disturbances and profound movement disorders. DRPLA patients experience dementia and movement disorders similar to HD, but also frequently develop a progressive ataxia. The SCAs are dominated by degeneration of cerebellar nuclei leading to progressive ataxia, although involvement of other systems is frequent. KD is characterized by slowly progressive motor neuron degeneration leading to muscle atrophy and weakness. The distinctive clinical picture seen in each of these diseases is a reflection of the selective vulnerability of different populations of neurons in each of these diseases despite widespread and overlapping expression of some of the disease proteins. The selective neurodegeneration in these diseases has long been enigmatic, since each disease protein shares the same gain-of-function mutation. Recent evidence strongly suggests that the context of the polyglutamine expansion is a strong determinant of this selectivity, causing the polyglutamine toxicity to be influenced by the normal interacting partners of the disease protein and, perhaps, partial loss-of-function of the disease protein.

Cellular and Molecular Regulation

Polyglutamine Aggregation

An important advance in understanding the toxicity of expanded polyglutamine was the recognition of

ubiquitinated neuronal inclusions in degenerating regions of the brain and spinal cord tissue from patients with these diseases (2). These inclusions consist of large accumulations of insoluble aggregates of the disease proteins (or polyglutamine-containing fragments) in association with a host of additional proteins. This insight was in agreement with known biochemical features of long polyglutamine tracts, the propensity to aggregate *in vitro* as anti-parallel β -strands termed 'polar zippers'. When the genetic basis for the polyglutamine diseases was established, it became apparent that there is a close correlation between the threshold CAG repeat length that leads to disease and the polyglutamine length that leads to aggregation in experimental systems. This observation formed the initial basis of the argument that aggregation of expanded polyglutamine is the acquired property underlying the toxic gain-of-function. While the primacy of aggregation in initiating the toxicity of the polyglutamine-containing proteins is now supported by substantial evidence, the nature of this toxicity and its target has remained elusive. Agents that prevent polyglutamine aggregation, such as Congo red or trehalose, have shown some effectiveness in ameliorating polyglutamine toxicity in mice. These results underscore the relationship of aggregation to toxicity and show the potential therapeutic value of preventing aggregation.

The Importance of Nuclear Localization in Polyglutamine Pathogenesis

Substantial evidence indicates that the nucleus is an important site of polyglutamine toxicity for at least some polyglutamine diseases. The first insight into this came from recognition that accumulation of mutant protein in the nucleus correlates with disease (Table 1). In some cases, disease-associated proteins are normally located in the nucleus. Others are normally located in the cytoplasm, but are redistributed to the nucleus where they accumulate and form inclusions during the disease process. Mutant forms of huntingtin, ataxin-3 and the androgen receptor all show this nuclear redistribution in disease. In addition to this circumstantial evidence, the importance of nuclear localization as a requisite step in polyglutamine pathogenesis is now supported by substantial experimental evidence from cell culture and animal models of HD, KD, DRPLA and SCAs 1, 3 and 7 where nuclear localization is required to initiate toxicity. It should be noted, however, that nuclear localization of mutant protein has not been observed in SCA2 and SCA6 (Table 1). In these diseases, mutant forms of ataxin-2 and ataxin-6 are found to accumulate in the cytoplasm and cell membrane, respectively. Any

Polyglutamine Disease, the Emerging Role of Transcription Interference. Table 1

Disease	Protein	Normal Location	Disease Location	Normal Function
Huntington's Disease	Huntingtin	cytoplasm and nucleus	cytoplasm and nucleus	unknown; implicated in regulation of transcription and vesicular transport
Kennedy's Disease	Androgen receptor	cytoplasm and nucleus	cytoplasm and nucleus	ligand-responsive transcription factor
Dentatorubro-pallidoluysian atrophy	Atrophin-1	nucleus	nucleus	unknown; <i>Drosophila</i> ortholog shown to be a corepressor of transcription
Spinocerebellar ataxia 1	Ataxin-1	cytoplasm	cytoplasm and nucleus	unknown; evidence of RNA-binding
Spinocerebellar ataxia 2	Ataxin-2	cytoplasm	cytoplasm	unknown
Spinocerebellar ataxia 3	Ataxin-3	cytoplasm	cytoplasm and nucleus	unknown; implicated in repression of transcription and proteolysis of polyubiquitinated proteins
Spinocerebellar ataxia 6	CACNA1A	plasma membrane	Cytoplasm and membrane	voltage-dependent calcium channel
Spinocerebellar ataxia 7	Ataxin-7	nucleus	nucleus	transcription factor; component of STAGA complex
Spinocerebellar ataxia 17	TATA-binding protein	nucleus	nucleus	transcription factor; component of basal transcription machinery

unifying hypothesis of polyglutamine pathogenesis will need to account for this difference in subcellular localization.

A Primary Role for Transcription Interference?

A wide array of altered biological processes have been described in model systems of polyglutamine disease and proposed as mechanisms to account for polyglutamine toxicity including impairment of the ►ubiquitin-proteasome system (UPS), altered synaptic function, disrupted axonal transport and induction of programmed cell death. While abnormalities in these systems may contribute to neuronal dysfunction or loss, evidence is scant that any of these pathways serve as primary targets for expanded polyglutamine. In contrast, a growing body of evidence suggests that interference with regulators of transcription may be a primary mechanism in polyglutamine disease, perhaps initiating a cascade of subsequent abnormalities. Among the early indications that transcription might be a proximal target of polyglutamine toxicity were expression profile analyses of mouse models of Huntington's disease and SCA1 showing that alterations to transcription are early events in neurodegeneration.

Elucidating the Normal Function of Disease Proteins Provides a Clue

The notion that transcription interference might be fundamental to polyglutamine toxicity is supported by the growing appreciation that many of the polyglutamine disease proteins normally function in regulation of transcription (Table 1). For example, the AR and TATA-binding protein are well-established transcription factors. Ataxin-7 was recently identified as a component of a multimeric ►histone acetylation complex SPT3-TAF-GCN5 acetylase (►STAGA), the homolog of the complex found to be impaired in a yeast model of HD (see below). Both ataxin-3 and atrophin-1 (mutated in SCA3 and DRPLA, respectively) have been implicated in regulating the function of ►histone deacetylases. Huntingtin augments transcription of brain-specific genes through interaction with the repressor element-1 transcription factor (►REST). If polyglutamine disease proteins are transcription regulatory factors, how might this relate to the toxicity of expanded polyglutamine? Recent evidence suggests that polyglutamine expansion results in partial loss of function for at least some proteins. For example, polyglutamine expansion in the androgen receptor results in a partial loss of androgen-responsive gene

expression. This is probably responsible for the feminization observed in KD and may also influence the range of neurons that are vulnerable to polyglutamine toxicity in this disease. Similarly, polyglutamine expansion is found to interfere with the ability of huntingtin to negatively regulate the neuronal repressor REST in model systems. This results in reduced expression of brain-specific genes such as brain-derived neurotrophic factor (►BDNF) (3). Reduced levels of this neurotrophic factor are reportedly found in HD brain and could conceivably contribute to the neurodegeneration seen in HD. This hypothesis is supported by the observation that conditional knockout of either BDNF or huntingtin in the mouse leads to progressive neurodegeneration, although not with overlapping patterns.

Can Gain of Function Lead to Loss of Function?

It is clear, however, that loss of disease protein function alone does not account for the degeneration observed in the polyglutamine diseases. A preponderance of evidence indicates that polyglutamine expansion endows the disease proteins with some toxic property. One simple proposal to explain this toxicity is sequestration and functional depletion of some limiting factor that is critical to neuronal function and survival. Thus, there have been extensive efforts to identify the cellular components that interact with deposited aggregates of polyglutamine-expanded disease proteins. Two classes of proteins stand out as most frequently and consistently found to interact with polyglutamine aggregates, 1) components of the UPS, and 2) an extensive list of transcription regulatory factors. The association of UPS components with polyglutamine aggregates (including ubiquitin, chaperones and subunits of the proteasome) has been interpreted as a cellular response to accumulating misfolded protein and an effort by the cell to eliminate the aggregates. The association with transcription factors, on the other hand, was unexpected. In some cases, the sequestered proteins represent normal binding partners of the disease protein. For example, the transcription factor CRX, which regulates retinal-specific gene expression, is sequestered and depleted by aggregates of polyglutamine-expanded ataxin-7. In a transgenic model of SCA7, loss of CRX activity is found and believed to contribute to the cone-rod dystrophy that occurs in this disease (4).

Among the first transcription-related factors found to interact with polyglutamine-expanded disease proteins was CREB-binding protein (►CBP). This transcription co-activator is sequestered within polyglutamine inclusions in numerous cell culture and animal models of different polyglutamine diseases and also found to colocalize with the intranuclear polyglutamine

inclusions observed in samples from patients with Huntington's disease, Kennedy's disease and DRPLA (5). CBP is a transcription co-activator that augments transcription through regulation of chromatin structure in the vicinity of target genes. CBP mediates transcriptional activation by a wide array of transcription factors including CREB, NF- κ B, AP-1 and p53, among others. The mechanism of CBP activity involves post-translational modification of ►histones by an intrinsic ►histone acetyltransferase (HAT) activity. The observation that CBP is sequestered by multiple disease proteins may relate to an affinity between the acetyltransferase domain of CBP and long polyglutamine stretches. A role for CBP depletion in polyglutamine disease is supported by cell culture studies demonstrating that polyglutamine toxicity is associated with reduced levels of histone acetylation as well as loss of CBP-dependent activation of transcription. Moreover, up-regulation of CBP was found to reduce polyglutamine cytotoxicity and prevent neurodegeneration in a *Drosophila melanogaster* model of polyglutamine disease.

Polyglutamine Disease and Aberrant Acetylation

A primary level of chromatin organization is the wrapping of DNA around cores of histone proteins to form nucleosomes. Post-translational modification of histones has emerged as an important epigenetic mechanism regulating gene expression leading to the recognition of a "histone code". Of the known histone modifications, acetylation is the longest recognized and perhaps best understood. Hyper-acetylation of histones is associated with transcriptionally active regions of chromatin, while hypo-acetylated histones are typically associated with transcriptionally silent chromatin domains. Homeostasis of histone acetylation is maintained by opposing enzyme activities; histone acetyltransferases (HATs), such as CBP, that attach acetyl groups to lysine residues in the tails of histones and histone deacetylases (HDACs) that remove these same acetyl groups. Acetylation of lysine residues in the amino-terminal tail of histones reduces their charge attraction for DNA, thereby reducing the degree of chromatin compaction and allowing greater accessibility for transcription factors.

The earliest evidence implicating aberrant histone acetylation as a mechanism of polyglutamine toxicity was provided by a study carried out by Hughes and colleagues prior even to recognition that many of the polyglutamine disease proteins are involved in regulating transcription (6). These investigators demonstrated that expression of expanded polyglutamine in the nucleus of *Saccharomyces cerevisiae* disrupted transcription in a pattern very similar to that seen in yeast carrying mutations in the Spt/Ada/Gcn5

acetyltransferase (►**SAGA**) complex that maintains transcriptionally active chromosomal regions by promoting histone acetylation. The authors proposed that expanded polyglutamine in the nucleus resulted in a deficiency in histone acetylation, leading to transcriptional dysregulation. As described above, ataxin-7 is a component of the homologous mammalian histone acetylase complex, STAGA.

The toxicity of expanded polyglutamine in cell culture is ameliorated pharmacologically by ►**HDAC inhibitors**, which restore levels of histone acetylation. Furthermore, polyglutamine-mediated neurodegeneration in *Drosophila* is prevented both by pharmacological treatment with HDAC inhibitors and by up-regulation of the expression of the histone acetyltransferase CBP. The most encouraging news comes from drug trials using transgenic mouse models of polyglutamine disease where pharmacological augmentation of histone acetylation with HDAC inhibitors was found to ameliorate neurodegeneration, reduce motor deficits and extend lifespan (7, 8).

Clinical Relevance

In the decade since the genetic basis of Huntington's and related diseases was determined, tremendous advances have been made in understanding the molecular basis of neurodegeneration. While the details remain far from clear, a promising target for therapeutic intervention has emerged. Independently and in parallel over the past decade, a novel class of drugs has been developed that can augment levels of histone acetylation. These drugs inhibit histone deacetylases and thus promote increased histone acetylation. Phase I and II clinical trials show that the histone deacetylase inhibitors have good safety and tolerability profiles in humans. The efficacy of these drugs in model systems from cell culture to invertebrates to transgenic mice gives reason to be optimistic that the era of targeted therapeutics for the polyglutamine diseases may be at hand. Indeed, a clinical trial of the HDAC inhibitors in Huntington's disease is currently planned in the US.

References

1. Zoghbi HY, Orr HT (2000) Glutamine repeats and neurodegeneration. *Annu Rev Neurosci* 23:217–247
2. Ross CA (1997) Intracellular neuronal inclusions: a common pathogenic mechanism for glutamine-repeat neurodegenerative diseases? *Neuron* 19:1147–1150
3. Zuccato C, Tartari M, Crotti A et al (2003) Huntingtin interacts with REST/NRSF to modulate the transcription of NRSE-controlled neuronal genes. *Nat Genet* 35:76–83
4. La Spada AR, Fu YH, Sopher BL et al (2001) Polyglutamine-expanded ataxin-7 antagonizes CRX function and induces cone-rod dystrophy in a mouse model of SCA7. *Neuron* 31:913–27

5. Nucifora FC Jr, Sasaki M, Peters MF et al (2001) Interference by huntingtin and atrophin-1 with CBP-mediated transcription leading to cellular toxicity. *Science* 291:2423–2428
6. Hughes RE, Lo RS, Davis C et al (2001) Altered transcription in yeast expressing expanded polyglutamine. *Proc Natl Acad Sci USA* 98:13201–13206
7. Hockly E, Richon VM, Woodman B et al (2003) Suberoylanilide hydroxamic acid, a histone deacetylase inhibitor, ameliorates motor deficits in a mouse model of Huntington's disease. *Proc Natl Acad Sci USA* 100:2041–204
8. Ferrante RJ, Kubilus JK, Lee J et al (2003) Histone deacetylase inhibition by sodium butyrate chemotherapy ameliorates the neurodegenerative phenotype in Huntington's disease mice. *J Neurosci* 23:9418–9427

Polymerase

Polymerases comprise a class of enzymes that catalyse the synthesis of nucleic acids on preexisting nucleic acid templates; assembling DNA from deoxyribonucleotides and RNA from ribonucleotides.

►**Nucleotide Excision Repair**

►**Reverse Transcriptase**

Polymerase Chain Reaction

►**PCR**

Polymorphism/Polymorphic

Definition

Polymorphism/polymorphic refers to the state where a gene exists in more than one version (allele), and where the rare allele can be found in more than 1% of the population. (A frequency less than 1% is called a rare variant). A polymorphism does not usually directly cause altered phenotypes, even when present in homozygous form.

- ▶ COPD and Asthma Genetics
- ▶ DNA Chips
- ▶ Epistasis in Cystic Fibrosis
- ▶ Genetic Epidemiology
- ▶ Genetic Predisposition to Multiple Sclerosis
- ▶ Hypothalamic and Pituitary Diseases Genetics
- ▶ Morbus Wegener
- ▶ Repeat Expansion Disease
- ▶ Rheumatoid Arthritis
- ▶ Schizophrenia Genetics

Polynucleotide Kinase

Definition

Polynucleotide kinase refers to an enzyme that transfers a phosphate group from an ATP donor molecule, to a polynucleotide (such as DNA), which lacks such a group at its 5' end.

- ▶ Shotgun Libraries

Polypeptide

Definition

A linear chain of amino acids joined head to tail via a peptide bond between the carboxylic acid group of one amino acid, and the amino group of the next amino acid. A polypeptide can be a protein or a subunit of a protein.

- ▶ Protein Databases
- ▶ Proteomics in Human-Pathogen Interactions

Polyploidy

Definition

Polyploidy is a condition in which cells or organisms contain more than two haploid chromosome sets.

- ▶ Gene Duplications
- ▶ *Xenopus* as a Model Organism for Functional Genomics

Polyprotein

Definition

Polyprotein describes the primary product of translation of the retrovirus gag, pro, pol, and env genes, which are subsequently cleaved to form the active components of the mature virion.

- ▶ Retroviruses

Polyps-and-Spots Syndrome

- ▶ Peutz-Jeghers Syndrome

Polysaccharide

- ▶ Oligosaccharide

Polytene Chromosomes

Definition

Polytene chromosomes that emanate in certain cells in Diptera are long thick chromosomes with alternating dense bands and diffuse interbands. They are produced by over-replication of large parts of the genome within one S phase.

- ▶ DNA Amplification

Population Stratification

Definition

Population stratification (also called population admixture) refers to differences in allele frequencies between cases, and controls due to systematic differences in ancestry rather than association of genes with disease. When there are different ethnic subgroups in an association study group this can result in false

positive or false negative associations. For example, a false positive result could occur if one ethnic group had a higher prevalence of COPD (based on unmeasured environmental factors) and a different frequency distribution of candidate SNP alleles. If the study population contains a mixture of ethnic groups, these alleles would appear to be associated with the disease although they may have no functional effects. It has been proposed that false positive associations due to stratification can be controlled by genotyping a few dozen unlinked genetic markers.

- ▶ COPD and Asthma Genetics
- ▶ Genomic Control

Porphyria Cutanea Tarda

Definition

Porphyria cutanea tarda describes an inherited or acquired chronic porphyria with skin symptoms and uroporphyrinogen decarboxylase deficiency in the liver.

- ▶ Acute Intermittent Porphyria

Porphyrias

Definition

Porphyrias are a group of hereditary disorders that involve defects in the heme metabolism, characterized by excretion of excess porphyrins in the urine and extreme sensitivity to light. The different forms of porphyrias are caused by mutations in one of the different specific enzymes of heme biosynthesis.

- ▶ Acute Intermittent Porphyria

Position Effect

Definition

Position effect (PE) defines a change in the expected level of transgene expression that is induced by nearby DNA elements or epigenetic effects from the surrounding chromatin.

- ▶ Transgene Silencing

Position Effect Variegation

Definition

Position Effect Variegation (PEV) refers to a repression of a transgene in some cells but not others. PEV is caused when the transgene is located near heterochromatin, with cell-to-cell differences in its ability to invade the transgene.

- ▶ Transgene Silencing

Positional Candidate Cloning

Definition

Positional candidate cloning defines a modified version of ▶positional cloning, in which direct screening of known candidate genes in a linked region is used instead of chromosomal walking.

- ▶ Atopy Genetics

Positional Cloning

Definition

Positional cloning is used when the biochemical nature of a disease is unknown. Marker genes not related to disease, physiology and genome-wide screens are the starting point for mapping the genetic components of the disease. The aim is first to identify the genetic region within which a disease-predisposing gene lies, and once this is found to localize the gene and determine its functional and biological role in the disease.

- ▶ Common Diseases Genetics
- ▶ *C. Elegans* as a Model Organism for Functional Genomics
- ▶ Genetic Epidemiology
- ▶ Large-Scale ENU Mutagenesis in Mice
- ▶ Morpholino Oligonucleotides, Functional Genomics by Gene 'Knock-down'
- ▶ Mutagenesis Approaches in the Zebrafish
- ▶ Mutagenesis Approaches in Yeast

Positive Reinforcer

Definition

All psycho-active drugs that are abused are positive reinforcers and can act as a discriminative stimulus.

Positive reinforcement is rewarding. Typical natural reinforcers are e.g. food, water and sex. Drugs of abuse usurp the brain's rewarding system.

► [Addiction, Molecular Biology](#)

Positron

Definition

A positron is the antiparticle of the electron. It is antimatter, has an electric charge of +1, spin of 1/2, and the same mass as an electron. When a positron annihilates with an electron, their mass is converted into energy in the form of two gamma ray photons.

► [Molecular Imaging Mod](#)

► [PET](#)

Positron Emission Tomography

JOERG VAN DEN HOFF

PET Center, Institute of Bioinorganic and Radiopharmaceutical Chemistry, Research Center Rossendorf, Dresden, Germany
j.van_den_hoff@fz-rossendorf.de

Definition

► [PET](#) (short for “► [Positron Emission Tomography](#)”) is a tomographic imaging technique that provides the ability to investigate cellular and molecular processes quantitatively *in vivo*, thus allowing truly functional imaging. Today, PET is firmly established in preclinical and clinical research as well as for clinical applications. The PET technique assesses the three-dimensional distribution of special radioactive substances (radio ► [tracers](#)) *in vivo*. These PET tracers are molecules containing short-lived ► [positron](#) emitting isotopes (► [labels](#)). Frequently utilized labels are ^{11}C , ^{13}N and ^{18}F , with half-lives of 20 m, 10 m, 110 m, respectively. These positron-emitting isotopes can substitute for the corresponding stable isotopes (e.g. ^{12}C , ^{14}N) in relevant biomolecules, leading to a wide class of PET tracers whose *in vivo* behavior is unaltered in comparison to their naturally occurring counterparts.

The label can also replace other elements in the molecular structure – for example the widely used ^{18}F can replace hydrogen in many organic compounds. This approach leads to PET tracers whose chemical properties deviate from systemic substances. In this

way it is possible to create tracers with desirable special properties such as enhanced trapping in the target regions. A notable example of this approach is [^{18}F]-2-fluoro-2-deoxy-D-glucose (^{18}F -FDG), which is currently the most widely used PET tracer for clinical oncological investigations.

Unlike X-ray computed tomography (CT) and magnetic resonance imaging (MRI) the primary signal detected with PET (namely the time dependent regional concentration of the radioactive tracers used) is always a ‘functional’ signal from the living organism, which is closely related to different phenomena (depending on the tracer used) such as blood flow, transport across diffusion barriers (e.g. glucose transport into the cell), substrate metabolism, protein synthesis, enzyme activity, receptor affinity and so forth.

At the same time, the “imaging agent”, i.e. the radioactive tracer, is applied in true tracer amounts (usually at the picomolar level), which guarantees that no effects of the tracer on the system under investigation are present.

Consequently, it is possible to obtain specific information concerning selected molecular processes with the PET imaging technique (‘molecular imaging’).

Description

Introductory Remarks

From its beginning as a novel experimental technique some 20 years ago, PET has evolved into an important tool in clinical practice as well as in basic research. The growing interest in the PET technique has been accompanied by major improvements in hardware and infrastructure – e.g. in tomograph performance (spatial resolution, sensitivity), isotope production, radiochemistry laboratories and computer hardware – as well as by advances in tracer development and quantitative evaluation of the resulting data. Recently, the field has seen the emergence of special brain tomographs (2.5 mm resolution) and dedicated ultra-high resolution animal scanners (1–2 mm resolution). Moreover, combined PET/CT systems have become available which integrate functional and anatomical imaging in one instrument.

In the field of drug development and evaluation PET is generally positioned in pre-phase I clinical trials. Due to the very high sensitivity of radio tracer based measurement techniques, PET investigations can be applied with a dose which is much smaller than the common starting dose for phase I trials (typically by a factor of about 1,000).

With pharmacokinetic measurements the distribution, metabolism and clearance of a potential drug can be determined. Pharmacodynamic measurements can determine the pharmacological effect of a new drug. Hence PET can play an important role in proof-of-concept

studies. Moreover, classical phase I studies (aiming at finding the maximum tolerated dose) are inappropriate for e.g. gene therapy, cell transplantation or anti-angiogenic drugs if they target only focal pathophysiology or exhibit only minimal normal tissue toxicity.

With the availability of ultra-high resolution animal tomographs, the potential of PET to rationalize and economize the earliest phases of drug development has increased considerably.

Mode of Operation

In PET investigations, the radioactive tracer is typically injected intravenously, but other application schemes (intra-arterial, inhalation) are possible – and might be necessary – depending on the properties of the tracer and/or the question addressed by the investigation.

The measurement of the regional tracer distribution in the PET method rests on the detection of the decay of individual atomic nuclei of the radioactive label used. Upon the decay of the radioactive label a positron is emitted from the atomic nucleus that is rapidly stopped in the surrounding tissue (mean ranges in water, 1.1 mm, 1.5 mm and 0.6 mm for ^{11}C , ^{13}N and ^{18}F , respectively). The positron then annihilates with an electron, accompanied by collinear emission of 2 photons, i.e. electromagnetic radiation, in opposing directions (annihilation radiation).

These high energy (511 keV) photons are detected by scintillation detectors arranged in a ring coaxial to the patient's body axis. The coincident detection of two photons (within a time window of a few nanoseconds) defines for each detected pair of photons a so-called line of response along which the primary decay of the radioactive label has occurred. Accepting only these coincident events yields the necessary collimation (i.e. directional information) that is required for subsequent tomographic image reconstruction of the tracer distribution.

To derive quantitative images, corrections for effects such as scattering and attenuation of the annihilation radiation in the traversed tissue have to be included in the data processing.

Inclusion of these corrections ensures that the resulting images directly yield the regional concentration of the applied tracer substance in absolute terms (usually units of Bq/ml, i.e. amount of radioactivity per unit volume, are used).

Data Evaluation Methods

A PET investigation basically yields the three-dimensional spatial distribution of a radiopharmaceutical as a function of time.

The most elementary approach to evaluation of this distribution is visual inspection. In many cases, changes in the regional distribution of the radiopharmaceutical as compared to the normal distribution

can be detected visually and can as such be useful for diagnosis, therapy evaluation or for the assessment of the effect of a drug. However, this approach represents a qualitative assessment of image contrast only, which is a serious drawback. For example, changes in image contrast may be due to a changed uptake in target tissue, reference tissue or both. Furthermore, global alterations of the uptake cannot be analyzed visually. Therefore, visual analysis is often accompanied by simple quantification methods such as regional tracer concentration calculations and relative uptake determination. In practice, these methods are adequate for many investigations, in particular for clinical diagnosis. However, there are also circumstances where the above methods are not adequate or not even applicable because the (time varying) tissue concentration of a radiopharmaceutical is determined by the subtle interplay between (time varying) plasma concentrations on the one hand and a number of parameters used to describe the kinetics of the tracer on the other. Consequently, the relation between the measured distribution and the parameters may be strongly non-linear and sometimes a one-to-one relation between single parameters and the distribution may not even exist. In these cases it is essential to employ more advanced data analysis methods (1).

Today, essentially all approaches to quantitative evaluation of PET investigations are utilizing (directly or implicitly) the concepts of [compartment modeling](#). The central postulate of compartment modeling is the assumption that the process/system under investigation can be decomposed into a (hopefully small) number of functional units, called compartments, which are interpreted as structureless pools containing the tracer in distinct states. This implies especially that, by definition, no concentration gradients are present within the compartments. It is important to realize that a compartment does not need to have an anatomical correlate. Rather, it might describe the tracer in a certain identifiable chemical modification. Taking into account the tracer principle, one usually ends up with pure first order kinetics, which corresponds to a mathematical description by first order differential equations. In the majority of cases, this description is adequate. However, in the special case of carrier-added experiments it may no longer hold.

The transport can be the physical transport from one physical space to another, e.g. from plasma to intracellular fluid or a biochemical transformation within a single physical space, e.g. the phosphorylation of FDG to FDG-6-PO₄ in the cell. Although each of these steps may be composed of many intermediate processes, only the rate-limiting step is of importance. For example, a tracer may be transported first from plasma to extracellular fluid, then into the intracellular fluid and finally into the nucleus or the mitochondria.

Nevertheless this can be modeled as a single step process provided that these subsystems approach steady state sufficiently rapid.

Summary

The PET method offers a set of advantageous properties, which are, in this combination, not provided by other means:

- Very high sensitivity with respect to applied amount of tracer substance, which is typically in the nanomolar or picomolar range. In comparison to MRI the sensitivity is higher by more than a factor of 10^6 .
- Simultaneous imaging of the whole field of view (for clinically used systems, typically 60 cm (transaxial) \times 15 cm (axial)) with nearly isotropic resolution in all three dimensions. The spatial resolution of the method is about 4–5 mm in all three dimensions in currently used clinical tomographs (1–2 mm with state of the art small animal systems).
- Dynamic, i.e. time dependent, imaging of the complete imaged volume with a time resolution down to a few seconds in favorable cases.
- Truly quantitative measurement of tracer accumulation in terms of regional concentration with good spatial resolution.
- Ability to quantify the observed phenomena (e.g. blood flow, metabolic turnover) in absolute terms.

Clinical Applications

Today PET is firmly established as a valuable diagnostic tool in oncology, neurology, and cardiology (2, 3, 4). The impact on other disciplines is less pronounced. PET investigations yield important information complementary to the more anatomically oriented tomographic techniques of X-ray computed tomography (CT) and magnetic resonance imaging (MRI).

In terms of application frequency, ^{18}F -FDG is today by far the most important PET tracer in clinical applications.

Oncology

The major clinical impact of PET in the field of oncology is due to the highly successful utilization of ^{18}F -FDG for assessment of tumor viability. The basis of the success of this approach is the fact that the majority of tumor entities exhibit a massively elevated energy metabolism accompanied by a corresponding increase in glucose consumption. FDG-PET directly assesses this parameter.

Successful applications can be noted especially in lung cancer, colorectal carcinoma, lymphoma, melanoma, head and neck cancer and certain brain tumors.

The near future will probably see further increased contributions of PET to problems of tumor therapy, e.g. biologically targeted radiation therapy for individually optimized treatment plans.

Neurology

Historically, neurological (and cardiological) applications of the PET method were the first that became relevant for clinical applications.

In neurology, the major clinical impact of PET is in dementia diagnosis (especially Alzheimer's disease), the search for epileptic foci, brain tumors (see above) and Parkinson's disease.

Cardiology

Here, the relevant clinical applications are quantitative assessment of myocardial viability and perfusion prior to therapeutic decisions.

Therapeutic Consequences

The utilization of PET has proven (or highly probable) benefits for a wide variety of indications (2, 3, 4). The nature and extent of these benefits are dependent on the stage of the disease already reached when the PET investigation is performed. In short the possible benefits might be summarized as follows:

- avoidance of unnecessary therapies. Examples are assessment of myocardial vitality with PET, which can help to decide the question whether to perform revascularisation, and avoidance of surgery for primary bronchial carcinoma in the presence of distant metastases detected with PET.
- early assessment of chemotherapy effectiveness.
- early assessment of surgery and radiation therapy effectiveness in tumor diseases. PET provides a means to decide whether a transition to systemic therapy (e.g. chemotherapy) is necessary.
- cancer of unknown primary. PET is frequently able to detect hitherto unknown primary tumors thus opening the way to local therapy approaches.

References

1. Willemsen ATM, van den Hoff J (2002) Fundamentals of Quantitative PET Data Analysis. *Curr Pharm Des* 8:1513–1526
2. Kuwert T, Bartenstein P, Grunwald F et al (1998) [Clinical value of positron emission tomography in neuromedicine. Position paper on results of an interdisciplinary consensus conference]. *Nervenarzt* 69:1045–1060
3. Reske S, Kotzerke J (2001) FDG-PET for clinical use. *Eur J Nucl Med* 28:1707–1723
4. Schwaiger M, Egert S, Brosius F (1996) [Scintigraphic evaluation of the ischemically damaged myocardium in patients with reduced left ventricular function. From clinical aspects to the cell]. *Z Kardiol* 85:453–468

Posterior Prevalence

Definition

Posterior prevalence designates the hierarchy of regulatory effects, whereby more posteriorly expressed

- ▶ [Hox](#) genes ultimately determine the specification of the segment in which the Hox genes are expressed.
- ▶ [Homeodomain Transcription Factors](#)

Postmeiotic Segregation

- ▶ [PMS](#)

Post-Mitotic

Definition

Post-mitotic refers to those cells that have left the cell cycle and no longer divide. Examples of post-mitotic tissues are muscle and brain.

- ▶ [Mitochondrial Myopathies](#)

Postnatal

Definition

The term postnatal refers to the time span after birth/delivery.

- ▶ [Bone and Cartilage](#)

Postsynaptic Density

Definition

The postsynaptic density is a highly condensed protein complex, close to the postsynaptic membrane, comprising of receptor proteins like ion channels and

G-proteins, structural (e.g. neurofilaments, PSD95 etc.) and regulatory proteins such as calmodulin, phosphokinases, phosphatase and phospholipases.

- ▶ [Neurons](#)

Post-Translational Modification

Definition

Post-translational modification is the chemical modification of a protein into the biologically active form following release from the ribosome. This includes removal and/or derivatization of specific amino acid residues, proteolytic cleavage, loss of signal sequences, and formation of disulfide cross-links. Other common post-translational modifications are phosphorylation, glycosylation, acetylation, hydroxylation, myristoylation, isoprenylation, and many other reactions, such as spontaneous oxidation or deamidation. All amino acid side chains, except those of alanine, glycine, isoleucine, leucine, and valine, can be modified. Posttranslational modifications may be causative for differences in the contents of a protein crystal or crystallization drop as compared to the original protein, especially over time.

- ▶ [3D Structure Determination by X-Ray](#)
- ▶ [Enzyme Catalyzed Post-translational Hydroxylation of Proteins](#)
- ▶ [Glycosylation of Proteins](#)
- ▶ [Mass Spectrometry: MS/MS](#)
- ▶ [Protein Databases](#)
- ▶ [Protein Microarrays as Tool for Protein Profiling and Identification of Novel Protein-Protein Interactions](#)
- ▶ [Recombinant Protein Production in Mammalian Cell Culture](#)
- ▶ [Recombinant Protein Expression in Yeast](#)

POU Transcription Factors

Definition

POU transcription factors comprise of a family of transcription factors that are responsible for mammalian development. POU is an acronym for Pit-1 that regulates expression of certain pituitary genes, Oct-1 (octamer factor) is widely expressed, Oct-2 is expressed in B lymphocytes and in certain areas of the brain, and Unc-86 protein functions in neuronal development in some species.

- ▶ [Hypothalamic and Pituitary Diseases Genetics](#)

PP2A

►Protein Phosphatase-2A

PPAR

►Peroxisome Proliferator-Activated Receptor

PPAR

Definition

PPAR protein is a member of the ►nuclear hormone receptor family. PPARs are ligand-activated transcription factors that increase transcription of target genes, by binding to a specific nucleotide sequence in the gene's promoter. Three different PPAR isotypes can be distinguished in vertebrates: PPAR α (NR1C1), PPAR β/δ , (NR1C2) and PPAR γ (NR1C3). Each of them is encoded in a separate gene and binds fatty acids and eicosanoids. It has been established that PPAR nuclear factors are key regulators of metabolism and energy homeostasis. They also play an important role in various signaling pathways (immunity, inflammation, apoptosis and cell differentiation). Evidence indicates that PPARs can also affect the pathogenesis and development of tumors.

PPAR α γ -Consensus Elements

Definition

PPAR α γ -consensus elements are transcription factor binding elements for the peroxisome proliferator-activated receptors (►PPAR) α and γ , respectively.

►Tangier Disease

PR

►Progesterone Receptor

PR (Protease)

Definition

PR (protease) refers to the retrovirus enzyme (product of the *pro* gene) that is responsible for cleaving the Gag, Pro, and Pol precursors to their final form, and altering the virion from an immature to a mature, infectious, state.

►Retroviruses

Prader-Willi and Angelman Syndromes

BERNHARD HORSTHEMKE

Institute for Human Genetics, University Clinic Essen, Essen, Germany

b.horsthemke@uni-essen.de

Definition

The Prader-Willi syndrome (PWS) and the Angelman syndrome (AS) are two distinct neurogenetic disorders that are caused by the loss of function of genes located within the chromosomal region 15q11-q13 (1). The two syndromes serve as a paradigm for diseases involving imprinted genes. ►Imprinting refers to an ►epigenetic process in which the male and the female germlines confer a specific mark onto certain chromosomal regions so that only the paternal or the maternal copy of a gene is active in somatic cells (2). The nature of the parental marks is unknown, but methylation of cytosine residues within the promoter region of genes as well as ►histone modifications play a major role in this process. Whereas PWS is caused by the loss of function of one or more paternally expressed genes, AS is caused by the loss of function of a maternally expressed gene.

Characteristics

PWS is characterized by neonatal ►muscular hypotonia, ►hypogonadism, ►hyperphagia and ►obesity, short stature, small hands and feet, ►sleep apnea, behavioural problems and mild to moderate mental retardation. As shown in Table 1, it is caused by the loss of genetic material from the paternal chromosome 15 [deletion, del(15)(q11q13)pat], the maternal origin of both chromosomes 15 [maternal ►uniparental disomy, upd(15)mat] or the presence of a maternal imprint on both chromosomes [imprinting defect]. All three

lesions lead to the lack of expression of imprinted genes that are active on the paternal chromosome only. At present, several such genes are known: *MKRN3*, *MAGEL2*, *NDN*, *SNURF-SNRPN* and more than seventy C/D box snoRNA genes (Fig. 1). The latter genes are located within introns of the *SNURF-SNRPN* transcription unit and encode small nucleolar RNAs (3). Unlike other snoRNAs, these snoRNAs lack any telltale complementarity to ribosomal RNA and may be involved in the modification of mRNAs. There is tentative evidence that the snoRNA genes and *NDN* play a role in PWS (see below).

AS is characterized by ►microcephalus, ►ataxia, absence of speech, abnormal ►EEG pattern, severe mental retardation and frequent laughing. Similar to PWS, the major lesions in AS are a deletion, uniparental disomy and an imprinting defect, but these lesions affect the maternal chromosome. Therefore it was concluded that AS involves a maternally expressed gene. Based on the finding of point mutations in patients with AS who do not have one of the above-mentioned lesions, *UBE3A* has been identified as the gene affected in AS (4, 5). Interestingly, *UBE3A* is imprinted only in the brain.

Cellular and Molecular Regulation
Chromosomal Mechanisms Leading to PWS and AS

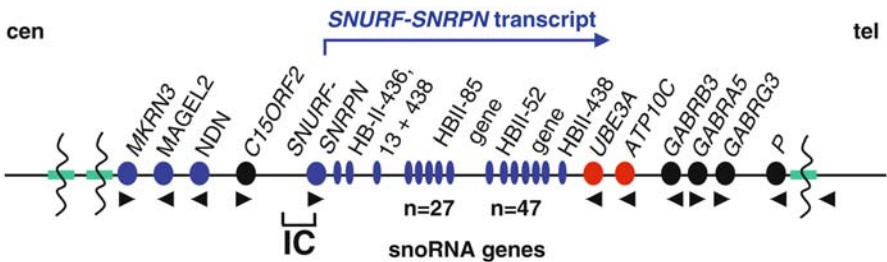
Deletions in PWS and AS comprise approximately 4 million base pairs and include most of the imprinted

domain plus several non-imprinted genes (Fig. 1). They do not differ in size between PWS and AS, but only in the parental origin. Most often, the deletions arise as a consequence of non-homologous crossovers between duplicated DNA sequences that flank this region. In a few patients, the deletion is the result of an unbalanced translocation. Because the PWS genes are silent on the maternal chromosome, a paternally derived deletion leads to a complete loss of function of these genes. Likewise, a maternally derived deletion leads to a complete loss of function of *UBE3A* in brain cells, because in this tissue *UBE3A* is silent on the paternal chromosome

Uniparental disomy arises in most cases from the mitotic correction of a meiotic error (6). During ►meiosis, the diploid set of chromosomes ($n = 46$) is reduced to a haploid set ($n = 23$). ►Non-disjunction of the homologous chromosomes 15 or the two sister chromatids during female meiosis results in an oocyte with two chromosomes 15 or no chromosome 15. In these cases, fertilization by a sperm with one chromosome 15 will lead to a zygote that is trisomic or monosomic for chromosome 15. These conditions are not compatible with normal development, but can be “rescued” by loss of one chromosome 15 from a trisomic cell or duplication of the paternal chromosome 15 in a monosomic cell. In two thirds of cases, one of the two maternal chromosomes will be lost from the trisomic cell. This will result in a normal set of

Prader-Willi and Angelman Syndromes. Table 1 Estimated frequency of genetic lesions and recurrence risk in PWS and AS

	del(15)(q11q13)	upd(15)	imprinting defect	gene mutation	other lesions
PWS	70% (pat)	28% (mat)	1%	unknown	0.1%
AS	70% (mat)	1% (pat)	4%	5%	20%
recurrence risk	<1%	<1%	0–50%	0–50%	unknown



Prader-Willi and Angelman Syndromes. Figure 1 Physical map of 15q11-q13. Blue circle, paternally expressed gene; blue ellipse; snoRNA gene; red circle, maternally expressed gene; black circle, biallelically expressed gene; green rectangle, duplicated DNA sequence; wavy line, deletion breakpoint; arrowhead, direction of transcription. cen; centromere; tel, telomere. The *SNURF-SNRPN* transcript is indicated by a blue line and serves as a host for the snoRNAs. For clarity, not all snoRNA genes are shown. (not drawn to scale).

chromosomes. If, however, the paternal chromosome is lost, the cell is left with two maternal chromosomes 15 [upd(15)mat]. Duplication of the paternal chromosome 15 in a monosomic cell will lead to upd(15)pat. Most often, a trisomic cell line is found only in the placenta (confined placental **▶mosaicism**), because only very few **▶blastomeres** contribute to the embryo. As the genes involved in PWS are silent on the maternal chromosome, upd(15)mat is associated with a complete loss of function of these genes. Likewise, upd(15)pat is associated with a complete loss of function of the *UBE3A* gene in brain.

Imprinting Control and Imprinting Defects in 15q11-q13

Imprinting in 15q11-13 is controlled by an imprinting centre (IC), which is located in the middle of the imprinted gene cluster and overlaps the *SNRPN* gene (7). It appears to consist of two elements. One IC element is required for the maintenance of the paternal imprint during early embryogenesis. A deletion of this element can be transmitted silently through the female germline, but leads to an incorrect, maternal imprint on the paternal chromosome when inherited from a male. A child with this chromosome will develop PWS, because the maternal imprint silences the PWS genes. The second element is required for maternal imprinting in the female germline. A deletion of this element can be transmitted silently through the male germline, but leads to an incorrect, paternal imprint on the maternal chromosome after transmission through the female germline. A child inheriting this chromosome will develop AS, because the paternal imprint silences the *UBE3A* gene. Imprinting defects can also occur without any DNA mutation. In fact, such epimutations are much more frequent than IC deletions. There is tentative evidence suggesting that assisted reproduction may increase the risk of imprinting defects.

Genes and Pathways Affected in PWS and AS

UBE3A encodes a ubiquitin-protein ligase that transfers ubiquitin to substrate proteins. Ubiquitin is a highly conserved 76 amino acid peptide that targets proteins for degradation. Although there is some evidence of a role for ubiquitin-mediated proteolysis during central nervous development, the pathogenesis of AS remains unclear.

In contrast to AS, the gene defect underlying PWS is less clear. As there is no patient with a point mutation in one of the paternally expressed genes, it is generally assumed that PWS results from the loss of function of two or more of these genes. Knock-out experiments in the mouse suggest that a deficiency for *NDN*, which is thought to play a role in cell cycle arrest in terminally differentiated neurons, may contribute to the respiratory abnormalities in individuals with Prader-Willi

syndrome through a suppression of the central respiratory drive (8).

In addition to *NDN*, one or more snoRNA genes may play a role in the pathogenesis in PWS. This hypothesis is based on the analysis of rare patients with a balanced translocation as well as the study of mouse models. Although the HBII-52 snoRNAs have an 18-nt phylogenetically conserved complementarity to a critical segment of serotonin 2C receptor mRNA and this receptor is known to play a role in the regulation of food intake, it is unclear whether this mRNA is the snoRNA target.

Clinical Relevance

The clinical diagnosis can easily be checked by probing differentially methylated sites for the presence of a methylated and an unmethylated allele, but further molecular studies are required to distinguish between the different lesions, or to search for *UBE3A* mutations in AS. In particular, the methylation test is an important diagnostic tool to identify patients with PWS among newborns with severe muscular hypotonia and among children with obesity and mental retardation.

PWS and AS are a paradigm for diseases involving imprinted genes. Other diseases in this category are the **▶Beckwith-Wiedemann syndrome**, **▶Silver-Russell syndrome** and transient diabetes mellitus. Each of these diseases is rare, but it is being suspected that epigenetic defects may also contribute to more frequent genetic disorders including complex diseases.

References

1. Nicholls RD, Knepper JL (2001) Genome organization, function, and imprinting in Prader-Willi and Angelman syndromes. *Annu Rev Genomics Hum Genet* 2:153–175
2. Reik W and Walter J (2001) Genomic imprinting: parental influence on the genome. *Nat Rev Genet* 2:21–32
3. Cavaille J, Buiting K, Kieffmann M et al (2000) Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc Natl Acad Sci USA* 97:14311–14316
4. Kishino T, Lalonde M, Wagstaff J (1997) *UBE3A/E6-AP* mutations cause Angelman syndrome. *Nat Genet* 15:70–73
5. Matsuura T, Sutcliffe JS, Fang P et al (1997) De novo truncating mutations in *E6-AP* ubiquitin-protein ligase gene (*UBE3A*) in Angelman syndrome. *Nat Genet* 15:74–77
6. Robinson WP (2000) Mechanisms leading to uniparental disomy and their clinical consequences. *Bioessays* 22:452–459
7. Buiting K, Saitoh S, Groß S et al (1995) Inherited microdeletions in the Angelman and Prader-Willi syndromes define an imprinting centre on human chromosome 15. *Nat Genet* 9:395–400
8. Ren J, Lee S, Pagliardini S et al (2003) Absence of *Ndn*, encoding the Prader-Willi syndrome-deleted gene *necdin*, results in congenital deficiency of central respiratory drive in neonatal mice. *J Neurosci* 23:1569–1573

Prader-Willi Syndrome

Definition

PWS is a congenital disorder (1:15000 newborns) that is caused by functional loss of imprinted genes on the maternal chromosome 15q11-q13 by different mechanisms (70% deletion of paternal 15q11-q13, 28% maternal uniparental disomy, and <2% imprinting center mutation on paternal allele). Signs include extreme hypotonia of the newborn, hypothalamic dysfunction, excessive eating from around age two years, uncontrolled weight gain and extreme obesity, hypogenitalism, cardiomyopathy and often death at 30–40 years. Mental retardation is variable and mild. The outcome can be improved by strict diet and growth hormone therapy.

► [Microdeletion Syndromes](#)

► [Prader-Willi and Angelman Syndromes](#)

when a clear distinction into stem or progenitor cells is not possible.

► [Neural Stem Cells](#)

Predictive Genetic Testing

► [Predictive Genetic Testing](#)

Predictive Testing and Genetic Counseling

ANNEMARIE SCHWAN, JÖRG T. EPPLEN

Human Genetics, Ruhr-University Bochum, Bochum, Germany

joerg.t.epplen@rub.de

Synonyms

Presymptomatic testing

Definition

Predictive diagnostics concerns the investigation of healthy persons for genetic changes that predispose to diseases later in life. There are different forms of predictive testing. Testing is used in

1. screening of newborns,
2. PND / PGD (► [prenatal diagnostics](#)/► [preimplantation genetic diagnostics](#)),
3. ► [multifactorial diseases](#),
4. hereditary cancer and
5. late-onset monogenic disorders.

Thus, diagnostics of hereditary diseases is not limited to genetic investigations *sensu stricto*, but sometimes it is also based on the results of different clinical examinations [e.g. diagnosis of ► [familial adenomatous polyposis](#) (FAP) after colonoscopy]. Nevertheless, genetic diagnostics is more comprehensive, allows virtually certain diagnostic statements for monogenic disorders and in addition differential diagnoses and subclassification of some disease entities.

Growing knowledge with respect to the pathophysiology and genetics of different disorders has led to new investigative possibilities, i.e. comparatively simple DNA tests. Besides diagnostics by applying genetic

Prechondrogenic

Definition

Prechondrogenic is a generic term describing the precursor status of a cell, which has the potential to become a chondroblast/chondrocyte.

► [Bone and Cartilage](#)

Precipitant

Definition

Different methods can be used to lower protein solubility and reach a supersaturated state. Most commonly, salts and solvents and polymers that compete with protein for water are used as precipitants.

► [3D Structure Determination by X-Ray](#)

Precursor Cell

Definition

Precursor cell is an umbrella term encompassing both ► [stem cells](#) and progenitor cells. It is normally used

tests for the diseased, the possibility has arisen of testing healthy persons for their risk of developing symptoms in later life. Genetic tests are offered to persons at risk for a genetic disease, who wish to determine whether or not they have inherited a mutated gene. Predictive genetic tests and the information about the respective results are invariably linked to (preceding) genetic counseling in an appropriate setting performed by specially educated experts.

From the perspective of the client, a principal differentiation can be made according to the following aspects in the field of predictive testing:

1. Are there therapeutic options or even effective treatment available after demonstration of a mutation in the genetic test? Do preventive therapeutic options (e.g. operations, nutritional changes) arise on account of a mutation in the predictive test?
2. Are any predictions possible concerning the clinical consequences?

Characteristics

Predictive diagnostics relate to the possibility of recognizing the predisposition to a disease before any clinical symptoms have occurred. In addition, the likelihood of the appearance of the disease can potentially be specified (with statistical precision). These diagnostics can potentially be performed even prenatally—and then have special difficulties, which must be considered and responded to during counseling. As indicated above, predictive diagnostics should take place exclusively after genetic counseling, except for the special cases of newborn screening.

After personal contact has been made initially and all general aspects are clarified, the patient should be informed about results of tests in a later session of counseling. The first counseling session preceding the initiation of predictive testing aims to discuss the counselee's self-determination and, therefore, it may influence and/or strengthen the autonomous decision for or against a genetic test. In order to allow completely self-determined action by the client, it is necessary to be able to judge the outcome of a gene test as well as the possibilities and risks and especially the individual consequences. Genetic counseling provides the basis for autonomous decisions by supplying pertinent information in an understandable form to the counselee (and his family). Thus, human genetic counseling comprises initially the pathophysiological basics of the disease and the individual genetic risk of the person. Furthermore, the diagnostic possibilities are explained and the importance of the test result is weighed. All in all, emphasis is put on the individual life situation and help is offered for consequent action (1). There are different forms of predictive testing that

should encompass different leads or strategies of counseling.

Clinical Relevance Newborn Screening

Newborn screening for inherited metabolic diseases is in a wider sense also predictive testing of persons within the scope of screening an entire population. Newborn screening aims at the earliest possible recognition of disorders so that intervention with effective treatment can prevent the most serious consequences of the disorder. The investigation is not primarily of a molecular genetic nature. Nevertheless, a pathological result leads to the diagnosis of a (clinically yet unapparent) genetic disease. Evidently, the consent to the investigation cannot be given by the person to be examined, but by his or her parents. There is a generally approved consensus that these investigations are only carried out if the disease in question—whenever untreated—causes severe health defects. At the same time the disease must be recognizable *via* a biochemical test before its outbreak and should be preventable by therapeutic measures. A pertinent example is ►**phenylketonuria**. A test result demonstrating the metabolic defect has effects on the parents and/or additional members of the family and their subsequent child planning. Therefore, additional counseling is needed concerning the specifics of the disease as well as further implications. At present, newborn screening programs have been implemented in nearly 30 countries for different diseases.

Prenatal Diagnostics (PND) and Preimplantation Genetic Diagnostics (PGD)

Obviously, predictive genetic diagnostics can already be carried out 'for the unborn child'. The intensity of genetic counseling before PND is dependent among other things on the progress of the pregnancy, the disease in question, the kind of investigation and/or the results to be expected or possible and finally the specific problems of the counselees and prospective parents. The contents and the progress of the counseling concerning a monogenic disease of a fetus with a high recurrence risk differs from chromosome diagnostics on account of the advanced age of the mother. These differences arise due to the variable impact of the results, the likelihood for a positive result and the preexisting general experience with the disease. Besides invasive PND, PGD is available in certain countries for a limited spectrum of genetic disorders. For all these predictive tests the approval is not given by the person to be examined herself or himself (embryo, fetus), but essentially by the mother or the parents instead. In contrast to newborn screening, no cures are at hand for the diseases in question, at least in the foreseeable future.

Predictive Diagnostics with Regard to Multifactorial Diseases

Multifactorial diseases are considered to be due to complex pathogeneses influenced by several to many genes and additionally by unknown environmental factors. Predisposing factors are known for several ►common diseases (either disease-favoring or protecting genetic variations). Therefore, nowadays the investigation of predisposing alleles within the scope of multifactorial diseases could already be offered for several diseases (e.g. morbus Crohn etc.). The magnitude of the relevance and the difficult interpretation of predisposing alleles in the context of further factors requires definitive statements about the limitations and most often rather 'soft' statements of predictive testing in cases of a multifactorial disease. These relations are generally quite complicated and need suitable mediation and interpretation of the results. In the future more knowledge about the predisposing factors will become available, particularly concerning diseases with a high frequency in the population (e.g. hypercholesterinemia, high blood pressure, coronary heart disease, but also morbus Alzheimer and parkinsonism). General doubts have been raised as to what extent this knowledge is helpful for an individual. Given basic scientific advances and the development of the population structure especially in western countries in the near future, the demand for such diagnostics may increase substantially.

Predictive Diagnostics for Cancer

A Mendelian disposition is presumed for ~5–15% of all human cancer diseases. Some of the responsible genes are already defined (e.g. *BRCAI*, *-II* for breast cancer; *MLH1*, *MSH2*, *-6* for colon cancer). During counseling it is initially clarified whether pedigree information and clinical findings are compatible with hereditary cancer in the family or if sporadic disease or a multifactorial disposition to cancer is more likely to be present in this family. Two specific features are notable during genetic counseling of persons at risk for cancer, which are involved in the test and the consequences. Relatives of cancer patients can be counseled while diseased relatives (index patients) must be involved obligatorily in testing. The necessity of the genetic investigation of sick relatives, so that for persons at risk a genetic testing can be offered, implies further family conflicts. On the other hand preventive examinations are possible against specific types of malignancies. Frequent and intensified clinical screening procedures (mammography, colonoscopy etc.) are recommended for persons with high risk. If a genetic change is known as the cause of the familial cancer disease, after predictive testing, classification into non-risk and (depending on the type of cancer) high-risk persons is possible and usually certain. High-risk

patients should undergo intensified cancer screening; non-risk patients only need the usual precautionary investigations. Carriers of predisposing mutations to cancer have, depending on the cancer type, different risks for disease manifestation. The cancer risk is dependent on the type of malignancy and the specific mutation. Female carriers of a *BRCA II* mutation have for example a cumulative risk of ovarian cancer of 27% by the age of 70 years. The cumulative risk for FAP is estimated to be around 93% before the age of 50 years. Definitive knowledge about patients bearing a hereditary risk factor or not conveys many potential problems for the tested person and also for the members of the family (2). Because of these aspects, several family members may need detailed counseling during which the medical facts and the specific psychic load factors are identified, discussed and taken care of appropriately.

Predictive Diagnostics of Late Onset Diseases

Predictive genetic diagnostics of hereditary late onset diseases, as for example ►chorea Huntington (HD), Huntington disease, lead to special problems (3). Today's knowledge about the cause of this genetic disease exceeds the present-day potential for therapy by far. In this and related fields, the gap in knowledge will certainly enlarge during the next few years. This discrepancy is an important factor that must be considered in predictive tests. Nevertheless, human genetic investigations also increase knowledge of the genetic background and they are essential for future causal therapy. The expertise from the genetic counseling of persons at risk for HD can be transferred to other late onset diseases.

Until 1993, 10 years after the genomic region bearing the HD mutation had been described, only indirect diagnostics were feasible. At least one affected family member was required, as well as as many as possible other direct relatives. Hence the phase of the transmission could be followed for the portion of chromosome 4 supposedly carrying the then unknown mutation. Since then direct tests can be performed with a reliability of >99% for the individual without the cooperation of any other family member. Nevertheless, knowledge about genetic testing of a person at risk is still at times not pertinent exclusively to the single person, but it permits (risk or carrier) statements about other members of the family. Before predictive diagnostics are performed, intense genetic counseling is mandatory, including different aspects. The complexity of the subject and the psychic load of the persons at risk usually require several counseling appointments. Together with the patient, the available theoretical knowledge is summed up for this devastating disease. The individually preexisting knowledge is based on the contact with the family member in question (*cave* e.g. genetic

heterogeneity etc.). During counseling, knowledge already available in the family is deepened and complemented by the counselor. In addition, the counselor refers to the disease symptoms, onset and course, possible treatments, the biological bases, inheritance and individual disease risk. Furthermore, genetic testing will be explained in detail with respect to the meaning of the results, i.e. their potential impact for the counselee and the family. The interpretation of negative and positive results has to be discussed beforehand. It must be differentiated clearly, whether a normal test result means a safe exclusion of later disease or not. The importance of genetic testing within the scope of personal and professional development and also in the family context is discussed in detail with the patient. The counseling should also be expanded to topics aside from the medical and psychological, such as social security and professional consequences.

A main focus of the counseling is the discussion of the psychic load that arises from definitive knowledge about the personal risk. Persons with high risk who are informed recently about their own disease risk often express the wish to receive a definitive result immediately. Many persons tend to ignore and/or deny the potentially arising load. Counseling serves, therefore, also for internalizing the conflict potentials that can arise from genetic testing. The conflicts can be of a personal, individual kind or they can encompass lifelong companions, siblings, children and parents. The result of a genetic test can also lead to social conflicts. The argument usually mentioned against a test is the concern about receiving a test result with demonstration of a mutation in the absence of therapy reversing or influencing progression of the disease. The most widespread argument for a test is the removal of unspecified uncertainty and help with decisions in respect to family planning.

After an initial counseling session or—whenever necessary several additional meetings—ample time to reflect and for conversations with members of the family is necessary for the persons at risk. The minimum time to reflect is normally one month. Expert psychological care is provided, too. After this time to reflect, the patient can get in touch with the counselor on his own initiative in order to furnish a blood sample. After at least another month, the test results may be disclosed to the clients, and they are supported thereafter. Either additional counseling sessions or at least telephone contact will be maintained for as long as required by the counselee.

Support measures after disclosure of the test results differ markedly, depending upon the needs of the individual counseled person. The psychological processes of coping with the DNA test result takes at least several (weeks to) months according to the experience of the many counselees who have passed through the

entire protocol. Exclusion of carrier-status does not always lead to immediate relief from suffering. Some non-carriers do not experience the expected relief about their test results at all. They may suffer from feelings like ‘survivor’s guilt’ and find it difficult to cope with the effects of the test results on the family system. Irrational debt feelings, e.g. regarding siblings carrying the disease-gene, are sometimes not simply dealt with by non-carriers. Another problem can develop if a person at risk was sure of being a carrier and had arranged their past life on this basis. Then the exclusion of a carrier-status can take their entire life planning into question and lead to temporary disorientation, before the new situation can be accepted and lived through in a fulfilled manner. With proof of a carrier status, partly displaced or shifted fears become an indelible certainty. These possible difficulties should be discussed with the persons at risk. Ideally genetic counseling should contribute to the fact that the tested person does not repent of having had the test—independently of the outcome of the test.

► **Hereditary Spastic Paraplegias**

References

1. Ad Hoc Committee on Genetic Counseling, American Society of Human Genetics. (1975) Genetic Counseling. *Am J Hum Genet* 27:240–242
2. Salkovskis PM, Rimes KA (1997) Predictive genetic testing: psychological factors. *J Psychosomatic Research* 43:477–487
3. Meiser B, Dunn S (2001) Psychological effect of genetic testing for Huntington’s disease: an update of the literature. *West J Med* 2001174:336–340

Preimplantation Genetic Diagnosis

Definition

Preimplantation genetic diagnosis describes the procedure following *in vitro* fertilization. One or two cells are removed from the developing embryo at 8 cell stage after fertilization and tested for the presence of a known mutation (single cell diagnostic). Embryos without the mutation are implanted into the uterine cavity to initiate a normal pregnancy. However, the molecular heterogeneity of many monogenic diseases requires a diagnostic strategy capable of detecting a range of mutations and compound genotypes. At present the procedure is not legal in all countries.

► **Heritable Skin Disorders**

► **Predictive Testing and Genetic Counseling**

Preinitiation Complex

Definition

Preinitiation complex is the complex of transcription factors and RNA polymerase that forms at a gene promoter DNA prior to the initiation of transcription.

- ▶ RNA Polymerase I
- ▶ Transcription Factors and Regulation of Gene Expression

Preintegration Complex

Definition

Preintegration complex (PIC) refers to a poorly understood structure, derived from the viral capsid that contains the viral DNA, integrase and probably other proteins. It is defined by its ability to directly correct integration of the viral DNA into added target DNA.

- ▶ Retroviruses

Premature Termination Codon

Definition

Premature termination codon (PTC) designates a termination (stop) codon that is located upstream of the normal termination codon.

- ▶ Repetitive DNA
- ▶ RNA Stability

Premenstrual Manifestation

Definition

Premenstrual manifestation is a special type of manifestation of acute intermittent porphyria in women with cyclic attacks (of abdominal and nerve pain) prior to their menses.

- ▶ Acute Intermittent Porphyria

pre-miRNA

Definition

Pre-miRNA is an intermediate precursor of microRNA, typically 60~70 nt in length, which is generated during animal miRNA biogenesis.

- ▶ MicroRNA

pre-mRNA

Definition

Pre-mRNA is the primary RNA transcript from genomic DNA before it is processed and altered into mature forms of mRNA by pre-mRNA splicing.

- ▶ Alternative Splicing
- ▶ Morpholino Oligonucleotides, Functional Genomics by Gene 'Knock-down'

Premutation

Definition

Premutation refers to the fragile-X mental retardation 1 (FMR1) gene of fragile X mental retardation syndrome. Carriers of premutation alleles (who have a trinucleotide repeat blood containing between 59 and 200 CCGs) of the FMR1 gene are often regarded as being clinically uninvolved, and do not typically have symptoms.

- ▶ Fragile X Syndrome

Prenatal Diagnostics

Definition

Prenatal diagnostics comprise all diagnostic procedures that are undertaken during pregnancy in order to investigate the health (and disease) of the growing fetus. Either non-invasive (ultrasound) or invasive strategies (e.g. chorionic villous biopsy or amniocentesis) may be pursued, depending on the needs of the pregnant woman.

- ▶ Predictive Testing and Genetic Counseling

Prenyl Group

Definition

A molecular fragment made up of isoprene units (see ► [Isoprenoid](#)).

► [Protein Prenylation](#)

Prenylation

Definition

Prenylation describes the posttranslational enzymatic modification of a cytosolic protein. A hydrophobic prenyl group (geranyl or farnesyl) is enzymatically attached to the C-terminus, and leads to rapid cellular redistribution and membrane translocation.

► [Peutz-Jeghers Syndrome](#)

► [Protein Prenylation](#)

Preproorexin

Definition

Preproorexin is the common precursor of orexin-A and orexin-B (► [Orexins](#)).

► [Narcolepsy](#)

Presomitic Mesoderm

Definition

Presomitic mesoderm characterises located unsegmented paraxial mesoderm, which is posterior to the most newly formed somite that generates somites segmentation.

► [Somitogenesis](#)

Presymptomatic Diagnosis

Definition

Presymptomatic diagnosis is based on results of predictive testing in persons who are at risk of

developing a specific heritable late-onset disorder, before manifestation of clinical symptoms.

► [Heritable Skin Disorders](#)

► [Predictive Genetic Testing](#)

Prevalence

Definition

Prevalence refers to the presence of a disease in a population at a certain time point, and is an epidemiological measure to determine the morbidity in a population.

► [Crohn Disease](#)

Prey

Definition

Prey designates a fusion protein that is composed of a transcriptional activation domain and a protein of interest X.

► [Two-Hybrid System](#)

PRF

► [Protein Research Foundation](#)

P-Rib-PP

Definition

P-Rib-PP stands for 5-Phosphoribosyl-1-pyrophosphate. It is an important precursor for de novo and salvage synthesis of purine and pyrimidine nucleotides.

► [Nucleotide Biosynthesis](#)

Primary Aldosteronism

Definition

Primary aldosteronism causes severe hypertension in humans (Conn's syndrome) with cardiac hypertrophy. It is characterized by high plasma aldosterone, and low renin-hypertension, with a variable degree of ►[hypo-kalemia](#) and ►[metabolic alkalosis](#).
►[Mendelian Forms of Human Hypertension and Mechanisms of Disease](#)

Primary Hyperparathyroidism

Definition

Primary hyperparathyroidism is a disorder of the parathyroid glands, in which one or more of the parathyroid glands are enlarged due to neoplasms (hyperplastic), are overactive, and secrete too much parathyroid hormone. Too much ►[parathyroid hormone](#) results in abnormally high levels of calcium in the blood (►[hypercalcemia](#)).
►[Hyper- and Hypoparathyroidism](#)

Primary Body Axis

Definition

Primary body axis defines the cartesian axis of the body plan, i.e. dorsal-ventral and anterior-posterior axis.
►[Drosophila Model of Cardiac Disease](#)

Primary Hypertension

►[Essential Hypertension](#)

Primary Cells

Definition

Primary cells are cells that are directly derived from normal embryonic or adult tissue, which are propagated in culture. These cells are considered to be genetically identical to cells in the tissue of origin.
►[Senescence](#)

Primary Structure

Definition

Primary structure refers to the amino acid sequence of a polypeptide chain. Of the four levels of protein structure, this is the most basic one. The primary structure determines the folding (tertiary structure) and ultimately the function of the proteins.
►[Classification of Active Centers](#)
►[Protein Databases](#)

Primary Constriction

Definition

Primary constriction defines a prominent cleft or constriction in metaphase chromosomes denoting the location of the centromere.
►[Centromeres](#)

Primitive Streak

Definition

Primitive streak designates a longitudinal cleft that is formed on the surface of the amniote early embryo by a convergence of cells. At the onset of gastrulation, epiblast cells migrate towards and into the streak, and in so doing acquire mesodermal cell fate.
►[Somitogenesis](#)

Primordial Germinal Cells

Definition

Primordial germinal cells (PGC) are proliferating diploid cells that are engaged into the germ line pathway, and migrate to the primitive gonad of the fetus to form germ cells.

► [Genomic Imprinting](#)

Principal Cell

Definition

Principal cells are the fundamental cells of an organ, which usually have a specific function. In the kidney, principal cells are a specific epithelial cell type of the collecting tubules responsible for water and sodium reabsorption.

► [Diabetes Insipidus, a Water Homeostasis Disease](#)

PRINTS

Definition

PRINTS is a database of protein fingerprints, consisting of multiple patterns to characterize an entire protein sequence. <http://bioinf.man.ac.uk/dbbrowser/PRINTS/>

► [Protein Databases](#)

Prion

Definition

Prions are the infectious agent of prion diseases, e.g. Creutzfeldt Jakob disease and BSE. The term is derived from “proteinaceous infectious particle”. According to the prion hypothesis, prions consist mainly or exclusively of a protein PrP^{Sc} and are free of nucleic acids.

► [Prion Diseases](#)

Prion Diseases

► [Protein Interaction Analysis: Chemical Cross-Linking](#)

Prion Diseases

HANS A. KRETZSCHMAR

Institute for Neuropathology, Ludwig Maximilian University, Munich, Germany

hans.kretzschmar@inp.med.uni-muenchen.de

Synonyms

Transmissible spongiform encephalopathies (► [TSE](#))

Definition

► [Prion](#) diseases are rapidly progressing, invariably fatal, neurodegenerative diseases associated with dementia and neurological deficits such as ataxia, visual disturbances and myoclonus. Histologically, nerve cell loss, spongiform change and various forms of prion protein deposits are found in the brain. They are a heterogeneous group of diseases that can be acquired, hereditary or idiopathic (sporadic). All prion diseases are experimentally transmissible with a relatively long incubation time and a comparatively short clinical duration. Prion diseases in humans include Creutzfeldt-Jakob disease, kuru and familial prion diseases (Table 1); prion diseases in animals are ► [scrapie](#) in sheep, ► [BSE](#) in cattle, chronic wasting disease in deer and others (Table 2).

Characteristics

Human prion diseases were shown to be transmissible after intracerebral inoculation of brain tissue from kuru and ► [CJD](#) patients into chimpanzees in the 1960s. The nature of the infectious agent of prion diseases has been the subject of numerous studies for many years. The assumption, which seemed natural in the early 1970s, that the agent must be a slow virus or a virino was challenged by the failure to detect viral nucleic acids and by the resistance of the agent to radiation, nucleases and other reagents that damage nucleic acids. In contrast, infectivity was closely associated with a protein, ► [PrP^{Sc}](#), and if any nucleic acid is to be associated with the infectious agent it cannot contain more than 50 nucleotides. The term “prion” was

Prion Diseases. Table 1 Prion diseases in humans

Disease Type	Disease
Idiopathic	Sporadic Creutzfeldt-Jakob disease (sCJD)
	Sporadic fatal insomnia (SFI)
Acquired	iatrogenic CJD (iCJD)
	(new) variant CJD (vCJD) Kuru
Genetic	Familial CJD (fCJD)
	Gerstmann-Sträussler-Scheinker syndrome (GSS)
	Fatal Familial Insomnia (FFI)

Prion Diseases. Table 2 Prion diseases in animals

Disease	Affected species
Scrapie	sheep, goat
Transmissible mink encephalopathy (TME)	mink
Chronic wasting disease (CWD)	Mule deer and elk (North America)
Bovine spongiform encephalopathy (BSE)	cattle
Feline spongiform encephalopathy (FSE)	cat (singular in puma and cheetah)
Exotic ungulate encephalopathy	kudu, nyala

proposed by Stanley Prusiner to distinguish the infectious pathogen from viruses and viroids. Albeit not formally proven, the prion hypothesis is supported by many lines of evidence.

Prions were originally defined as small, proteinaceous infectious particles that resist inactivation by procedures that modify nucleic acids, and they have been redefined as a proteinaceous particle that lacks nucleic acid (2). The change in conformation of a normal protein encoded by the host genome, the cellular isoform of the prion protein (PrP^{C}), into an altered isoform, the scrapie isoform of the prion protein (PrP^{Sc}), is the core of this hypothesis. The term PrP^{Sc} is used for the isoform of the prion protein that is closely associated with infectivity and that is part and parcel of the prion. Whether PrP^{Sc} is the only necessary constituent of prions is at present unknown. According to the prion hypothesis, the infectious agent, the prion, would require PrP^{C} molecules for its propagation, and organisms devoid of PrP^{C} should not be susceptible to

prion diseases. Indeed, this has been shown experimentally by the resistance of PrP gene knockout mice ($\text{Prnp}^{-/-}$ mice) to scrapie.

Cellular and Molecular Regulation

Human PrP^{C} is a glycoprotein of 253 amino acids before cellular processing. There is an 85–90% homology to prion proteins of other mammalian species. PrP^{Sc} is a membrane protein expressed mainly in neurons, but also in astrocytes and a number of other cells. It has an N-terminal signal sequence of 22 amino acids, which is cleaved off the translation product. Twenty-three terminal amino acids are removed when glycosylphosphatidylinositol (GPI) is attached to serine residue 230. Mature PrP^{C} is attached to the cell surface by this GPI anchor and undergoes endocytosis and recycling. It seems, however, that PrP may exist in alternative membrane topologies (ctmPrP and ntmPrP), whose implications for the function of PrP^{C} and its role in pathogenesis are just beginning to be elucidated. There are two N-glycosylation sites that are glycosylated differently in different human CJD variants. The N-terminal moiety of the protein contains an octapeptide repeat, (PHGGGWGQ) $\times 4$, which has been suggested to function in copper binding.

Whereas PrP^{C} is found on the surface of many cell types in all mammals and birds studied so far, PrP^{Sc} is generated from PrP^{C} in a posttranslational process and is closely associated with infectivity. In terms of the prion hypothesis, it is part and parcel of the infectious agent, the prion. NMR structural studies have shown that the C-terminal half of PrP^{C} contains a two-stranded antiparallel β -sheet (S1 and S2) and three α -helices, whereas the N-terminal moiety is thought to have no definite structure in aqueous solution. PrP^{C} and PrP^{Sc} seem to differ mainly in their folded structures. PrP^{Sc} purified from hamster brain consisted of 42% α -helical and only 3% β -sheet structure, whereas PrP^{Sc} purified from scrapie-infected hamster brain is composed of 30% α -helix and 43% β -sheet. PrP^{Sc} shows increased protease resistance, is insoluble in aqueous solution and tends to form fibrils that show birefringence after binding of Congo red.

The human PrP gene (*PRNP*) has a simple genomic structure and consists of two exons and a single intron 13 kb in length. The entire protein-coding region is located in exon 2. In families with inherited prion diseases, a large number of different point mutations and insertion mutations have been described in the open reading frame of *PRNP*. The insertional mutations are situated in the N-terminal half of the protein in an octapeptide repeat region, whereas the point mutations cluster in the central and C-terminal regions of the protein.

The common polymorphism at amino acid position 129 of the prion protein, where humans carry a methionine

(M) or valine (V), clearly influences susceptibility to the sporadic and iatrogenic types of prion diseases (Fig. 1) and, furthermore, determines in part the phenotype of the sporadic as well as of some inherited prion diseases. Studies have revealed a marked overrepresentation of homozygotes (mainly for methionine) at this position in cases of sporadic CJD compared to the normal population. CJD homozygotes at codon 129 also show a higher susceptibility to iatrogenic CJD and a shorter incubation time as well. There is also a strong correlation of codon 129 genotype and clinicopathological phenotype.

The Conversion Process of PrP^C to PrP^{Sc}

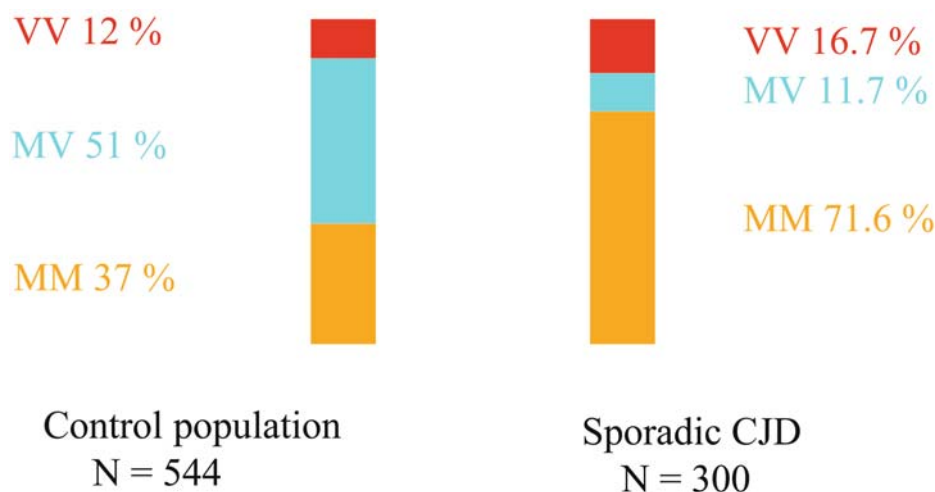
Conversion of PrP^C to PrP^{Sc} seems to be a late posttranslational process, which in scrapie-infected cells occurs after PrP^C has reached its normal location on the cell surface or even later during ►endocytosis. Why this is such a rare event and how PrP^{Sc} triggers the conversion of PrP^C are not understood. In the nucleation model, PrP^C and PrP^{Sc} are in equilibrium. PrP^{Sc} is stable only when it adds onto a seed or PrP^{Sc} aggregate, a process that has been compared to crystal formation. The spontaneous formation of an initial PrP^{Sc} aggregate (seed) would be a very rare event. Once the seed has been formed, the ensuing addition of PrP monomers could follow at a fast rate. In contrast, the refolding model holds that PrP^C is unfolded and that the conversion process consists of a refolding of the molecule under the influence of a PrP^{Sc} molecule. In this process, a high activation energy barrier must be overcome and chaperones and an energy source may be required. The two hypotheses, which have also been named the Lansbury and the Prusiner mechanisms, are by no means mutually exclusive.

Clinical Relevance

Prion Diseases in Animals

Scrapie, a disease that naturally occurs in sheep and goats, has been known for more than 200 years and was the first prion disease to be shown to be infectious in 1936 by two French scientists, Cuillé and Chelle. Affected sheep clinically show abnormal behavior, such as excessive scraping as well as trembling and ataxia and other motor disturbances. There are no known hereditary prion diseases in animals; however, allelic variations in the ovine PrP^C sequence exert a strong influence on susceptibility to natural and experimental scrapie. Although the disease has been known for centuries and its infectious nature was recognized more than 60 years ago, the exact mode of natural transmission is not known. Maternal transmission seems to be an established fact, but there are also reports of transmission by keeping sheep on pastures that were previously occupied by scrapie-infected flocks. Scrapie is experimentally transmissible to many mammalian species. Epidemiological studies show that scrapie is not apparently transmissible to humans by the oral route.

BSE is a disease of cattle that was first described in Great Britain in 1986. It is estimated that since then almost 1 million animals have been infected. The mean incubation time is about 5 years; therefore, most animals did not manifest disease because they were slaughtered at age 2 or 3 years. Nonetheless, more than 160,000 affected animals were diagnosed and killed over the years. It is now clear that BSE was spread by feeding cattle with contaminated meat and bone meal (MBM) prepared from the offal of sheep, cattle and pigs. However, the origin of BSE is still obscure. It is now thought that changes in the hydrocarbon extraction



Prion Diseases. Figure 1 Distribution of the genotypes at codon 129 of the human prion protein gene (*PRNP*) in sporadic CJD and the normal population (controls).

method, including temperature changes, that were made in the rendering of offal in the late 1970s allowed the infectious agent to survive the manufacturing process and pass from sheep to cattle. Because by strain typing and PrP^{Sc} banding pattern in Western blots all scrapie strains tested so far are different from BSE, an alternative hypothesis, i.e., transmission of pre-existing natural BSE at low incidence as a consequence of the preceding changes in the rendering process, cannot be dismissed. BSE is experimentally transmissible to many species and seems to have passed the species barrier to humans.

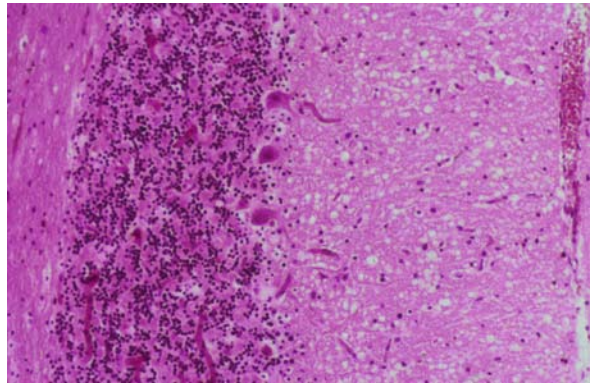
In addition, a number of other prion diseases exist, some of which are of unknown origin such as chronic wasting disease in mule deer and elk, and others are transmitted by contaminated offal from sheep or cattle such as transmissible mink encephalopathy, feline spongiform encephalopathy and exotic ungulate encephalopathy.

Prion Diseases in Humans

CJD and Gerstmann-Straussler-Scheinker syndrome (GSS) were first described as neurodegenerative diseases in the 1920s, whereas kuru was first reported in 1957. The exceptional nature of GSS as a hereditary disease that is experimentally transmissible to laboratory animals was discovered in 1981. In 1989, the first mutation of the prion protein gene was identified in a GSS family. Neuropathology has played a lead role in defining the various entities that are now known as human prion diseases and is still of particular importance in routine diagnosis. The classical neuropathological changes consist of (1) spongiform degeneration (Fig. 2), (2) PrP^{Sc} deposition, (3) neuronal loss, and (4) astrocytic gliosis. PrP^{Sc} deposition is found as diffuse delicate accumulations in the gray matter in areas of high synaptic density and in a distribution similar to that of synaptophysin. This type of PrP^{Sc} deposition has therefore been called synaptic. Other forms are described as pericellular, perivacuolar, in kuru plaques that are visible in routine H&E stains and plaque-like, i.e. in small plaques that are only visible after immunohistochemical staining of PrP^{Sc}.

Kuru

Kuru was first described as a deadly neurodegenerative disease affecting the Fore people in the eastern highland of New Guinea. This disease was mainly characterized by ataxia and predominantly affected women and children. William Headlow, a veterinary pathologist, noted the neuropathological similarities of scrapie and kuru and suggested that infection experiments be performed in apes. Gajdusek and Gibbs were successful in transmitting the disease to chimpanzees in 1966. Later, ritualistic cannibalism of deceased clan members was identified as the mode of transmission in



Prion Diseases. Figure 2 Spongiform change (small vacuoles in the neuropil) in the molecular layer of the cerebellum in a Creutzfeldt-Jakob disease case. H&E stain.

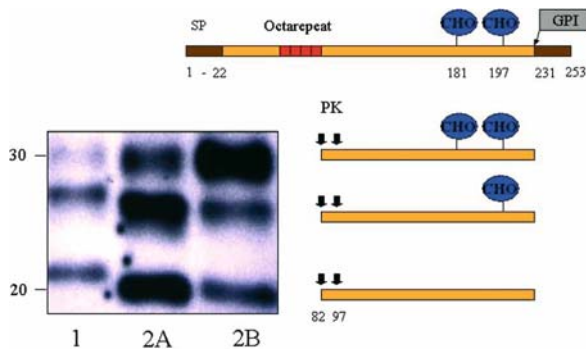
the Fore population. After cannibalistic practices ceased, the disease disappeared. Kuru incubation times of more than three decades were reported.

Sporadic (Idiopathic) Creutzfeldt-Jakob Disease

Sporadic CJD (sCJD) most often affects patients in their 60s. It usually presents with dementia and various neurological signs such as myoclonus, ataxia, pyramidal and extrapyramidal signs, EEG changes (periodic sharp wave complexes) and the 14-3-3 protein in the CSF. It runs a relentless course, leading to death usually within 6 months. A definite diagnosis can only be made by neuropathological or biochemical examination of the brain.

Parchi and co-workers described two different molecular types of PrP^{Sc} associated with distinct clinical and pathological phenotypes in sCJD (1) (Fig. 3). The two types of PrP^{Sc} are distinguished by their different physicochemical properties, particularly their appearance on Western transfers after digestion with proteinase K. The full spectrum of sCJD variants was defined in a large series of 300 CJD patients and was compared with molecular and pathological phenotypes. Six distinct types of sCJD were described by molecular analysis. Almost 90% of sCJD patients were homozygous at codon 129, the vast majority being MM homozygotes, whereas only half of the normal population in Europe and the United States is homozygous. PrP^{Sc} types 1 and 2 were found to be associated with all *PRNP* genotypes. However, there was a strong association between PrP^{Sc} type 1 and MM homozygosity as well as between PrP^{Sc} type 2 and VV or MV patients.

Sporadic CJD is often thought to arise spontaneously. This would be compatible with the encountered worldwide incidence of 1 case per million per year and would be an attractive model in terms of the prion



Prion Diseases. Figure 3 Western blot analysis of human prion proteins. Shown are the primary structure of PrP^C and the protease-resistant core of PrP^{Sc} as well as their appearance on Western blots. The insert on the left side shows an original Western blot with the three PrP^{Sc} bands after pK digestion of two sporadic CJD cases (PrP^{Sc} type 1 with the unglycosylated PrP^{Sc} at 21 kD, PrP^{Sc} type 2 with the unglycosylated PrP^{Sc} at 19 kD) and an vCJD case in which the three PrP bands correspond in size to PrP^{Sc} type 2 but show a different glycosylation pattern.

hypothesis. However, the sporadic origin is by no means proven.

Hereditary Prion Diseases

About 10–15% of all prion disease cases are hereditary. A large number of different point mutations and insertion mutations of the prion protein gene have been identified in familial prion diseases. Insertion mutations represent additional repeats of the N-terminal Cu²⁺-binding octapeptide. Depending on their clinical and neuropathological characteristics, familial prion diseases are designated as familial CJD, Gersmann-Sträussler-Scheinker syndrome (GSS), or fatal familial insomnia (FFI).

Familial CJD often is indistinguishable from sporadic CJD and is most often associated with the E200K mutation and less often with D178N-129V, V180I, R208H, V210I, M232R and insertion mutations. GSS is thought to be distinguishable from CJD by the predominance of ataxia, whereas dementia and myoclonus are more prominent in CJD. Neuropathologically, GSS is quite distinct and is characterized by large, multicentric PrP-containing amyloid plaques; spongiform change is variable. It is found most often in families with the P102L mutation and less often with P105L, A117V, F198S, D202N, Q212P, Q217R and insertion mutations. Fatal familial insomnia (FFI) often presents with insomnia and dysautonomia and later shows signs of ataxia, dysarthria, myoclonus and pyramidal tract dysfunction. Neuropathologically, FFI is characterized by neuronal loss and astrocytic gliosis preferentially affecting the ventral anterior and medial

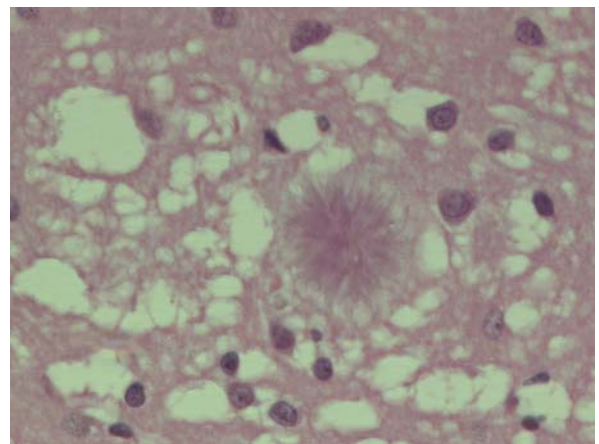
dorsal nuclei of the thalamus and the inferior olive. FFI is caused by a D178N mutation of *PRNP* associated with a methionine codon at position 129 of the same allele.

Iatrogenic CJD

The first possible person-to-person transmission of CJD was reported in a recipient of a corneal transplant from a donor with CJD in 1974. Other modes of accidental iatrogenic transmission were reported in the following years, including the use contaminated EEG depth electrodes, neurosurgical instruments, cadaveric pituitary-derived gonadotropin and human growth hormone (hGH) and dura mater grafts. Transmission by contaminated hGH has raised major concern because of the large number of possibly affected persons. Worldwide more than 120 hGH-associated cases have been reported to date, whereas 8000 persons received hGH between 1963 and 1985 in the United States alone. The affected patients commonly present with gait abnormalities and ataxia, whereas dementia is a late manifestation and is mild. The incubation time in these cases has been estimated to be 12 years or longer.

New Variant CJD and the Transmission of BSE to Humans

A new variant of CJD was described in 10 patients in the United Kingdom in 1996 (3). These patients had a mean age of 29 years and presented with psychiatric disturbances, whereas signs more typical of CJD developed later in the course of disease. Neuropathology at autopsy was exceptional, showing extensive depositions of PrP^{Sc} in various areas of the brain in a fashion that had only been described in hereditary disease before 1996. In addition, there were florid plaques (Fig. 4) with a central PrP accumulation and



Prion Diseases. Figure 4 A typical florid plaque consisting of delicate strands and surrounded by confluent vacuoles in a case of nvCJD. H&E stain (by courtesy of Dr. James Ironside, Edinburgh, Scotland).

surrounding vacuoles that had not been seen in human prion disease before. From 1996 to date, 137 cases (October 2003) of this variant were identified in the United Kingdom, 6 were found in France and 1 in Italy. Information on the clinical appearance has been compiled, and a definition for suspect ►vCJD cases is now available. All vCJD patients to date have been methionine homozygotes at codon 129 of the prion protein gene; the youngest patient was 15 years old and the oldest was 73 years old at death. It is not possible at present to define the clinical and pathological features to be expected in codon 129 valine homozygotes or in heterozygotes that are likely to be observed in the future. vCJD has shown significantly new features in the pattern of extracerebral deposition of PrP^{Sc} in the tonsils, lymph nodes, spleen and appendix. This has raised concern that blood cells may also harbour the infectious agent and that the disease might be spread by blood transfusions. Although this suspicion is unsubstantiated to date, measures have been taken to prevent the spread by blood transfusion in various countries.

It is hypothesized that vCJD cases have been caused by the consumption of food or other products containing large amounts of the BSE agent (BSE prions). This hypothesis is strengthened by epidemiological and experimental findings. The appearance of vCJD 10 years after BSE in the country with the highest incidence of BSE is highly consistent with this hypothesis. The banding pattern of PrP^{Sc} in vCJD has been shown to be different and resembles the pattern in BSE. Although BSE and vCJD show significant differences pathologically, upon transmission to genetically homogeneous animals (inbred mice) they elicit practically identical patterns, whereas in these strain typing experiments all tested scrapie strains and sporadic CJD cases were different. These findings have been confirmed in transgenic animals expressing bovine PrP.

If we assume that vCJD is caused by transmission of the BSE agent to humans, the available epidemiological data are insufficient to predict the number of cases that must be expected to occur in the future.

References

1. Parchi P, Giese A, Capellari S et al (1999) Classification of sporadic Creutzfeldt-Jakob disease based on molecular and phenotypic analysis of 300 subjects. *Ann Neurol* 46:224–233
2. Prusiner SB (1998) Prions. *Proc Natl Acad Sci USA* 95:13363–13383
3. Will RG, Ironside JW, Zeidler M et al (1996) A new variant of Creutzfeldt-Jakob disease in the UK. *Lancet* 347:921–925

Pristanic Acid

Definition

Pristanic acid is a branched-chain fatty acid (2,6,10,14-tetramethylpentadecanoic acid) that is present in dietary components, notably meat. Pristanic acid is elevated in certain peroxisomal disorders.

►Peroxisomal Disorders

PRL

►Prolactin

Pro

Definition

Pro is the retrovirus gene that encodes the virion protease, PR.

►Retroviruses

Proband

Definition

A proband is the family member through whom a family's medical history comes to light. The proband may also be called the index case.

►COPD and Asthma Genetics

►Familial Dilated Cardiomyopathy

Processed Pseudogene

Definition

Processed pseudogene (Retropseudogene) refers to DNA sequences that are very similar to known genes.

Processed pseudogenes lack introns, contain a poly-A tail, and are framed by short direct repeats, suggesting that an mRNA was copied into DNA that was then integrated into the genome. The vast majority of processed pseudogenes are non-functional.

- ▶ [Repetitive DNA](#)
- ▶ [Transposons](#)

Processivity

Definition

Processivity describes the number of nucleotides that a DNA polymerase incorporates per binding to DNA. A processive DNA polymerase can synthesize several thousand nucleotides per binding event.

- ▶ [Replication Fork](#)

ProcheckNMR

Definition

ProcheckNMR refers to a computer program available through the Protein Data Bank (PDB) that checks the validity of a protein structure determined with NMR. The program varies features, such as the consistency with the allowed regions of the Ramachandran Plot.

- ▶ [3D Structure by NMR](#)

ProDom

Definition

ProDom is a comprehensive set of protein domain families automatically generated from the SWISS-PROT and TrEMBL sequence databases.

- (▶ <http://prodes.toulouse.inra.fr/prodom/current/html/home.php>)
- ▶ [Protein Databases](#)

Profilin

Definition

Profilins are small globular, cytoplasmic actin binding proteins that are expressed in all eukaryotic cells. They are primarily involved in G-actin sequestering and prevention of polymerisation. They interact with protein containing polyproline stretches. The binding to G-actin is negatively regulated by PIP2 at the plasma membrane.

- ▶ [Actin Cytoskeleton](#)

Progenitor Cell

Definition

Progenitor cells are the progeny of ▶ [stem cells](#). They are characterized by a limited ability ▶ [for selfrenewal](#) and a lineage-restriction in their developmental potential.

- ▶ [Neural Stem Cells](#)

Progesterone Receptor

Definition

Progesterone receptor (PR) is the transcription factor of the nuclear receptor superfamily which is expressed under the control of the ▶ [estrogen receptor](#) (ER) α , and activated by progestins as its ligands.

- ▶ [Breast Cancer](#)

Prognosis

Definition

Prognosis is the forecast as to the outcome of a specific disease, usually expressed as the probability of 5 year survival in the case of a cancer. This forecast is done following internationally accepted guidelines based on staging, i.e. volume of the tumor and depth of invasion, and grading, i.e. proliferation and dedifferentiation.

- ▶ [Cancer Invasion and Metastasis: Cellular, Molecular and Clinical Aspects](#)

Programmed +1/-1 Frameshifting

Definition

An event occurring in mRNAs at which the ribosome, without disrupting continued protein elongation, shifts from decoding in the frame in which it initiated (the zero frame), into the open reading frame shifted 1 nucleotide in the downstream (3') [+1] and upstream (5') [-1] direction, respectively.

- ▶ Genetic Code
- ▶ Translational Frameshifting, Non-standard Reading of the Genetic Code

Programmed Cell Death

- ▶ Apoptosis

Programmed Translational Frameshifting

- ▶ Translational Frameshifting, Non-standard Reading of the Genetic Code

Progressive External Ophthalmoplegia

Definition

Progressive external ophthalmoplegia is a partial or complete paralysis of the muscles that move the eye.

- ▶ Mitochondrial Myopathies

Progressive Supranuclear Palsy

Definition

Progressive Supranuclear Palsy (PSP) is a sign or symptom (cause) of parkinsonism, which is not

responsive to dopamine replacement therapy, and which causes eye movement abnormalities, classically an upward gaze palsy.

- ▶ Parkinson's Disease: Insights from Genetic Causes

Projections

Definition

Projections, in the context of microscopic imaging, refer to two-dimensional images generated by adding the information of a three-dimensional object along the optical axis of the microscope.

- ▶ Cryo-Electron Microscopy: Single-Particle Reconstruction

Prolactin

Definition

Prolactin is a peptide hormone secreted by the anterior pituitary. Its primary function is to initiate and maintain lactation.

- ▶ Hypothalamic and Pituitary Diseases Genetics

Proline Isomerase

- ▶ Peptidyl Prolyl cis/trans Isomerases

Prolyl Isomerase

- ▶ Peptidyl Prolyl cis/trans Isomerases

Prometaphase

Definition

Prometaphase defines a stage of mitosis when the nuclear envelope is disrupted, and chromosomes start

to associate with spindle microtubules and begin movements that will lead to their alignment at the cell equator.

- ▶ Cell Division
- ▶ Centromeres
- ▶ Mitotic Recombination

Promoter

Definition

Eukaryotic gene transcription is driven by specific promoter sequences. The promoter is the regulatory site on DNA to which the enzyme RNA polymerase can bind, initiate the transcription of DNA into RNA, and control cell-type specific expression of the gene. A typical eukaryotic promoter contains the site of transcription initiation (Inr), an A+T rich sequence motif ("TATA box") located some 25 bp upstream of Inr, and one or more binding sites for DNA binding transcription factors, located at -50 to -100 bp from the site of initiation. Another type is the G+C-rich so-called CpG island that usually does not contain a TATA box and contains factor binding sites both upstream and downstream of the (multiple, scattered) start sites of transcription.

- ▶ Catalytic RNA
- ▶ Core Promoters
- ▶ DNA-based Vaccination
- ▶ Enhancer
- ▶ *In Vivo* Imaging of Transgenic Mice with Fluorescent Protein Expression
- ▶ Mouse Genomics
- ▶ NF- κ B Pathway
- ▶ Recombinant Protein Production in Mammalian Cell Culture
- ▶ RNA Polymerase I Transcription
- ▶ RNA Polymerase II Transcription
- ▶ RNA Polymerase III
- ▶ Splicing
- ▶ Transcription Factors and Regulation of Gene Expression
- ▶ Two-Hybrid System

Promoter Clearance

Definition

Promoter clearance designates a regulatory step, which ensures that the RNA polymerase, along with a complement of general transcription factors, have

traversed 50 to 60 nucleotides after initiation before it is fully engaged in elongation.

- ▶ Core Promoters

Proneural Transcription Factors

Definition

Proneural Transcription Factors refer to genes that specify clusters of potential neuroblasts in the *Drosophila* ectoderm and were originally called "proneural genes". These genes encode transcription factors of the ▶ [basic-helix-loop-helix](#) (bHLH) family. In vertebrates, the term proneural refers to the function of bHLH factors in inducing neurogenesis.

- ▶ Neural Development

Pronuclear (Micro)Injection

Definition

Pronuclear injection is the injection of recombinant DNA (or other substance) into the pronucleus of fertilized zygotes (at embryonic development day 1 in the mouse). It is a standard procedure to generate transgenic mice. Microinjected DNA integrates randomly into the genome of the host animal. Integration sites and numbers of integrated copies of the transgene are normally unpredictable. Thus, every resulting founder animal may have a unique phenotype and needs to be tested separately for transgene expression.

- ▶ Mouse Genomics
- ▶ Transgenic and Knockout Animals

Proof Reading Mechanism

Definition

If a tRNA is mischarged, the wrong amino acid can be removed by hydrolysis, giving the tRNA a second chance for correct aminoacylation. This system prevents the incorporation of the incorrect amino acid and thereby increases the fidelity of translation.

- ▶ tRNA

Proofreading Exonuclease

Definition

Proofreading exonuclease is an enzyme that recognizes and cleaves a mismatched primer terminal during DNA replication.

► Replication Fork

Prophase

Definition

Prophase defines the initial stage of mitosis, when chromatin condenses and appears as individual chromosomes.

► Cell Division
► Centromeres
► Mitotic Recombination

Prophylactic Immunization

Definition

A prophylactic vaccine is given to patients before exposure to the pathogen.

► DNA-based Vaccination

Prophylaxis

Definition

Prophylaxis refers to a treatment given for the prevention of a disease; for example, people with bleeding disorders are regularly given blood clotting factors to prevent bleeding episodes.

► Hereditary Hemostatic Defects and Recombinant Proteins for Treatment

PROSITE

Definition

A database that currently describes over 1,300 protein families and domains. Most of these families or domains are detected using one or more sequence patterns. PROSITE uses a single consensus pattern to

annotate each family of sequences. <http://us.expasy.org/prosite/>

► Protein Databases

Prostaglandins

Definition

Prostaglandins are cyclooxygenase metabolites of arachidonic acid.

► Nuclear Hormone Receptors: Regulators of Gene Transcription and Cellular Signaling

Prosthetic Groups

Definition

Some functionality of proteins cannot be realised by the standard (proteinogenic) amino acid residue itself. In such cases, they can reversibly bind co-enzymes or bind strongly (often covalently) to chemical compounds that enable or support the function, and are thus referred to as prosthetic groups.

► Classification of Active Centers

Protease

Definition

Protease is an enzyme that degrades proteins by breaking peptide bonds.

► Cell Polarity

Protease Inhibitors

► Protease and Inhibitors

Proteases and Inhibitors

DORA CAVALLO-MEDVED, BONNIE F. SLOANE
Department of Pharmacology, Wayne State University
and Barbara Ann Karmanos Cancer Institute, Detroit,
MI, USA
bsloane@med.wayne.edu

Definition

Proteases, also termed peptidases, are defined as ►enzymes that break down proteins by catalyzing the ►hydrolysis of ►peptide bonds. Peptidases consist of two groups of enzymes: 1) endopeptidases that cleave peptide bonds within the protein and 2) exopeptidases that remove amino acids (►Amino Acids: Physico-chemical Properties) sequentially from either the N or C-terminus. Endopeptidases are also termed proteinases and are divided into five mechanistic classes (1), which are characterized below.

Proteolysis is ubiquitous in all cells and elementary in numerous biological processes such as cell growth and death, blood clotting, matrix remodeling and immune defense. In addition, proteases are involved in the pathogenicity of several diseases including cancer, arthritis, HIV, and Alzheimer's and as such these enzymes are important targets for drug design. Proteolytic activity can be divided into two different categories: 1) limited proteolysis, in which a protease cleaves a limited number of peptide bonds of a target protein usually leading to the activation or maturation of the formerly inactive protein and 2) unlimited proteolysis, in which a protease degrades a protein into its amino acid constituents. Regulation of protease activity is important in maintaining homeostasis and is accomplished by inhibitors, compartmentalization of the enzyme and activation cascades.

Characteristics

Proteases are classified by: 1) their site of action and 2) the nature of their active site residues and reaction mechanisms. A complete list of proteases can be found in the MEROPS database (2) and in a review by Puente and coworkers (3). Here we will focus in general on the five major classes of endopeptidases: serine, cysteine, aspartic, metallo and threonine proteases.

Serine Proteases

Serine proteases are among the earliest studied enzymes and are involved in a number of physiological processes including digestion, blood clotting and proteolytic activation cascades (1). Traditionally, serine proteases have been divided into two distinct families: 1) the chymotrypsin family, which includes mammalian enzymes such as chymotrypsin, trypsin and elastase, and 2) the subtilisin family, which includes the bacterial enzyme subtilisin. Recently three additional families have been identified: 1) the wheat serine carboxypeptidase II family, 2) the human cytomegalovirus serine proteinase family and 3) the proteolytic component of ClpP of the ATP-dependent protease family. Although the general 3D structures of the five families are different, they do contain the same active site geometry, which includes an active serine residue,

and their catalytic process proceeds *via* the same mechanism.

Cysteine Proteases

Cysteine proteases are categorized into two superfamilies: 1) the papain family of cysteine proteases and 2) the family of enzymes related to interleukin 1 β converting enzyme (ICE) (4). Although both the papain and the ICE superfamilies contain a cysteine residue in their active sites, these two superfamilies share no sequence homology. The papain superfamily consists of plant proteases such as papain, which is the archetype and the most studied member of this family, bromelain, mammalian lysosomal cathepsins and the cytosolic calpains. The involvement of cysteine proteases in biological processes is observed in apoptosis by caspases and ICE, in tumorigenesis by cathepsin B and in membrane signaling by calpains.

Aspartic Proteases

There are two families of aspartic proteases: 1) the pepsin family that includes digestive enzymes such as pepsin and chymosin, lysosomal cathepsin D and the processing enzyme renin and 2) viral aspartic proteases including the HIV protease (1). The crystal structure of aspartic proteases shows that these enzymes are bilobed molecules with the active site, which contains two aspartic acid residues, located between two homologous lobes. Aspartic proteases are involved in a number of biological processes. Pepsin hydrolyses acid-denatured proteins in the stomach, renin has a hypertensive action *via* the renin-angiotensin system, HIV protease is essential for viral maturation and cathepsin D is involved in tumorigenesis.

Metalloproteinases

Metalloproteinases are classified into four families: 1) matrix metalloproteinases (MMP) (5); 2) disintegrin and metalloproteinases (ADAM); 3) angiotensin-converting enzymes (ACE) and 4) neprilysins (NEP) (1). Although each family of enzymes differs widely in their sequences and structure most contain a catalytically active zinc atom. In some cases, cobalt or nickel replaces the zinc atom without loss of activity. MMPs are the best-studied metalloproteinases. Most MMPs have a common domain structure that includes pro, metalloprotease and hemopexin-like domains. These enzymes are involved in embryonic development, tissue remodeling, arthritis, cancer and cardiovascular disease. ADAMs contain a common domain structure that includes pro, metalloprotease, disintegrin-like, cysteine-rich, transmembrane and cytoplasmic domains. These enzymes function in fertilization and development, membrane protein shedding and signal transduction. ACEs are cell surface proteins and are

important regulators of blood pressure homeostasis, immunity, reproduction and neuropeptides. NEPs are cell surface enzymes that are involved in degradation of brain enkephalins and amyloid β peptides

Threonine Proteases

The most studied member of the threonine proteases is the 26S proteasome, often termed “the proteasome” (6). The ▶proteasome is a multiprotein complex consisting of a 20S core particle that is associated with one or two 19S regulatory particles. The 20S core particle is composed of four stacked rings: two outer α subunits rings and two inner β subunits rings. The β subunits are responsible for the enzymatic activity that is associated with N-terminal threonine residues. In general, proteins degraded by the proteasome are tagged ATP-dependently by the ubiquitin-conjugation system. Although the proteasome was originally thought to be a recycler of misfolded or damaged proteins, recent evidence indicates that the proteasome is also involved in the cell-cycle, cell survival, signal transduction and antigen processing.

Molecular Interactions

The molecular interaction between a protease and a substrate is described using a well-established nomenclature. The amino acid residues of the polypeptide substrate are considered to bind to the subsites of the enzyme's active site. Hence, the peptide substrate amino acid residues are termed “P” and the protease subsites are termed “S”. The scissile bond in the peptide substrate is the bond that is cleaved by the protease. The amino acid residues of the N-terminal side of the scissile bond are numbered P3, P2, P1 and those residues on the C-terminal side are numbered P1', P2', P3', etc. Thus, the P1 and P1' residues are those adjacent to the scissile bond. The subunits on the protease that interact with the substrate are numbered complementary to the substrate binding residues (i.e. S3, S2, S1, S1', S2', S3'). As mentioned above, the residues in the active site and the catalytic mechanisms differ among the families of proteases.

Serine Proteases

Serine proteases contain the three amino acid residues histidine, aspartic acid and serine in their active sites. These form a catalytic triad and are essential in the catalytic process. ▶Catalysis by serine proteases follows two steps: 1) formation of the acyl enzyme intermediate between the substrate and the essential serine residue of the protease, which results in the cleavage of the scissile bond and 2) ▶deacylation in which the acyl-enzyme intermediate is hydrolyzed by a water molecule to release the peptide. In the first step the ▶hydroxyl group of serine residue is the attacking

▶nucleophile, whereas the water molecule acts as the attacking nucleophile in the second step, and the histidine residue acts as a general base in both steps.

Cysteine Proteases

Cysteine proteases contain essential cysteine and histidine residues in their active sites that play roles similar to serine and histidine, respectively, in serine proteases. Catalysis by cysteine proteases is similar to that of serine proteases following a two-step process: 1) formation of an acyl enzyme intermediate between the substrate and the essential cysteine residue, which results in the cleavage of the scissile bond and 2) deacylation resulting in the release of the peptide. The ▶thiolate ion of the cysteine residue acts as the attacking nucleophile in both steps and is stabilized through the formation of an ion pair with the neighboring ▶imidazolium group of the histidine residue.

Aspartic Proteases

Aspartic proteases are bilobed molecules that contain one aspartic acid residue in each of their lobes. These aspartic residues are in close proximity in the active molecule and contribute to the catalytic diad of the enzyme. Catalysis by aspartic proteases differs from that of serine and cysteine proteases in that activated water molecules in the protease's active site directly attack the substrate's scissile bond. The nucleophilic attack is achieved by two simultaneous proton transfers: 1) from a water molecule to the diad of the carboxyl groups and 2) from the diad to the carbonyl oxygen of the substrate thus resulting in the cleavage of the substrate.

Metalloproteinases

Metalloproteinases contain one zinc atom in their active site that is bound to two histidines and one glutamic acid. Like aspartic proteases, metalloproteinases use an activated water molecule as the attacking nucleophile. The zinc-bound water molecule attacks the carbonyl group of the scissile bond leading to the formation of a non-covalent tetrahedral intermediate. The transfer of the glutamic acid proton to the departing amine group breaks down this intermediate, resulting in the cleavage of the substrate.

Threonine Proteases

The proteasome is the least understood of the proteases with respect to its molecular interactions. It is a threonine protease with three different active sites corresponding to: 1) trypsin-like, 2) chymotrypsin-like and 3) peptidylglutamyl ▶peptide hydrolase or caspase-like activities. These active sites are located on subunits β 1, β 2 and β 5, respectively, of the 20S core

particle. Each subunit contains an N-terminal threonine that functions as the attacking nucleophile. It is speculated that lysine and glutamate/aspartate residues in the active sites may play roles similar to histidine and aspartate in serine proteases. Proteins that are degraded by the proteasome must first be tagged with ubiquitin, which is recognized by the 19S regulatory particles. Multi-ubiquitination of target proteins is carried out by a cascade of enzymes, e.g., E1, E2 and E3.

Regulatory Mechanisms

Proteases are involved in a number of biological processes and as such need to be stringently regulated. Posttranslational mechanisms of protease regulation include compartmentalization, high substrate specificity, activation cascades and inhibitors.

Spatial control and compartmentalization of proteases can limit protease activity to a subcellular localization. For example, cathepsins are predominately restricted to **lysosomes**, which have an acidic environment that is favorable for activity and stability of the enzymes. Similarly, the association of proteases with the plasma membrane as in the case of membrane-type MMPs limits the activity of these proteases to substrates that are in contact with the cell surface. Compartmentalization can also occur within the protease itself. For example, the proteasome is designed as a narrow cylinder with its active sites on the inside of the cylinder, with entrance to the interior monitored by the regulatory particles that cap the ends of the cylinder. Substrate specificity is critical to the functional activity of the enzyme. The residue in the P1 position of the substrate usually determines the binding and cleavage of the substrate to the protease. Residues in P3, P2, P1', P2', etc. sites are also important in influencing substrate binding and protease specificity. For example, chymotrypsin, trypsin and elastase essentially share identical catalytic triads yet their specificity for substrates differs. Specific proteolytic activity is demonstrated by caspases, which are cysteine proteases that cleave substrates at their aspartic acid residues. Another example of specificity is collagenases (MMP-1, MMP-8, MMP-13 and MMP-14), the only MMPs that can efficiently degrade interstitial collagens I, II and III at specific sites in their triple-helical domains.

Activation cascades are important processes that regulate various proteolytic events. Activation may involve the conversion of an inactive proenzyme to its active form by the cleavage of the proregion. This process can be accomplished by autocatalysis or may require the activity of other proteases. For example, procathepsin B can be activated by cathepsin D or can autoactivate itself. Activation cascades involve the activation of one protease leading to the successive activation of other proteases. An example is seen on the surface of cancer cells where active cathepsin B will

convert pro-urokinase plasminogen activator to active urokinase plasminogen activator that in turn converts plasminogen to active plasmin. Plasmin can activate proMMPs to active MMPs, which degrade extracellular matrix proteins during tumor invasion.

Protease inhibitors are probably the most effective regulators of proteolytic enzymes (7). They can bind tightly to proteases and produce inactive complexes. Inhibitors can be found both inside and outside cells and can function both irreversibly and reversibly depending on the specific enzyme-inhibitor complex. The specificity of inhibitors is determined by the catalytic mechanism and the structural conformation in and around the active site of the protease. Commonly, the most effective protease inhibitors are those that mimic specific substrates. Like proteases, endogenous inhibitors are often grouped into families and/or classes. For example, serpins, cystatins and tissue inhibitors of metalloproteinases are common inhibitor families against serine, cysteine and matrix metalloproteases, respectively. The role of protease inhibitors is important since changes in inhibitor concentrations alone can alter proteolytic activity within the cell.

Synthetic inhibitors are usually small peptide molecules that are often used as experimental tools in deciphering the roles of one or more proteases within a system. There are both broad spectrum inhibitors such as E64, which irreversibly inhibits most papain-like cysteine proteases, and selective inhibitors such as CA074, which inhibits cathepsin B. Selection of an inhibitor for a protease is challenging, considering that proteolytic pathways often involve the degradation of common protein substrates. For example, the calpain inhibitor, calpain inhibitor I, can also inhibit proteasomal activity. A listing of endogenous and synthetic protease inhibitors can be found in the MEROPS (2) databases.

In conclusion, it is evident that the role of proteases is integral to the growth and survival of living systems. Aberrant regulation of proteases and their inhibitors has been shown to result in the onset of various diseases and pathological conditions. The development of more selective inhibitors will facilitate our understanding of the biological and/or pathological functions of these proteases.

References

1. Barrett AJ, Rawlings ND, Woessner JF (1998) Handbook of Proteolytic Enzymes. Academic Press, London
2. MEROPS merops.sanger.ac.uk
3. Puente XS, Sánchez LM, Overall CM et al. (2003) Human and mouse proteases: a comparative genomic approach. *Nat Rev Genet* 4:544–548
4. Chapman HA, Riese RJ, Shi G-P (1997) Emerging roles for cysteine proteases in human biology. *Annual Reviews in Physiology* 59:63–68

5. Visse R, Nagase H (2003) Matrix metalloproteinases and tissue inhibitors of metalloproteinases. Structure, function and biochemistry. *Circulation Research* 92:827–839
6. Adams J (2003) The proteasome: structure, function, and roles in the cell. *Cancer Treatment Reviews* 29 (Supplement 1): 3–9
7. Mykles DL (2001) Proteinase families and their inhibitors, *Methods in Cell Biology* 66:247–287

Proteasome

Definition

The proteasome, also known as prosome or macropain, is a large cytoplasmic protein complex composed of multiple proteins, which are responsible for degrading proteins that have been marked for destruction by ubiquitination or by some other means. The proteasome is characterized by its ability to cleave peptides with arg, phe, tyr, leu and glu adjacent to the leaving group at neutral or slightly basic pH. It has an ATP-dependent proteolytic activity, and may also catalyze basal processing of intracellular antigens. The proteasome is composed of the cylindric 20S core particle (20S proteasome) and a 19S cap particle, which docks onto both ends of the 20S proteasome to yield the 26S proteasome. The 26S proteasome is thought to be the biologically active unit in the cell. Proteasomes are found in the cytoplasm and in the nucleus.

- [Adherens Junctions](#)
- [Limb Girdle Muscular Dystrophies](#)
- [Proteases and Inhibitors](#)
- [Repeat Expansion Diseases](#)
- [Ubiquitination](#)

Protein

Definition

Proteins (also referred to as polypeptides) are biological macromolecules consisting of the naturally occurring amino acids. Proteins often adopt a defined three-dimensional structure that is necessary to exert a particular function, as for example, enzymatic catalysis or specific recognition of other biologically relevant molecules.

- [Protein Databases](#)
- [Protein-Ligand-Interaction by NMR](#)

Protein A

Definition

Protein A is a 42 kDa protein from the cell wall of *S. aureus*. It binds with high affinity to the Fc domain of several Immunoglobulin G classes and subclasses.

- [Surface Plasmon Resonance](#)

Protein and Membrane Transport in Eukaryotic Cells

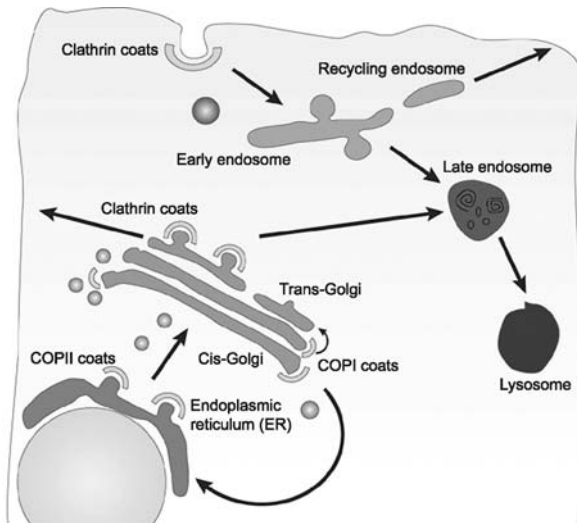
MICHAEL KRAUSS, VOLKER HAUCKE

Institute for Chemistry-Biochemistry, Free University, Berlin, Germany

vhaucke@chemie.fu-berlin.de

Definition

Eukaryotic cells contain distinct compartments that are characterized by specific protein and membrane compositions. To establish compartment identity and to allow for communication between the different organelles, transport processes take place that ensure the delivery of proteins to their proper destination. Two principal mechanisms of transport need to be distinguished. The first one is used to translocate nascent polypeptides into or across the membrane of a compartment through a proteinaceous channel. This holds true for proteins initially destined for the ► [endoplasmic reticulum](#) (ER), mitochondria, chloroplasts, ► [peroxisomes](#) or the ► [nucleus](#). In this essay we will focus exclusively on the second type of sorting process, which transports proteins and lipids in vesicles or tubules along the secretory and the endocytic pathways (Fig. 1). The transport along both routes is predominantly mediated by membrane-bounded vesicles or tubules, which bud from donor compartments and subsequently fuse with a specific acceptor compartment. In the secretory pathway, proteins are initially transported from the ER to the *cis*-Golgi. They are passed through the Golgi stacks by cisternal progression until they finally end up in the *trans*-Golgi network (TGN). From the TGN, transport can proceed into two different directions. The first branch leads to the plasma membrane where transport vesicles can fuse by constitutive or regulated exocytosis. The second branch delivers proteins to the late endosomes and lysosomes. The endocytic pathway serves to internalize nutrients, pathogens, growth factors and recycling receptors from the plasma membrane. These cargo



Protein and Membrane Transport in Eukaryotic Cells. Figure 1 Transport routes along the secretory and the endocytic pathway in eukaryotic cells. During the secretory pathway newly synthesized proteins are initially transported from the ER to the *cis*-Golgi. After passage through the Golgi stacks they reach the *trans*-Golgi network, from where they can be delivered to the plasma membrane or the endosomal/lysosomal system. The endocytic pathway serves to internalize cargo proteins from the plasma membrane. These are initially transported to endosomes. From here some are recycled back to the cell surface, whereas others are forwarded to late endosomes and lysosomes for degradation. The best characterized transport vesicles are those that bud with help of COPI, COPII or clathrin coats (taken with kind permission from T. Kirchhausen (2000), *Nat Rev Mol Cell Biol* 1:197–198).

molecules are initially delivered to endosomes; from here some return to the plasma membrane, whereas others are transported to lysosomes for degradation.

Characteristics

To preserve the identity of individual compartments during the dynamic exchange of proteins and lipids, the generation of vesicles at different donor membranes and the fusion of vesicles with the appropriate target membranes must be highly selective. Accordingly, membrane transport within the secretory and endocytic pathways employs specific sets of transport vesicles. Despite the inherent complexity, the majority of transport events make use of similar overall principles. Vesicle formation is generally initiated by the binding of individual coat proteins to the surface of the parent membrane. Coat proteins serve two main functions. First, they recognize **▶sorting signals** within the cytosolic domains of cargo proteins and thereby ensure incorporation of specific cargo proteins into the nascent vesicle. Second, due to their propensity to assemble

into higher ordered multimeric complexes on the donor membrane, coat proteins have the ability to deform the underlying lipid bilayer. Components of the coat may drive membrane curvature until a coated bud is formed which eventually pinches off. After fission the released vesicle is rapidly uncoated and targeted to its appropriate acceptor membrane. Following an initial docking step, vesicle-bound v-**▶SNARE proteins** tie up in a complex with their cognate t-SNARE proteins in the target membrane. Formation of a SNARE complex is essential to bring vesicle and target membranes into close apposition, thus allowing fusion of the two lipid bilayers. Finally, the SNARE complex is disassembled and – together with other components of the transport machinery – recycled back to the initial donor compartment.

Molecular Interactions Vesicle Formation

At least three different types of transport vesicles have been described (1). These can be distinguished on the basis of the different types of coat complexes which assemble onto the respective donor membranes and have been termed COPII, **▶COPI** and **▶clathrin** coats (Fig. 1). Although coat formation appears to follow similar principles, significant mechanistic details between the different types of coats have emerged. COPII-coated vesicles bud from specialized regions of the ER devoid of ribosomes, termed ER exit sites and transfer newly synthesized proteins to the *cis*-Golgi (1, 2). Coat formation at the ER is initiated when a guanine nucleotide exchange factor (GEF), Sec12p, recruits and activates Sar1p, a member of the **▶Ras superfamily of small GTP binding proteins**. Membrane translocation of Sar1p induces recruitment of a complex of Sec23p and Sec24p proteins and stimulates the association of Sec23p with SNARE proteins needed for later fusion events. The ternary complex consisting of Sar1p and Sec23/24 in turn associates with a complex of Sec13p and Sec31p proteins to form the complete COPII coat structure. Indeed, Sar1p, Sec23/24 and Sec13/31 together comprise the minimal machinery necessary for vesicle budding. A large protein bound to the cytosolic side of the ER, Sec16p, interacts with both Sec23/24 and Sec13/31 complexes; it appears to serve as an anchor defining ER exit sites and seems to coordinate coat formation. Membrane proteins containing sorting signals are incorporated into the coat by direct binding to the Sec23/Sec24 complex. These sorting signals are diverse and can contain hydrophobic, di-acidic or di-basic motifs. Soluble cargo proteins may associate with members of the p23/p24 family of transmembrane proteins, which form a complex with Sec23p. After fission of the vesicle the release of all coat constituents is triggered by hydrolysis of GTP bound to Sar1p.

ARF1 (ADP ribosylation factor 1), another small GTPase similar to Sar1p plays a central role in initiating budding processes at Golgi membranes. ARF1 is myristoylated at its amino terminus; specific exchange factors catalyze loading of ARF1 with GTP and thereby induce a conformational change that triggers exposure of the myristoyl anchor and subsequent membrane insertion. Membrane-bound ARF1 recruits distinct coats to different regions of the Golgi. One of them is COPI or coatomer, a complex of seven proteins, that functions in retrograde transport from the *cis*-Golgi to the ER and between Golgi cisternae. COPI recognizes proteins that expose a basic sorting signal (KKXX or KKKXX, where X is any amino acid) in their carboxy terminal domain. COPI coated vesicles also incorporate KDEL receptors which retrieve escaped luminal proteins to the ER. These soluble proteins are tagged with a retrieval signal at their carboxy terminus (KDEL). After fission, hydrolysis of bound GTP triggers release of ARF1 and the residual coat components from the vesicle.

In addition to its role in recruiting COPI to the *cis*-Golgi, ARF1 triggers coating with clathrin at the *trans*-Golgi network (TGN). Membrane binding of clathrin is mediated by AP1 ►adaptor protein complexes and ►GGA proteins (3). AP1 adaptors are heterotetramers formed by two large, one medium and one small subunit that shuttle cargo between the TGN and the endosomal system. A large number of membrane proteins bear sorting signals (tyrosine-based YXX Θ or dileucine motifs, with Θ being an amino acid with a bulky hydrophobic residue) in their cytosolic tails, which are recognized by AP1, including the ►mannose 6-phosphate receptor (MPR), which itself associates with soluble enzymes destined for lysosomes. The large subunits of AP1 adaptor complexes associate with clathrin and accessory proteins. Clathrin triskelia polymerize into polygonal lattices forming a shell around the adaptor-coated bud. GGA proteins seem to have partially overlapping or complementary functions with AP1 as they share common binding partners including clathrin and some accessory proteins. In addition, GGAs have been implicated in the transport of cargo proteins (including the MPR but also other transmembrane proteins) from the TGN to endosomes. The precise interplay of these two clathrin adaptors remains to be elucidated. Distinct regions of the TGN serve as binding sites for non-clathrin associated AP4 adaptors which may fulfill important sorting functions in polarized cells.

Clathrin-coated vesicles also originate at the plasma membrane and transport cell-surface proteins to endosomes (4). In this case clathrin recruitment is mediated predominantly by AP2 adaptor complexes that have a heterotetrameric organization analogous to AP1 and associate with overlapping sets of sorting signals.

Besides AP2 additional clathrin binding proteins have been identified that recognize a more limited set of cargo proteins, like ARH and Dab2 which interact selectively with the cytoplasmic tail of the LDL receptor, β -arrestin which binds to G-protein coupled receptors or epsin which contains a ubiquitin interacting motif and thus might help to internalize mono-ubiquitinated cell surface receptors. Each of these proteins can bind to clathrin, but also displays affinity for AP2. Thus, these proteins might function independently of AP2 or cooperate with AP2 during assembly of the clathrin coat. The AP2/clathrin coat may contain additional adaptor proteins such as AP180/CALM, which appears to regulate the size of the coated vesicle. A number of accessory proteins have been identified which serve to assist vesicle formation and, together with clathrin and its adaptors, form a dynamic network of complex protein-protein and protein-lipid interactions. One of these factors is dynamin, a large GTPase mediating fission of clathrin/AP2 coated-vesicles. Dynamin assembles into oligomeric rings around the neck of a coated vesicle. GTP hydrolysis induces a conformational change within dynamin thereby driving scission of the vesicle. Other accessory proteins, namely the ►molecular chaperones Hsc70 and auxilin, mediate uncoating of AP2/clathrin coated vesicles.

Alternative processes by which membrane and proteins are internalized from the plasma membrane are macropinocytosis and ►phagocytosis (4). Both are actin-dependent processes that serve to internalize large volumes of extracellular fluids or particulate material, respectively. Neither of the two processes recognizes specific cargo. Accordingly, macropinocytosis and phagocytosis are not mediated by coats.

Coat recruitment to endosomes is less well understood. Different adaptors have been localized to endosomes. One of them is AP3, which might be recruited to the membrane in response to the activation of an ARF protein. Whether AP3 co-assembles with clathrin in a coat is still a matter of debate. Another coat protein recruiting clathrin onto endosomal membranes is Hrs, which recognizes mono-ubiquitinated membrane proteins and segregates them into the lysosomal pathway.

Membrane Fusion

After fission the free vesicle is rapidly uncoated and targeted to the appropriate acceptor membrane. In most cases this process is assisted by a member of the Rab family of small GTPases. More than 60 different Rabs are encoded by the mammalian genome; they localize to specific intracellular compartments and might thus contribute to organelle-specific sorting events (5). Rab proteins cycle between a cytosolic GDP-bound inactive state and a GTP-bound activated state, in which they associate with membranes by insertion of a geranylgeranyl anchor. Upon GTP loading, Rab proteins interact

with a set of specific effector molecules residing in the acceptor membrane thereby docking the vesicle to the target bilayer. Rab effectors often form larger complexes by associating with proteins that might help to temporally and spatially coordinate discrete transport events. A prominent example of such an effector complex in yeast is the ►**exocyst**, which is assembled stepwise onto Rab-associated membranes and promotes docking of secretory vesicles at the plasma membrane. It is important to note that a single Rab (i.e. Rab5 in endosomes) can function in multiple stages of vesicular traffic including budding, docking and fusion. In addition, some Rabs have been shown to interact with actin- and myosin-based motor proteins, implicating Rabs in the transport of vesicles from their site of formation to the site of fusion.

The actual fusion of the vesicle requires the tight association of so-called SNARE proteins (6). They form a large family of mostly transmembrane proteins, which reside in the donor as well as in the acceptor membrane. Their functional amino terminal domains contain a conserved α -helical stretch, the SNARE motif, which is exposed into the cytosol. After Rab-assisted docking of a vesicle to the membrane one helix of the monomeric v-SNARE and three helices of the oligomeric t-SNAREs assemble into a coiled-coil structure forming a tight four helix bundle, the *trans*-SNARE complex. As only distinct combinations of v- and t-SNAREs can engage into a stable complex, the specificity of this reaction has been proposed to control the site of vesicle fusion, although the formation of non-physiological SNARE-complexes has been observed *in vitro*. The zippering of SNAREs moves vesicle and target membranes into close apposition and may provide the free energy necessary for subsequent membrane fusion. Members of a small family of soluble proteins, Sec1p/Munc18-like (SM) proteins associate with Rab effectors and/or SNARE proteins and appear to couple Rab-mediated membrane docking with SNARE complex formation. Since the SNARE complex is extremely stable, its disassembly depends on the activity of the soluble ATPase NSF. NSF in concert with NSF-attachment proteins (SNAPs) promotes disassembly of post-fusion *cis*-SNARE complexes.

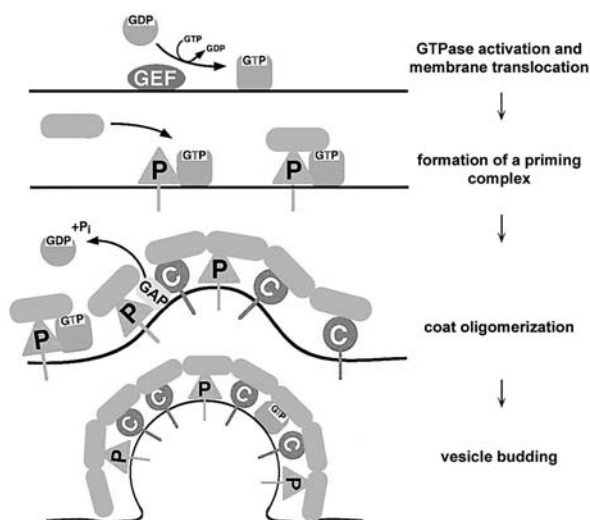
One of the best-studied fusion events is the exocytosis of ►**synaptic vesicles** (SVs) at the neuronal plasma membrane. Rab3 has been shown to regulate targeting and docking of SVs to the active zone. In addition, deletion studies have demonstrated that Rab3 functions at a post-docking step during fusion. The SNARE complex that tethers SVs to the plasma membrane is made by the v-SNARE synaptobrevin and two t-SNAREs, syntaxin 1 and SNAP-25, with SNAP-25 contributing two α -helices to the complex. Several additional proteins, like synaptotagmin I, complexins,

►**Munc18**, ►**Munc13** and RIM, bind to these individual SNAREs or the assembled SNARE complex and are thought to regulate docking, priming or fusion of SVs.

Regulatory Mechanisms

As outlined above members of the ARF and the Rab families of small GTPases are important regulators of protein and membrane transport. Accordingly, to ensure the selectivity of distinct transport events the activity cycles of both types of GTPases need to be tightly controlled by guanine nucleotide exchange factors (GEFs) and GTPase-activating proteins (GAPs) (1). In this respect the activity of Sar1p is regulated by the exchange factor Sec12p mentioned above which restricts COPII coat formation to the ER and by the coat constituent Sec23p which simultaneously serves as a GAP, inactivating Sar1 upon completion of the coat. Similarly, in agreement with a role of ARF1 in COPI and clathrin recruitment to the Golgi, a number of ARF1-specific GEFs have been identified that reside at the Golgi membranes (Fig. 2). Together with COPI an ARF1-specific GAP becomes incorporated into the coat which associates with ligand-bound KDEL receptors. This mechanism links ARF1 activity with cargo recognition. GAP activity appears to be modulated by membrane curvature; this ensures that ARF1 hydrolyzes GTP after fission of the coated vesicle. The activation of Rab proteins appears to be more complex (6). In the inactive state Rabs reside in the cytosol bound to a GDP-dissociation inhibitor (GDI). The activity of a GDI displacement factor (GDF) facilitates dissociation of GDI and membrane recruitment of Rab. The localization of Rabs to the appropriate membrane is ensured by the activity of specific GEFs. Once activated, Rabs engage in effector complexes that covalently link the Rab-associated membrane with the cognate acceptor membrane. After fusion, GAPs catalyze GTP hydrolysis on the Rab and Rab-GDP is recycled into the cytosol by associating with a Rab-GDI.

Another important regulatory mechanism is contributed by membrane lipids, in particular ►**phosphoinositides**. Different species of phosphoinositides have been identified that seem to be enriched in distinct cellular membranes. Phosphatidylinositol 4,5-bisphosphate (PI4,5P₂) and phosphatidylinositol 4-phosphate (PI4P) are generated at the plasma membrane and at the TGN, respectively. Phosphatidylinositol 3-phosphate (PI3P) and 3,5-bisphosphate (PI3,5P₂) are abundant in endosomes. Several proteins involved in budding or fusion processes contain domains that selectively interact with distinct phosphoinositides and thereby contribute to the correct localization of these proteins. Accordingly, regulators of homotypic docking and fusion of endosomes and the endosomal mono-ubiquitin binding



Protein and Membrane Transport in Eukaryotic Cells.

Figure 2 Schematic overview of distinct steps that lead to coated vesicle formation. Coat formation is initiated when a specific guanine exchange factor (GEF) triggers activation of a small GTPase and promotes its translocation to the donor membrane. Together with membrane factors which act as primers (P) (these can be specific proteins or certain phospholipids) the membrane-bound GTPase subsequently recruits individual coat components to form a priming complex. Components of the coat also associate with cargo proteins (C), which expose appropriate signal sequences in their cytoplasmic tails. Coat oligomerization leads to a deformation of the underlying lipid bilayer. Cargo molecules are concentrated in the emerging coated bud. GTPase-activating proteins (GAPs) become incorporated into the coat. They are activated before or after fission of the coated vesicle and trigger release of the GTPase into the cytoplasm (adapted from S. Springer et al. (1999), *Cell* 97:145–148).

adaptor Hrs specifically associate with PI3P, whereas AP1 and AP2 adaptors display binding specificity for PI4P and PI4,5P₂. Interestingly, the kinases involved in the generation of the latter lipid species have turned out to be ARF effectors, suggesting a feed-forward loop driving coat formation. As recruitment of AP2 to the plasma membrane does not seem to be directly supported by a specific GTPase, PI4,5P₂ might provide a trigger to initiate coat formation (4). Indeed, many accessory proteins involved in clathrin-mediated endocytosis from the plasma membrane display binding specificity for PI4,5P₂, including AP180/CALM, epsin and dynamin. Correspondingly, the inositol phosphatase synaptojanin 1 which hydrolyses PI4,5P₂ was found to affect the stability of AP2/clathrin coats. By a yet unidentified mechanism PI4,5P₂ also enhances Ca²⁺-triggered exocytosis

of SVs at the neuronal plasma membrane. One potential effector protein is **▶synaptotagmin I** which combines both Ca²⁺- and PI4,5P₂-binding properties.

References

1. Kirchhausen T (2000) Three ways to make a vesicle. *Nature. Rev Mol Cell Biol* 1:187–198
2. Bonifacino JS, Glick BS (2004) The mechanisms of vesicle budding and fusion. *Cell* 116:153–166
3. Robinson MS, Bonifacino JS (2001) Adaptor-related proteins. *Curr Opin Cell Biol* 13:444–453
4. Conner SD, Schmid SL (2003) Regulated portals of entry into the cell. *Nature* 422:37–44
5. Pfeffer S (2003) Membrane domains in the secretory and endocytic pathways. *Cell* 112:507–517
6. Jahn R, Lang T, Sudhof TC (2003) Membrane fusion. *Cell* 112:519–533

Protein Array

▶Protein Microarrays as a Tool for Protein Profiling and Functional Proteomics

Protein Biosynthesis

Definition

Protein biosynthesis refers to **▶translation**.

▶Fragile X Syndrome

Protein Chemical Modification

▶Protein Interaction Analysis: Chemical Cross-linking

Protein Chips

▶Protein Microarrays as a Tool for Protein Profiling and Functional Proteomics

Protein Complex

Definition

Protein complex denotes the association of identical or different polypeptide chains, called subunits, by noncovalent interaction which leads to stable oligomeric proteins or protein complexes.

► [Affinity Chromatography and *In Vitro* Binding \(Beads\)](#)

Protein Conformational Diseases

► [Defective Protein Folding Disorders](#)

Protein Conjugation

► [Protein Interaction Analysis: Chemical Cross-linking](#)

Protein Crystallization for X-Ray

JOCHEN MUELLER-DIECKMANN,
MANFRED S. WEISS, MATTHIAS WILMANN
EMBL Hamburg Outstation, c/o DESY, Hamburg,
Germany
wilmanns@embl-hamburg.de

Definition

Before the three-dimensional structure of a protein molecule can be determined by X-ray crystallographic methods, the protein molecules have to be assembled into a regular and periodic three-dimensional lattice – a crystal. This process is called protein crystallization. The principles of protein crystallization as well as many experimental and technical approaches to it will be discussed in the following description.

Description

Introduction

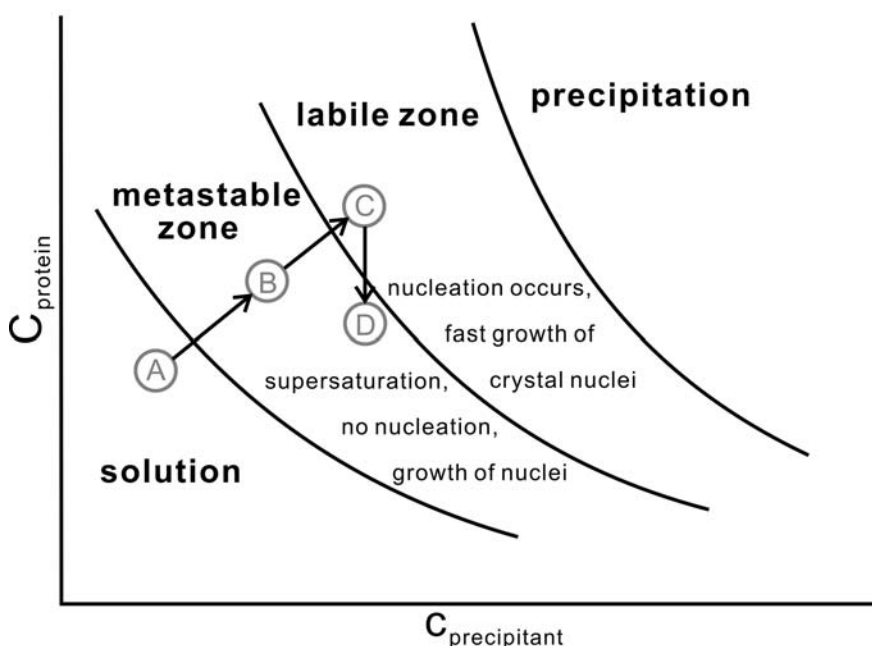
The three-dimensional structure of a biological macromolecule constitutes not only the basis for investigating

its function but can also serve as a template to determine how and where small molecule ligands bind to it. Such an approach has proven to be extremely useful for the identification and characterization of putative lead compounds that affect the function of a macromolecular target and which, at some point, may lead to the development of new drugs. Of the two techniques, which can provide three-dimensional structural information, nuclear magnetic resonance (NMR) and X-ray crystallography, the latter contributes about 85% of all structures deposited in the ► [Protein Data Bank](#). Furthermore, X-ray crystallography is basically without limitations as long as the macromolecular targets to be investigated are properly folded. Although many steps in NMR and X-ray crystallography have been greatly improved during recent years, there are still common and specific bottlenecks. Common to both methods, and other structural and imaging techniques as well, is the need to purify mg amounts of pure samples. Specific to X-ray crystallography is the need to grow crystals diffracting to sufficient resolution (typically 3 Å or better). Presently available crystallization methods and associated technologies are outlined in this review.

Principles

The process of crystallization of any substance can be defined as a change of its state of aggregation. For proteins, in almost all cases such a process is defined by the change from an aqueous protein solution to the solid, crystalline state. In order to trigger this transition, the crystalline phase needs to be thermodynamically more stable than the protein in solution. To reach such conditions one may change the environmental parameters either to lower the thermodynamic stability of the solution phase or to increase that of the crystalline phase. In practical terms, these two effects are normally not distinguished.

The process of crystallization may be divided into two principle processes, the formation of crystal nuclei (nucleation) and the growth of these nuclei to crystals (crystal growth). A simple and schematic phase diagram of an aqueous protein/precipitant system is shown in Fig. 1. In the solution zone at low protein concentration c_{Protein} and low precipitant concentration $c_{\text{Precipitant}}$, the solution is thermodynamically the most stable phase. If the protein concentration, the precipitant concentration or both are increased, the solubility line is crossed and the meta-stable zone is reached. In this zone, the solution becomes supersaturated and hence is not the most thermodynamically stable phase anymore. Unless crystal nuclei are present, e.g. from a seeding experiment, no crystals will grow because new nuclei cannot form. Increasing the protein and/or the precipitant concentration further, the third zone, referred to as the labile zone, is reached. Both



Protein Crystallization for X-Ray. Figure 1 Schematic phase diagram of an aqueous protein/precipitant mixture. The precipitant can be any chemical compound that is able to reduce the thermodynamic stability of the protein solution phase.

crystal nucleation and crystal growth take place in this zone. At even higher protein and/or precipitant concentrations, the precipitation zone is reached. In this zone, the protein precipitates too rapidly to allow the formation and growth of crystal nuclei.

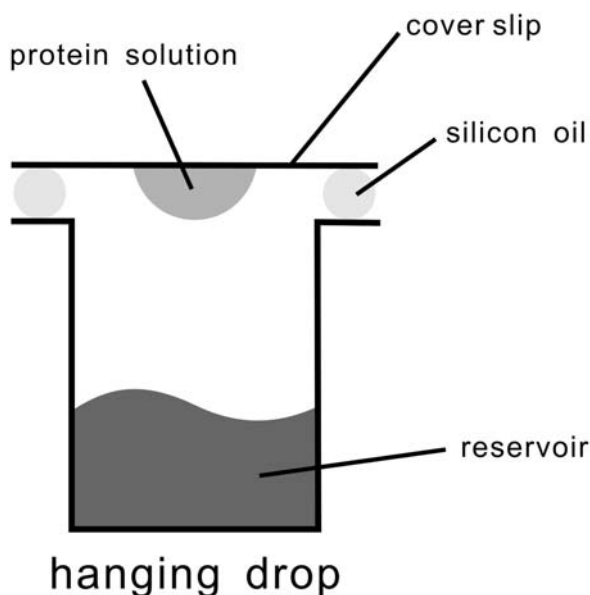
The aim of a crystallization experiment is the formation of only a few crystals, which then grow large enough (>0.1 mm in their smallest dimension) for subsequent analysis by X-ray crystallography. In order to achieve this, the experiment should start at some point A in the stable solution zone, then slowly move across the meta-stable zone (point B) into the labile zone (point C) in the phase diagram. The whole process needs to be on an appropriate time scale to allow for sufficient time for the formation of crystal nuclei. Once nucleation has occurred, a course that leads the system back into the meta-stable zone (point D) is desirable. Such a process would ensure the formation of only a few nuclei and subsequent crystal growth without competition from the formation of more nuclei.

Techniques

In order to move from the solution zone into the labile zone (Fig. 1) two methodologically distinct methods can be applied, the batch method and the diffusion-based method. In the batch method, the protein and the precipitant solution are mixed and placed in a sealed container so that no further transport process takes place. Super-saturation needs to be achieved (point C in

Fig. 1) at the outset of the experiment. The formation of the first nuclei and their subsequent growth into crystals will subsequently lower the protein concentration, leading to a path in the phase diagram towards point D. In the diffusion-based methods, the crystallization experiment starts in the stable solution zone (point A in Fig. 1). Depending on the experimental design, there may be some differences in the path of crystallization. Nowadays, the hanging-drop set-up (Fig. 2) is the most popular and widespread approach. In this technique, the hanging drop contains the protein and some precipitant in a defined volume ratio while the reservoir solution only contains the precipitant. The solutions are separated by a vapor phase and the whole system is sequestered by sealing it with silicon oil or grease. The system is initially not in thermodynamic equilibrium because of the differences between the precipitant concentration in the reservoir and that in the drop. Eventually equilibrium is achieved by the diffusion of water and other volatile components (if present) from the drop to the reservoir. Hence, both the protein and the precipitant concentration of the drop will increase, moving the protein solution from the solution zone (point A), into the meta-stable (point B) and then to the labile zone (point C).

Other diffusion-based techniques are the sitting-drop technique, which also makes use of diffusion across the vapor phase, dialysis, which utilizes diffusion across a membrane, diffusion across an oil barrier, diffusion



Protein Crystallization for X-Ray. Figure 2 The hanging-drop vapor diffusion experimental set-up. Such an experiment can be carried out using commercially available cell culture plates with 24 or 96 depressions.

across a phase boundary when the protein solution and the precipitant solution are placed next to each other without mixing them or diffusion through gels.

Screening

There are many physicochemical parameters affecting the solubility of a protein. Just to name a few, there are the ionic strength of the solution, the nature of the precipitant, the pH, the temperature, the presence or absence of certain chemicals such as detergents etc. Since it is impossible to predict which combination of parameters will yield a phase diagram and a path through the phase diagram that leads to crystal formation and since the total number of combinations of parameters exceeds by far what is realistically doable, the whole parameter space cannot be sampled completely. This required the development of a number of efficient screening protocols.

Usually, crystallization experiments are performed immediately after the biological macromolecule has been purified. The experiment begins with a solution of the purified biological macromolecule at fairly high concentration (1.5–200 mg/ml) in order to be close to the solubility limit. It is also essential that the purity level is as high as possible (>95%). This includes both the chemical purity (are other molecules still present in the mixture?) and the conformational purity (does the molecule to be crystallized occur in different conformational states?).

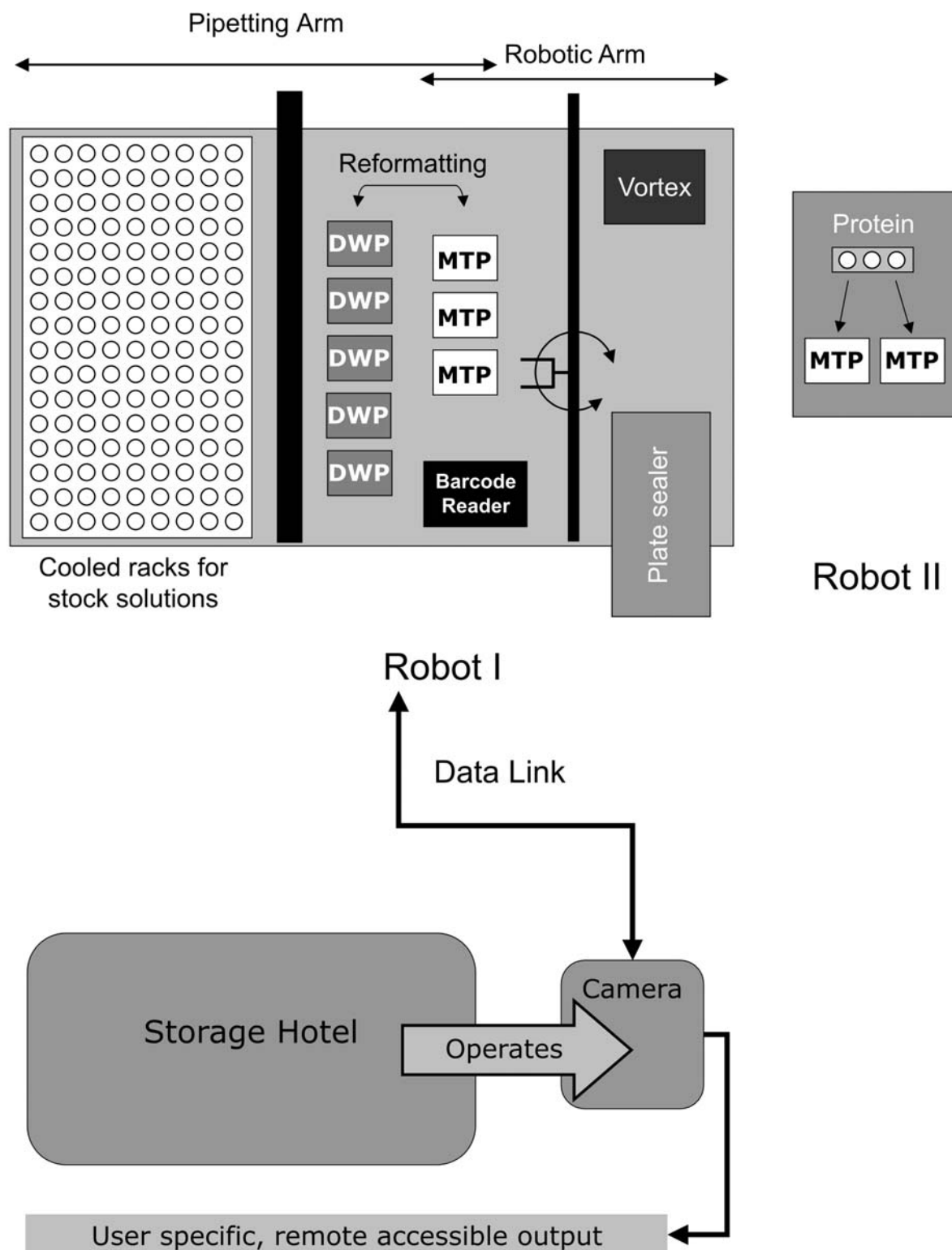
The most common sampling protocols include the grid screening approach of McPherson (3), the footprint screening approach of Stura (1) and the random sampling approach (4). A very popular approach is also the sparse matrix protocol of Jancarik and Kim (2), mainly because the first commercially available screens were designed after it. The incomplete factorial approach pioneered by Carter (1) has also been used occasionally. Over the past ten years or so, many researchers have designed their own screens based on the original ideas of Jancarik and Kim (2), mixed with their own expertise. Many of these screens are now commercially available from different companies (Hampton Research, Emerald Biostructures, Molecular Dimensions, Jena Bioscience and possibly others as well). Most of them comprise about 100–300 single experiments. Considering that for each experiment about 5–100 μg of the biological macromolecule (depending on the concentration needed to achieve super-saturation) is needed, the initial amount of pure protein to be prepared is in the mg range.

The content of the various screens has also been the subject of many discussions. From a purely statistical point of view, no assumptions about the crystallization conditions should be made in the absence of any prior knowledge. However, this is not realized in most commercially available screens, which are formed from empirical or even anecdotal data of previous successes in some cases. The assumption is that they will be more likely to produce crystals from samples that have not been crystallized yet. However, it can be shown that this is not correct (5). The best chance for success in cases where no prior knowledge exists is the uniform distribution of crystallization conditions in the N -dimensional crystallization hyperspace. Here, N represents the number of variable crystallization parameters, which is essentially unlimited. Such a distribution is achieved through the purely random design of experiments and is realized in the computer program CRYSTOOL (5).

A large number of crystallization conditions for biological macromolecules can be found in the Biological Macromolecule Crystallization Database.

The Future of Protein Crystallization

In the past decade, the sequencing of entire genomes from many organisms has led to a dramatic revolution in biology. Although this wealth of information that has emerged from the availability of a huge number of sequences certainly contributes enormously to the understanding of life on a molecular level, it has been realized that it is in no way sufficient to explain the functioning and inner workings of the cell. In order to proceed in this direction, the three-dimensional structures of all molecules in the cell (proteins, RNA, DNA and polysaccharides) have to be known (► [structural](#)



Protein Crystallization for X-Ray. Figure 3 Layout of a typical HT-crystallization platform. Robot I performs the liquid handling up to the preparation of reservoir filled MTPs (see text). Additional functionalities include the sealing of DWPs and MTPs after their completion, the vortexing of the sealed DWPs after cocktail preparation and the registration of barcodes. All DWPs and MTPs are barcoded for record keeping and relevant information is stored in a relational database, which is administered by the software running Robot I. A data link with the camera software allows the display of the reservoir and protein information alongside the crystal images and crystallization specific data, like temperature, date of setup or experiment type.

genomics). However, despite numerous improvements in X-ray crystallographic methodology, it is still a slow technique when compared to other techniques such as gene sequencing. The main reason for this is the necessity of obtaining relatively large amounts of very pure sample and of growing crystals of it. New and faster procedures have to be developed in order to overcome this problem. Here, some of the recent developments to improve and accelerate the process of crystallization of biological macromolecules are outlined and discussed.

Since neither the ability to crystallize nor specific crystallization conditions for biological macromolecules can be predicted, a large number of different molecular constructs have to be scrutinized with respect to a large number of different crystallization solutions in order to identify appropriate crystallization conditions. An aggravating restriction to this approach is the fact, that there is usually only a limited amount of very pure and structurally homogenous biological sample available. The obvious response to the latter is to decrease the volumes of sample solution in crystallization experiments. This led to the miniaturization of the hardware necessary to set up a large number of experiments and to the developments of robotics which could perform these experiments with high precision and reliability. The term nano-crystallization has been coined to refer to the crystallization drops smaller than 1 μ l. The current limit seems to be at drop volumes of about 50 nl.

Setting up a very large number of crystallization experiments (300–1500 per construct and per temperature) is a very repetitive task and as such predestined for automation. As a consequence, the processes of crystallization, crystal visualization and handling rely heavily on automated methods. Such methods have been developed both in academic institutions and in commercial enterprises. Currently, there are a few academic facilities available, to which external users can submit samples for high-throughput (HT) crystallization free of charge, provided they allow the experimental results to be stored anonymously for subsequent statistical analysis. An increase in sites offering such crystallization services (for hire or free of charge) for the general user community is to be expected in the near future. Combined with automated structure solution suites, this offers the possibility of quick and relatively inexpensive structure determination to non-expert users.

All HT crystallization platforms consist of a combination of liquid handling devices, storage and retrieval and visualization systems. Within the last 2–3 years there has been a huge increase in suppliers providing these modules. The precise selection of equipment and the extent of its inter-connection depend largely

on the goals and throughput demands of the individual laboratories. A schematic outline of a HT-crystallization platform is shown in Fig. 3. The various components perform different tasks. Robot I begins with stock solutions and prepares the crystallization cocktails in volumes of 1–2 ml in deep-well plates (DWPs). In a second step, smaller volumes (about 100 μ l) are taken from the DWP and pipetted into micro-titerplates (MTPs), which are the ones finally used for the crystallization experiment. These MTPs are then handed over to Robot II, which pipettes a small volume (about 100 nl) from the reservoir into the crystallization depression. Afterwards, the same volume of protein solution is pipetted into the crystallization depression, also by Robot II. Now, the MTP is sealed and placed into the storage hotel. The storage hotel retrieves the MTPs on a user-defined schedule and delivers them to the visualization system, which is a computer controlled camera mounted on an xy-table. The visualization system combines the information from Robots I and II with respect to the reservoir composition and the information on the protein with the imaging information and writes everything into a remotely accessible (e.g. web based) database.

With the advent of such high degrees of automation of crystallization experiments, new bottlenecks appeared. Probably the most serious, is the accumulation of an unprecedented number of experiments, which have to be evaluated. This has traditionally been done by experienced crystallographers. Naturally, these experimenters also decided *in situ* which experimental results to consider in the absence of crystal producing experiments and which parameters to vary in future experiments, based on all experiments examined. In the future, this task has to be done by software programs. The urgency for the development of such software is recognized and under way in commercial and academic settings.

References

1. Ducruix A, Giegé R (1992) In: Ducruix A, Giegé R (eds) Crystallization of nucleic acids and proteins, Oxford University Press, New York
2. Jancarik J, Kim S-H (1991) Sparse matrix sampling: a screening method for crystallization of proteins. *J Appl Cryst* 24:409–411
3. McPherson A (1990) Current approaches to macromolecular crystallization. *Eur J Biochem* 189:1–23
4. Segelke BW, Rupp B (1998) Beyond the sparse matrix screen: a web service for randomly generating crystallization experiments. *ACA Meeting Series* 25:78
5. Segelke BW (2001) Efficiency analysis of sampling protocols used in protein crystallization screening. *J Crystal Growth* 232:553–562

Protein Crystallography

- Protein Crystallization for X-Ray
- Structure-Based Drug Design
- X-Ray Crystallography

Protein Data Bank

Definition

The Protein Data Bank (PDB; ►<http://www.rcsb.org/pdb/>) was established at Brookhaven National Laboratories in 1971 as an archive for biological macromolecular crystal structures, and is now the single worldwide archive of structural data of biological macromolecules.

- MAD Phasing
- Protein Databases
- Structural Genomics: Structure-to-Function Approaches

Protein Databases

MIN-SEOK KWON^{1,2}, SANG YUN CHO¹,
YOUNG-KI PAIK^{1,2}

¹Yonsei Proteome Research Center, Biomedical
Proteome Research Center, Seoul, Korea

²Department of Biochemistry, Yonsei University,
Seoul, Korea
paiky@yonsei.ac.kr

Definition

A ►protein database is a collection of data that has been constructed from physical, chemical and biological information on sequence, ►domain structure, function, three-dimensional structure and protein-protein interactions. Collectively, protein databases may form a protein sequence database. It is therefore important to use appropriate protein databases which can 1) analyze and store data pertaining to protein science and 2) facilitate usage of analytical software available to the scientific community. Protein databases can generally be divided into two types. The first type is a universal database, which covers the proteins present in all known biological species. The second type is a specialized database, as described here, which

deals with the proteins belonging to a specific group or family of proteins of certain species (1). Each protein database can be further classified into more specialized categories according to the type of information sought.

Characteristics

Recently, massive numbers of data sets have been accumulated due to the rapid development of various high throughput analytical techniques and tools. This increases data redundancy and brings even more difficulties in data analyses. To standardize data formats and minimize redundancy, there have been worldwide efforts, as follows:

- 1) Protein data format standardization through the Human Proteome Organization (HUPO) Proteomics Standards Initiatives (PSI). Due to the tremendous increases in uncontrolled vocabulary and differing formats (4), HUPO launched PSI in 2001 to standardize informatics tools and to promote integration between databases, using ontology-based, controlled vocabulary and ►XML format. The PSI (►<http://psidev.sourceforge.net>) aims to define community standards for data representation in proteomics to facilitate data comparison, exchange and verification. PSI currently focuses on two limited, feasible domains of proteomics data standardization, mass spectrometry and protein-protein interactions. In addition, there is a clear need for a modular, integrative standard for functional ►genomics data. Intensifying these efforts in this domain is the major challenge for the immediate future.
- 2) Provision of a highly curated protein index, IPI (International Protein Index). Use of different identification systems may lead to increasing redundancy between protein databases. To minimize this redundancy, the ►EBI research team has been working to establish a new protein index that allows minimal overlaps between species (human, rat and mouse).
- 3) Construction of controlled vocabulary. The Gene Ontology Consortium (GOC) has produced a controlled vocabulary that can be applied to all organisms even as knowledge of gene and protein roles in cells is accumulating and changing. The GOC provides three structured networks of defined terms to describe gene product attributes. The GOC has a hierarchical structure that employs three principles, molecular function, biological process and cellular components. Use of ►gene ontology not only makes the role of proteins clearer but also minimizes protein redundancy.
- 4) Use of human readable and machine readable file format. XML represents an object-oriented file format and contains hierarchical structure, making

it suitable for expression of biological data and easy parsing of each program. This is a primary reason why many protein databases use XML file format.

Categories of a Protein Database

Because protein datasets are being created from various experimental groups, it would be desirable to provide appropriate databases to meet their needs. Currently there are several types of protein database available to the public. Although commercial protein databases are not described here, it is presumed that many private protein databases are operating *via* paid subscriptions. Protein databases can be further classified into more specialized categories according to the type of information sought.

- Protein sequence database
- Protein structure database
- Protein-protein interaction database
- Protein pattern and profile database
- 2D PAGE database
- Metabolic and pathway database
- Signaling pathway

Protein Sequence Database

A protein sequence is composed of 20 different amino acids; this sequence is termed the **primary structure** of a protein. The primary amino acid sequence is not only related to the three dimensional structure and function of a protein but also to the protein identity. This type of protein database, which collects amino acid sequences of proteins and related information, is termed a protein sequence database. Protein sequences can be determined experimentally or translated from nucleotide sequences found in nucleotide libraries.

Examples

- **Entrez Proteins** (Protein Database in Entrez at **NCBI**) has been compiled from a variety of sources, including Swiss-Prot, PIR, **Protein Research Foundation PRF**, PDB and **translations** from annotated coding regions in **GenBank** and **RefSeq**. This database can retrieve protein sequence records by accession number, author name, organism, gene/protein name and a variety of other text terms. For retrieval of large data sets, **Batch Entrez** is available. Entrez Proteins also includes **BLink** (BLAST Link), a feature that displays **BLAST** search results for every protein sequence in the Entrez Proteins data domain. For access, follow the BLink link displayed beside any hit in the results of an Entrez Proteins search.
- Swiss-Prot is a highly informative cross-referenced, protein-sequence library available through the Internet. This protein sequence database provides a high level of **annotations** (such as a description of the

function of a protein, its domain structure, **post-translational modifications** and variants) and a high level of cross-linked integration with about 60 external databases. This site is jointly managed by the **Swiss Institute of Bioinformatics** (SIB) and EBI (European Bioinformatics Institute). As of 7 Feb 2006, Swiss-Prot contains 207,132 sequence entries.

- TrEMBL is a computer-annotated supplement of Swiss-Prot that contains all the translations of **EMBL** (European Molecular Biology Laboratory) nucleotide sequence entries not yet integrated in Swiss-Prot. 2,605,584 TrEMBL sequence entries are available as of 7 Feb 2006.
- PIR provides the protein sequence database (PSD) of functionally annotated protein sequences, which grew out of the Atlas of Protein Sequence and Structure (1965–1978). PIR has been incorporated into an integrated knowledge based system of value-added databases and analytical tools. Since 2002, PIR-PSD has been integrated into UniProt knowledgebase.
- There are many specialized protein sequence databases. Although the total size of such databases is small, they are well classified according to specific subjects and therefore may be very practical for biologists in specialized fields. In many cases, this type of database provides not only protein sequence data on the specific organism or family, but also related information such as annotation data, pathway data and structural information.

Protein Structure Databases

Protein structure determines function, given that the specificity of active sites and binding sites depends on the precise three-dimensional conformation. **Nuclear magnetic resonance** (NMR) and **X-ray crystallography** are the most important techniques in determining protein conformation. Protein structure databases contain information related to protein three-dimensional structure and **secondary structure** obtained from analyses by X-ray crystallography and NMR.

Examples

- PDB, a typical, worldwide repository for processing and distribution of 3-D structure data on large molecules of proteins, nucleic acids and other biological macromolecules, is operated by the Research Collaboratory for Structural Bioinformatics (RCSB). Using the Protein Structure Explorer interface, the molecular structure, PDB files and file headers can be viewed as HTML and downloaded. A significant number of proteins are present in PDB for a variety of reasons, either because they have been crystallized under different conditions or by different groups or because they were subjected to

site-directed mutagenesis and different entries have been deposited for each of the variants. Common molecular structure visualization program tools such as ►RasMol (►<http://www.openrasmol.org/>) and ►Swiss-Pdb Viewer (►<http://kr.expasy.org/spdbv/>) are used to see 3-D protein structures. As of June 2005, 31,217 structures have been deposited in PDB, 85% are from X-ray crystallographic studies and 15% are from NMR studies.

- CATH refers to a database in which proteins are classified at the level of class, architecture, topology and homologous superfamily. This database provides a good lexicon of protein structure descriptions with a new classification hierarchy. Structures within each H-level are further clustered on sequence identity.
- The DALI domain dictionary is based on the automatic classification of protein domains by sequencing identity. This is constructed by clustering protein neighbors within an abstract file space.
- DSSP (Dictionary of Secondary Structure in Proteins) is a database containing derived information on secondary structure and solvent accessibility for protein 3-D structures stored in PDB.
- PSSH (protein sequence-to-structure homologies) are derived from HSSP2, an improved version of the ►HSSP (homology-derived secondary structure of proteins) database. Whereas each HSSP entry lists all protein sequences related to a given 3D structure, PSSH is the inverse, with each entry listing all structures related to a given sequence.
- SCOP (structural classification of proteins) is a simple resource database in which known proteins are grouped by their secondary structure properties (e.g. coiled coil, all-alpha, all-beta and mixed alpha-beta structures).
- Swiss-3D image is a database of high-quality annotated images of biological macromolecules with known 3-D structure. The goal of this image collection is to provide non-expert users with essential structural information about any particular protein.

Protein-Protein Interaction Databases

A protein-protein interaction database is constructed on the basis of protein-protein interaction information obtained from yeast two-hybrid, co-purification, affinity column chromatography, *in vitro* binding and IP/coIP methods.

Examples:

- BIND (biomolecular interaction network database) is a complex database, which contains interaction, molecular complex and pathway records. This database will span the complexity of interaction

gathered through experimental studies of biomolecular interactions.

- DIP (database of interacting proteins) catalogs experimentally determined interactions between proteins. It combines information from a variety of sources to create a single, consistent set of protein-protein interactions. The data stored within the DIP database were compiled manually by expert curators and also automatically using computational approaches that utilize knowledge about protein-protein interaction networks extracted from the most reliable core subset of DIP data.
- MINT (molecular interactions database) is a database designed to store functional interactions between biological molecules (proteins, RNA, DNA). Beyond cataloguing the formation of binary complexes, MINT was conceived to store another type of functional interaction, namely enzymatic modifications of one of the partner molecules. At present, MINT focuses on experimentally verified protein-protein interactions. Both direct and indirect relationships are considered.

Protein Pattern and Profile Databases

►Motifs can be detected in protein, DNA and RNA sequences, but the most common use of motif-based analysis is the detection of sequence motifs corresponding to structural or functional features in proteins. The use of protein sequence patterns or profiles to determine the function of proteins is an essential tool of sequence analysis (1, 5).

Examples:

- InterPro, an integrated documentation resource of ►protein families, domains and functional sites, was created in 1999 for amalgamation of the major protein signature databases into one comprehensive resource. ►PROSITE, ►Pfam, ►PRINTS, ►ProDom, ►SMART and ►TIGRFAMs have been manually integrated and curated and are available in InterPro for text- and sequence-based searching. The latest release of InterPro (10.0 version) contains 11,972 entries describing 8,597 families, 3,079 domains, 228 repeats and 20 post-translational modifications (3).

2-D PAGE Databases

As ►proteome research progresses rapidly, databases containing 2-D gel profiling and annotation may have a significant impact on protein mapping of any biological sample. 2-D PAGE databases consist of gel image data obtained by scanning the 2-DE and textual information on gel spots with regard to molecular mass (M.W.), isoelectric point (pI), status information on the identified spot and cross-reference links.

Examples:

- SWISS-2D PAGE database is a federated 2-D gel database containing data on proteins identified on various 2-D PAGE and SDS-PAGE reference maps and provides a link to other related databases through active, online hypertext. With Swiss-Prot one may locate proteins on 2-D PAGE maps or display any region of a 2-D PAGE map where one might expect to find a protein.
- YPRC-PDS (Yonsei Proteome Research Center – Proteome Database System). With this ►[laboratory information management system](#), one can store, retrieve and analyze various data, including two-dimensional electrophoresis (2-DE) images and associated spot information obtained during studies of hepatocellular carcinoma (HCC) (2).

Metabolic Pathway Databases

Metabolic databases provide descriptive information on enzymes, biochemical reactions and metabolic pathways.

Examples:

- ENZYME is not a metabolic database in the strictest sense, as it is a repository of information relative to ►[enzyme](#) nomenclature. It is primarily based on the recommendations of the Nomenclature Committee of the International Union of Biochemistry and Molecular Biology (IUBMB) (NC-IUBMB 1992). ENZYME is used as the nomenclature source for enzyme names and reactions by most metabolic databases as well as by other biomolecular databases.
- The BioCyc collection of databases provides electronic reference sources on the pathways and genomes of different organisms. Currently, detailed organism-specific databases are available for 14 species. In addition, the MetaCyc metabolic pathway database contains literature-derived metabolic pathway data for 160 species.
- KEGG (Kyoto Encyclopedia of Genes and Genomes) is an ambitious effort to computerize the current knowledge of molecular and cellular biology in terms of the information pathways that consist of interacting molecules or genes. KEGG consists of four types of data, pathway maps, molecule tables, gene tables and genome maps. It builds a functional map representing metabolic and regulatory pathways.
- WIT (what is there) is a metabolic pathway reconstruction resource, that is the curators of WIT are attempting to reconstruct complete metabolic pathway models for organisms whose genomes have been completely sequenced.
- PathDB. Although it contains almost the same information as KEGG, PathDB enables users to find

any set of interrelated reactions. It also contains a pathway visualization interface that enables inspection of any selected pathway.

Signaling Pathway Databases

This signaling pathway database is to stimulate complementary research in individual laboratories and to facilitate access to necessary information on biological signaling pathways. This database can be classified into the following areas, depending on the format, for it contains both graph and tree-type data structures.

Examples:

- TRANSPATH is an information system on gene-regulatory pathways and an extension module to the TRANSFAC database system. Because this database provides tree, graph and hand drawn images, one can perform pathway queries using graph traversal algorithms.
- CSNDB (cell signaling network database) provides a tool for searching the pathway related to molecules through pathway finder, from pathways in human cells. This database displays starting- and end-points of each pathway and data representing collection of reactions, including nodes.
- SPAD (signaling pathway database) displays as an image the signaling pathway being used most frequently for a growth factor, cytokine, hormone or stress. Images can also be shown in hyperlink.

Clinical Relevance

A bioinformatician usually predicts the function of human proteins using the ►[homology](#), similarity, motif search and structure similarity of annotated proteins of other species. This may be useful for designing experiments to determine protein functions. Both EBI and SIB provide highly curated information on human proteins through HIS (Human Proteomics Initiative). This has also been applied to screen diagnostic marker proteins using a search program that is protein sequence-based mass spectrometry.

References

1. Bairoch A (1997) Proteome Database. In: Wilkins MR, Williams KL, Appel RD (eds) et al Proteome Research: New Frontiers in Functional Genomics. Springer, Berlin Heidelberg New York, pp 93–129
2. Cho YS, Park K-S, Shim JE et al (2002) An integrated proteome database for two-dimensional electrophoresis data analysis and laboratory information management system. *Proteomics* 2:1104–1113
3. Baxevanis AD (2003) The Molecular Biology Database Collection: 2003 update. *Nucleic Acids Res* 31:1–12

4. Boguski MS, McIntosh MW (2003) Biomedical informatics for proteomics. *Nature* 422:233–237
5. Gibas C, Jambeck P (2001) Developing Bioinformatics Computer Skills. O'Reilly, Pub Co, pp 133–327

Protein Degradation

Definition

Protein degradation refers to regulated digestion by proteinases and removal of protein (►[proteolytic degradation](#)).

- Repeat Expansion Diseases
- Two-Hybrid System

Protein Disulfide

Definition

Protein disulfide is a covalent bond between the side chains of two cysteine residues of a protein.

- [Protein Disulfide Bonds](#)

Protein Disulfide Bonds

ERVIN WELKER, MAHESH NARAYAN,
HAROLD A SCHERAGA

Baker Laboratory of Chemistry and Chemical Biology,
Cornell University, NY, Ithaca, USA
has5@cornell.edu

Definition

Proteins are long polymers of amino acids, and are essential for all biochemical processes within the cell. Most proteins can perform their various functions only when folded, i.e. if they have adopted a specific, stable structure. Many secretory proteins use disulfide(s) that is/are covalent bond(s) between the side chains of two cysteine residues in order to ensure extra stability. It is achieved by imposing distance and angle constraints between the C^β and S^γ atoms of the joined cysteine residues, thus destabilizing the unfolded state by

reducing its entropy. Disulfide bonds may also stabilize the folded state enthalpically through favorable local interactions, e.g., by contributing to the packing of a local cluster of hydrophobic residues.

►[Protein disulfide](#) bonds are formed in the endoplasmic reticulum of the cell both co- and post-translationally. This composite process by which a reduced, unfolded protein recovers both its native disulfide bonds (disulfide-bond regeneration) and its native structure (conformational folding) is called oxidative folding. This process is facilitated *in vivo* by folding-catalysts such as protein disulfide isomerase (PDI)(1). Nevertheless, our basic understanding of the oxidative folding process comes from *in vitro* studies of multiple disulfide-containing single-domain proteins (2, 3, 4, 5). Protein disulfide bonds are also used as a tool to study a variety of conformational properties of proteins (6).

Characteristics

Disulfide-Bond Reactions

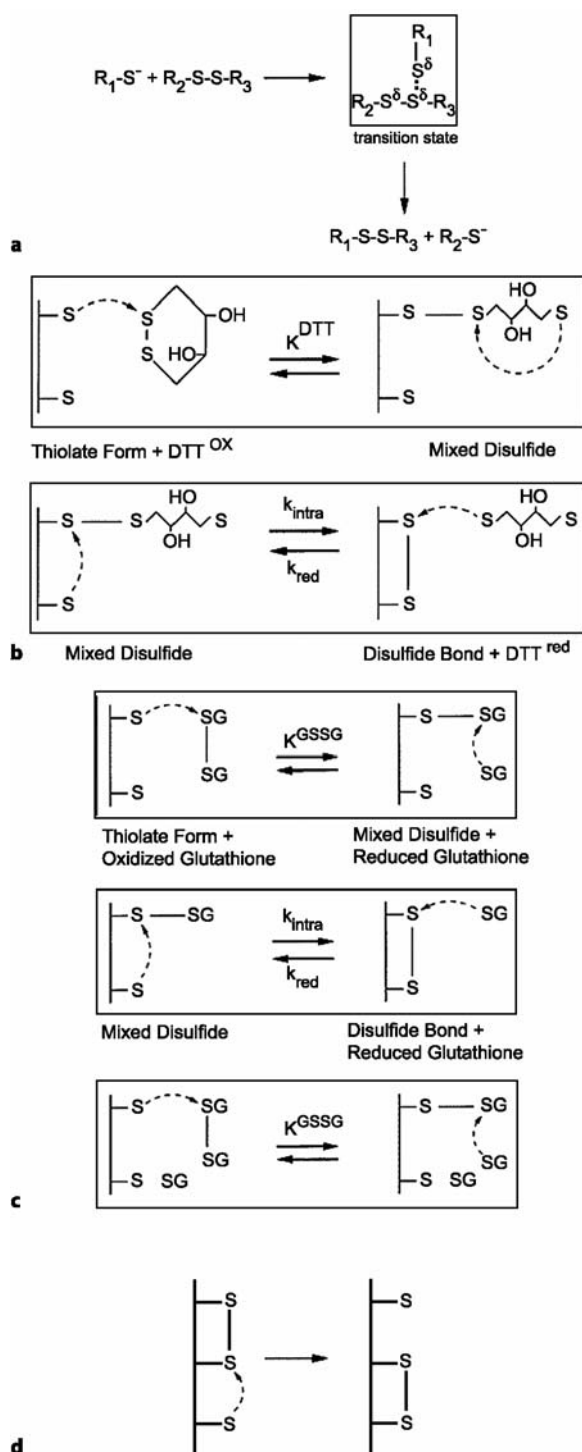
Oxidation, reduction and reshuffling are based on a single chemical reaction, thiol/disulfide exchange, in which a thiolate anion approaches a disulfide bond closely enough to react (Fig. 1).

Factors Affecting the Course of Disulfide-bond Regeneration

The course of oxidative folding is affected by three general structural factors that have been identified from *in vitro* oxidative folding studies and model systems, namely the proximity, reactivity and accessibility of the disulfide bonds and thiol groups. The proximity of two reactive groups (defined as their effective intramolecular concentration) is determined by the propensity of the backbone to bring the two groups into juxtaposition; in unfolded species, this proximity is largely determined by the loop entropy, although enthalpic interactions may contribute significantly as well. The reactivity of the two groups depends on the fact that most disulfide reactions occur through thiol/disulfide exchange and that only the thiolate (not the thiol) form is reactive; hence, changes in pH and in the local electrostatic environment may affect the rate of disulfide-bond reactions. However, in structured intermediates, the most critical factor seems to be the accessibility of the thiol groups and disulfide bonds. Thiol/disulfide exchange reactions can occur only when a thiol and a disulfide bond come into contact; hence, burial of the disulfide bond or the thiol by the protein structure prevents their contact and blocks the reaction.

The Stages of Oxidative Folding

Accordingly, the formation of stable tertiary structure that protects the native disulfide bonds is a key event in the oxidative folding of proteins, because it stabilizes



Protein Disulfide Bonds. Figure 1 Chemistry of disulfide bond reactions. (a) In thiol/disulfide exchange, a thiolate anion R_1S^- displaces one sulfur of a disulfide bond R_2SSR_3 . Protein disulfide bonds are formed and reduced by two such thiol/disulfide exchange reactions with a redox reagent, the first of which involves the formation of a *mixed disulfide bond* between the protein and the redox reagent. For completeness, the reactions

such bonds by making them inaccessible to the redox reagent and to protein thiols (in structured intermediates). Indeed, the rate of intramolecular **▶disulfide reshuffling** is sufficiently high to ensure that no native disulfide bond would survive much longer than a few minutes in an unstructured disulfide intermediate with a free thiol group, unless it was buried in protective tertiary structure. Thus, the rate-determining step in oxidative folding is often the formation of a disulfide-containing intermediate with stable tertiary structure. Accordingly, the oxidative folding process may be divided into pre- and post-folding stages. The prefolding stage is characterized by rapid interconversions of unstructured **▶disulfide species** by disulfide reshuffling, at rates determined largely by loop entropy and the conformational biases of the polypeptide chain (given a pH around the physiological values). By contrast, the postfolding stage is characterized by folded, native-like disulfide species, in which the reaction rates may deviate strongly from the corresponding rates in unstructured disulfide species.

Prefolding Stage

The prefolding stage of oxidative folding of multiple-disulfide-containing proteins consists of disulfide reactions among unstructured disulfide species. The reduced protein is successively oxidized to populate ensembles of species with each ensemble having the same number of disulfide bonds, which establishes a pre-equilibrium (steady-state-like condition) among the ensembles (i.e., the rates of formation and consumption of a particular ensemble are the same). The disulfide species within each unstructured ensemble likewise reach a quasi-equilibrium distribution, because the disulfide reshuffling within the ensembles is generally more rapid than the redox reactions between the ensembles [the fully oxidized ensemble (scrambled species) is an exception because it has no free thiols to facilitate disulfide reshuffling]. Hence, the distribution of disulfide species within each ensemble is relatively insensitive to the redox conditions and the relative concentrations of the other unstructured ensembles. These pre-equilibrium and quasi-equilibrium conditions may be observed qualitatively in the HPLC elution profiles of the unstructured ensembles, each of which is relatively unchanged throughout the regeneration.

are illustrated with (b) cyclic (DTT^{red}/DTT^{ox}) and (c) linear (GSSG/GSH) redox reagents. (d) Thiol/disulfide exchange reactions can also occur intramolecularly; e.g., a protein thiolate group may attack a disulfide bond of the same protein, leading to *disulfide reshuffling*.

Folding-Coupled Regeneration Steps

When a disulfide species acquires a critical number of native disulfide bonds that are enough to stabilize the protein sufficiently under the given conditions, it folds conformationally. There are alternative interpretations for the mechanism of these folding-coupled regeneration steps that are the subject of current research (7). Through these folding-coupled regeneration steps, all the unstructured species constituting the pre-equilibrium distribution of ensembles convert to a structured one, and/or to the native protein.

Postfolding Stage

The postfolding stage consists of disulfide reactions that take place in structured species. Such intermediates can significantly alter the subsequent regeneration by burying reactive groups in stable tertiary structure, thereby inhibiting their rearrangement by thiol/disulfide exchange. Two kinds of structured intermediates are distinguished, disulfide-secure, and disulfide-insecure. Productive intermediates tend to be disulfide-secure, meaning that their structural fluctuations preferentially expose their thiol groups, while keeping their disulfide bonds buried. By contrast, dead-end species tend to be disulfide-insecure, in that their structural fluctuations expose their disulfide bonds in concert with their thiol groups. Under typical oxidative folding conditions, ►**disulfide-secure species** preferentially oxidize to the native protein, whereas ►**disulfide-insecure species** preferentially reshuffle to the unstructured ensemble. Obviously, the ratio between reshuffling and oxidation depends on the redox conditions as well; at very low concentrations of an oxidizing agent, even disulfide-secure species will preferentially reshuffle, albeit at slow rates. However, the distinction between disulfide-secure/disulfide-insecure is a conformational one that is independent of the redox conditions, depending only on the relative free energies for structural fluctuations that expose the thiol groups and disulfide bonds. Stable conformational structure in productive disulfide-secure species seems to be the critical factor accelerating the regeneration of the native protein. This regeneration becomes drastically slower if it is carried out under conditions that destabilize the tertiary structure of these intermediates. The rapid regeneration can be restored if other agents (e.g., phosphate) are added to restabilize the conformational structure.

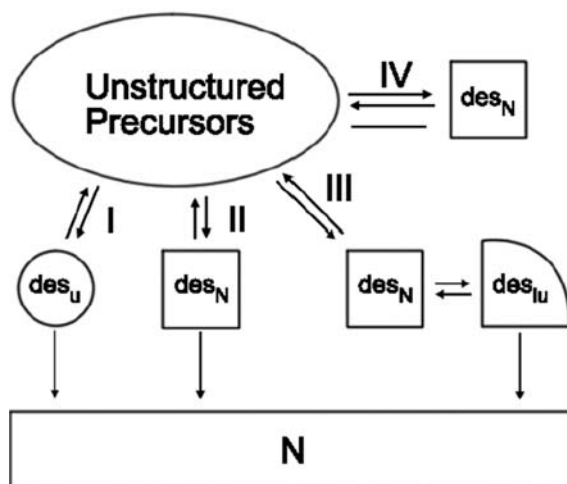
Generic Pathways of Oxidative Folding of Single Domain Protein

Four generic pathways of oxidative folding can be discerned from oxidative folding studies, depending on whether any disulfide intermediates are structured, whether the structured disulfide species are metastable (i.e., bury both their disulfide bonds and thiol groups,

thereby inhibiting them from subsequent reactions) and whether the ►**metastable disulfide species** are disulfide-secure or -insecure (Fig. 2).

Protein Disulfides as a Tool to Study Conformational Properties of a Protein

Since disulfide-bond reactions are inhibited by the burial of reactive groups (such as protein disulfide



Protein Disulfide Bonds. Figure 2 Four generic types of oxidative folding pathways for a typical disulfide-containing protein. For simplicity, we assume that only des species are capable of folding to a native-like structure (indicated by the N subscript). The U and Iu subscripts indicate that the corresponding species are unfolded or locally unfolded, respectively (although such species may have some conformational order, as observed in ribonuclease A). In pathway I, the native protein N regenerates directly from an unfolded precursor des_U (a des_U pathway). Pathway II involves the formation of a folded disulfide-secure intermediate, des_N that is not metastable; des_N is oxidized directly to the native state (a des_N pathway). Pathway III likewise involves a folded intermediate des_N that oxidizes preferentially to the native protein but is metastable (a metastable des_N pathway). To become oxidized to the native protein, this metastable intermediate must undergo a local unfolding step (des_{IU}). Finally, pathway IV involves a folded, metastable disulfide-insecure species des_N that preferentially reshuffles rather than being oxidized to the native protein (a metastable dead-end pathway). The rate-determining steps of pathways I and II are associated with the formation of the structured species, while that of the metastable dead-end pathway IV is associated with conformational unfolding of the metastable species. In this respect, pathway III is intermediate between pathways II and IV; the rate-determining step may be either the formation or the (local) unfolding of the structured intermediate, depending on the conditions and the δG of local unfolding required to oxidize the metastable species.

bonds, ►mixed disulfide bonds, and protein thiol groups) in stable tertiary structure, disulfide-bond reactions can monitor the accessibility of the reactive groups and indirectly assay the stability of the protective tertiary structure, similar to H/D exchange and proteolysis experiments. The rate of disulfide-bond formation also depends on the proximity of the two cysteine residues, defined here as the probability of their sulfur atoms coming within the distance required for thiol/disulfide exchange. This rough assay of inter-residue distances can be used to characterize secondary, tertiary, and quaternary structure in proteins and complements other distance-measuring methods such as fluorescence resonance energy transfer (►FRET) and the nuclear Overhauser effect (NOE) of NMR spectroscopy. As mentioned previously, disulfide-bond reactions are also sensitive to the reactivities of the thiolate and disulfide groups, which depend on the pH of the solvent and the pK_a values of the thiol groups involved and, more generally, on the electrostatic environment of the reactive groups.

Disulfide Methodologies

The instantaneous and quantitative blocking of the thiols are critical technical issues in all protein disulfide studies. Improper blocking has been the origin of many contradictory results in the field. There are basically three blocking methods in use, namely alkylation of thiols, “freezing” of the reaction by dropping the pH, and reversible blocking techniques that utilize reagents such as 2-aminoethylmethanethiosulfonate (AEMTS). While alkylation of the thiols is an irreversible blocking technique, it is not generally sufficiently rapid to compete with the fast intramolecular disulfide reshuffling. Thus, it is successfully applied to block disulfide-secure intermediates, where this reshuffling does not take place on the time scale of blocking and to block unstructured ensembles where the blocking does not disturb the equilibrium distribution of the rapidly interconverting species. Quenching of the reaction by lowering the pH or by addition of a reversible blocking reagent such as AEMTS (in excess of all free thiols) can occur sufficiently rapidly to freeze all disulfide reactions; however, extra precautions must be taken to avoid subsequent rearrangements during further processing of the sample.

References

1. Gilbert HF (1994) In: Pain RH (ed) *Mechanisms of Protein Folding*. W. H. Freeman, New York, pp 104–136
2. Creighton TE (1997) *Biol Chem Hoppe-Seyler* 378:731–744
3. Dadlez M (1997) *Acta Biochim Polon* 44:433–452
4. Schmid FX, Frech C, Scholz C et al (1996) *Biol Chem* 377:417–424

5. Narayan, M, Welker, E, Wedemeyer, WJ et al (2000) *Acc Chem Res* 33:805–812
6. Wedemeyer, WJ, Welker, E, Narayan, M et al (2000) *Biochemistry* 39:4207–4216
7. Welker, E, Wedemeyer, WJ, Narayan, M et al (2001) *Biochemistry* 40:9059–9064

Protein Domains

JÖRG SCHULTZ

Department of Bioinformatics, Biocenter, Würzburg University, Würzburg, Germany
joerg.schultz@biozentrum.uni-wuerzburg.de

Definition

When the first structures of proteins were solved by ►X-ray crystallography, it became clear, that especially larger proteins could consist of different structural subunits, which were named domains. Surprisingly, some of these domains were found in otherwise unrelated proteins. This observation gained importance with the rise of automated sequencing and the parallel invention of algorithms for the detection of remote homologues. These developments facilitated the detection of domains in a plethora of different proteins, indicating that the shuffling of domains was an important mechanism in the evolution of proteins. Therefore, domains cannot only be seen as structural units, but also as units of evolution. With a size of 30–400 amino acids, most domains can fold independently around a relatively hydrophobic core, further underlining their role as independent building blocks. Although structure determination is an important part of the understanding of a domain family, nowadays domains can be discovered in proteins using sequence only. With the rise of large sequencing projects, this turned out to be a powerful tool to fight one of their major problems, namely the step from sequence to function. Biochemical analysis showed that domains of the same family detected in different proteins perform similar functions. As therefore domains also represent units of function, detecting a domain in a protein sequence can give the first clues about its function. In summary, domains are evolutionarily conserved units of structure and function.

Description

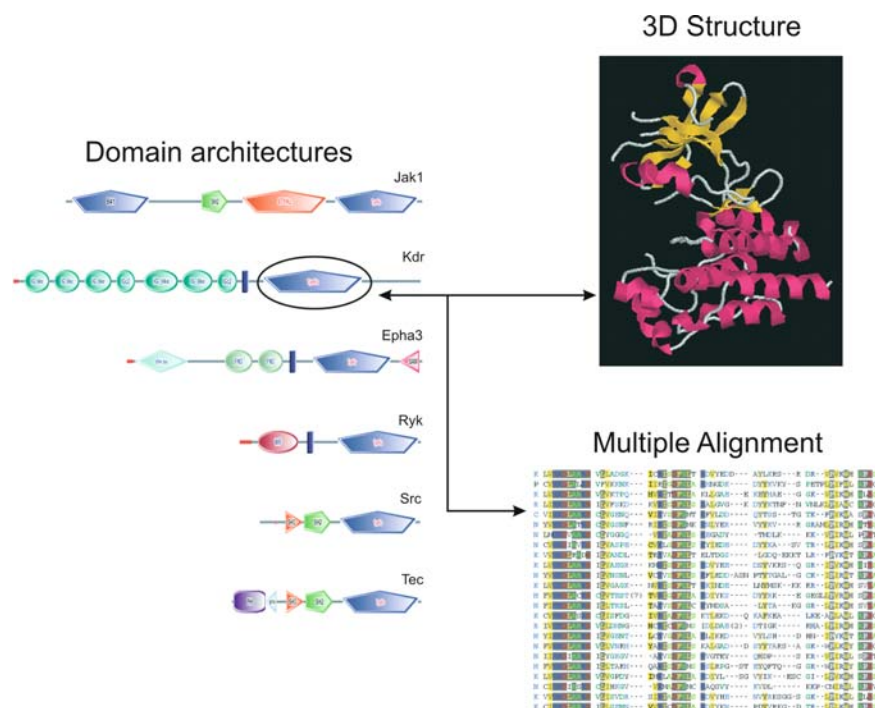
One of the basic questions in sequence analysis is how one could predict functional features of a protein using sequence data only. To address this problem, one makes use of evolution. If two sequences are sufficiently

similar, one can assume that they have originated from a common ancestor, they are homologs. In this case, sequence similarity implies a similar structure and, thereby, a similar function. If the function of one of the homologous proteins is known, one can transfer this knowledge to the other proteins, predicting their function. Obviously, the detail of prediction is highly correlated with the degree of sequence similarity and care has to be taken that one does not end up with overpredictions. A crucial difficulty in this process arises from multi-domain proteins, as here the homology is restrained to defined parts of the proteins. This is illustrated by the proteins depicted in Fig. 1, for example Src and EphA3. Both contain a tyrosine kinase domain (TyrKc), thus the functional aspect of phosphorylating other proteins is conserved between them. But, other regions of the proteins do not show any similarity. In this example, EphA3 is a receptor protein kinase with a transmembrane region and extracellular domains, whereas Src is cytoplasmic. Transferring functions like the binding of ephrins from EphA3 to the Src kinase would obviously be wrong. This example should illustrate, that a careful analysis of the domain content of a protein is a crucial step in the prediction of

its function and should precede any more general homology searches.

Whereas in the example above domains can be used to understand the detailed molecular function of a protein, they can, on the large scale, provide a first functional classification of proteins. Within the analysis of the human genome, different domain families were grouped into broader functional classes and proteins containing these domains were assigned into these classes. It turned out that humans, compared to *Drosophila melanogaster* and *Caenorhabditis elegans*, have more proteins involved in cytoskeleton, defence and immunity, and transcription and translation. These expansions can be related to vertebrate physiology (1).

As outlined above, domains are not only units of function but also of evolution. Their continuous re-use in different proteins leads to the question of their evolutionary origin. Using sensitive methods for the detection of even very divergent members of a domain family (see below) this question was addressed by searching in different bacterial, archaeal and eukaryotic genomes. It was found, that there is a fundamental difference between catalytic and non-catalytic



Protein Domains. Figure 1 Different representations of domains – The domain architecture view shows the whole protein and all domains found within. It illustrates the re-use of the tyrosine kinase domain (TyrKc) within a different context. The multiple sequence alignment of the TyrKc family (only a section is shown here) highlights sites that are conserved throughout all members of the family. It also shows sites that are in corresponding positions within the three dimensional structure. The structure shown here is of the kinase domain of the human endothelial growth factor receptor 2.

domains. Whereas the former seem to have been already present in the last common ancestor, the non-catalytic domains have mainly been evolved in the eukaryotes. The analysis also revealed that a substantial number of domains has been horizontally transferred from eukaryotes into bacteria, further underlining the re-usability of domains as fundamental building blocks in protein evolution (2).

Identification of New Domains

The importance of domains for the evolution, structure and function of a protein raises the question as to how they can be detected within a protein sequence in the first place. In the following, the general procedure will be outlined; detailed descriptions, hinting at many pitfalls, can be found in different reviews (3, 4).

Within the process of predicting the function of a protein, one usually starts with a ►homology search against all currently known protein sequences. The most commonly used tool here is the ►BLAST program suite (►<http://www.ncbi.nlm.nih.gov/BLAST>). In addition to finding ►orthologs of the protein in other species and additional ►paralogs, one might find a partial homology to other proteins. As proteins sharing one domain are homologous only in the region of the domain, the homology found might be the first sign of the existence of an independent domain in the query protein. To further validate this assumption, it is useful to restrain the search to the region detected in the partial homology. After these searches have uncovered the first members of the domain family, more sophisticated methods should be used to uncover additional divergent members of the family. Here, one usually follows two directions. First, one can use a newly detected homologue as the starting point for a further search. This reciprocal search on the one hand strengthens the homology, if the original sequence is found. On the other hand, one will find additional sequences not detected in the first search, caused by an asymmetry in the BLAST algorithm. Obviously, these can be the starting points for an additional round of searches further gathering remote homologues. The second direction uses information gained by combining multiple members of a domain family. The most straightforward approach here is to perform ►PSI-BLAST searches. These automatically build a multiple sequence alignment of all homologues found. In the next search, this alignment rather than a single sequence is used to identify new members, which then are iteratively added to the search alignment. The same principle is used in hidden-Markov-model (►HMM) searches. HMMs are a statistical representation of an alignment, including preferences of amino acids at specific sites and gaps/insertions, which can be used to search against sequence databases (a frequently used implementation is HMMer, which can be found at

►<http://hmmer.wustl.edu>). Retrieved novel sequences have to be added to the alignment, which will then be used as a starting point for a next iteration of HMM searches. Although this procedure will be laborious, the high sensitivity of HMM searches will result in the detection of more members of the new domain family. In general, combinations of the methods mentioned will lead to the best description of the domain family. A cornerstone in this process is the generation of a reliable ►multiple sequence alignment as it allows the highlighting of features common to the whole domain even when sequence identity is low. It is not only the basis for further homology searches; careful interpretation of the alignment will give the first hints about the structure and possible functional residues of the domain. It has to be noted that although the quality of alignment programs is still improving, manual optimisation of the alignment will be needed to gain maximal information.

Having extracted as many members of a domain family as possible, the next step is the elucidation of the function of the domain. Here, one has to rely on previously published experimental characterisations of proteins containing this domain. These might report mutations of sites within the domain or even deletion experiments. Experimental results obtained with single proteins might hint at the function of the domain family.

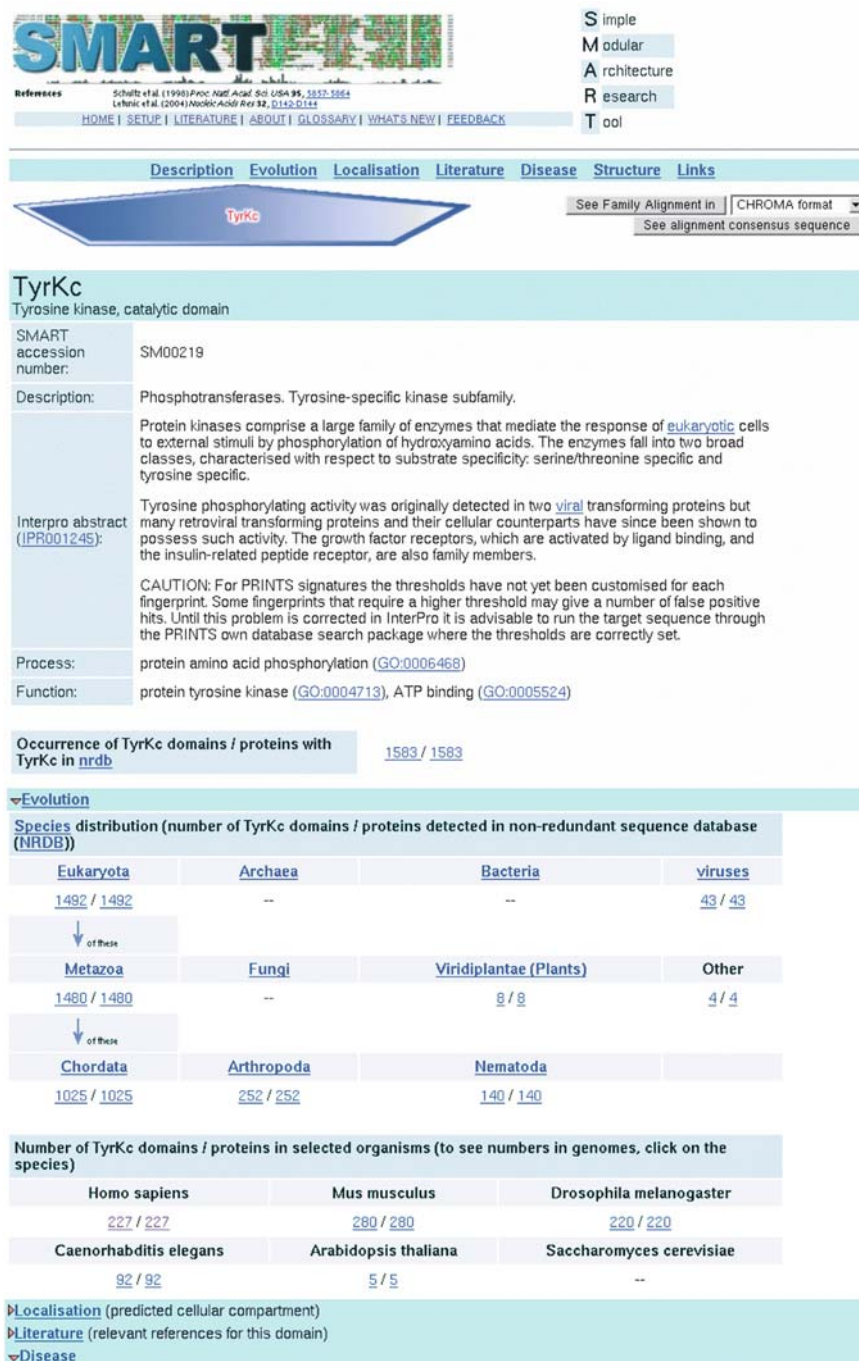
Domain Databases

As outlined above, the detailed description of a domain family is laborious and contains many pitfalls. Fortunately, there exist databases of domain families and Web servers, which allow the automatic identification of domains in given sequences. With the common goal of domain identification, different resources are based on different approaches. The cores of many of them are multiple sequence alignments of representative members of the domain family. As their quality is crucial for the sensitivity and selectivity of domain identification, most databases rely on manually curated alignments. Based on them, Pfam (►www.sanger.ac.uk/Software/Pfam), SMART (►databasesmart.embl-heidelberg.de) and ►TIGRFAM (►www.tigr.org/TIGRFAMs) use HMMs for the actual detection of domains. These databases provide for each of the domain families a manually adjusted threshold for the distinction between members and false positives. Other databases like Blocks (blocks.fhcrc.org) use shorter, ungapped alignments. A different approach, used by CDD (►www.ncbi.nlm.nih.gov/Structure/cdd/cdd.shtml), is the automatic, intrinsic generation of these alignments by storing profiles generated by PSI-Blast searches. Possibly the most straightforward, but also less accurate method is the identification by domain specific patterns as used by Prosite (►www.expasy.org/prosite). In addition to these original domain databases, there exist

metadatabases, which combine the results of searches within different domain databases, for example InterPro (www.ebi.ac.uk/interpro). In general it is advisable to search in several of these databases, as the type of detected domains and the sensitivity may differ.

A great advantage of all these databases is that they offer additional information about each domain. In the

following, the SMART domain database will be briefly discussed as an example of the domain annotations provided by these databases (Fig. 2). In the case of the SMART database, manual annotation is combined with automatically generated information. Within the manual annotation a short description of the function of the domain family is given and references to publications



SMART

References: Schultze et al. (1998) *Proc Natl Acad Sci USA* 95, 5852-5864; Lohr et al. (2004) *Nucleic Acids Res* 32, D142-D144

HOME | SETUP | LITERATURE | ABOUT | GLOSSARY | WHAT'S NEW | FEEDBACK

Simple
Modular
Architecture
Research
Tool

Description Evolution Localisation Literature Disease Structure Links

TyrKc

See Family Alignment in CHROMA format
See alignment consensus sequence

TyrKc
Tyrosine kinase, catalytic domain

SMART accession number: SM00219

Description: Phosphotransferases, Tyrosine-specific kinase subfamily.

Interpro abstract (IPR001245): Protein kinases comprise a large family of enzymes that mediate the response of *eukaryotic* cells to external stimuli by phosphorylation of hydroxyamino acids. The enzymes fall into two broad classes, characterised with respect to substrate specificity: serine/threonine specific and tyrosine specific.

CAUTION: For PRINTS signatures the thresholds have not yet been customised for each fingerprint. Some fingerprints that require a higher threshold may give a number of false positive hits. Until this problem is corrected in InterPro it is advisable to run the target sequence through the PRINTS own database search package where the thresholds are correctly set.

Process: protein amino acid phosphorylation (GO:0006468)

Function: protein tyrosine kinase (GO:0004713), ATP binding (GO:0005524)

Occurrence of TyrKc domains / proteins with TyrKc in nrdb 1583 / 1583

Evolution

Species distribution (number of TyrKc domains / proteins detected in non-redundant sequence database (NRDB))

Eukaryota	Archaea	Bacteria	viruses
1492 / 1492	--	--	43 / 43
↓ of these			
Metazoa	Fungi	Viridiplantae (Plants)	Other
1480 / 1480	--	8 / 8	4 / 4
↓ of these			
Chordata	Arthropoda	Nematoda	
1025 / 1025	252 / 252	140 / 140	

Number of TyrKc domains / proteins in selected organisms (to see numbers in genomes, click on the species)

Homo sapiens	Mus musculus	Drosophila melanogaster
227 / 227	280 / 280	220 / 220
Caenorhabditis elegans	Arabidopsis thaliana	Saccharomyces cerevisiae
92 / 92	5 / 5	--

Localisation (predicted cellular compartment)

Literature (relevant references for this domain)

Disease

Protein Domains. Figure 2 Continued

SwissProt sequences and [OMIM](#) curated human diseases associated with missense mutations within the TyrKc domain.

Protein	Disease
(SRS)(SMART)	(OMIM:602216): Peutz-Jeghers syndrome (OMIM:175200):
Hepatocyte growth factor receptor precursor (EC 2.7.1.112) (Met proto-oncogene tyrosine kinase) (c-met) (HGF receptor) (HGF-SF receptor). (SRS)(SMART)	(OMIM:164860): Renal cell carcinoma, papillary, familial and sporadic (OMIM:605074): Hepatocellular carcinoma, childhood type (OMIM:114550):
Protein kinase C, gamma type (EC 2.7.1.37) (PKC-gamma). (SRS)(SMART)	(OMIM:176980): PROTEIN KINASE C, GAMMA; PRKCG (OMIM:601777): Cone dystrophy, progressive (OMIM:600179): Leber congenital amaurosis, type I (OMIM:204000): Cone-rod dystrophy 6 (OMIM:601777):
Retinal guanylyl cyclase 1 precursor (EC 4.6.1.2) (Guanylate cyclase 2D, retinal) (RETGC-1) (Rod outer segment membrane guanylate cyclase) (ROS-GC). (SRS)(SMART)	(OMIM:602337): Brachydactyly, type B1 (OMIM:113000): Robinow syndrome, autosomal recessive (OMIM:268310):
Tyrosine-protein kinase transmembrane receptor ROR2 precursor (EC 2.7.1.112) (Neurotrophic tyrosine kinase, receptor-related 2). (SRS)(SMART)	(OMIM:147670): Leprechaunism (OMIM:246200): Rabson-Mendenhall syndrome (OMIM:262190): Diabetes mellitus, insulin-resistant, with acanthosis nigricans
Insulin receptor precursor (EC 2.7.1.112) (IR) (CD220 antigen). (SRS)(SMART)	

Protein Domains. Figure 2 Annotation of the tyrosine kinase domain as given by the SMART domain database.

concerning its detection, structure and function are provided. The automatically generated information includes not only a summary of all proteins containing this domain, but also their phylogenetic breakdown. This allows a fast overview of the evolution of the domain but also of the spread within defined taxonomic subgroups. Furthermore links to all structurally characterised members are provided. A more recent addition includes outlinks to the [►OMIM](#) (online Mendelian inheritance in man) database, which characterises genes involved in diseases. Starting with a domain of interest, all genes linked to diseases with missense mutations in this domain can be retrieved.

In summary, current domain databases not only allow the fast and easy identification of domains within a sequence, they also offer a wide variety of additional information about the domains found. When analysing a new protein sequence, it is therefore advisable to search against these databases before performing more traditional homology searches by e.g. BLAST.

From Domain to Function

As noted above, one of the major reasons for analysing the domain architecture of a protein is to gain insight into its function. Indeed, after the identification of the first domains, it was assumed that all members of a domain family perform more or less identical functions. With ongoing sequencing and improvement of similarity search algorithms, more divergent members of a family were unravelled. Experimental characterisation revealed that the ‘one domain – one function’ hypothesis had to be rejected. An impressive example might be the PTB (phosphotyrosine binding) domain. The first characterised, name-giving members of this

domain family bound other proteins on phosphorylated tyrosines. Further experiments revealed that other members of the domain family could bind proteins on unphosphorylated tyrosines or even in a tyrosine independent manner. Only recently, PTB domains were found that do not bind proteins at all but phospholipids. A similar scenario can be established for many other domains. Therefore, to go the step from domain to function, further characterisation is needed. One current approach uses phylogeny to identify subgroups with similar functions within one domain family. Another identifies key functional sites and tries to classify members of a domain family based on the amino acid at this site.

Complementary to the detection that one domain family can perform multiple functions is the observation that different domain families have similar functions. In addition to the aforementioned PTB domain, the SH2 domain and subfamilies of the tyrosine phosphatases can work as phosphotyrosine adaptor domains. Similarly, FHA, BRCT, WW and WD40 can bind to phosphorylated serines and threonines. The combination of these modules in different proteins allows the building up of a complex and changing interaction network, crucial for regulation of cellular processes (5).

[►Classification of Active Centers](#)

[►Functional Assays](#)

References

1. Lander ES, Linton LM, Birren B et al (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860–921

2. Ponting CP, Aravind L, Schultz J et al (1999) Eukaryotic signalling domain homologues in archaea and bacteria. Ancient ancestry and horizontal gene transfer. *J Mol Biol* 289:729–45
3. Bork P, Gibson TJ (1996) Applying motif and profile searches. *Methods Enzymol* 266:162–84
4. Bateman A, Birney E (2000) Searching databases to find protein domain organization. *Adv Protein Chem* 54:137–57
5. Pawson T, Nash P (2003) Assembly of cell regulatory systems through protein interaction domains. *Science* 300:445–452

Protein Expression Profiling

Definition

Protein expression profiling designates a global analysis of protein expression status in a given cell or tissue (on, off, strong, weak, etc). Usually it is performed with 2-dimensional gel electrophoresis or mass spectrometry.

► [Proteomics in Human-Pathogen Interactions](#)

Protein Family

Definition

Protein family indicates a group of proteins, sharing regions with conserved amino acid sequences and tertiary structure, called domains, and performing similar functions. These proteins are encoded by sets of duplicated genes, called gene families.

► [Cancer Invasion and Metastasis: Cellular, Molecular and Clinical Aspects](#)

► [Protein Databases](#)

► [Protein Domains](#)

Protein Fold

Definition

Protein fold classifies and simplifies the three-dimensional (tertiary) structure of a protein. For large proteins, which are often assembled from a number of globular domains, it is often easier to characterize their tertiary structure by classifying the domain folds of these independent units.

► [Protein Folding](#)

► [Structural Genomics: Structure-to-Function Approaches](#)

Protein Folding

MATTHIAS P. MAYER

Center for Molecular Biology Heidelberg (ZMBH),
Heidelberg, Germany
m.mayer@zmbh.uni-heidelberg.de

Definition

Proteins are only functional when their amino acids are arranged in a defined three-dimensional structure. Protein folding is the process that leads from a non-native arrangement of the amino acids of a polypeptide chain, to the distinct spatial structure characteristic for the native protein. Thus, folding of newly synthesized polypeptides converts the linear, genetically stored information into a 3D-structure, which is subsequently subject to evolutionary selection. Protein folding is therefore the final part of gene expression after transcription, RNA-processing and translation. Protein folding processes are often, but not always, linked to translation occurring either cotranslationally, domain-wise or immediately posttranslationally. Proteins destined for posttranslational translocation across a biological membrane, for example for import into mitochondria, plastids, the ER or other organelles, unfold upon transition through the membrane, and refold in the interior of their target compartment. In the case of chloroplasts, proteins targeted for the lumen of the thylacoids fold and unfold repetitively until they have reached their final destination.

Protein folding processes comprise, in a wider sense, any formation and rearrangement of secondary structure elements and/or protein domains. The refolding of stress denatured proteins and large conformational changes in proteins due to ligand binding, protein-protein interactions or posttranslational modifications can, therefore, also be considered protein-folding processes (1).

Characteristics

Energy Landscapes and Folding Funnels

The entire information necessary to reach the native conformation is encoded in the amino acid sequence of a polypeptide, as originally demonstrated by Anfinsen, who was able to refold a denatured protein *in vitro* to

the native, active state. Thus, the two central questions of protein folding are: how the primary sequence of a polypeptide encodes the final spatial arrangement, and in which way the native structure is reached.

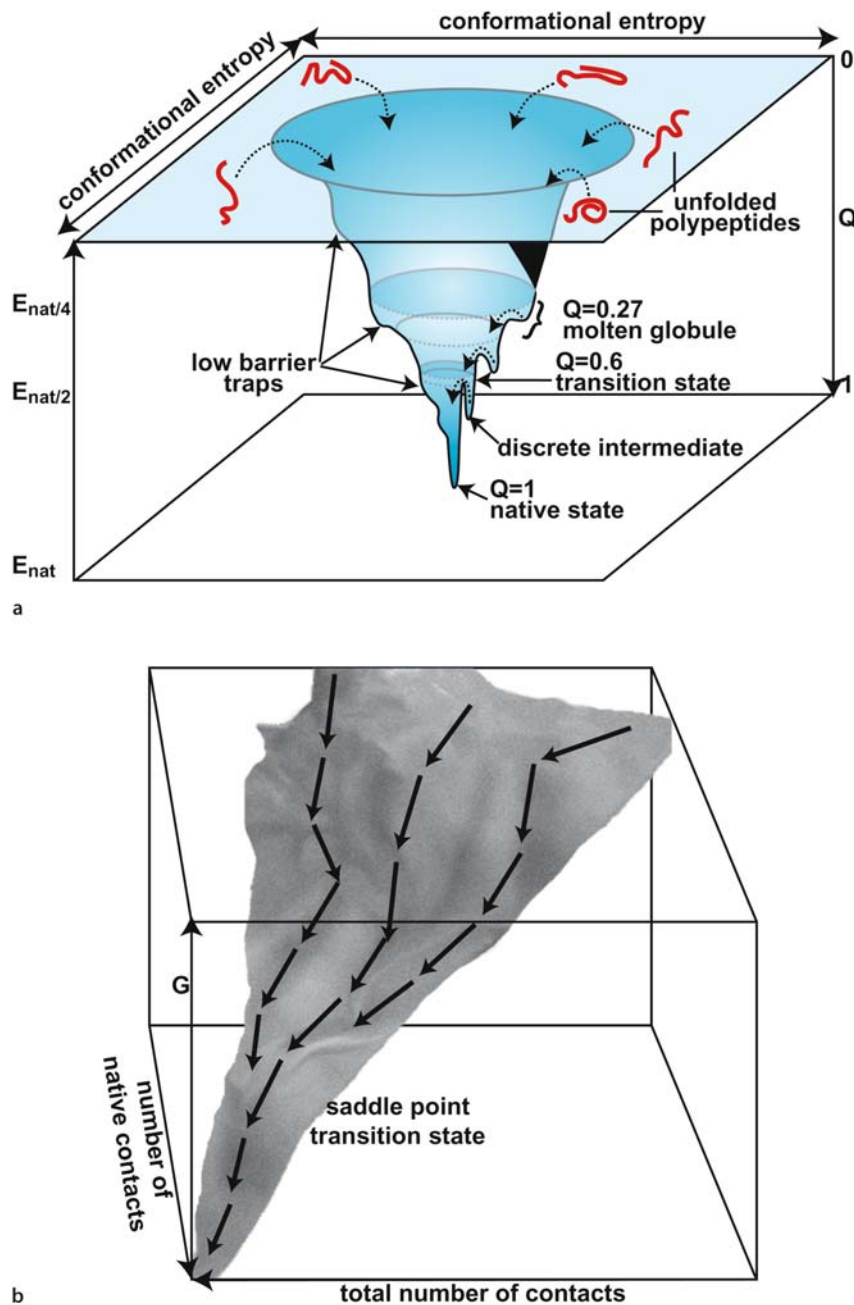
There is still very rudimentary understanding about the first question, which goes little beyond the propensity of residues to favor certain types of secondary structure. The prediction of the protein fold, based solely on the amino acid sequence, is therefore not possible so far. Much more progress has been made concerning the way the final structure is reached. Native states of isolated protein molecules almost always correspond to the most thermodynamically stable structures. Consequently, polypeptides fold by following a folding path that leads down a slope of free energy to the native state. However, since the ensemble of unfolded molecules of a protein comprises of a vast variety of conformations, there is no mandatory succession of partly folded states on the way to the native state. This is best illustrated in the form of a folding funnel, which is a concept for understanding the mechanism of self-organization of proteins (Fig. 1). The width of the funnel signifies the conformational entropy; the depth indicates the decrease of enthalpy during protein folding. At the top of the funnel, the protein exists in a large number of random states that have relatively high enthalpy and high entropy. The drive for maximization of entropy keeps the protein as random as possible at the top of the funnel, while the tendency for minimization of enthalpy tries to drag the protein down the funnel. The decrease in enthalpy is caused by an increase in the fraction of native-like contacts Q , defined as the number of native-like contacts in a given folding state divided by the total number of pair wise contacts in the native state. The polypeptide chains proceed folding by a random search of many conformations that are accessible and preferably downhill of the folding funnel. The wall of the folding funnel is not smooth because some of the non-native-like contacts may have a significant stabilization enthalpy that needs to be overcome in order to continue folding. The change in free energy can be calculated from the change in enthalpy and entropy during the folding reaction, which is usually plotted in a 3D-diagram versus the total number of contacts and the number of native-like contacts, resulting in an energy surface or energy 'landscape' (2, 3). If the shape of the energy landscape is favorable, the number of accessible conformations that needs to be sampled is greatly restricted and the protein is able to fold rapidly. This energy landscape that guides protein folding is encoded in the amino acid sequence, sculpted by the environment, and consequently subject to natural selection. In evolution, proteins were not only optimized for functionality but also folding efficacy and protein

stability, and natural proteins are a compromise between these, sometimes conflicting demands.

Stages of Protein Folding

Three major models were proposed to explain the early events in protein folding. The simple framework model suggests that the secondary structure elements form first, based on the intrinsic properties of the amino acids in the sequence of the polypeptide chain. The secondary structure elements subsequently collide to form tertiary contacts. The alternative nucleation-condensation model suggests that a short region of the polypeptide chain forms a transient secondary structure that acts as a template for the condensation of additional structural elements from adjacent segments of the polypeptide chain. The third alternative, the hydrophobic collapse model, assumes that the hydrophobic residues conglomerate nonspecifically to minimize solvent exposure. The hydrophobic collapse is followed by rearrangements to the final native structure. The actual mechanism of protein folding most likely involves some or all of these processes, and there is evidence for each mechanism from folding studies of different proteins.

These early processes lead to a compaction of the polypeptide chain, whereby the loss of conformational entropy is partly counter balanced by the increase in entropy of the water that is released from the surface of hydrophobic side chains when they pack together into a hydrophobic core. The first compact state formed is called the molten globule, which stands for an ensemble of conformations that still lacks a large number of native contacts and secondary structure elements. In this highly dynamic state, inherent fluctuations in the conformation of the partly folded polypeptide chain enable even residues that are highly separated in the amino acid sequence to come into contact with each other. As, on average, native-like contacts are more stable than non-native interactions, they will be more persistent and the polypeptide proceeds towards the native state. The bottleneck of folding, the transition state, is the formation of the native topology stabilized by contacts of key residues. Mutational replacement of these residues greatly slows down protein folding, demonstrating the importance of these residues for the folding process (4). The ensemble of transition states continues folding towards the native state. Under native conditions, folding of small single domain proteins usually appears to be a highly cooperative process. However, in some proteins intermediates may form that slowly convert into the native state. Multidomain proteins fold largely independently in each domain, with the docking of the domains as a relatively late process.



Protein Folding. Figure 1 Folding funnel and energy landscape. (a) A schematic representation of a protein folding funnel depicting the changes in enthalpy and conformational entropy during protein folding. Unfolded polypeptides (red) have a high degree in conformational freedom and a high enthalpy and are therefore found in the entropy plane at the top of the folding funnel. During folding the polypeptide chains form contacts, whereby native-like contacts are more stable and persistent. As the fraction of native-like contacts (Q) increases the protein loses enthalpy and entropy as a result of the concomitant compaction. By jumping small energy barriers the polypeptide passes through the molten globule and the transition states to the native state at the bottom of the folding funnel. (b) A schematic representation of an energy landscape which depicts the change of Gibbs free energy ΔG as a function of total number of contacts and number of native-like contacts. Unfolded polypeptides start at the top and follow a different path (black arrows) down the energy slope towards the native state.

Peptide Bond cis/trans Isomers

The peptide bond has a considerable double bond character, imposing an energy barrier for rotation of around 60 kJ mol^{-1} for Xaa-Xaa bonds (Xaa, any amino acid except proline), and ca. 80 kJ mol^{-1} for Xaa-Pro bonds. Peptide bonds are thus planar, and the flanking $\text{C}\alpha$ atoms can be either in the trans or the cis conformation. Due to steric hindrance, the cis conformation is disfavored by 14 to 17 kJ mol^{-1} for Xaa-Xaa bonds, and by only 2 kJ mol^{-1} for Xaa-Pro bonds. As a consequence, the cis conformation only occurs in the unfolded state in a negligible fraction of the Xaa-Xaa bonds, but in 10 to 30% of all Xaa-Pro bonds, if enough time for equilibration of the conformational states is allowed. Therefore, [▶peptidyl-prolyl-cis/trans-isomerization](#) is in many cases rate-limiting for folding, and a cause for the occurrence of intermediates that slowly convert into the native state.

About 5 to 7% of all Xaa-Pro bonds in native proteins are in the cis conformation, and of 1435 non-redundant protein structures in the Brookhaven Protein Database, 43% contain at least one cis proline. For these proteins, the trans to cis isomerization of the Xaa-Pro bonds is necessary before folding can be completed. The cis conformation of Xaa-Xaa peptide bonds also occurs in native proteins, though only 43 cis peptide bonds were found in 571 proteins. In these cases, a trans to cis isomerization of these non-proline peptide bonds may retard protein folding, and be the cause for the formation of a folding intermediate.

Cis peptide bonds decelerate folding reactions, but they nevertheless do occur frequently in folded proteins, as mentioned above, mainly before proline and occasionally before other amino acid residues. Cis prolines are very well suited to introduce tight turns into loop structures. More importantly, however, prolyl isomerization increases the energetic barriers of folding, and thus not only folding but also unfolding is decelerated, thereby stabilizing the native state of a protein. Non-prolyl cis peptide bonds are strongly destabilizing, because they introduce strain into the peptide backbone. They are found preferentially near the active sites of enzymes, where strained conformations might be important for catalytic activity.

Molecular Interactions

Protein Folding in a Crowded Environment

Folding processes inside the cell are much more complex than the refolding of denatured model proteins *in vitro*. Proteins are synthesized *in vivo* at the ribosomes in a vectorial manner from the N to the C terminus. In contrast to *in vitro* experiments, which are performed with diluted proteins ($0.1 \text{ g}\cdot\text{l}^{-1}$) and complete polypeptide chains, nascent polypeptides emerging from the ribosome do not contain the complete information necessary for folding. In addition,

the cytosolic concentration of macromolecules including ribosomes, nucleic acids and proteins, is enormously high ($340 \text{ g}\cdot\text{l}^{-1}$). In this crowded macromolecular environment, exposed hydrophobic amino acids of nascent polypeptides and folding intermediates may interact inappropriately leading to misfolding and aggregation. While, as a consequence of this problem, cytosolic proteins should fold as fast as possible, the folding of proteins destined for another compartment should be delayed in order to allow rapid unfolding for translocation through the membrane. Due to Brownian motion and thermal vibrations, even native proteins are always in danger of spontaneously unfolding and thereby losing their active structure. This feature of proteins is probably the evolutionary price for conformational flexibility, which is essential for protein function. For most proteins there are only small energy barriers between the native and the misfolded state. A number of proteins are specifically thermo labile, and their folding status is even more susceptible to changes in the cellular environment. Stress conditions, like a sudden increase in temperature, can therefore lead to unfolding, aggregation, or degradation of many proteins (3).

To optimize cellular protein folding, protective systems have developed in the course of evolution. These systems consist of families of highly conserved proteins, the so-called [▶molecular chaperones](#) (5). Chaperones are found in high concentrations in all cells, from bacteria to humans. They guide a large variety of folding processes throughout the life cycle of proteins, from synthesis to degradation. For example, they assist the *de novo* folding of proteins or form repair machines for misfolded or even aggregated proteins, and are therefore especially important for the survival of cells during stress situations. Since heat shock can induce the synthesis of many chaperones, those are also called [▶heat shock proteins](#) (Hsps). The name of each of the chaperone families is derived from the molecular weight of the corresponding main representative (for example, Hsp70: a protein with a molecular weight of 70 kDa). Table 1 summarizes the structure and function of the most important chaperone families and lists some of their prokaryotic and eukaryotic members, as well as their cochaperones (regulatory and cooperating proteins). In addition to these more general chaperones, there are also very specific chaperones that are specialized for very few or even a single substrate (e.g. FimC in *E. coli* is specialized for the folding of the pili subunits FimA, FimF, FimG and FimH; Hsp47 in the ER is specialized for collagen).

The minimum definition for a chaperone is the ability to prevent the aggregation of its substrate proteins. This function is called the “holder” function and is generally energy independent. Chaperones that are able to refold

Protein Folding. Table 1 Major families of chaperones: structure and function

Chaperone family	Structure	ATP	Examples		Cochaperones	Functions
			prokaryotic	eukaryotic		
sHsp	8-24-mer	-	IbpA, IbpB	Hsp27 crystallines		<ul style="list-style-type: none"> - Prevention of aggregation of heat denatured proteins - Binding to inclusion bodies - Prevention of aggregation of heat denatured proteins - Component of the lens of the vertebrate eye
Hsp60	14-mer	+	GroEL		GroES	<ul style="list-style-type: none"> - <i>De novo</i> protein folding - Prevention of aggregation of heat denatured proteins
	16-mer	+		CCT/TRiC		<ul style="list-style-type: none"> - <i>De novo</i> folding of actin and tubulin and some other proteins
Hsp70	monomer	+	DnaK		DnaJ, GrpE	<ul style="list-style-type: none"> - <i>De novo</i> protein folding - Prevention of aggregation of heat denatured proteins - Solubilization of protein aggregates together with ClpB - Regulation of the heat shock response
				Hsp70, Hsc70	J domain protein, Bag, Hip, Chip, Hop, HspBP1	<ul style="list-style-type: none"> - <i>De novo</i> protein folding - Prevention of aggregation of heat denatured proteins - Solubilization of protein aggregates together with Hsp104 - Regulation of the heat shock response - Regulation of the activity of folded regulatory proteins
Hsp90	dimer	+	HtpG			<ul style="list-style-type: none"> - Tolerance to extreme heat shock
				Hsp90	Hop, Chip, p23, p50 ^{cdc37} , FKBP51/52 Cyp40, PP5,	<ul style="list-style-type: none"> - Stress tolerance - control of folding and activity of more than 120 native proteins including transcription factors and kinases - regulation of the heat shock response
Hsp100	6-mer	+	ClpB			<ul style="list-style-type: none"> - Thermotolerance - Disaggregation together with Hsp70 (Dna K)
			ClpA			<ul style="list-style-type: none"> - Proteolysis together with ClpP peptidase
				Hsp104		<ul style="list-style-type: none"> - Thermotolerance - Disaggregation together with Hsp70

denatured proteins to the native state are called “folder” chaperones. The folder function is generally dependent on ATP hydrolysis and the regulation by cochaperones. Protein folding in the cell is not only assisted by chaperones, but also by so-called folding catalysts like peptidyl-prolyl-cis/trans-isomerases (PPIase), which catalyze the cis/trans isomerization of Xaa-Pro bonds, and by protein disulfide isomerases (PDI), which reduce disulfide bonds and reoxidize free cysteines, thereby assisting the formation of correct disulfide bonds.

Families of Chaperones

sHsp

The family of small heat shock proteins consists of ATP-independent chaperones that are characterized by their homology to the α -crystalline domain. This signature domain is extended at the N and C terminus by unrelated domains of variable length to give a molecular weight of 10 to over 40 kDa. sHsps form oligomeric spheres of 8 to 24 subunits, which at least in some cases dissociate into dimers upon temperature increase. By binding to unfolded proteins, these sHsp dimers coaggregate with their substrates. Such coaggregates of sHsps and misfolded proteins are more efficiently dissolved by the action of Hsp70 chaperones in cooperation with Hsp100 chaperones, than pure aggregates of misfolded proteins alone. At least one member of the sHsp family, Hsp27, is involved in the regulation of programmed cell death.

Hsp60

The family of Hsp60 proteins, also called chaperonins, consists of two distantly related subfamilies of ATP-dependent oligomeric chaperones. The Group I chaperonins, which exist in bacteria (GroEL), mitochondria and chloroplasts (Cpn60), form a barrel out of two stacked rings with seven identical subunits of 57 kDa each. GroEL/Cpn60 cooperates with the cochaperones GroES/Cpn10 (ca. 10 kDa), which also forms a seven-member ring and acts as a lid on the GroEL barrel. Unfolded proteins and folding intermediates bind to the central cavity of one of the two rings, and ATP and GroES binding to the substrate-containing ring leads to an enclosure of the substrate, which then may fold isolated from aggregation promoting influences in the cytosol. ATP- and GroES-induced conformational changes in the substrate-enclosing ring are proposed to promote conformational changes in the substrate, including global unfolding and global compaction, which further promote the folding to the native state. GroEL has a large variety of substrates (10–15% of *E. coli* proteins), most of them below 60 kDa, the size-limit for enclosure in the GroEL cavity. Substrates that are too big for an enclosure in GroEL under the GroES dome are folded in trans, meaning that GroES binds to the empty and

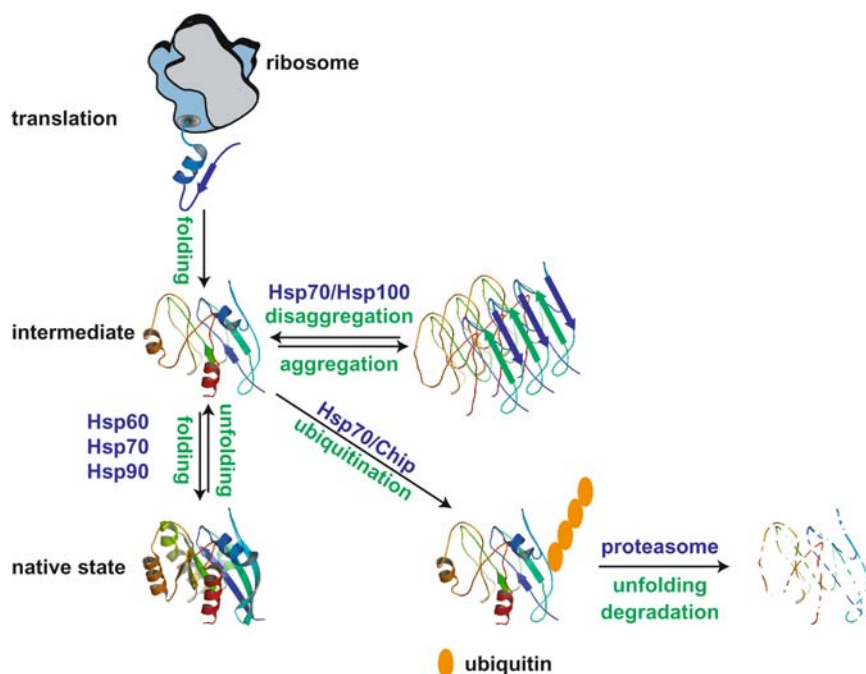
not the substrate-containing ring triggering folding, and substrate releases through the cooperativity of the two rings.

The Group II chaperonins, which exist in the eukaryotic cytosol (CCT/TRiC) and in archaeans (thermosome), also form a barrel out of two stacked rings, but with eight different subunits this time. These subunits have a built-in lid, and therefore do not cooperate with a lid-forming cochaperone. CCT is believed to work in a GroEL analogous way. The main substrates for CCT in the eukaryotic cytosol are actin and tubulin, but a number of other substrates have also been identified.

Hsp70

The 70 kDa heat shock proteins are central components of the cellular chaperone network, and are involved in a large variety of protein folding processes (6). These processes include *de novo* folding of newly synthesized polypeptides; refolding of spontaneously unfolded or stress misfolded proteins; disaggregation of protein aggregates; translocation of polypeptides across biological membranes, where Hsp70s often act on both sides of the membrane; assembly and disassembly of oligomeric protein complexes, e.g. clathrin uncoating or assembly and disassembly of virus capsids; control of activity and stability of many regulatory proteins including transcription factors and protein kinases; and regulation of the heat shock response. The highly conserved Hsp70 proteins are found in almost all organisms and are essential for all eukaryotic organisms, where they reside in all major cellular compartments. In *E. coli*, three distinct Hsp70s are found called DnaK (major form), HscA (specialized for iron-sulfur-cluster proteins) and HscC (unknown function). In yeast and higher eukaryotes, 10 to 15 Hsp70s plus 3 to 4 Hsp70-related proteins with higher molecular weight (Hsp110 and Hsp170s) exist. In the cytosol there are at least one Hsp70 (called Hsc70, heat shock cognate), which is constitutively at high levels, and one Hsp70, which is stress induced. Pathogenic processes like cancer often lead to a continuously high level of the stress inducible Hsp70 form. In the human cytoplasm, three Hsp110 are found called Hsp105, Apg-1 and Apg-2. The Hsp70 in the endoplasmic reticulum is called Bip (Immunoglobulin **B**inding protein) and the Hsp170 is named Orp150 (**O**smotic shock regulated protein). The mitochondrial Hsp70, mortalin, is also found in the cytoplasm, often as a sign of ongoing pathological processes.

All functions of Hsp70 proteins are based on an ATP driven cycle of transient binding and release of a short peptide stretch within the substrate polypeptide. This cycle is regulated by cochaperones of the family of J-domain proteins (Hsp40s) and nucleotide exchange factors. The family of J-domain proteins, which has



Protein Folding. Figure 2 Protein folding in the cell. A newly synthesized protein starts folding either cotranslationally while still attached to the ribosome or posttranslationally. Folding proceeds generally through intermediary states towards the native state. Due to the exposure of hydrophobic side chains in the crowded environment of the cell intermediates may interact with each other and aggregate. Proteins, which have completed folding, may also unfold again and subsequently aggregate due to natural or mutation-induced instability, physiological or non-physiological modifications, or environmental stress. Chaperones of the Hsp60, Hsp70 and Hsp90 families assist de novo folding and refolding of denatured proteins. In bacteria, yeast and plants Hsp70 chaperones cooperate with Hsp100 chaperones in disaggregation of protein aggregates and subsequent refolding to the native state. Hsp70 also interacts with the ubiquitin ligase Chip, which leads to ubiquitination of misfolded proteins and ensuing unfolding and degradation by the proteasome.

increased tremendously in the course of evolution (*E. coli*, 6; yeast, 20; humans, 44), couples substrate binding by their Hsp70 partner protein with ATP hydrolysis, therefore acting as a targeting factor. Four families of nucleotide exchange factors for Hsp70s have been discovered so far: the prokaryotic GrpE (Mge1 in mitochondria), the family of Bag proteins in the cytosol of higher eukaryotes, the family of HspBP1 proteins in the cytosol and ER (Bap), and the Hsp170 protein Lhs1 in the yeast ER, which acts as nucleotide exchange factor for the ER-resident Hsp70 Kar2. Other known interacting cochaperones of Hsp70 proteins are the Hsp70-Hsp90-organizing protein Hop, the ubiquitin ligase Chip, and the Hsp70-interacting protein Hip.

Hsp90

The 90 kDa heat shock proteins are highly abundant and essential ATP-dependent chaperones in all eukaryotes (7). Mammalian cells contain four full-length Hsp90 proteins, the constitutive Hsc90 β , the heat inducible Hsp90 α , both in the cytosol, the ER-localized Grp94 (gp96), and the mitochondrial Hsp75 (Trap1). All four proteins consist of an N-terminal ATPase

domain, a middle domain implicated in substrate binding, and a C-terminal dimerization domain. In addition, an Hsp90 relative, Hsp90N, without N-terminal ATPase domain has been found. The interaction of the cytosolic Hsp90s with substrate proteins is regulated by a plethora of cochaperones and cooperating chaperones. These are either involved in substrate loading onto Hsp90, like the Hsp70 chaperone and the Hsp70-Hsp90 organizing protein Hop, or in the regulation of Hsp90's ATPase cycle and substrate release, like Aha1, p23, and the immunophilins, FKBP51, FKBP52 and cyclophilin-40. The interaction of Hsp90 with protein kinases often involves the protein p50^{cdc37}. In addition, the protein phosphatase PP5 and the ubiquitin ligase Chip interact with Hsp90. Hsp90 forms the core component of a dynamic multimeric chaperone complex, which in cooperation with Hsp70 interacts with more than 120 natively folded proteins, many of which are involved in signal transduction processes, regulation of cell homeostasis, cell cycle, differentiation and apoptosis. These substrate proteins, which include numerous transcription factors and kinases, are converted by the chaperone

complex into conformational states that permit rapid activation through signaling events, including post-translational modifications (phosphorylation) and interactions with specific ligands such as hormones. Many of the protein kinases and transcription factors, which are substrates of the Hsp70-Hsp90 chaperone machinery, are implicated in physiological and pathophysiological processes like tumorigenesis, neurodegenerative diseases, inflammation, autoimmunity, viral and bacterial infections, ischemia/hypoxia, and ageing. Hsp90 and Hsp70 are, therefore, prime targets for pharmacological intervention.

Hsp100

The Hsp100 chaperones (Clp proteins in prokaryotes) belong to the superfamily of AAA+ proteins, which also includes the 'ATPase associated with a variety of cellular activities' (AAA) proteins. Characteristic of this superfamily is a considerable sequence homology in their 'AAA' domains, which are important for ATP hydrolysis and oligomerization. Generally, AAA+ proteins form ring-shaped homohexamers, which drive the assembly and disassembly of macromolecular complexes by ATP-dependent remodelling of their substrates. ClpB in *E. coli*, Hsp104 in yeast and Hsp101 in plants cooperate with Hsp70 proteins to dissolve protein aggregates and reactivate the misfolded proteins. Other members of the Hsp100 family associate with peptidases and unfold proteins for degradation. Hsp100 proteins of the ClpB/Hsp104 subgroup are essential for the development of thermotolerance and survival at extremely high temperatures. Homologues of Hsp104/ClpB have not so far been found in the cytosol of metazoans.

Regulatory Mechanisms

Regulation of Protein Conformation

For many proteins, evolution has not selected for the most stable conformation because conformational flexibility is essential for protein function. Some proteins, especially those that act as molecular switches in signal transduction pathways, and also potentially dangerous proteins like the caspase activated DNase (CAD), are particularly unstable, either unfolding spontaneously or existing in more than one native conformation. These proteins are generally already bound by Hsp70 chaperones during translation at the ribosomes, and are subsequently handed over to the Hsp90 chaperone complex or a specific inhibitory binding protein (ICAD in the case of CAD). In both cases, the substrate protein is kept in an inactive state but is poised for activation. This regulation of protein conformation is best investigated for the steroid hormone receptors. The current hypothesis assumes that Hsp70 induces conformational changes in the hormone receptor, thereby opening up the ligand-binding site

and subsequently transferring the receptor onto the Hsp90 chaperone complex, which holds it in an open state ready for hormone binding. With a certain half-life (ca. 5 min) this complex is thought to decay, whereupon the substrate is rebound by Hsp70 and the activation cycle restarts. Hormone binding stabilizes the receptor in the active conformation, leading to the dissociation from the chaperone complex, followed by dimerization, entry into the nucleus and transcriptional activation of the target genes. In the absence of the chaperones, the hormone binding activity of the receptor is very low, underlining the importance of the induced conformational changes.

Global Regulation of Protein Folding

The cellular protein folding system is a robust self-sustaining system, where the need for folding assistants is regulated by the total amount of unfolded and misfolded proteins. In a process generally known as heat shock or stress response, an elevated level in unfolded proteins leads to an increase in transcription of heat shock genes, and a subsequent rise in the total amounts of chaperones and proteases to assist refolding or degradation of hopeless cases. Key players in this regulation are Hsp70 and Hsp90 proteins, which in the eukaryotic cytosol bind to the heat shock transcription factor ►**HSF**, and keep it in a monomeric form probably by regulating its conformation. This interaction is thought to be a dynamic binding and release cycle, and an increase of unfolded proteins competes with HSF for binding to Hsp70 leading to an increase in free and active HSF, which in turn transcriptionally induces an increase of the Hsp70 and Hsp90 levels, until enough free Hsp70 and Hsp90 is available to sequester HSF again. In the mammalian ER, the Hsp70 protein Bip binds to the ER stress sensors IRE1, PERK, and ARF6. Like in the cytoplasm, unfolded polypeptides compete with the stress sensors for binding to Bip, and the free stress sensors activate the so-called unfolded protein response (UPR), leading to an inhibition of general translation into the ER and an increase in transcription and translation of ER resident chaperones.

►**Defective Protein Folding Disorders**

►**Recombinant Protein Production in Mammalian Cell Culture**

References

1. Buchner J, Kiefhaber T (2005) Protein Folding Handbook. WILEY-VCH Verlag, Weinheim
2. Dinner AR, Sali A, Smith LJ, Dobson CM and Karplus M (2000) Understanding protein folding via free-energy surfaces from theory and experiment. Trends Biochem Sci 25:331–339
3. Dobson CM (2003) Protein folding and misfolding. Nature 426:884–890

4. Lindorff-Larsen K, Rogen P, Paci E, Vendruscolo M, Dobson CM (2005) Protein folding and the organization of the protein topology universe. *Trends Biochem Sci* 30:13–19
5. Young JC, Agashe VR, Siegers K, Hartl FU (2004) Pathways of chaperone-mediated protein folding in the cytosol. *Nat Rev Mol Cell Biol* 5:781–791
6. Mayer MP, Bukau B (2005) Hsp70 chaperones: Cellular functions and molecular mechanism. *Cell Mol Life Sci* 62:670–684
7. Wegele H, Muller L, Buchner J (2004) Hsp70 and Hsp90—a relay team for protein folding. *Rev Physiol Biochem Pharmacol* 151:1–44

Protein G

Definition

Protein G refers to a cell surface protein from the group G streptococci. It binds to the Fc domain of Immunglobulin G classes and subclasses, but with a different affinity profile as [▶protein A](#).

▶Surface Plasmon Resonance

Protein Homology

Definition

Protein homology designates similarities of amino acid sequences between two proteins.

▶Leucine Zipper Transcription Factors: bZIP Proteins

Protein Interaction Analysis, Variations of the Yeast Two-Hybrid System

IGOR STAGLJAR

Institute of Veterinary Biochemistry and Molecular Biology, University of Zurich, Zurich, Switzerland
stagljar@vetbio.unizh.ch

Definition

Besides protein identification and quantification of differences in protein profiles of cells, tissues and

organs, the characterization of physical protein-protein interactions (so-called “interactive proteomics”) is being recognized as a key objective of proteomics research. One of the best tools available to analyze protein-protein interactions is the yeast two-hybrid system (YTH). The YTH system has spun-off several modified versions that provide additional utility to academic and pharmaceutical research, such as the repressed transactivator (RTA) system and the split ubiquitin membrane yeast two-hybrid method (MbYTH).

Characteristics

Protein-protein interactions play a fundamental role in the regulation and execution of almost all cellular processes. One of the most widely used techniques to detect protein-protein interactions is the YTH. It takes advantage of the finding that many eukaryotic [▶transcription factors](#) can be divided into two functionally distinct domains that mediate DNA binding and transcriptional activation. In the classical yeast two-hybrid approach, a [▶bait](#) is constructed by fusing a protein X to the DNA-binding domain (DBD) derived from a transcription factor and a [▶prey](#) is constructed by fusing a protein Y to the activation domain (AD) of a transcription factor. The bait and prey fusions are co-expressed in yeast, where the interaction of X and Y leads to the reconstitution of a functional transcription factor. Reconstitution of the transcription factor is measured by assaying the activity of [▶reporter genes](#). Commonly, auxotrophic markers that can be selected for are used in combination with the *lacZ* gene encoding bacterial β -galactosidase. *HIS3* and *LEU2* allow selection of interactions by monitoring growth on selective plates lacking histidine or leucine respectively, whereas *lacZ* can be easily measured using a colorimetric assay. The ability to screen large cDNA libraries quickly for proteins that interact with a given protein of interest is the major advantage of the YTH system over other, biochemical methods utilized for the identification of protein-protein interactions (1). A major limitation of the YTH method is that interaction between bait and prey occurs in the nucleus. Consequently, transmembrane proteins cannot be efficiently studied by the YTH method because they tend to form aggregates when expressed in the nucleus in their full-length form. In addition, interactions between many membrane proteins are dependent on post-translational modifications, such as glycosylation and disulfide bond formation, that take place within the endoplasmic reticulum but not in the nucleus. Lastly, [▶transcriptional activators](#) cannot be used as ‘bait’ proteins in the YTH assay as they can constitutively activate the reporter genes even in the absence of protein interactions.

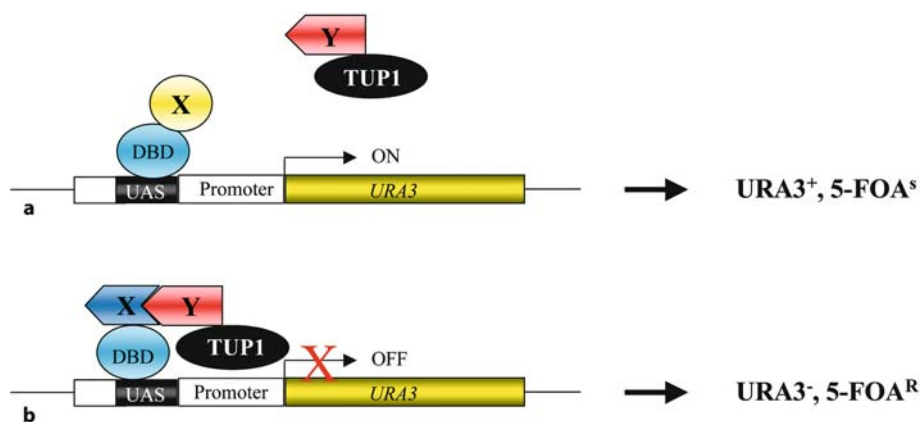
The Repressed Transactivator (RTA) System

In the past few years, two alternative YTH-based technologies have been developed to address the above-mentioned technical difficulties associated with the conventional YTH method. These include the repressed transactivator (RTA) system (2), and the split ubiquitin membrane yeast two-hybrid method (MbYTH) (3). RTA technology (Fig. 1) was developed to address limitations of the standard YTH method, which precludes use of transcription activators as baits for identification of interacting proteins. In the RTA system, the activator 'bait' is fused to the GAL4 DNA-binding domain and the 'prey' is fused to the repressor domain of the yeast TUP1 protein. Interaction of the activating bait fusion with the prey-TUP1 fusion represses expression of the reporter gene, which is generally detected using counter selection (2, 4). Hirst et al. validated their RTA system by demonstrating its ability to detect interactions between the mammalian basic helix-loop-helix proteins MyoD and E12 and between the c-Myc oncoprotein and the Bin1 tumor suppressor. They also used the RTA assay to screen for novel proteins interacting with the activation domain of the VP16 transcriptional activator (2). In addition, the RTA approach has recently been used to identify a novel interacting partner of the human androgen receptor (AR) called DDC (L-dopa decarboxylase) (4). Thus, the RTA system should prove useful for identifying co-activators and regulators of transcription from any model organism. A drawback of this

technology is that, like any other genetic system, it identifies a large number of false positives and it detects binary interactions.

The Split-Ubiquitin Membrane Yeast Two-Hybrid System

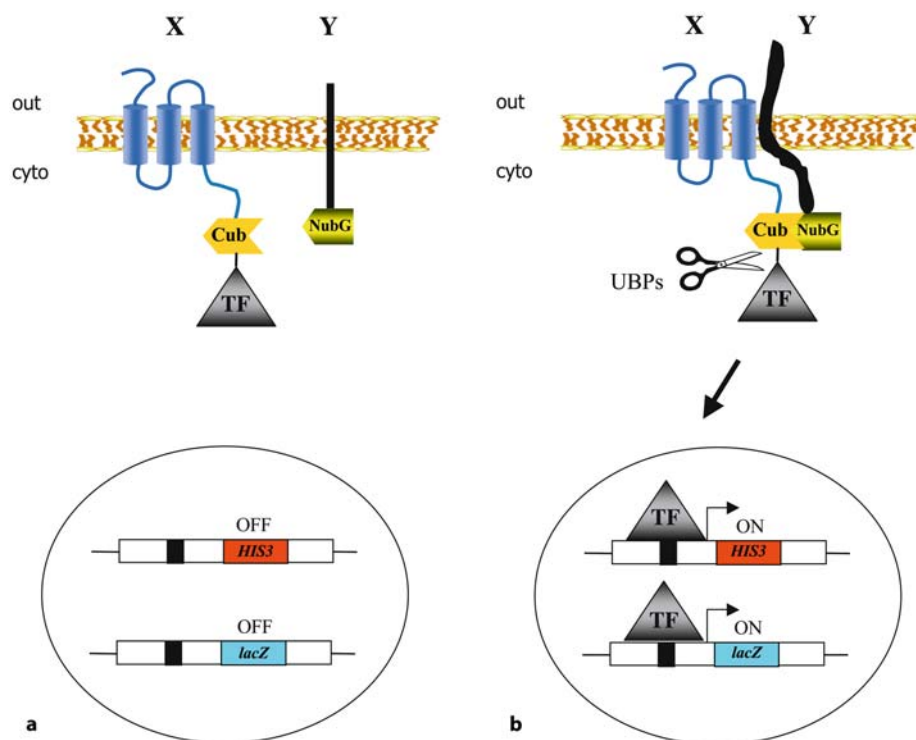
An alternative YTH approach that is applicable to studying membrane proteins is the split-ubiquitin membrane yeast two-hybrid system (MbYTH) (3, 5). MbYTH uses the split-ubiquitin approach in which the reconstitution of two ubiquitin halves is mediated by a specific protein-protein interaction. Like the YTH, where a transcription factor is reassembled upon interaction of two test proteins, the split-ubiquitin system consists of two fragments of ubiquitin that are brought together by interacting proteins. ► **Ubiquitin** (Ub) is a small, highly conserved protein that is attached to lysine residues of proteins in order to tag the proteins for proteasomal degradation. Ubiquitin-tagged proteins are recognized by ubiquitin-specific proteases (UBPs) that cleave after the C-terminal (Gly 76) residue of Ub and the first residue of the target protein, allowing release of the protein for degradation by the 26S proteasome. Johnsson and Varshavsky (6) found that native ubiquitin could be split into an N-terminal (Nub) and a C-terminal (Cub) half. The two halves retain a basic affinity for each other and spontaneously reassemble to form quasi-native ubiquitin. If a reporter protein is fused to the C-terminus of Cub, it will be cleaved off by UBPs upon assembly of the Nub and Cub moieties. A point mutation in the N-terminal



Protein Interaction Analysis, Variations of the Yeast Two-Hybrid System. Figure 1 Repressed transactivator (RTA) method, a two-hybrid system for transactivator bait proteins. Like the conventional yeast two-hybrid system, a bait protein of interest (X) is fused to a GAL4 DNA binding domain (DBD). However, the prey protein (Y) is created by fusion of the cDNA library with a TUP1 repressor protein. (a) If there is no interaction between the bait X and the TUP1-prey Y fusion protein, the bait activator allows constitutive expression of the *URA3* reporter gene. This permits the growth of yeast cells in the absence of uracil, but causes sensitivity to 5-FOA, which is converted into a toxic product through the enzyme encoded by the *URA3* gene. (b) In a situation where there is an interaction between X and Y, TUP1 will cause repression of the *URA3* expression, which can be detected by growth of yeast in the presence of 5-FOA. Therefore, viability of the reporter strain (or colony growth) on 5-FOA media represents a positive interaction between the bait and prey protein.

domain of ubiquitin (NubG) abolishes the affinity of the two halves for each other, such that NubG and Cub fail to refold into split-ubiquitin when coexpressed in yeast. However, if the two ubiquitin halves are fused to the interacting proteins X and Y, this interaction brings the NubG and Cub moieties close enough together to reconstitute quasi-native Ub, resulting in the release of the reporter protein by the UBPs. In the MbYTH assay (Fig. 2), an artificial transcription factor (TF) consisting of the bacterial LexA protein and the *Herpes simplex* VP16 transactivator protein is fused to the Cub moiety. An integral membrane protein (X) is expressed as a fusion to the Cub-LexA-VP16 reporter cassette, with this cassette attached to either the N- or C-terminus of the transmembrane protein, depending on the orientation in the membrane of this protein. The second protein under investigation (Y), either another

transmembrane protein or a cytoplasmic protein, is expressed as a fusion to NubG. If interaction between the X and Y proteins occurs, a split-ubiquitin molecule can be reconstituted, leading to the proteolytic release of the transcription factor to activate a reporter gene (2, 5, 7). Thus, the reassociation event initiated by the protein interaction is converted into a transcriptional output that can be easily detected. This assay has been used to investigate the influence of mutations on the assembly of fragments of presenilin (►PSEN1/PSEN2) (a protein implicated in ►Alzheimer's disease), to characterize the interaction between the yeast α 1,2-mannosidase Mns1p and Rer1p in the endoplasmic reticulum and to study intra- and inter-molecular interactions between plant sucrose transporters (7). Moreover, Thaminy et al. (5) have recently successfully adapted the MbYTH system for prey library screening and have identified three novel



Protein Interaction Analysis, Variations of the Yeast Two-Hybrid System. Figure 2 The split-ubiquitin membrane yeast two-hybrid system (MbYTH), a yeast-based genetic assay for characterization of membrane protein interactions. Here, the membrane bait protein of interest X is fused to the C-terminal domain of ubiquitin (Cub) followed by the artificial transcription factor TF, and the potential interacting prey protein Y is fused to the mutated N-terminal domain of ubiquitin (NubG). Note that prey Y can also be a cytosolic protein. (a) If both proteins (X; Y) do not interact, TF is not cleaved from Cub and remains associated with the membrane. In this case, reporter genes *HIS3* and *lacZ* remain silent, which renders cells unable to grow on medium lacking the amino acid histidine and colorless when exposed to X-gal. (b) If both membrane proteins X and Y interact, the re-associated NubG and Cub are recognized and cleaved by the ubiquitin-specific proteases (UBPs) (scissors), liberating the TF, which is transported into the nucleus. This in turn activates the *lacZ* and *HIS3* reporter genes resulting in β -galactosidase activity and histidine synthesis, respectively.

interacting partners of the mammalian ErbB3, a receptor tyrosine kinase involved in the regulating of proliferation and differentiation in many tissue types. Thus, MbYTH allows rapid and sensitive characterization of proteins associated with a particular full-length transmembrane protein of interest and is generally applicable to most transmembrane proteins of any organelle. Like the RTA technology, the major disadvantage of MbYTH is that it identifies a large number of false positives, presumably through interactions between proteins that do not normally occur *in vivo*. In addition, this technology cannot be applied to study membrane protein complexes, as it detects mostly binary interactions.

Clinical Relevance

Transcriptional activators play key roles in establishing the development and activation stages of cells by mediating responses to inter- and intra-cellular signals. Accordingly, transcriptional activators represent approximately 20% of the protein coding genes of the human genome. Similarly, proteins associated with membranes total approximately one third of all proteins in a typical eukaryotic cell and execute a variety of essential cellular tasks that include cell signaling, transport of membrane-impermeable molecules and cell adhesion. Thus, characterization of novel interacting partners of transcriptional activators and membrane-associated proteins is an important parameter in understanding their function and will have a great impact on biological research and pharmaceutical target discovery. In particular, its application in linking disease pathways and tractable targets, elucidating the function of targets and refining strategies to modulate protein function will accelerate pharmaceutical R&D. The studies on protein interactions will thus provide predictors of toxicity and may be used to develop biomarkers as fingerprints for particular disorders or cellular responses.

References

1. Fields S, Song O (1989) A novel genetic system to detect protein-protein interactions. *Nature* 340:245–246
2. Hirst M et al (2001) A two-hybrid system for transactivator bait proteins. *Proc Natl Acad Sci USA* 98:8726–8731
3. Stagljar I, Korostensky C, Johnsson N (1998) A genetic system based on split-ubiquitin for the analysis of interactions between membrane proteins *in vivo*. *Proc Natl Acad Sci USA* 95:5187–5192
4. Wafa LA et al (2003) Isolation and identification of L-dopa decarboxylase as a protein that binds to and enhances transcriptional activity of the androgen receptor using the repressed transactivator yeast two-hybrid system. *Biochem J* 375:373–383
5. Thaminy S, Auerbach D, Arnoldo A et al (2003) Identification of novel ErbB3-interacting factors using the split-ubiquitin membrane yeast two-hybrid system. *Genome Res* 13:1744–1753
6. Johnsson N, Varshavsky A (1994) Split ubiquitin as a sensor of protein interactions *in vivo*. *Proc Natl Acad Sci USA* 91:10340–10344
7. Stagljar I, Fields S (2002) Analysis of membrane protein interactions using yeast-based technologies. *Trends Biochem Sci* 27:559–563

Protein Interaction Analysis: Chemical Cross-Linking

OWEN W. NADEAU

Department of Biochemistry and Molecular Biology,
Kansas University Medical Center, Kansas City, KS,
USA

onadeau@kumc.edu

Synonyms

Protein conjugation; bioconjugation; protein chemical modification; covalent coupling

Definition

Protein interaction analysis is one of the main fields of functional proteomics, covering an array of traditional protein chemical and newly developing techniques that center primarily either on identifying components or determining mechanisms of protein-protein interactions in both short-lived and long-lived protein complexes. The stability of a complex is determined, in part, by the surface contact area or protein interface formed between interacting proteins. Surface contact areas are generally unique for interacting pairs of proteins, with some proteins containing several protein-binding domains that permit multiple contacts in large protein complexes. ►**Chemical cross-linking**, an extension of protein chemical modification, is an established technique used to join two or more of the proteins in a complex covalently using multi-functional chemical compounds, termed ►**cross-linking reagents**. These reagents generally contain two or more functional groups that target reactive amino acid side chains on adjacent peptides located within or in close proximity to the protein interface of a protein complex. The architecture and chemical potential of the protein interface are important determinants in complex formation and have a direct impact on the type of cross-linking reagent that can be used successfully to probe the interaction.

Characteristics

Chemical cross-linking provides essentially a snapshot of binding events or other physiological processes that

require the interaction of two or more proteins. Cross-linking reagents generally sample only a few of the possible interactions that can occur throughout surface contact regions between two interacting proteins; such regions may vary in area from approximately 500–5000 Å². Although cross-linking provides information for only a fraction of the potential contacts that are possible between proteins in a complex, it can be used to determine the spatial organization of proteins in a complex, to measure the distance between interacting surface residues and to detect either small structural rearrangements or global conformational changes in proteins. Resurgence in the use of chemical cross-linking has occurred with the advent of recent advances in mass spectrometric techniques and the development of large protein data bases; the combined power of these methodologies significantly increases the potential for determining protein-protein interactions with increasingly smaller amounts of starting material.

Proteins as Chemical Reactants

Numerous physiological processes are mediated through protein-protein interactions, including enzyme catalysis, ►[muscle contraction](#), receptor-ligand binding, antibody-antigen immunocomplex formation and many others. Both structural and chemical factors (hydrogen bonding, hydrophobic and electrostatic forces) that facilitate protein-protein interactions also influence chemical cross-linking of protein complexes (1). The chemical nature of the protein is another important determinant in bioconjugation reactions. Proteins, which are also considered to be reactants in cross-linking reactions, are linear polymers containing different combinations of amino acids, each with side chains of differing chemistries varying in charge, shape, size and polarity. Cross-linking generally occurs between appropriate ►[nucleophilic](#) amino acid side chains (Table 1) at distinct sites on the protein where both limited steric constraints and the chemical environment complement the size and chemistry of the cross-linking reagent. Intramolecular cross-linking occurs only when proximal residues on one protein target are covalently linked. Chemical coupling of amino acid side chains on adjacent proteins in a complex, termed intermolecular cross-linking, can occur either at sites forming the interface or at sites in close proximity to the contact surface. In some protein complexes, over 30% of the surface area of a monomer can be subsumed by the protein interface. This comprises numerous contact sites with varying hydrophobic and/or hydrophilic potentials, partially influencing the type of cross-linking reagent that can access the sites. Protein interfaces also contain water-filled cavities, which can accommodate hydrophilic (polar) cross-linkers of appropriate size. Many enzymes have substrate-binding sites or hydrophobic patches that can

bind hydrophobic reagents. Phosphorylase kinase, a regulatory enzyme complex containing sixteen subunits, has several high affinity binding sites for the hydrophobic cross-linker, phenylenedimaleimide (Table 1), that are in close proximity to the protein interfaces between its regulatory α and β and catalytic γ subunits (2).

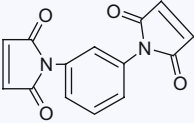
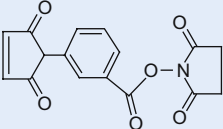

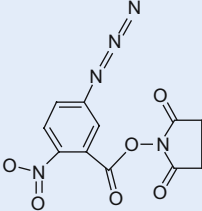
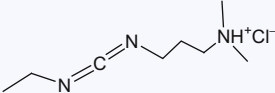
Environment of Chemical Cross-Linking

Although the chemical environment of a protein has a direct impact on its interactions with potential partners, the complexity of its environment determines the level of information that can be gained in a cross-linking experiment (3). Cross-linking either *in vitro* or *in vivo* forms the basis for the two most basic approaches to cross-linking.

In the *in vivo* approach, proteins of interest are cross-linked directly within cells or tissues. ►[Hydrophobic cross-linkers](#), such as formaldehyde (Table 1), which are able to penetrate cell membranes, are generally required for this approach. The primary advantage of cross-linking *in vivo* is that it has the potential to couple proteins in their native environment and is less likely to generate non-specific conjugates for a given group of interacting proteins, particularly if reagents with short cross-linking spans are used. The major disadvantage of this technique however, is that it is very difficult to control conditions of cross-linking within the cell or target any specific protein or protein complex. Once the cross-linker is introduced into the cell, it is free to interact with numerous potential targets, which complicates interpretation of the data.

Cross-linking *in vitro* permits greater control over the conditions of cross-linking for a given protein; in addition, the number of potential interacting partners can be adjusted, as well as the concentration of reactants, pH, ionic strength, temperature, time of cross-linking and the type of cross-linkers used. Thus both non-polar and water-soluble cross-linkers can be used to probe hydrophobic and hydrophilic sites respectively on target proteins. As opposed to the *in vivo* method, the level of purity for many proteins of interest can be manipulated, effectively reducing the number of potential conjugates that can be formed, so that ►[cross-linking patterns](#) can be more easily interpreted. The primary disadvantage of the *in vitro* technique is that proteins are modified under non-physiological conditions. Detergents and other reagents used to disrupt cells can have a deleterious affect on protein function, altering both intrinsic intermolecular interactions and extrinsic near-neighbor interactions (3). Moreover, many proteins are compartmentalized in cells, potentially limiting their accessibility to other cellular proteins, which are targeted by non-specific cross-linking only under non-physiological conditions.

Protein Interaction Analysis: Chemical Cross-Linking. Table 1 Cross-linkers

Structure and Name	Type	Characteristic Solubility	Reactive Groups	Selectivity
 1. N,N'-m-Phenylenebismaleimide	Homo-bifunctional	Hydrophobic	Maleimide	Sulfhydryl
 2. m-Maleimidobenzoyl-N-hydroxysuccinimide ester	Hetero-bifunctional	Hydrophobic	Maleimide/ N-hydroxysuccinimide ester	Sulfhydryl/ amine
 3. Formaldehyde	Mono-functional	Hydrophilic	Carbonyl	Broad/ amine
 4. N-5-Azido-2-nitrobenzoyloxy-succinimide	Hetero-bifunctional: Photo-reactive	Hydrophobic	N-hydroxysuccinimide ester/azide	Amine
 5. 1-Ethyl-3-[3-dimethylaminopropyl] carbodiimide hydrochloride	Zero-length	Hydrophilic	Carbodiimide	Carboxyl/amine

Cross-Linking Reagents

Two comprehensive treatises by Wong and Hermanson cover both the mechanisms and actions of most of the cross-linking reagents that are commercially available (4, 5). Several basic types of cross-linkers will be covered herein. Cross-linking reagents can be separated into two basic classes on the basis of their solubility in either organic solvents and/or water and are termed hydrophobic or **hydrophilic cross-linkers**, respectively. Hydrophobic cross-linkers are often introduced into aqueous environments with carrier solvents such as acetonitrile or acetone. The most commonly used cross-linkers are bifunctional reagents (Table 1); these compounds contain two reactive groups, separated by a spacer group. The reactive groups react with two neighboring amino acid side chains, covalently joining them at a distance defined primarily by the length of the spacer group. Both the geometry of cross-linking and

the solubility of the reagent are determined, in part, by the spacer group. **Cleavable cross-linkers** possess modified spacer groups that can be readily cleaved by oxidizing reagents, reducing reagents or bases. Homo-bifunctional and heterobifunctional cross-linkers are bifunctional reagents that have identical and different reactive groups, respectively. Photoactivatable cross-linkers are heterobifunctional reagents that combine both chemically active and photoreactive groups, permitting an extra layer of control in the cross-linking process. For example, the protein of interest can first be labeled with the chemically reactive group, purified to remove excess cross-linker and then cross-linked to a subsequently added target by exposing the complex to activating wavelengths of light.

Zero-length cross-linkers primarily activate the carboxyl-containing side chains of aspartate and glutamate, forming reactive intermediates that are

targeted by proximal ϵ -amine side chains of lysine. Ultimately, the two amino acids are coupled directly through an amide bond, without any residual, intervening atoms from the cross-linking reagent. The use of formaldehyde, a reagent traditionally used for fixing tissues, approximates zero-length cross-linking by forming a methylene bridge ($-\text{CH}_2-$) between two proximal lysine (or other nucleophilic) side chains, covalently joining reactants at distances between 2 to 3 Å. Other amino acid side chains are also cross-linked by this reagent, including those of tyrosine, histidine and arginine.

Detection of Cross-Linked Proteins

Cross-linking is not a stand-alone technique. In order to detect the formation of a conjugate and determine the identity of its integral components, cross-linking must be coupled with methods that enable visualization, purification and detection of cross-linked proteins. Rapid screening of cross-linked proteins is commonly carried out using polyacrylamide gel electrophoresis (▶PAGE), which is a technique used to separate proteins on the basis of their size. ▶Two-dimensional polyacrylamide gel electrophoresis techniques, which are capable of resolving proteins based on their charge and size, are employed for analysis of conjugates formed in more complex mixtures of proteins. The proteins are visualized by staining in polyacrylamide gels. There are many traditional forms of chromatography that can be used to purify cross-linked proteins from complex mixtures of proteins, including ▶size exclusion chromatography, ▶affinity chromatography, ▶hydrophobic interaction chromatography and ▶high performance liquid chromatography (HPLC). These methods often require large quantities (nanomolar amounts) of starting material to provide enough of the purified conjugate of interest for further analysis. A more popular method for analyzing conjugates combines 2D-PAGE and mass spectrometric techniques. Cross-linked proteins are first resolved and visualized by 2D-PAGE. Appropriate samples are then excised from the gel and digested directly in the gel matrix with proteases or chemical reagents that cleave proteins after specific amino acids, generating mixtures of peptides that are characteristic of the interacting proteins. The peptide digests are then analyzed by mass spectrometric techniques on the basis of their charge to mass ratios. The masses measured for peptides in the digest are then compared against those masses predicted for families of peptides that result from the cleavage of potential protein candidates with a specific cleavage reagent. Integral proteins that are cross-linked in the isolated complex are identified from best matches generated by an appropriate predictive program; sites of cross-linking on specific peptide stretches between interacting partners are determined by similar methods.

There are many Web sites (ExPASy, SWISS-PROT, NCBI and other data bases) that contain search engines and predictive programs for these forms of analyses. Analysis of fmole quantities of peptides is now possible with modern mass spectrometric techniques, significantly enhancing the potential for detecting protein-protein interactions by cross-linking.

Clinical Relevance

Protein-protein interactions underlie many critical cellular processes that are important in both the normal functioning of cells and disease states, including ▶prion diseases and ▶Alzheimer's diseases. Chemical cross-linking has proven invaluable in determining the identity, spatial organization and sites of interaction between the protein components of numerous enzyme and structural protein complexes, further defining their role in physiological processes. Additionally, chemical cross-linking is used by industries in both the areas of therapeutics and diagnostics to couple proteins with different functions. For example, enzymes catalyzing specific reactions can be tethered to proteins that bind specific targets, facilitating either their detection or destruction.

References

1. Archakov AI, Govorun VM, Dubanov AV et al (2003) Protein-protein interactions as a target for drugs in proteomics. *Proteomics* 3:380–391
2. Nadeau OW, Sacks DB, Carlson GM (1997) Differential affinity cross-linking of phosphorylase kinase conformers by the geometric isomers of phenylenedimaleimide. *J Biol Chem* 272:26196–26201
3. Nadeau OW, Carlson GM (2002) Chemical cross-linking in studying protein-protein interactions. In: Golemis E, Adams P (eds) *Protein-Protein Interactions*. Cold Spring Harbor Laboratory Press, New York, pp 75–91
4. Wong SS (1993) *Chemistry of Protein Conjugation and Cross-linking*. CRC Press, Boca Raton
5. Hermanson GT (1996) *Bioconjugate Techniques*, Academic Press, San Diego

Protein Interaction Analysis: Phage Display

ZOLTÁN KONTHUR¹, JÖRN GLÖKLER^{1,3},
KARL SKRINER²

¹Department of Vertebrate Genomics, Max Planck Institute of Molecular Genetics, Berlin, Germany

²Department of Rheumatology and Clinical Immunology, Charité University Medicine Berlin, Berlin, Germany

³RiNA GmbH, Berlin, Germany

konthur@molgen.mpg.de
 gloekler@ma-network.com
 karl.skriner@charite.de

Synonyms

Bacteriophage display; bacteriophage surface display; phage surface display

Definition

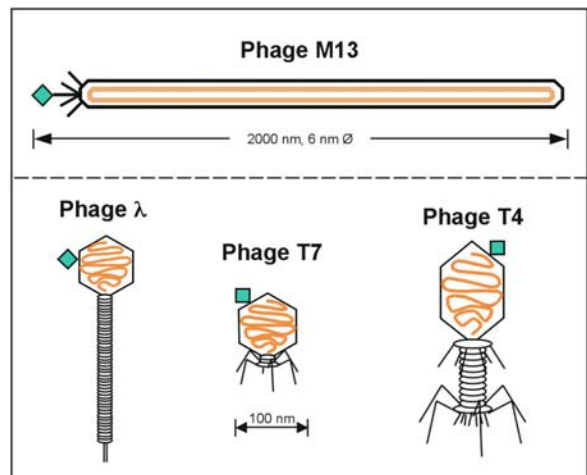
Phage display is a technology that allows the identification of interaction partners *in vitro*. The interaction can be of any type, protein-protein, DNA- or RNA-protein or compound-protein interaction. The key feature of phage display is allowing a (poly-) peptide molecule of any type to be presented on the surface of a **bacteriophage** by fusion to one of the virion's capsid proteins. The gene corresponding to the fusion molecule is incorporated into the phage genome inside the capsid and hence results in the physical linkage between the genotype and the phenotype. By cloning different cDNA-fragments into the phage genome, large libraries of up to 10^{10} different molecules can be generated and presented on the capsid surfaces. By applying these libraries to a selection process based on affinity purification, individual binders for individual target molecules can be enriched, while non-binders are discarded. Simply sequencing the cloned DNA fragment integrated into the phage genome reveals the identity and amino acid sequence of the binder.

Characteristics

The most prominent surface display technology is based on the filamentous bacteriophage M13, but other – alternative – phage display systems, such as the phages λ , T7 or T4 are also available (Fig. 1). All bacteriophages used have in common that they infect only Gram-negative *Enterobacteriaceae* and are used in conjunction with different laboratory *Escherichia coli* strains.

The viral coat of M13 contains only five different proteins, one major and four minor proteins in ~2700 and 5 copies, respectively. By now, fusions to all five coat proteins have been exploited for phage display, resulting in mono- or multi-valent displays of combinatorial libraries (1).

The most common application is amino-terminal fusion to the coat proteins pVIII or pIII. Fusion to pVIII is primarily suitable for peptides of ~8 amino acids only, since larger inserts hamper the phage capsid assembly process. In contrast, pIII tolerates larger polypeptide inserts of up to 300 amino acids but only if phage infectivity is not disturbed. To overcome size limitations and to present proteins as large as 86 kD, **phagemid** vectors are in use, which need **helper**



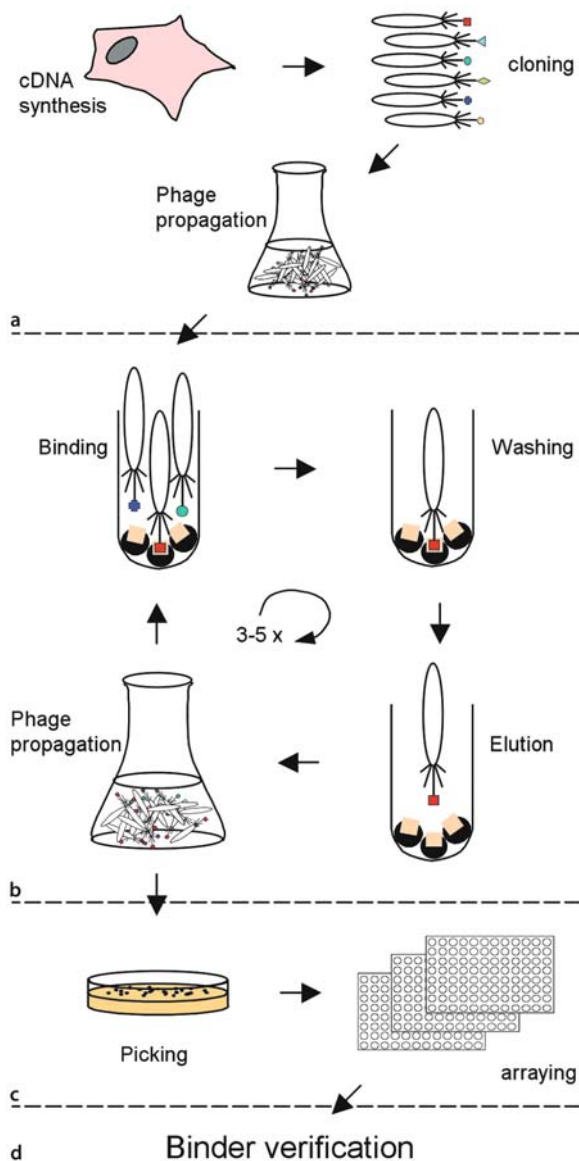
Protein Interaction Analysis: Phage Display.

Figure 1 Bacteriophages used for surface display. M13 phages are most commonly used, but the number of applications for alternative systems is steadily growing. Green squares represent the foreign polypeptide displayed as a fusion protein to one of the coat proteins on the surface of the phage.

phage superinfection. The helper phage contributes all the wild type proteins necessary for phage assembly and infectivity and therefore the desired recombinant peptide/protein is presented on the phage surface as a mixture with wild type proteins (1).

In an affinity driven *in vitro* selection process called **biopanning** (Fig. 2), phage display allows the screening of these large recombinant libraries for specific binders against virtually any type of target molecule, such as proteins, peptides, nucleic acids or chemical compounds. The selection is usually carried out on immobilized target molecules attached to a solid support material, e.g. plastic surfaces of immunotubes or microtitre plates or by directed, tag-dependent attachment of the selection targets to magnetic beads. After incubation of the library with the target, unbound phage particles are washed away and specifically bound virus particles are used for infecting *E. coli*. The propagated phage population is applied to the next round of biopanning. After several rounds, single clones of binders are isolated and tested individually for specific interaction.

Thus, phage display enables the rapid isolation of desired molecules from hugely diverse libraries. However, the system has some obvious limitations. Not all eukaryotic proteins can be expressed in *E. coli* and not all proteins can be presented on the phage surface due to biological limitations, e.g. failure to traverse the membrane. Also, there are no post-translational modifications, which may play a role in naturally occurring protein-protein interactions, carried out on these proteins.



Protein Interaction Analysis: Phage Display.

Figure 2 General scheme of phage display technology. (a) Generation of a phage display library by mRNA extraction, cDNA cloning into a phagemid vector and phage propagation. (b) Biopanning: affinity-driven enrichment of target-specific binders during several rounds of selection. (c) Generation of arrayed target-specific libraries suitable for high throughput screening by picking individual clones into microtitre plates. (d) Evaluation of selected binders by biochemical means such as ELISA, dot-blot, surface plasmon resonance (SPR) or by DNA hybridization on DNA microarrays.

There are three types of phage display libraries commonly used. The libraries consist either of recombinant **antibody fragments** originating from different species including human, random peptide

sequences or tissue or organism specific cDNA expression products.

In conclusion, phage display allows for the identification of artificial ligands and of natural binding partners specific for any given target molecule (2).

Clinical Relevance

With currently more than 1800 references in PubMed, phage display is a successful technique applied to various fields of protein engineering and detection of protein-protein interactions. Here, a short overview is given of the clinically most relevant fields of application, namely antibody engineering, allergen and autoantigen discovery and finding marker molecules in cancer.

Phage Display Derived Antibodies as Global Monitoring Tools, Diagnostics and Therapeutics

Antibodies or fragments thereof are the most commonly used biological reagents for the specific detection of proteins and hence they are the most prominent and promising candidates for a fast application in proteome-wide analyses. Especially after the completion of sequencing efforts such as the human genome project, the focus is steadily shifting to post-genomic research – the analysis of the gene expression products. For successful investigation on the whole proteome level, many individual proteins have to be characterized – in parallel – in respect of function, localization, post-translational state or just simply amount. Antibody arrays are believed to fulfill all the necessary requirements for a single multiparameter experiment best and therefore, large numbers of antibodies are needed. To overcome limitations of traditional immunization and hybridoma approaches (e.g. time, cost, quantity and reproducibility), biomolecular diversity selection methods, such as phage display are advantageous. Phage display is fast and very cost-effective and monoclonal antibody fragments are simply obtained. Furthermore, the process of phage display selection has been automated and utilizing this system, specific monoclonal antibodies can be generated against virtually all types of target molecules, peptides (i.e. epitopes), proteins, small signaling molecules or chemical compounds (3). By now several studies on antibody arrays with phage display derived antibodies have been reported and it has been shown that phage display derived antibody fragments can be used for protein expression profiling.

Generally, three types of antibody libraries can be distinguished, immunized, naïve and synthetic (4). The main difference between these libraries is the origin of diversity. Immunized libraries are generated from antibody genes amplified from an immunized individual (e.g. cancer patients, patients after infection etc.), naïve libraries are generated from germline antibody

genes of several individuals (i.e. without immunization), while in synthetic libraries the variability is introduced by DNA mutagenesis. The mutagenesis can be site-directed by degenerate primers/oligonucleotides or unspecific. Until the recent introduction of transgenic mice with fully human Ig repertoires, phage display was the only way of obtaining human monoclonal antibodies. An advantage of phage display derived antibodies is the possibility of generating fully human antibodies against human proteins, which can be used for *in vivo* diagnostic or therapeutic purposes and which have no or reduced immunogenicity and show no human anti-mouse antibody effect (HAMA). To date, the majority of commercially available therapeutic antibodies are still ►chimeric or ►humanized. However, the first fully human phage display-derived antibody (HUMIRA®) for the treatment of ►rheumatoid arthritis is on the market and a number of others, still in clinical trials, are to follow.

Display of Complex cDNA Expression Product Libraries from Tissues and Whole Organisms

In recent years, the numbers of applications utilizing cDNA expression product libraries displayed on phage surfaces are steadily increasing. The general concept became available in the mid-nineties with the development of vectors allowing the cloning of cDNA libraries as C-terminal fusions to M13 phage coat proteins. Carboxy terminal fusion is desirable since cDNA inserts obtained after poly(A)-priming and reverse transcription always contain translation stop codons and prevent the synthesis of hybrid coat proteins in N-terminal fusion vectors. In M13 however, the most commonly used phage coat protein for display (pIII) presents its N-terminus to the solvent and the integrity of its C-terminus is believed to be obligatory for efficient phage assembly. The vector pJuFo overcomes this problem by utilizing the strong natural interaction of leucine zippers to indirectly display cDNA expression products on the phage surface. For this purpose, the leucine zipper domain of c-jun is cloned at the 5' terminus of the phage coat protein pIII and the leucine zipper domain of c-fos is cloned as an N-terminal fusion to the cDNA expression product to be displayed. Both domains are expressed from the same phagemid vector and during phage assembly the jun-decorated pIII heterodimerizes with a fos-decorated cDNA product in the periplasm of the host, establishing the physical link between geno- and pheno-type (1). Meanwhile, cDNA libraries were also displayed as C-terminal fusions with pVI on M13, gpD in phage λ and gp10B in phage T7, but these types of cDNA display vectors have been less frequently used. A comprehensive list of cDNA libraries displayed on phage can be found in reference 2.

The main clinically relevant applications of cDNA phage display libraries have been the identification of natural binders to antibodies derived from patients suffering from ►allergy, ►autoimmune diseases or ►cancer.

Phage Display in Allergy: Identification of Allergens

Allergies are one of the most common health threats in industrialized countries. According to the World Health Organization, as much as 20% of the Western population suffers from type-I allergic symptoms such as allergic rhinitis, conjunctivitis or asthma characterized by an immunoglobulin E (IgE) response against normally innocuous antigens (allergens). Despite the gradually increasing knowledge about the mechanisms underlying allergic reactions, only a little is known about the repertoire of allergenic proteins provoking these responses. Typically, allergenic sources such as moulds, foods and mites contain complex mixtures of IgE-binding molecules which all need to be determined to allow the production of standardized allergen panels for successful diagnosis and allergy typing or subtyping, as well as application for desensitizing.

By cloning cDNA libraries of allergenic sources into phage display vectors and sequential enrichment on IgE purified from patient serum, many allergens have been identified (5), for example, allergens from birch, peanut, house dust mite, celery, wheat and different moulds.

A very prominent example of a highly complex allergenic source is the mould *Aspergillus fumigatus*. Due to the complex nature of both the cDNA expression library and the patient's IgE, the identification of large numbers of allergens demands that large numbers of individual clones be tested. This could only be achieved by combining phage display with genomic DNA array technology. Hence, DNA arrays were generated by picking and spotting ~13,400 individual clones of an enriched *A. fumigatus* phage displayed cDNA expression library. Sequential DNA hybridization of the arrays with already known allergens of the mould and randomly chosen clones from the array resulted in an additional 67 novel allergens being identified, adding up to a total of 81 IgE-binding proteins in *Aspergillus fumigatus*. Knowledge of large numbers of allergenic molecules is clinically relevant, since the occurrence of sensitization against different allergens can vary greatly between different individuals and because allergens with lower frequency (minor allergens) have been found to be cross-reactive structures in many allergenic sources (e.g. profilin in birch and in peanut). Many of the allergens have been recombinantly expressed and evaluated for their diagnostic performance. Recombinant allergens were found to show superior specificity in comparison with allergen extracts and hence, allow for elucidating patient- and disease-specific sensitization patterns.

Further information and references to allergens can be obtained from the database of allergens (► www.allergens.org) maintained by the Allergen Nomenclature Sub-Committee of the International Union of Immunological Societies.

Phage Display in Autoimmunity: Identification of Autoantigens

Autoimmune disorders affect 3% of the population in Western countries and are characterized by an impaired immune system leading to self-reactive autoantibodies and finally, as a consequence, to tissue destruction. In spite of the clinical importance of these diseases, little is known about their aetiopathogenesis. In recent years, considerable progress has been made in understanding immune function with regard to response specificity (e.g. major histocompatibility complex and T-cell activation). However, the underlying dysregulation leading to autoimmunity is still largely unknown. As studies with identical twins have shown, next to genetic factors environmental factors are of great importance for disease manifestation and onset (6). These factors are presumably also infectious events, which – non-specifically or stochastically – lead to the inflammatory cascade. For example, systemic autoimmune rheumatic diseases are characterized by the intense activation of the humoral immune system. Here, autoantibodies and immune complexes dominate the pathogenic and clinical picture. Autoimmune disease entities are characterized by the initial activation of the (auto) immune system leading to the inflammatory cascade (initiation phase), the chronification and perpetuation of the inflammatory processes at certain sites (transformation phase) and finally the destruction of the target tissues resulting in irreversible organ damage (effector phase).

Recently, the focus has been on establishing disease-specific autoreactivity patterns, since the diagnosis of the more than 80 autoimmune disorders is primarily based on clinical symptoms. To support the diagnosis however, specific and sensitive serological tests are essential and their availability can lead to an early diagnosis of the disease. Early diagnosis and treatment is important in order to reduce disease specific tissue destruction, for example of the joint in rheumatoid arthritis and the gut lining in celiac disease. Unfortunately, for most disorders, no conclusive serological markers are available, since in many diseases the currently identified autoantigens are mostly abundant, highly conserved and/or modified and able to form big complexes often associated with RNA and situated in apoptotic blebs.

In the pursuit of novel autoantigens, the application of cDNA expression product libraries displayed on phage

is becoming more and more popular, since phage display retains a high degree of variability. This is an important feature, as pathogenetically relevant autoantigens may differ markedly from patient to patient and the number of cells expressing them may be low. Hence using phage display, autoantigens can be found which would have been missed using different methods such as cell separation techniques, even when cells were pooled.

For example, a melanocyte cDNA library displayed on phage was used to identify new autoantigens in vitiligo, a common pigmentation disorder of the skin. Furthermore, phage display was used in thyroid-specific autoimmune disorders, such as Graves' disease and Hashimoto's thyroiditis, for autoantigen discovery as well as for the investigation of antigens in autoimmune diabetes, Sjögren's syndrome, multiple sclerosis, systemic lupus erythematosus and rheumatoid arthritis. However, except for rheumatoid arthritis, none of the autoimmune disorders have yet been investigated with the combined phage display and DNA array technology described for *Aspergillus fumigatus* allergy above. By applying this technology to other autoimmune disorders, various new autoantigens will be discovered. In future, the application of these methods might allow early diagnosis of numerous autoimmune diseases.

Phage Display in Cancer: Identification of Marker Molecules and Potential Therapeutic Targets

In cancer research, phage display has just started to be used for the elucidation and identification of marker and therapeutic target molecules. In the first instance, phage display peptide libraries were applied to *in vivo* screening in mice and man for the identification of peptide sequences that specifically attach to tumor vasculature, so-called homing peptides. These peptides can potentially be used for the specific delivery of anti-cancer drugs to tumor tissues (7). Another aspect in cancer research is the use of cDNA expression product phage display libraries for the discovery of marker molecules. For example, a phage-displayed cDNA library from an invasive ductal breast carcinoma tumor was used to investigate the humoral response of breast cancer patients, identifying a number of immunogenic antigens.

Different cancer tissues or cell lines derived cDNA expression product phage display libraries, e.g. breast carcinoma, prostate cancer and colorectal cancer cell line are available. Also available are libraries from tissues frequently associated with cancer, such as brain, leukocytes, kidney and lung. The screening of these libraries will undoubtedly identify a panel of marker molecules applicable in the serological diagnosis of different cancers in the future.

References

1. Cramer R, Rhyner C, Flückinger S et al (2005) Identification of natural protein-protein interactions with cDNA libraries. In: Sidhu SS (ed) *Phage Display in Biotechnology and Drug Discovery*. Marcel Dekker Inc, New York
2. Konthur Z, Cramer R (2003) High-throughput applications of phage display in proteomic analyses. *DDT: TARGETS* 2:261–270
3. Konthur Z, Walter G (2002) Automation of phage display for high-throughput generation of antibody fragments. *DDT: TARGETS* 1:30–36
4. Hust M, Dübel S (2004) Mating antibody phage display with proteomics. *Trends Biotechnol* 22:8–14
5. Rhyner C, Weichel M, Flückinger S et al (2004) Cloning allergens via phage display. *Methods* 32:212–218
6. Krenn V, Petersen I, Häupl T et al (2004) Array Technology and proteomics in autoimmune diseases. *Path Res Pract* 200:95–103
7. Zurita AJ, Arap W, Pasqualini R. (2003) Mapping tumor vascular diversity by screening phage display libraries. *J Control Release* 91:183–186

Protein Interactions

Definition

Protein interactions describe the formation of non-covalent reversible complexes between proteins, usually mediated by specific macromolecular recognition sites. In a wider sense, it also comprises the mutual influence of protein molecules on their dynamics, due to either attractive or repulsive forces, which may be mediated by a variety of sources, including the solvent, electrostatic, van der Waals or hydrophobic interactions, or steric repulsive forces.

- ▶ [Analytical Ultracentrifugation](#)
- ▶ [Protein Interaction Analysis: Chemical Cross-Linking](#)
- ▶ [Protein Interaction Analysis: Phage Display](#)

Protein Kinase

Definition

Protein kinase refers to an enzyme that catalyses the transfer of a phosphate group from a donor molecule (usually ATP) to an amino acid residue of a target protein.

- ▶ [Cap-Independent Translational Control](#)
- ▶ [NF-κB Pathway](#)
- ▶ [Receptor Serine/Threonine Kinases](#)
- ▶ [Receptor Tyrosine Kinases](#)

Protein Ligand Interactions Studied by X-Ray

MILTON T. STUBBS II

Institute for Biotechnology, Martin Luther University Halle-Wittenberg, Halle (Saale), Germany
stubbs@biochemtech.uni-halle.de

Definition

Knowledge of the binding mode of a ligand for its target protein – hydrogen bonding patterns, hydrophobic, electrostatic and van der Waals' interactions – can greatly facilitate the drug discovery process. X-ray crystallography is at present the only technique that can deliver information in atomic detail for both target protein and ligand simultaneously. Increasingly, crystallographic experiments involving state of the art technologies accompany the progress of potential drug candidates. Such experiments are helping us to understand the driving forces behind protein-ligand complex formation, which in turn will aid computational approaches to drug design.

Description

Rationale

The ultimate goal of any pharmaceutical research is the launching of a new drug. Currently, research efforts in the pharmaceutical industry are restricted to around 500 drug targets, most of which are proteins. The sequencing of the human genome has revealed a total of some 35,000 genes encoding for proteins. Clearly, genomics will have a massive impact on determining new drug targets and developing novel therapies.

The route from target identification to drug is long and arduous, however – it may take up to two decades to get a new compound on to the market. Even if a large number of compounds are available from natural and synthetic sources (such as compounds synthesised for other drug design efforts), it is not an easy task to find out which ones might be useful for a particular indication. This requires time consuming and expensive screening of the compound libraries. The screening process can be divided into three basic stages: primary screening of the whole or part of the library, resulting in a 'hit', modification of the hit to yield compounds of suitable pharmacological and physico-chemical properties (a 'lead') and finally refinement of the lead in biological assays to provide a drug candidate for entry into clinical trials.

While high throughput screening and combinatorial chemistry are ideally suited to generating initial hits, translation of these into leads and drugs is fraught with

difficulties. Knowledge of the detailed binding mode of hit compound to target protein can greatly enhance this process. The three dimensional structure of a protein–ligand complex allows rationalisation of the existing experimental data and facilitates further evaluation and modification of the compounds in question. The full potential of structure-based techniques is obtained when integrated into an iterative process, the drug design cycle (Fig. 1). Structural information can be achieved *via* computational methods (1), through NMR (2) or by means of X-ray crystallography (3) (Fig. 2). This last method is the subject of this chapter.

Characteristics of Protein Crystals

Protein crystals are highly ordered molecular arrays, formed by equilibration of supersaturated solutions under special conditions of pH, temperature and salt and/or solvent concentration. In the case of protein–ligand interactions, crystallisation conditions for the target molecule are usually already known. Nevertheless, each protein has its own peculiar crystallisation conditions, which can change significantly through, for example, a change in purification protocol, protein source (e.g. it is extremely rare for homologous proteins from different species to crystallise under the same conditions) or complex partner. This should always be borne in mind when attempting to repeat a published crystallisation procedure; even when the chosen protein construct is identical, it is always advisable to vary the conditions systematically to

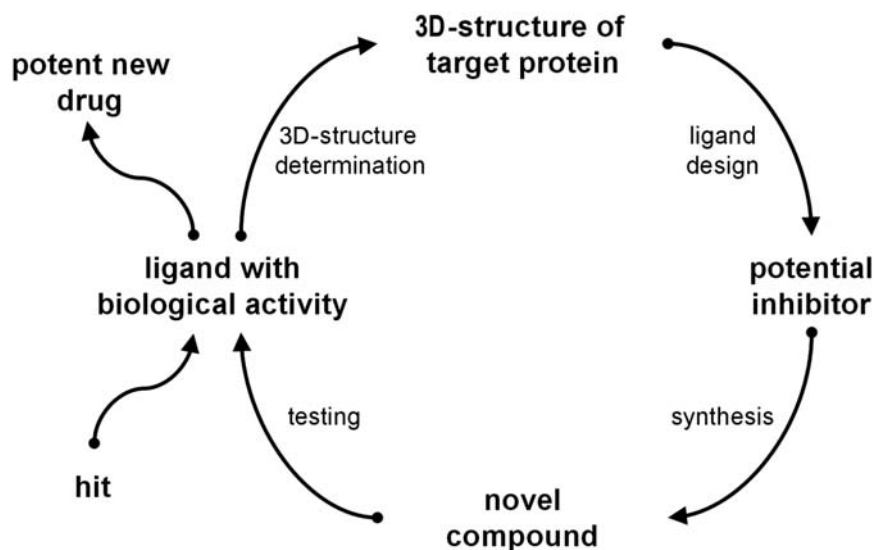
maximise the chance of obtaining diffraction quality crystals.

Contrary to an often-held misconception, protein crystals are highly solvated (Fig. 3); up to 80% solvent content has been reported. Thus, it is possible to diffuse substrates, ligands and inhibitors into protein crystals, so that in favourable circumstances one can follow catalytic reactions and determine ligand-binding modes at atomic [resolution](#) accurately.

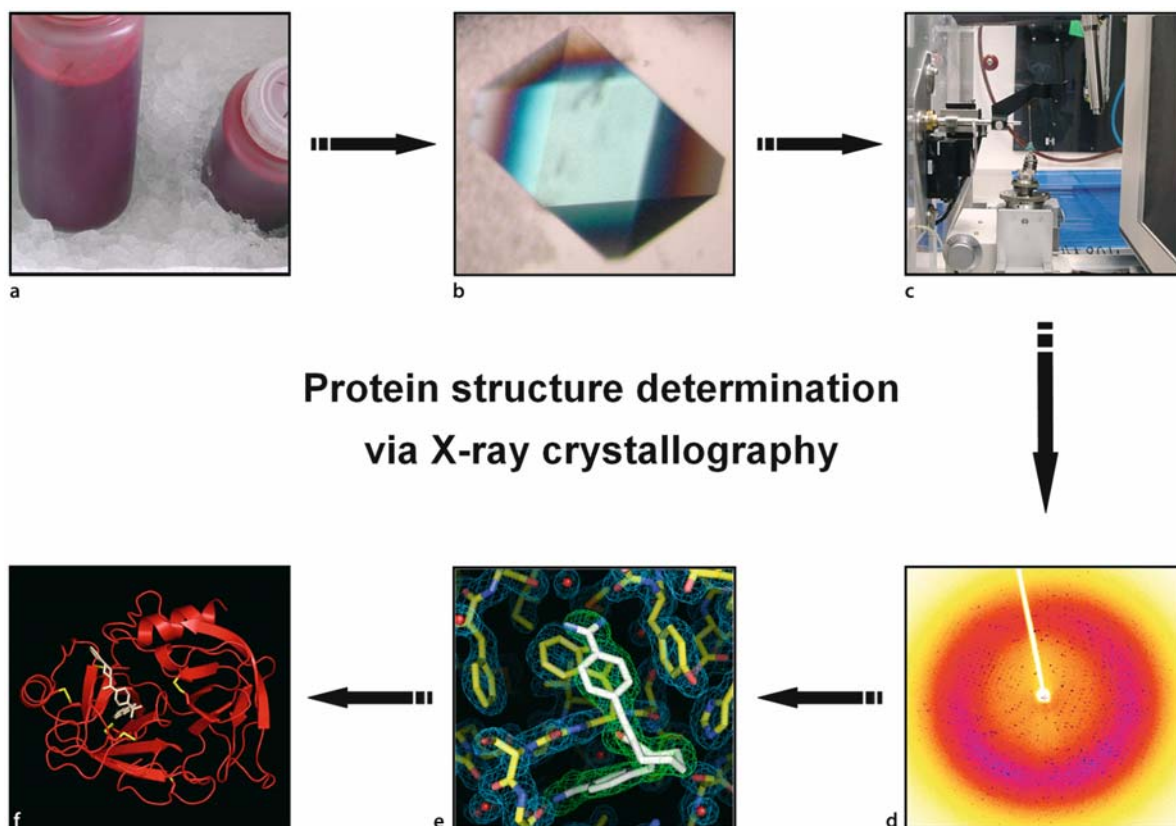
Protein–Ligand Interactions in Crystals: Co-Crystallisation vs. Soaking Experiments

Before a high resolution X-ray experiment can be carried out, it is necessary to introduce the chosen ligand into a crystal of the target protein. The most obvious way to do this is to co-crystallise the ligand with the protein. The ligand is incubated at a typical concentration of 1–10 mM together with the target protein prior to crystallisation. As noted above, such a small change in the protein preparation may lead to changes in crystallisation behaviour; it is therefore advisable to carry out a small scale screening (e.g. variation of the pH by ± 0.5 units, adjustment of precipitant concentrations). Ligands to be analysed often exhibit poor solubility, particularly under the often high salt conditions prevailing in the crystallisation experiment; this can be alleviated through addition of small concentrations of DMSO.

Co-crystallisation can consume large quantities of target protein. Far more convenient (and cost effective)



Protein Ligand Interactions Studied by X-Ray. Figure 1 The structure-based drug design cycle. Starting with a 'hit' or a 'lead' from biological screening, the structure of the compound is determined in the presence of the target protein. This allows analysis of the interactions made between the ligand and protein, and suggests strategies for modifying the test compound. More complete biological testing of these new generation compounds (including e.g. animal models) results in new leads whose binding mode can often differ from that of the parent compound. It is therefore necessary to go through this cycle in an iterative manner until a novel drug candidate is achieved.



Protein Ligand Interactions Studied by X-Ray. Figure 2 (a) Starting with sufficient quantities of protein, (b) crystals may be obtained given the correct biochemical conditions. (c) The crystals are measured using an X-ray camera to produce (d) a diffraction pattern. Evaluation and phasing of the diffraction pattern yields (e) an electron density whose interpretation results in (f) the three-dimensional structure of the complex under investigation.

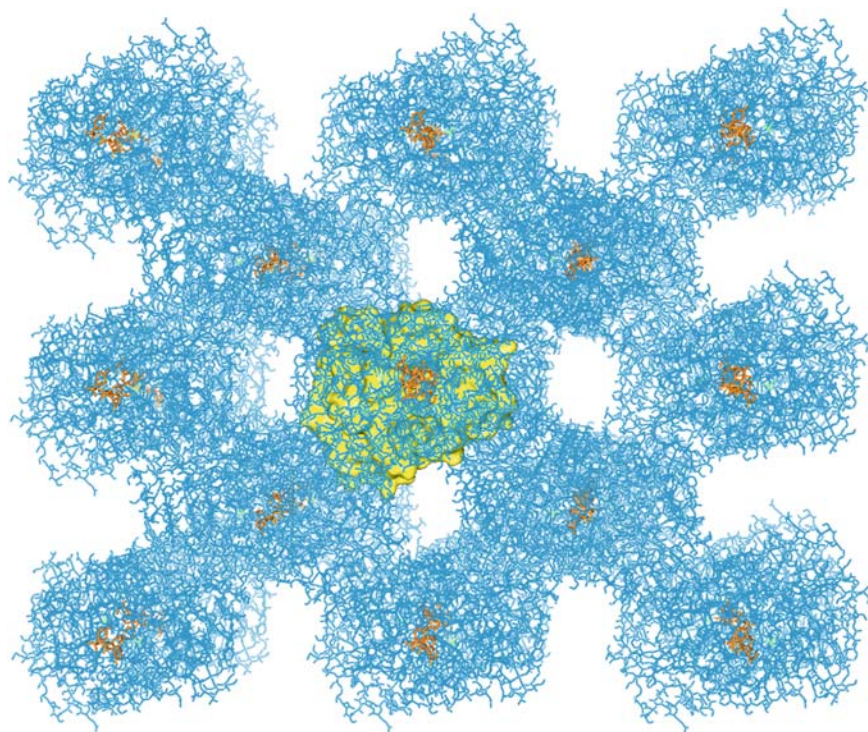
is the establishment of a crystal soaking system. Prerequisite for such an approach is a crystal form in which the binding site of interest is free both chemically (absence of covalently bound or high affinity ligand) and crystallographically (i.e. the binding site faces towards the bulk solvent, Fig. 3). A suitable crystal is introduced into a harvesting solution containing the chosen inhibitor, which is then free to diffuse through the crystal. The harvesting solution is a modified form of the crystallisation buffer, usually containing a higher concentration of precipitant to compensate for the missing protein and serves to stabilise the crystal. As in the co-crystallisation experiments, it is possible to increase the solubility of the ligand through addition of DMSO. A practical alternative is the chemical cross-linking of the crystals using e.g. glutaraldehyde (5), allowing buffer exchange to more favourable conditions.

From Crystal to Electron Density Map

The collection and processing of X-ray data (structure factors) from protein-ligand complex crystals follow

standard crystallographic procedures. As changes in the diffraction pattern produced by the inclusion of the ligand are very small, the resulting **electron density** is very sensitive to the quality of the data. It is therefore necessary to collect a very complete data set with good merging statistics. Experience shows that it is useful to use a **resolution** cut-off where the diffraction pattern is still strong ($I/\sigma(I) > 3$ in the outermost resolution shell).

Having collected the data, the **phase problem** must be solved; this is covered elsewhere in this volume. In most cases, the structure of the target is known prior to the study of protein ligand interactions, so that the complex can be solved using difference Fourier techniques. It is then sufficient to submit the known model to a rigid body refinement prior to calculating a new electron density. Occasionally, however, it is necessary to resort to molecular replacement methods, for example when co-crystallisation results in a new crystal form. This involves positioning the known model in the new crystal cell to maximise the fit between the calculated and observed structure factors.



Protein Ligand Interactions Studied by X-Ray. Figure 3 Protein crystals are ordered molecular arrays; one protein molecule (trypsin, in yellow) is depicted here together with some of its nearest neighbours (4). As can be readily seen, large solvent channels are present through the crystal; the solvent content of protein crystals ranges typically between 30% and 80%. It is therefore possible to diffuse small molecule ligands such as substrates or inhibitors into the crystal. If the ligand-binding site faces these solvent channels (also known as ‘bulk solvent’), then soaking the crystal in an inhibitor solution will allow the inhibitor molecules to reach the active sites (shown in orange), facilitating rapid structure determinations of a large number of compounds.

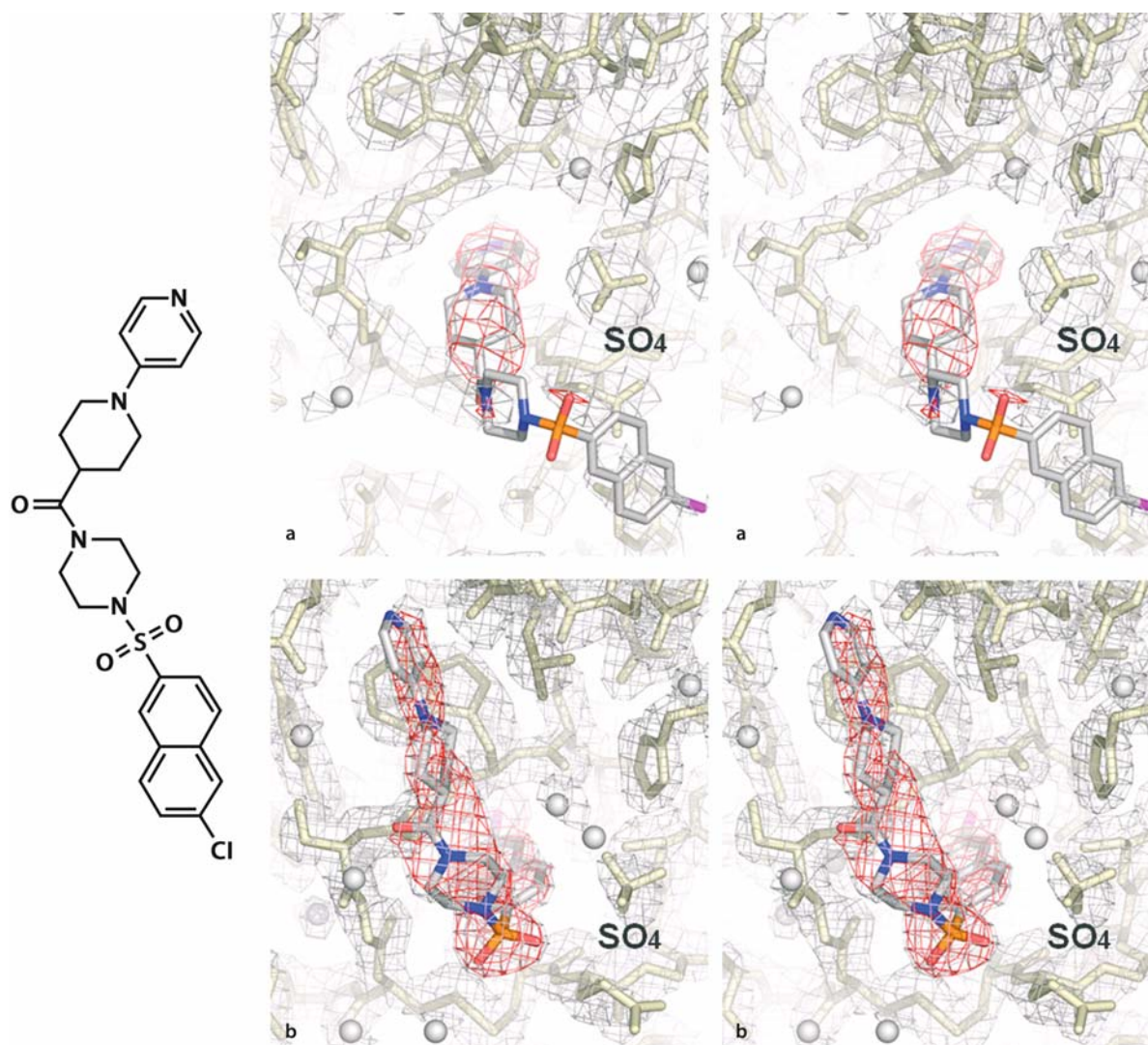
The quality of the resulting electron density map is not only dependent on the quality of the experimental data, it is also very sensitive to the molecular model used. Every atom of the model contributes throughout the diffraction pattern and misplaced atoms therefore have an effect on the whole of the electron density. It is therefore extremely important to model the electron density correctly throughout the structure – not simply in the neighbourhood of the ligand. To this end, it is good practice to remove all non-covalently bound molecules (for the most part solvent molecules) from the model prior to map calculation. Unless the density for the ligand is absolutely clear (Fig. 4), it pays to check the density of the whole protein model and all solvent molecules, rebuilding where necessary. The final model should be refined in the usual way, monitoring the crystallographic R-factor $\Sigma|F_{\text{obs}} - F_{\text{calc}}| / \Sigma|F_{\text{obs}}|$ and model geometry.

Problems Associated with X-Ray Protein-Ligand Complexes

Unfortunately, there are no objective criteria for determining whether the ligand density has been

correctly modelled – the R-factor is rather insensitive to inclusion of the few atoms of the inhibitor. The interpretation of the ligand electron density is therefore highly subjective; a series of representative examples are given below.

One oft quoted criterion for the quality of a structure is that of resolution. While ambiguous electron density becomes harder to interpret at lower resolution (see Figs. 4, 5 for examples), this does not mean that low resolution studies are of no value. If the data are well phased (i.e. the model fits the density well) and the ligand has a well defined binding mode, then a 3 Å structure can be as ‘correct’ as one at high resolution. The greatest problem to be dealt with when studying protein-ligand interactions using X-ray crystallography is breaks in the electron density (Figs. 4, 5). Unlike interpreting electron density for the protein, where the covalent connections between individual atoms allow an ‘educated guess’ in regions of ambiguity, the only guides available to map the ligand are the shape of the electron density, chemical knowledge of the ligand and inspection of the protein environment. Ambiguities may arise from a variety of causes. If a region of the

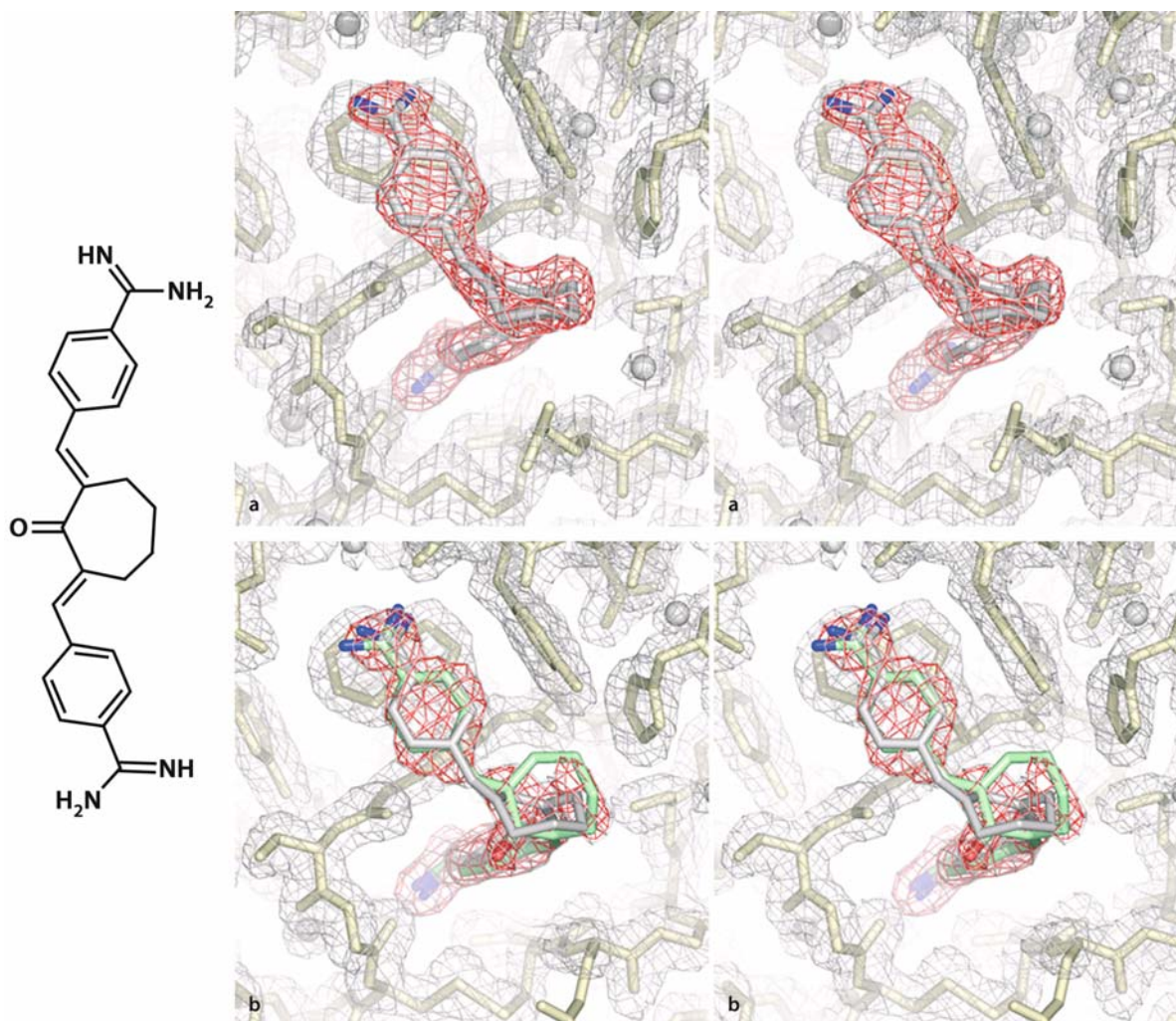


Protein Ligand Interactions Studied by X-Ray. Figure 4 (a) Problems in interpreting protein-ligand interactions occur when the electron density is incomplete. In this case, the form of the density within the binding pocket clearly shows the presence of the piperidinyl/pyridinyl group in the specificity pocket of trypsin. The absence of density outside the binding pocket is a result of a lack of preferred contacts for the chloronaphthyl group. Despite the limited resolution of 3 Å, the quality of the density allows ready interpretation (note the definition of the sulphate ion) as the model (and thereby the phases) are well characterised. (b) X-ray data for the same inhibitor in a different crystal form reveals electron density for the complete ligand at a resolution of 2.1 Å. In contrast to 4(a), the chloronaphthyl group enters the specificity pocket, while the piperidinyl/pyridinyl moieties occupy a secondary binding site specific for the structurally related coagulation proteinase factor Xa. The alternative binding modes are less a result of different crystal packing effects than of alternative crystallisation conditions; crystals corresponding to 4(a) grow at a pH of 7, whilst 4(b) crystals grow at pH8 (6). The binding mode (b) could be confirmed in rat trypsin variants designed to mimic the binding pocket of factor Xa (7). This as well as Figs. 5, 6 are in stereo; focusing the left eye on the left image and the right eye on the right one will result in being able to observe the structure in three dimensions.

ligand makes no preferred contacts with ordered atoms, then the electron density becomes weak and cannot be interpreted (Fig. 4a). Alternatively, there may be incomplete binding of the compound, resulting in an overlap of densities from the ligand and from partially bound solvent molecules. An extension of this problem

is the presence of several species that bind, as in racemic or enantiomeric mixtures (Fig. 5).

Having built an atomic model, the resulting structure must be checked for validity. It is becoming apparent that the concept of 'the' structure is a gross oversimplification. However, the presence of electron



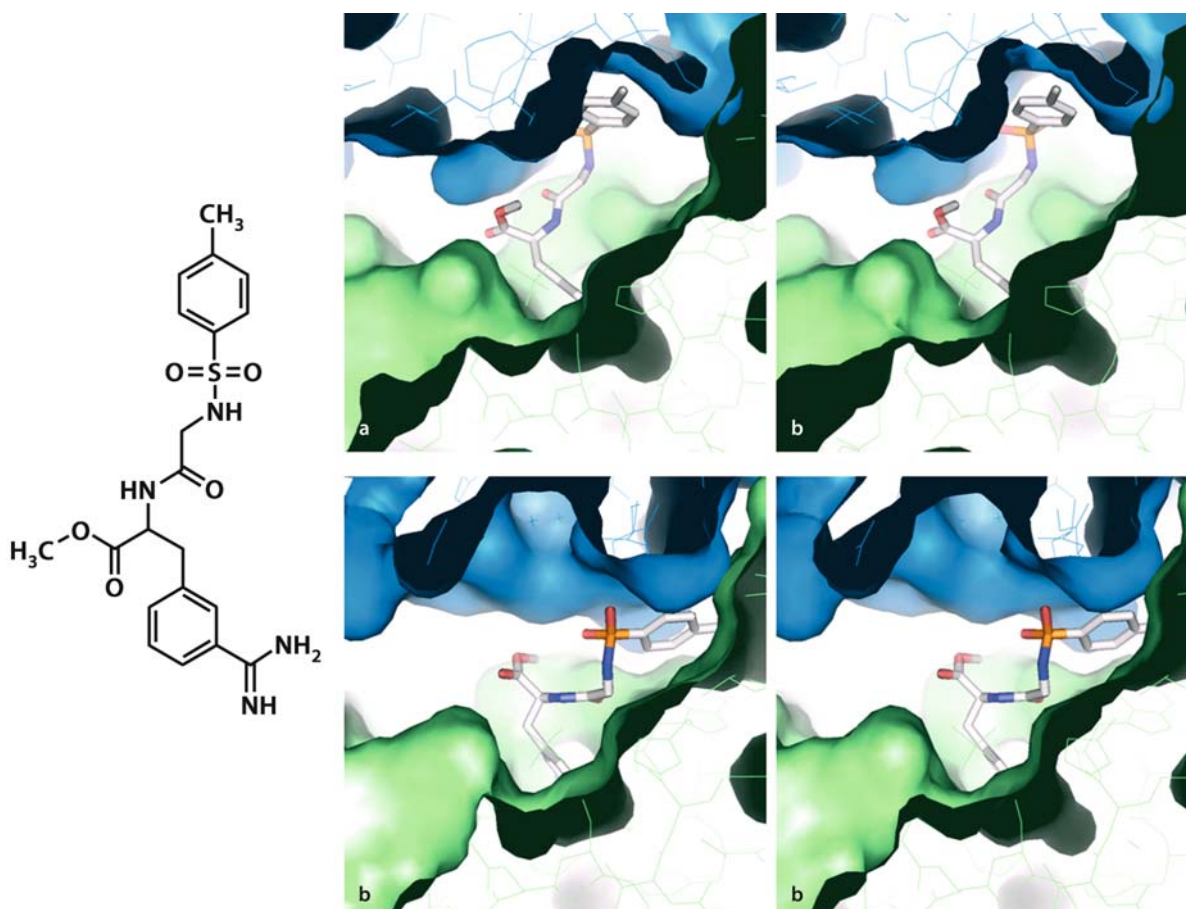
Protein Ligand Interactions Studied by X-Ray. Figure 5 (a) Electron density of the bis-benzamidino inhibitor in co-crystals with a trypsin variant at 1.8 Å resolution reveals clearly the (Z,Z)-configuration of the ligand, apparently favoured in this crystal form. (b) Soaking of the same inhibitor into a different crystal form indicates the presence of both (Z,Z) (white) and (E,Z) (green) configurations at a resolution of 2.0 Å (5). As witnessed here, the presence of several ligands and/or multiple binding modes can result in ambiguities in the electron density that are often difficult to interpret; in most cases only the dominant binding species can be identified correctly. This can be a particular problem with ligands of low occupancy with partially bound solvent molecules.

density proves that the observed structure must exist under certain circumstances. One striking example is provided by the pH dependent ligand-binding mode detected in trypsin crystals (Fig. 4) (6). Thus the conditions of the crystallographic experiment should be taken into account in interpreting the results. Furthermore, the crystal environment itself can affect the observed interaction (Fig. 6), with neighbouring molecules influencing the conformation and interactions taken up by the ligand. In such cases, it appears reasonable to assume that the binding of the compound is weak; nevertheless, multiple binding modes have been observed for strong inhibitors (unpublished

results). Such multiple crystal forms are proving highly useful as a tool for studying protein ligand interactions.

Therapeutic Consequences and Clinical Relevance

The unravelling of the human genome, as well as those of important human pathogens, promises identification of novel drug targets for the treatment of disease. However, mere knowledge of a candidate gene sequence is insufficient for the production of new therapeutic agents. Detailed information on structural aspects of protein-ligand interactions will have important repercussions for the translation of the one dimensional



Protein Ligand Interactions Studied by X-Ray. Figure 6 The binding of this inhibitor is strongly influenced by packing effects. In crystal form (a) (that also depicted in Fig. 3), the basic benzamidine function of the ligand binds to the active site in its 'target' molecule (green), whilst the terminal tosyl group nestles against the neighbouring symmetry related molecule (blue). In crystal form (b), the tosyl group is sandwiched between its 'target' and a symmetry molecule. This suggests that the tosyl group contributes little to binding, at least under the conditions prevailing in the crystal (8).

genome data into novel pharmaceuticals. While considerable progress has been made using crystallography as a tool – for example, the use of protein crystallography for screening compound mixtures (9) – our understanding of the biophysical foundations underlying these non-covalent interactions is still rather rudimentary. For example, it is still not possible to predict the affinity of a ligand accurately from the structure alone (10), particularly in cases where target flexibility is an issue (8). Nevertheless, considerable advances are to be anticipated in the near future, forming a solid basis for computational methods in drug design.

►Protein Crystallization for X-Ray

References

1. Klebe G (2000) Recent developments in structure-based drug design. *J Mol Med* 78:269–281
2. Pellechia M, Sem DS, Wüthrich K (2002) NMR in drug discovery. *Nature Rev Drug Disc* 1:211–219
3. Blundell TL, Jhoti H, Abell C (2002) High throughput crystallography for lead discovery in drug design. *Nature Rev Drug Disc* 1:45–54
4. Renatus M, Bode W, Huber R et al (1998) Structural and functional analyses of benzamidine-based inhibitors in complex with trypsin: Implications for the inhibition of factor Xa, tPA and urokinase. *J Med Chem* 41: 5445–5456
5. Rauh D, Klebe G, Stürzebecher J et al (2003) ZZ made EZ: Influence of inhibitor configuration on enzyme selectivity. *J Mol Biol* 330:761–770
6. Stubbs MT, Reyda S, Dullweber F et al (2002) pH-dependent binding modes observed in trypsin crystals – Lessons for structure-based drug design. *Chem Bio Chem* 3:246–249
7. Reyda S, Sohn C, Klebe G et al (2003) Reconstructing the binding site of factor Xa in trypsin reveals ligand-induced structural plasticity. *J Mol Biol* 325:963–977

8. Rauh D, Klebe G, Stubbs MT (2004) Understanding protein-ligand interactions: The price of protein flexibility. *J Mol Biol* 335:1325–1341
9. Nienaber VL, Richardson PL, Klighofer V et al (2000) Discovering novel ligands for macromolecules using X-ray crystallographic screening. *Nat Biotechnol* 18: 1105–1108
10. Dullweber F, Stubbs MT, Musil D et al (2001) Factorising Ligand Affinity: A Combined Thermodynamic and Crystallographic Study of Trypsin and Thrombin Inhibition. *J Mol Biol* 313:593–614

Protein Localization

►Immunochemical Methods, Localization

Protein Microarrays as a Tool for Protein Profiling and Functional Proteomics

ULRIKE KORF
Molecular Genome Analysis, DKFZ, Heidelberg,
Germany
u.korf@dkfz.de

Synonyms

Protein array, antibody array, protein chip, antibody chip

Definition

Protein microarrays contain a defined set of proteins immobilized at high density in a geometric pattern on a solid surface, typically a glass ►microscope slide. Each single sample on this solid surface can be addressed in a similar way to the organization of samples in a multiwell plate.

Ideally the complete ►proteome of an organism, either as the proteins themselves or as highly specific capture molecules (Table 1) directed against all proteins of the proteome, is represented in a protein microarray. Both types open the possibility of monitoring time-resolved dynamic events in a cell, tissue or organism and are expected to impact biological and biomedical knowledge. Protein microarrays are envisaged as overcoming the well-known technical demands of ►standard proteome research technologies by allowing detailed studies with less material, higher throughput, better reproducibility and in less time.

According to the classification introduced by Kodadek, protein microarray applications can be divided into two principal types (1). Microarrays containing antibodies to assess the concentration of proteins are called protein detection arrays and typically antibodies are immobilized as capture agents. This type of protein microarray corresponds to the technology employed for ►RNA expression profiling. Nevertheless, as RNA expression-profiling data do not necessarily correspond to the actual protein concentration in a cell and RNA levels do not deliver information on functional aspects, protein detection microarrays are a necessary and valuable adjunct, allowing determination of the actual protein concentration in a cell. The functional properties of a large collection of immobilized recombinant proteins are usually investigated by the second category of protein microarray, the protein-function

Protein Microarrays as a Tool for Protein Profiling and Functional Proteomics. Table 1 Comparison of different capture agents for protein microarrays¹

Capture Agent	Affinity	Specificity	Production time	Amount starting material	Automation	Immobilization
Monoclonal antibody	medium-high	high	very long	high	no	difficult
Polyclonal antibody	high	medium-high	long	high	no	difficult
Phage library (Fab/scFv)	medium	high	short	low	yes	easy
Antibody-like scaffold	medium	high	short	low	yes	easy
Aptamer	medium-high	medium-high	short	low	yes	easy
Imprinted polymer	low-medium	medium	short	high	potentially	n.a.

array. Both principal types of protein microarrays present unique ways to screen biochemical properties at the proteome level.

Characteristics

Fabrication of Protein Microarrays

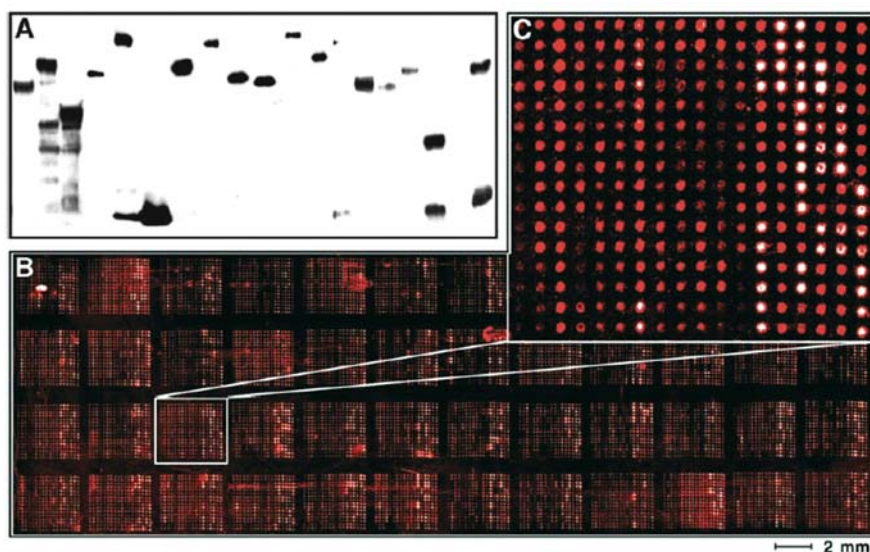
The fabrication of protein microarrays relies on methods and instruments originally developed for production of [▶DNA arrays](#). In order to generate an addressable pattern on the microarray surface a high-precision instrument, a microarray spotter, which delivers a tiny droplet of the dissolved protein (sub-nl volume) is employed as a printing device. This results in the formation of microscopic spots approximately 200–250 μm in diameter. The distance between spots varies between 300–500 μm . In principal, proteins can be arrayed at high spatial density (1000 spots/ cm^2) and a standard microscope slide (25 mm \times 75 mm) can accommodate thousands of samples in parallel as shown in Fig. 1. The droplet can be delivered to the surface by direct contact between the tip and the solid support. In surface contact printing, the pins are dipped into the sample solution, resulting into sample uptake into [▶split pins](#) or transfer of a small volume to the tip of a solid pin. The volume delivered to the microarray surface by direct contact depends on parameters such as pin type, sample fluidity and slide surface. Drop delivery in a non-contact fashion using piezo-element based arrayers is accomplished in a more gentle way with higher precision. Up to now the use of

piezo-element based arrayers has been restricted by the comparatively low throughput.

Proteins can be immobilized on the solid surface by covalent attachment or by absorption. Covalent attachment requires the activation of the surface of a glass microscope slide with a functionalized [▶silane reagent](#). Alternatively a thin pad can be attached to the glass slide, trapping proteins into a 3-dimensional structure. Frequently, applied materials are ultra-thin polyacrylamide gel pads or nitrocellulose layers, but carbohydrate-based matrices have also been employed to generate a 3-dimensional coating. A 3-dimensional coating or a pad is assumed to accommodate the tertiary structure of proteins better than will a regular glass surface. However, in order to maintain the native character of immobilized proteins, the effect of the solid surface on the stabilization of the proteins has to be examined for each experimental set-up. Table 2 gives an overview on the different surface coatings provided by commercial suppliers.

Methods for Detection

Proteins captured on the protein microarray need to be visualized for detection. Methods such as [▶chemiluminescence](#), [▶fluorescence](#) and [▶radioactivity](#) facilitate the quantification of proteins and were successfully adapted to the microarray surface. Each method has advantages in terms of [▶sensitivity](#) and signal-to-noise ratio. Nevertheless, instruments employed for detection of microarray spots have to offer a resolution matching the small spot size.



Protein Microarrays as a Tool for Protein Profiling and Functional Proteomics. Figure 1 Protein microarray accommodating 5800 unique GST fusion proteins from yeast. Proteins were spotted in duplicate and were probed with anti-GST antibody for visualization. The enlarged image in the upper right corner shows one of 48 blocks of the proteome chip. The inserted figure in the upper left corner shows Western blot analysis of 19 representative GST fusion proteins. Picture taken from [6].

Protein Microarrays as a Tool for Protein Profiling and Functional Proteomics. Table 2 Overview on solid supports for protein microarrays

Structure	Surface	Slide	Company
3-D	Polyacrylamide	Hydrogel	Perkin Elmer
3-D	Hydrogel	AccelIR8	Schott Nexterion
3-D	Nitrocellulose	FAST	Schleicher&Schuell
3-D	Nitrocellulose	Vivid	PALL
3-D	Hydrophilic polymer	Code-Link	Amersham Surmodics
3-D	Dendrimeric polymer	ProteoBind	VBC Genomics
2-D	Coatings with functionalized silane	Various types	Greiner, Schott/Quantifoil

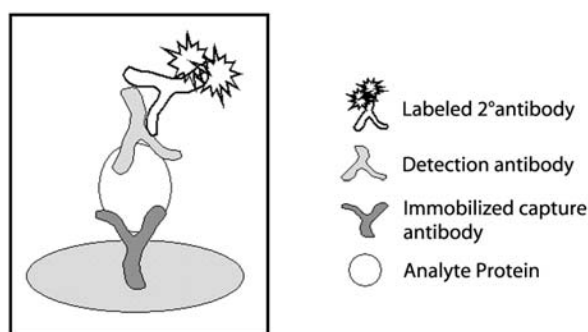
Fluorescence-based detection is currently the most popular type of detection in microarray experiments because standard microarray readers rely on fluorescence detection to generate the experimental read-out. Therefore a huge collection of reagents is commercially available, which includes pre-labeled secondary antibodies and labeling reagents. Chemiluminescence, widely applied in ►Western blotting, although highly sensitive has drawbacks in terms of ►dynamic range and compatibility with multiplexing. However, certain types of enzymatic activity, e.g. ►kinase activities are best assessed by the incorporation of radioactive label into substrate proteins and subsequent exposure to a ►phospho-imaging plate. In addition, signal amplification based on enzymatic methods is used in protein microarray detection; for example the signal intensity has been magnified by adapting the rolling circle detection method (►RCA) to the microarray format (2) and by the application of the ►TSA method.

Applications

Protein-Detection Microarrays

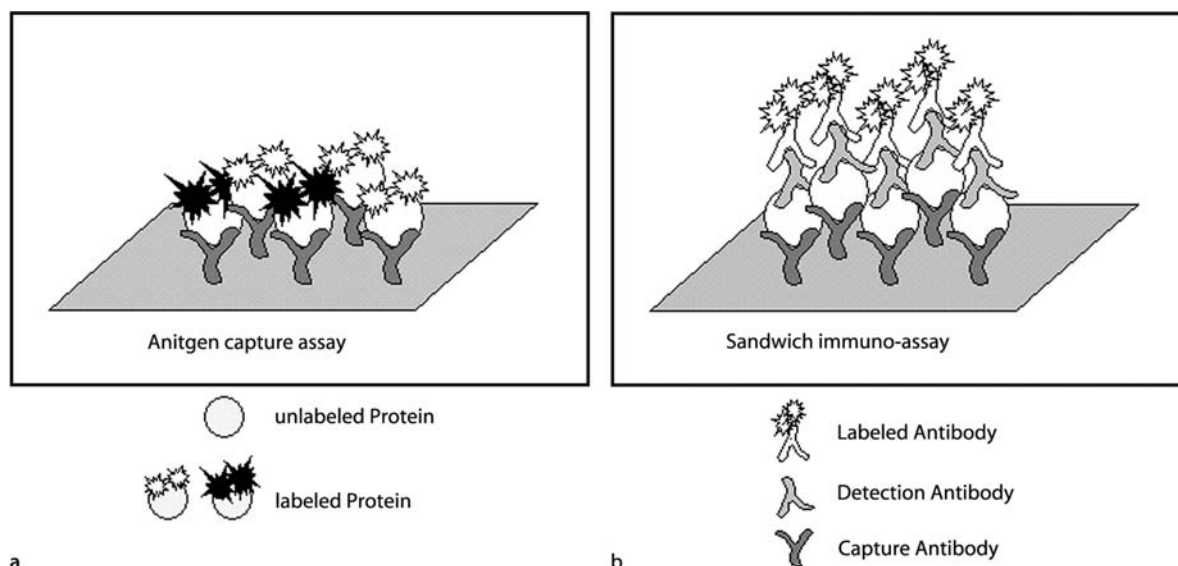
A reliable method for estimating protein concentration over several orders of magnitude still presents a challenge. A specific protein has to be detected among 10,000s of other different proteins. This challenge is multiplied by the fact that the copy number of proteins varies between a few copies per cell and 100,000s copies per cell. In addition, proteins, unlike DNA, are not very stable. Many proteins quickly denature upon lysis and undergo degradation. Furthermore, proteins are ►posttranslationally modified, e.g. by ►phosphorylation, and distinguishing between different modified protein species has to be accomplished as well.

Antibodies have been widely used for the detection of proteins. The antibody-based detection of analytes was first described in 1929 and this technique was greatly advanced by the development of enzyme-linked immunosorbent assays (ELISA) in which an antibody is immobilized to a solid support to capture the



Protein Microarrays as a Tool for Protein Profiling and Functional Proteomics. Figure 2 ELISA principle.

corresponding antigen from a complex mixture of proteins. Standard ELISA techniques rely on two antigen-specific antibodies. The first antibody is immobilized onto the carrier surface while the second antibody is employed to detect the captured analyte. The immuno complex is commonly quantified using a secondary antibody that results in a readable and quantifiable signal. This mode of detection *via* a sandwich antibody complex (Fig. 2) is called indirect detection. In the late 1980s Roger Ekins opened the way to miniaturization of immune assays by the introduction of the “microspots” concept (3). According to Ekins, extremely sensitive multianalyte detection can be achieved by immobilizing antibodies in extremely small spots to a solid support. The critical point is, that in contrast to standard ELISA technology, analyte consumption can be neglected and accurate measurements are possible even at low protein concentrations. The degree of ►antibody cross reactivity and the signal-to-noise ratio are crucial for accurate protein quantification. By combining the ideas of multiplexing and increased sensitivity, the development of antibody microarrays for quantification purposes became the subject of



Protein Microarrays as a Tool for Protein Profiling and Functional Proteomics. Figure 3 (a) Antigen capture assay performed by labeling of sample proteins with a fluorescent dye. (b) Sandwich immunoassay employing two independent antigen-specific antibodies.

intense research. Protein-detection microarrays can be thought of as multiplexed ELISAs, requiring a set of highly specific antibodies. Two different strategies were developed for the detection of captured proteins on the protein microarray.

Antigen Capture Assay

The antigen capture approach requires introduction of a detectable label into all proteins of the sample. In general, a suitable marker is introduced into all proteins by covalent attachment to functional groups of specific amino acid side chains. Chemically activated fluorescent dyes are frequently used as directly readable marker. In this case, the signal-intensity is proportional to the amount of sample bound to the capture agent. Since the direct detection does not involve a signal amplification step, it is therefore of limited sensitivity (1 ng/ml). Alternatively, a small molecule ▶**hapten**, which can subsequently be recognized with a corresponding antibody, can be used. This hapten-based approach opens access to further signal amplification. The antigen capture assay is typically employed for direct comparison of two different samples and therefore requires the introduction of two different markers (indicated by a black star and white star in Fig. 3a) that can subsequently be distinguished. The read-out is ratiometric and for this reason only relative information on protein concentration is gained. The need to label the protein may present a problem: Some proteins are labeled preferentially on their antigenic epitopes and may thus escape detection. In this case the sandwich immunoassay presents an alternative.

Sandwich Immunoassay

The sandwich approach does not require the protein to be labeled and significantly lower detection levels can be achieved. In addition to the capture antibody, a second antigen-specific antibody (detection antibody) directed against a different epitope of the protein is used (Fig. 3b). The fact that independent and highly specific antibodies are required as affinity agents for each protein has slowed down the broad application of this type of assay. In sandwich immuno assays the sample preparation is less demanding and not prone to batch-to-batch variation. This approach will probably dominate microarray based immunoassay techniques. All antibody-based strategies are complicated by the fact that antibody specificity is less restricted than expected and in general foreign epitopes on different proteins will be recognized. Especially in a multiplexed assay format, this can be problematic with regard to signal specificity. Because the recognition of contaminating proteins by two independent antibodies is less likely, the application of two independent antibodies guarantees a higher degree of signal specificity. Microarray-based sandwich immunoassays have been most successfully employed for the quantitative detection of cytokines in body fluids.

Protein-Function Microarrays

The protein microarray technology provides a platform to study *in vitro* protein function on a ▶**genome**-wide or a system-wide basis. Immobilized proteins can be assayed in parallel under many different conditions and with many different samples. Functional protein arrays

are envisaged to be a promising tool for drug discovery. Thus, due to the inherent potential of process automation, protein microarrays present an ideal tool for the study of biological systems.

Whole proteome microarrays are difficult to achieve for higher organisms and have only been realized for the yeast proteome. The greatest obstacle so far has been to produce a high number of pure ►recombinant proteins in parallel. This step initially requires the ►cloning of ►cDNAs into appropriate ►expression vectors. In addition to a simple cloning strategy, protein function arrays covering the full genome require access to normalized ►cDNA libraries.

Protein-Protein Interactions

In functional genomics, protein microarrays are employed to screen a defined collection of proteins for interactions with a certain protein employed as a probe. This technique provides a controlled *in vitro* basis to study protein function on a system-wide or genome-wide basis. In the latter case, the protein microarray contains an ordered library of recombinant proteins representing the complete proteome of an organism. However, studying a limited set in a low-throughput fashion has proven to be successful for the characterization of unknown proteins.

The advantage of this approach is that conditions such as pH, ionic strength, temperature and state of posttranslational modification can be controlled in the microarray experiment. On a genome-wide basis, this approach has been realized for yeast. Pioneering work has been carried out by analyzing the entire proteome of the budding yeast *S. cerevisiae* for kinase substrates of all yeast kinases (4). Microscope slides were spotted with 5,000 different proteins cloned from yeast and expressed with an epitope tag for detection. The microarray was incubated with fluorescently labeled calmodulin to probe for calmodulin binding proteins. Based on the sequence analysis of all calmodulin-binding proteins, a novel binding site for calmodulin was identified that was not detected by a yeast-two-hybrid screen. The outcome puts this microarray-based approach next to established technologies employed to probe for protein-protein interactions. The analogous protein microarray representing the human proteome has not been produced. The complex ►exon-intron structure of the human genome makes the task of converting this information into functional recombinant proteins more challenging. In addition, the human genome is 6–7-fold larger than the yeast genome. However, 30,000 genes of the human genome are represented on large filter sets and bacterial protein expression can be induced on these filters (5).

Protein-Small Molecule Interactions

A proteome array can be screened with drugs to identify proteins that interact with the investigated

compound. This was first achieved with the yeast protein microarray and demonstrated the usefulness of protein microarrays in probing for interactions between proteins and small molecules. Interaction of a specific binding protein with several small molecules with dissociation constants in the micro-molar to nano-molar range was demonstrated. The same array was also employed to probe for lipid binding proteins. In a different experiment, binding of radioactive GTP to microarray-immobilized GTP-binding proteins was performed. A dose response analysis for one of the proteins on the array revealed that results were close to published data (6).

Immune Profiling and Antibody-Specificity Profiling

Protein microarrays spotted with well-known autoantigens can be used to screen serum for the presence of autoantibodies. For certain types of ►autoimmune disease, with well-known autoantigens, the use of microarray technology can increase throughput and reduce sample size. Novel autoantigens can be identified by employing proteome-wide protein microarrays. Recent reports have identified tumor specific autoantibodies in cancer patients. This may lead the way to the identification of new molecular markers useful for the early detection of certain tumors.

Clinical Relevance

Protein microarray technology has an impact on clinical diagnostics. The analysis of cytokines has been multiplexed and commercial kits are available for parallel detection of cytokines. Compared to standard ELISA technology, the sample consumption during analysis is insignificant and detection therefore volume-independent.

In addition, protein microarrays will soon influence the diagnostics of autoantibodies, screening for autoantigens, identification of tumor-specific antibodies and the characterization of antibody specificities in general.

►Chip Technologies, Basic Principles

References

1. Kodadek T (2001) Protein microarrays: prospects and problems. *Chem Biol* 8:105–115
2. Schweitzer B, Wiltshire S, Lambert J et al (2000) Inaugural article: immunoassays with rolling circle DNA amplification: a versatile platform for ultrasensitive antigen detection. *Proc Nat Acad Sci USA* 97:10113–10119
3. Ekins R, Chu F, Biggart E (1990) Multispot, multianalyte, immunoassay. *Ann Biol Clin* 48:655–666
4. Zhu H, Klemic JF, Chang S et al (2000) Analysis of yeast protein kinases using protein chips. *Nat Genet* 26:283–289
5. Bussow K, Nordhoff E, Lubbert C et al (2000) A human cDNA library for high-throughput protein expression screening. *Genomics* 65:1–8
6. Zhu H, Bilgin M, Bangham R et al (2001) Global analysis of protein activities using proteome chips. *Science* 293:2101–2105

Protein Misfolding Disorders

► Defective Protein Folding Disorders

Protein Phosphatase-2A

Definition

Protein phosphatase-2A (PP2A) is a heterotrimeric serine-threonine phosphatase that is involved in Wnt signaling and many other signaling pathways. It catalyses the cleavage of phosphate groups, e.g. the conversion of fructose-1, 6-biphosphate in fructose 6-phosphate at release of inorganic phosphate.

► Wnt/Beta-Catenin Signaling Pathway

Protein Prenylation

MARK D. DISTEFANO, LEILA N. ALBERS,
JU-HUA XU
Department of Chemistry, University of Minnesota,
MN, USA
distefan@chem.umn.edu

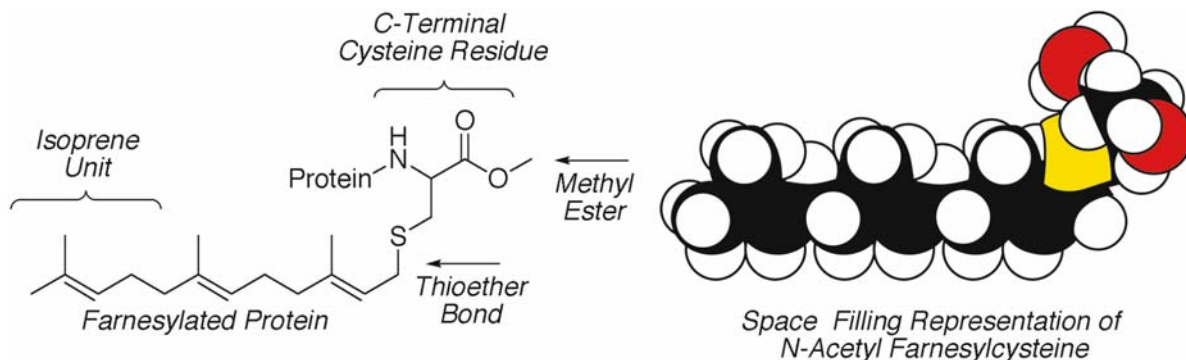
Definition

Protein prenylation is a post-translational modification involving the attachment of a 15 or 20 carbon

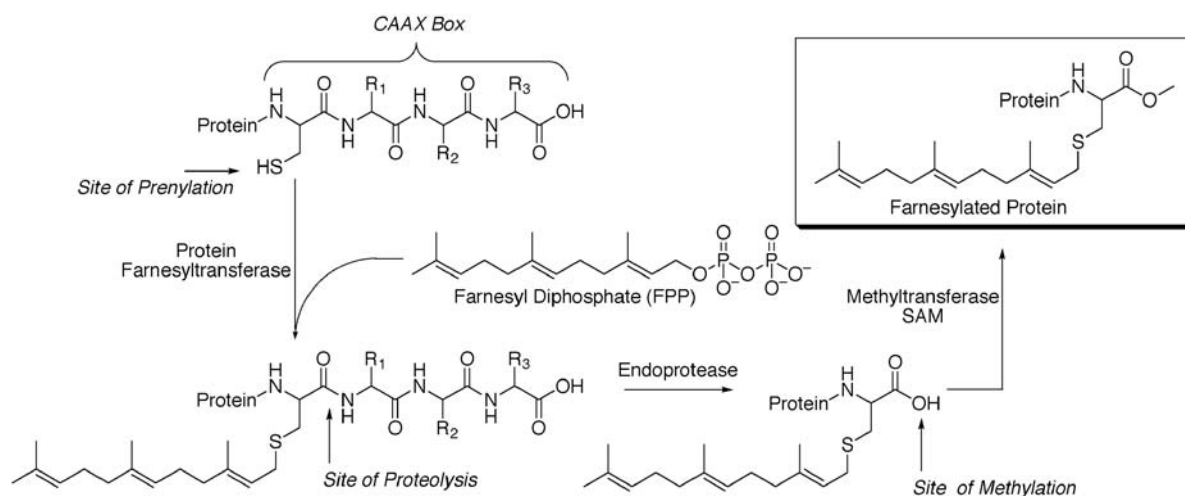
►isoprenoid group to a protein. In each case the ►prenyl group is attached to a specific cysteine amino acid residue near the C-terminus of the protein *via* a ►thioether linkage. The ►farnesyl (C15) group consists of three repeating isoprene subunits, while the ►geranylgeranyl group (C20) consists of four of such subunits. Two chemical representations of a farnesyl group attached to the C-terminal cysteine of a protein are shown below in Fig. 1. Protein prenylation is a three-step process that involves attachment of the isoprenoid followed by proteolysis and ►methyl esterification of the C-terminal prenylated cysteine. In most cases, this modification is essential for the proper function of the targeted protein. Protein prenylation occurs only in eucaryotes and is of particular interest because it is found in proteins involved in signal transduction pathways that regulate critical cellular functions including cell growth and proliferation. The enzyme ►Ras is farnesylated and is an example of such a protein. Since specific mutations in *ras* genes result in oncogenic variants that lead to uncontrolled cell proliferation, molecules that interfere with the prenylation and subsequent processing of Ras can stop the growth of cells that contain transforming *ras* mutants. Such compounds may have utility as anticancer drugs and there are currently a number of candidates in clinical trials for this purpose.

Characteristics

Protein prenylation was first discovered in 1979 by Kamiya in studies of fungal mating factors; these are short polypeptides involved in signaling between different cells. Although originally considered to be a biochemical oddity, it is now clear that protein prenylation is widespread in eucaryotes and is of critical importance for a variety of proteins involved in oncogenesis, secretion, nuclear structure and signal transduction (1). A large number of prenylated proteins



Protein Prenylation. Figure 1 Examples of structural representations prenyl groups. Right: conventional line-angle representation of a farnesylated protein. Left: space filling representation of N-acetyl farnesyl cysteine. Colors: black (carbon), white (hydrogen), red (oxygen), blue (nitrogen) and yellow (sulfur).



Protein Prenylation. Figure 2 Pathway for the process of protein prenylation exemplified with farnesylation.

have been identified and the number is increasing rapidly. Farnesylated proteins include Ras proteins, lamins, transducin, mating factors, rhodopsin kinase and cGMP phosphodiesterase- α . Geranylgeranylated proteins include numerous members of the Ras superfamily of small **G-proteins** and many others; this latter modification is the more common of the two. There are also examples of proteins that can be modified with more than one type of prenyl group. It has been estimated by growth of mammalian cells on ^3H -mevalonic acid, the precursor for all isoprenoids, that as many as 5% of all proteins are isoprenylated. This prevalence, coupled with the central role that many of these modified proteins play in cellular signaling, underscores the significance of this post-translational modification.

Enzymology of Prenylation Process

Prenylation is the first reaction in a three-step maturation process that involves initial attachment of the prenyl group to a specific cysteine residue present near the C-terminus of a protein *via* the formation of a thioether bond. This reaction is catalyzed by a class of enzymes called protein prenyl transferases that use either farnesyl diphosphate or geranylgeranyl diphosphate as the source of the prenyl group. There is at least one protein farnesyl transferase and two different protein geranylgeranyl transferases located in the cytosol of all eucaryotic cells. Following this event, the resulting prenylated proteins manifest increased hydrophobicity that targets them to the cell membrane, where the residues C-terminal to the prenylated cysteine are removed by an endoprotease to yield a polypeptide possessing a prenylcysteine at its C-terminus. The final step in the prenylation process involves the methylation of the **anionic C-terminal**

carboxylate to generate a neutral methyl ester. That reaction is catalyzed by a methyltransferase that utilizes S-adenosyl methionine (SAM) as the methyl group donor. The resulting mature prenylated proteins have a high affinity for the cell membrane due to the presence of the nonpolar prenylcysteine methyl ester positioned at their C-terminus; a summary of the overall prenylation process is shown in Fig. 2. In contrast to other lipid modifications of proteins such as palmitoylation, prenylation is not reversible. Instead, it appears that prenylated proteins are degraded by proteases down to prenylated cysteines, which are then oxidatively deprenylated to free cysteine and the corresponding prenyl aldehydes.

Protein Prenyl Transferases

Of the enzymes that participate in the process of protein prenylation, the protein prenyl transferases are the best understood. Protein farnesyl transferase (PFTase) and protein geranylgeranyl transferase (PGGTase) types I and II have been cloned from a number of species ranging from yeast to mammals. In an effort to understand how certain proteins are targeted for prenylation, considerable effort has been devoted towards determining the primary sequence elements that dictate prenylation. Several classes of signals for prenylation have been described. The most general motif is the **CAAX box** which represents the carboxy-terminal four residues of the prenylated protein. Proteins terminating in CAAX may be either farnesylated or geranylgeranylated depending on the identity of the terminal X residue. In addition to the simple CAAX motif, other classes of geranylgeranylation signals exist that are considerably more complex in their behavior. PFTase and PGGTase I that recognize the CAAX motif are composed of two subunits, α and

β . These heterodimeric proteins share a common α subunit but differ in their β subunits. The type II (prenylates a non-CAAX sequence) PGGTase has been shown to be a heterotrimer composed of a similar $\alpha\beta$ core together with a third type of subunit termed an escort protein. These enzymes are zinc containing metalloenzymes that use coordination of the substrate-derived thiol from cysteine with Zn(II) to form a more [▶nucleophilic](#) Zn(II)-thiolate complex. Steady state kinetic analyses of the mammalian and yeast prenyl transferases clearly indicate that these enzymes catalyze prenyl group transfer *via* sequential pathways in which the isoprenoid binds first followed by protein substrate binding. [▶Peptides](#) as short as four residues that contain a CAAX box can substitute for the peptide substrate; this has greatly facilitated studies with these enzymes. A number of x-ray crystal structures of PFTase alone and in complexes with isoprenoids and peptide substrates have been solved. These structures have proved to be particularly useful for understanding the substrate specificity of this enzyme. A representation of the overall PFTase structure and a close up view of the active site are shown in Fig. 3 (2).

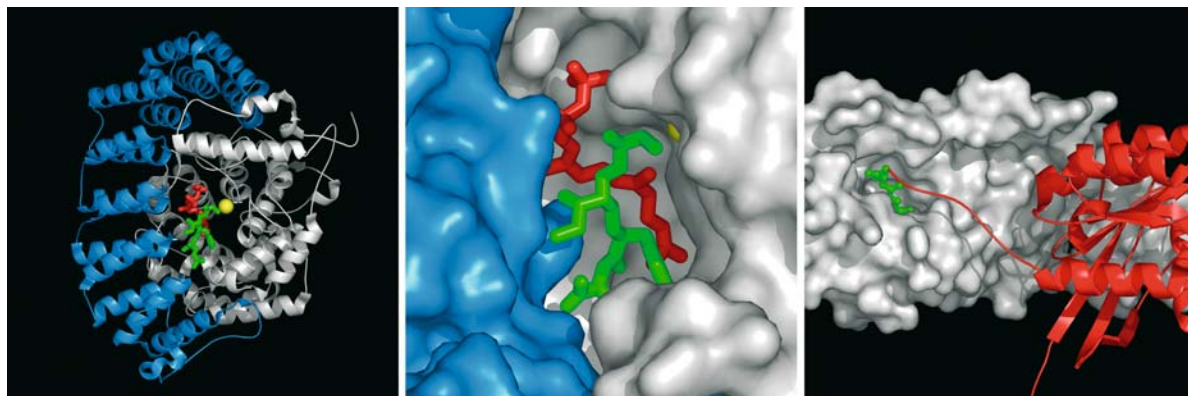
Medicinal Chemistry of Protein Prenylation

A major motivation for research on protein prenylation came from the observation that Ras proteins are farnesylated and that inhibition of Ras prenylation reverses the cancer promoting effects of oncogenic ras mutants. That discovery suggested that the development of PFTase inhibitors ([▶FTIs](#)) might lead to the creation of a new class of highly specific anticancer drugs distinct from the more common but less specific DNA damaging agents (3). In response to this, chemists have synthesized a diverse array of molecules that

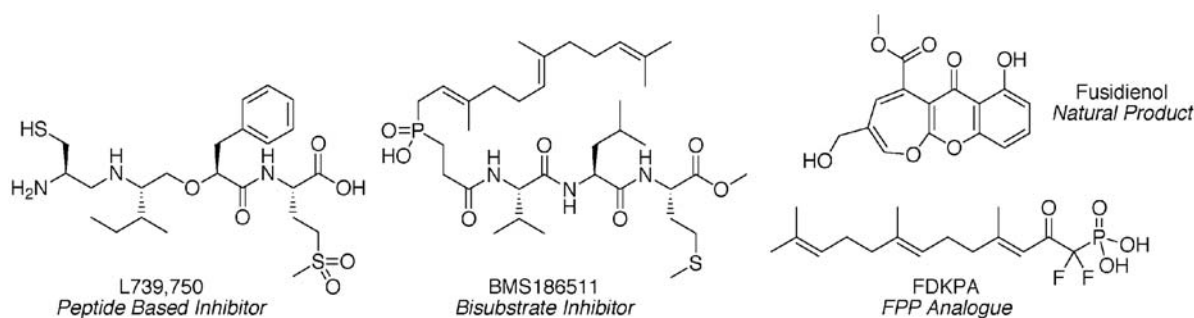
inhibit PFTase (4, 5). These compounds can be broken down into several classes including [▶peptidomimetics](#) that incorporate features of the CAAX box substrate described above, isoprenoid mimics that include elements of the farnesyl group and bisubstrate inhibitors that combine features of both. Natural products that inhibit PFTase activity have also been isolated. Representative structures for each of these classes are given in Fig. 4. To date, the peptidomimetic class has shown the greatest promise and has been investigated the most. Interestingly, the molecules in this class have been developed using a variety of approaches. Some have been developed from CAAX box peptides while others have been created by grafting key structural features from CAAX box peptides onto rigid molecular scaffolds. High throughput screening of large libraries of existing drug collections has also yielded useful leads. Using similar approaches, inhibitors of PGGTase have also been developed. In most cases the above inhibitors were initially evaluated using *in vitro* enzyme assays, although many have been examined in cell based systems and a few have progressed to animal testing and human clinical trials. A limited amount of work has also been done on designing inhibitors that target the enzymes involved in the proteolysis and methylation of prenylated proteins, although the utility of such compounds for medical purposes is currently unclear.

Function of Protein Prenylation

There are currently two distinct roles for protein prenylation in biological systems. One of these is to increase the hydrophobicity of proteins and cause them to associate with the cell membrane by insertion of the isoprenoid into the lipid bilayer. Such targeting allows



Protein Prenylation. Figure 3 Some structural features of proteins involved in the process and function of prenylation. Left: structure of PFTase showing the alpha subunit (blue), beta subunit (white), FPP (red), clinical candidate drug L-739,750 and Zn(II) (yellow). Center: close up view of the active site of PFTase showing the bound drug (same color scheme as Left). Right: complex between RhoGDI (white) and CDC42 (red) showing the C-terminal geranylgeranyl group (green) from CDC42 occupying an extensive binding pocket present in RhoGDI.



Protein Prenylation. Figure 4 Representative examples of different types of farnesyl transferase inhibitors (FTIs).

them to interact with other proteins involved in signal transduction pathways. For Ras proteins, this appears to be the function of prenylation. Under normal conditions farnesylated Ras associates with the membrane; however, in cells treated with PFTase inhibitors, the (unfarnesylated) Ras localizes to the cytosol and is thus incapable of interacting with its requisite signaling partner proteins. A second role for prenylation is based on the idea that there are specific prenyl group receptors or binding sites on other proteins. It has been shown that prenylation is required for protein-protein interactions in a number of cases. One of the clearest examples of this type of interaction occurs in the complex between Rho-GDI and CDC42 whose structure has been solved by x-ray crystallography. CDC42 is a geranylgeranylated protein involved in the regulation of cytoskeletal assembly while Rho-GDI is a modulator of CDC42 activity (6). The structure of the complex of these two proteins (Fig. 3, right) clearly shows the geranylgeranyl group from CDC42 enveloped within a lipid-binding cavity present in Rho-GDI. Thus, at least in some cases, it appears that prenyl groups play an important role in stabilizing protein-protein interactions, although how commonly this occurs is presently not clear.

Clinical Relevance

Prenylation and Tumor Biology

As noted above, the initial motivation for the development of FTIs stemmed from the observation that inhibition of farnesylation of Ras oncoproteins stops the growth of certain tumor cells. Since Ras proteins participate in a number of signaling pathways, it is not surprising that cancer cells expressing oncogenic Ras exhibit a broad range of phenotypic and biochemical effects when treated with FTIs; while it is clear that treatment of FTIs prevents the farnesylation of Ras, the mechanism by which this event causes growth arrest or cell death is unclear. FTIs have been implicated as inhibitors of angiogenesis and the cell-cycle but the precise molecular mechanisms underlying those events

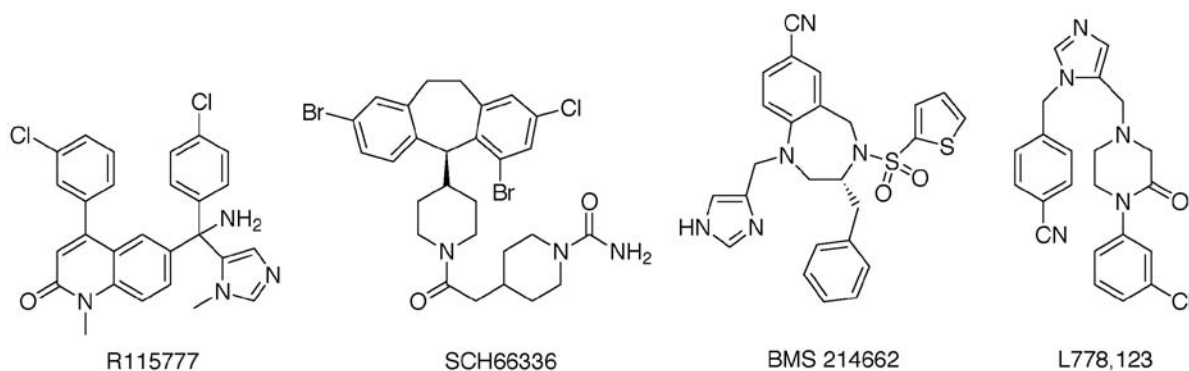
remain controversial. A significant complication in studies of FTI action is that these compounds also act on tumor cell lines lacking Ras mutations. While some of these cell lines are characterized by oncogenic mutations in proteins that act upstream from Ras and can signal through normal Ras, others cause transformation *via* Ras independent pathways. Given that there are more than 20 farnesylated proteins in a mammalian cell, this raises the question of whether FTIs act, at least in part, by inhibiting the prenylation of other key proteins; efforts to identify such proteins are in progress.

Clinical Trials of FTIs for Cancer Treatment

Currently, a number of FTIs are being examined in human clinical trials (7, 8). The structures of four of such drug candidates are shown in Fig. 5. These compounds have been used to treat a broad range of advanced human cancers in Phase 2 and Phase 3 trials. Unfortunately, the results of these experiments have not been promising. While the drugs have been generally well tolerated with little discernable toxicity, no significant positive results have been observed when these drugs have been administered alone. However, the fact that FTIs act by a mechanism (inhibition of protein prenylation) that is distinct from other types of chemotherapeutic drugs suggests that these new molecules could have significant application in combination therapy approaches. Accordingly, a large number of trials involving FTIs administered in combination with drugs such as 5-fluorouracil, gemcitabine, cisplatin, paclitaxel and several others have been performed. FTIs have also been used in concert with radiation therapy. The results of these combination treatments have been more positive, including several cases of partial or complete remission or disease stabilization.

Future Directions for Prenylation Research

Given that clinical trials are still in their early stages and that much needs to be learned about the timing and administration of FTIs in combination modalities, it is



Protein Research Foundation. Figure 5 Structures of FTIs in human clinical trials.

too soon to know whether FTIs will be useful agents for cancer therapy. However, it is noteworthy that the discovery and subsequent biochemical studies of protein prenylation have led to new ideas about how cancer and other human diseases might be treated. The ubiquitous occurrence of protein prenylation in fungi, protozoans, mammals and viruses coupled with the complex biochemistry of prenylation that involves a number of enzymes and prenyl group receptor proteins suggests that there are a plethora of possible targets for drug development. Research on the chemistry and biology of protein prenylation is an exciting area and will remain so for many years.

References

1. Zhang FL, Casey PJ (1996) Protein prenylation: molecular mechanisms and functional consequences. *Ann Rev Biochem* 65:241–269
2. Long SB, Hancock PJ, Kral AM et al (2001) The crystal structure of human protein farnesyltransferase reveals the basis for inhibition by CaaX tetrapeptides and their mimetics. *Proc Natl Acad Sci USA* 98:12948–12953
3. Leonard DM (1997) Ras farnesyltransferase: a new therapeutic target. *J Med Chem* 40:2971–2990
4. Qian Y, Sebt SM, Hamilton AD (1997) Farnesyltransferase as a target for anticancer drug design. *Biopol* 43:25–41
5. Gibbs RA, Zahn TJ, Sebolt-Leopold JS (2001) Non-peptidic prenyltransferase inhibitors: diverse structural classes and surprising anti-cancer mechanisms. *Curr Med Chem* 8:1437–1465
6. Hoffman GR, Nassar N, Cerrione RA (2000) Structure of the Rho family GTP-binding protein Cdc42 in complex with the multifunctional regulator RhoGDI. *Cell* 100:345–356
7. Karp JE, Kaufmann SH, Adjei AA et al (2001) Current status of clinical trials of farnesyltransferase inhibitors. *Curr Opin Oncology* 13:470–476
8. Brunner TB, Hahn SM, Gupta AK et al (2003) Farnesyltransferase inhibitors: an overview of the results of preclinical and clinical investigations. *Cancer Res* 63:5656–5668

Protein Research Foundation

Definition

The Peptide Institute, Protein Research Foundation, is collecting information related to amino acids, peptides and proteins, and making databases. (► <http://www.prf.or.jp/en/>)

► [Protein Databases](#)

Protein Self-Association

Definition

Protein self-association refers to the self-assembly of proteins to larger units (typically non-covalent), and defines the reversible formation of multimeric units of protein through protein interactions, generating monomeric and oligomeric states in an association-dissociation equilibrium.

► [Analytical Ultracentrifugation](#)

Protein Species

Definition

A protein species is defined by its chemical structure. Each modification, such as methylation, removal or addition of an amino acid, leads to a new protein species.

► [Two-Dimensional Gel Electrophoresis](#)

Protein Structure Prediction

VOLKHARD HELMS

Center for Bioinformatics, Saarland University,
Saarbrücken, Germany
volkhard.helms@bioinformatik.uni-saarland.de

Definition

The expression *ab initio* protein structure prediction is used for theoretical or computational approaches that aim at predicting the three-dimensional structure of a folded protein from its amino acid sequence alone. Although significant progress has been made in this area over the past fifteen years, this task is still considered a grand challenge project. *Ab initio* structure prediction actually involves two difficult parts: (a) sampling the large accessible conformational space of a polypeptide chain to identify putative low energy conformations, and then (b) picking the most favourable structure among those in terms of its **free energy**.

Due to the availability of more and more experimentally determined protein structures deposited in the protein data base (**www.rcsb.org**), alternative knowledge based methods such as **homology modelling** or threading are nowadays considered more reliable than *ab initio* methods, as soon as sequence similarity to proteins with known structure can be exploited. Also, these methods provide results much faster than *ab initio* methods.

Nonetheless, *ab initio* structure prediction has its own value because (i) certain protein folds may forever escape experimental structure determination, (ii) because it is now being estimated that up to twenty percent of all proteins may be fully or partially unstructured and (iii) because some of the *ab initio* methods also allow the investigation of the folding pathways of proteins.

Description

According to Anfinsen all of the information necessary for a protein to fold to the native state resides in the protein's amino acid sequence. In the absence of large kinetic barriers in the free energy landscape, this suggests that the native conformations of most proteins are the lowest free energy conformations for their sequences.

According to the “new view of protein folding” (1) that became popular during the late 1990's folding does not proceed as a naïve random search of the polypeptide chain on a “golf-course” model of the energy surface.

Instead, the free energy surfaces are in general funnel-shaped with the native, folded state at the bottom. The **protein folding** process can thus be thought of as a diffusional motion on such surfaces. The important implication for structure prediction is that this characteristic property allows one to concentrate the sampling efforts on favourable regions of the conformational space.

Successful structure prediction requires a free energy function sufficiently close to the true potential for the native state to be at one of the lowest free energy minima, as well as a method for searching conformational space for low energy minima. *Ab initio* structure prediction is challenging because current potential functions have limited accuracy and the conformational space to be searched is vast. Therefore, many methods use reduced representations, simplified potentials and coarse search strategies in recognition of this resolution limit.

Brute-Force Sampling Without Bias: Molecular Dynamics

Molecular dynamics simulations solve Newton's equation of motion to follow the atomic motions in a molecular system. They are the method of choice to describe the equilibrium dynamics of folded proteins. When combined with an efficient continuum description of the surrounding solvent, they are now able to generate microsecond long simulations within a few days of computing time. This is sufficient to study the folding of isolated secondary structure elements, of small peptides and of fast folding single domain proteins.

Alternatively, the group of Vijay Pande (Stanford) has attracted nearly one hundred thousand volunteers who downloaded their simulation software and computed short folding trajectories on their home PCs (2). This model of dividing a grand task into many small tasks is perfectly adapted to the paradigm of distributed computing. Assembling the individual results into a big picture then allows characterisation of the folding landscape of fast-folding proteins up to 60 residues in length. These calculations belong to the world's largest computational efforts to date.

Another promising approach to increase the folding efficiency is based on multicanonical multi-replica simulations where many simulations of the same system are run in parallel at different temperatures. At regular intervals, the systems may exchange particle velocities, thus enhancing the likelihood of the systems overcoming energy barriers.

Smart Sampling: Use Information About Preferred Conformations

As mentioned above, it is usually not necessary to exhaustively search the huge conformational space of a

polypeptide chain. “Rosetta” (3) is one of the most successful modern methods for structure prediction. It uses a precomputed set of preferred conformations of nine residue long sequence stretches extracted from the ►protein database. An optimisation algorithm then randomly assigns these conformations to short segments of the target sequence and scores their suitability by an energy function. The main advantage of methods like this one over simpler Monte Carlo schemes is that relatively large structural changes can be accomplished by a few optimisation trials and again only favourable regions of the conformational space need to be explored.

Smart Sampling: Concentrate on Important Degrees of Freedom

Another way of reducing the computational complexity is to restrict the conformational search to the torsional degrees of freedom. In doing so, it becomes advantageous to use techniques such as internal coordinate Monte-Carlo (4) that provide the advantage of a seven-fold reduction of degrees of freedom over conventional optimisation in Cartesian coordinates. Successful applications include optimisation of protein loops or of terminal segments of proteins that remained unresolved in experimental structure determinations.

Structure Prediction of Membrane Proteins

Membrane proteins make up about 30% of the human genes and are drug targets for about half of the known drugs. Despite their great significance, only about thirty structures of membrane proteins have been determined so far, compared to 25,000 structures of soluble proteins. The reasons for this are of a technical nature; membrane proteins don’t like to crystallize in three dimensions etc. Therefore, applying molecular modelling to generating structural models of membranes is a highly rewarding area.

In the hydrophobic bilayer environment, the backbone atoms of the polypeptide chain have no other alternative for satisfying their hydrogen bonding capacities than folding into alpha-helices (or beta-sheets for proteins in the outer membranes of bacteria). Therefore, structure prediction of the transmembrane parts can be considered as a task of simply assembling a set of seven or more alpha helices in more or less straight conformations.

Recent crystal structures, e.g. of the CIC channel, revealed that this simple picture is not generally true, but it may still apply to a number of classes of membrane proteins like ►G-protein coupled receptors.

Structural Refinement of Coarse Models

After generating an atomic model, ►molecular mechanics methods (energy minimisation and ►molecular

dynamics simulation) can be used to energetically relax strained regions of the modelled structure. This is often done in a hierarchical manner, e.g. first relaxing the side chains while keeping the modelled backbone fixed and then relaxing everything. Care should be taken about using a suitable environment (aqueous solvent or membrane environment).

Clinical Applications

Can, in this (post-)genomic era, *ab initio* protein structure predictions play a practical role in the drug discovery process? Yes. Although it is not always possible to deduce a protein’s function from its three-dimensional structure, low to moderate resolution structures may be suitable for predicting the biochemical/enzymatic function of a protein by matching to a known active site template extracted from some other previously solved structure. Using such template matching, one can suggest in many cases what the protein does or suggest a small number of experiments to establish its function.

Also, if one has a known inhibitor, one can use these predicted low-resolution structures to predict where it binds (5). Even if it turns out that one cannot predict the best lead molecule – if one could guarantee that such a lead is in the top 100 or 1,000 compounds, this could accelerate the drug discovery process and reduce its cost.

The clear direction for the future is to elucidate the role that proteins play in cellular pathways in order to pick and choose those that are likely to be “druggable.” Preliminary indications are that structure prediction can assist in the automated assignment of proteins to known pathways – a first step in this process.

►Protein Folding

References

1. Dobson CM, Šali A, Karplus M (1998) Protein Folding: A Perspective from Theory and Experiment. *Angew Chem Int Ed* 37:868–893
2. Pande VS, Baker I, Chapman J et al (2003) Atomistic protein folding simulations on the submillisecond time scale using worldwide distributed computing. *Biopolymers* 68:91–109
3. Simons KT, Kooperberg C, Huang E et al (1997) Assembly of protein tertiary structures from fragments with similar local sequences using simulated annealing and Bayesian scoring functions. *J Mol Biol* 268:209–225
4. Abagyan RA, Totrov MM, Kuznetsov DA (1994) ICM: a new method for protein modeling and design: Applications to docking and structure prediction from the distorted native conformation. *J Comp Chem* 15:488–506
5. Vaidehi N, Floriano WB, Trabanino R et al (2002) Prediction of structure and function of G protein-coupled receptors. *Proc Natl Acad Sci USA* 99:12622–12627

Protein Synthesis Control

►Cap-Independent Translational Control

Protein Tags

KAY TERPE

Proteins and Nucleic Acids, IBA GmbH, Göttingen, Germany
terpe@iba-go.com

Definition

A polypeptide or protein fused to a protein of interest allowing a single step purification of the protein of interest on a specific matrix is termed a protein or ►**affinity tag**. Protein tags are mainly used in the field of protein purification, detection and localization and are therefore an important tool for protein binding studies and for protein complex analysis. Most vectors and ►**expression systems** designed to produce the tagged fusion protein enable recombinant protein production in prokaryotic or eukaryotic cell systems for cost efficient ►**large scale protein production**.

Characteristics

Affinity or protein tags share the following features: (a) one-step adsorption purification, (b) a minimal effect on tertiary structure and biological activity of the protein of interest, (c) easy and specific removal to retrieve the native protein, (d) simple and accurate assay of the recombinant protein during purification and (e) applicability to a number of different proteins. The most frequently used systems are: Arg-Tag, calmodulin-binding peptide, cellulose-binding domain, DsbA, c-myc-tag, glutathione S-transferase, FLAG-tag, HAT-tag, His-tag, maltose-binding protein, NusA, S-tag, SBP-tag, Strep-tag and thioredoxin (1). Other not so frequently used tags are Protein A, Protein G, Z domain, Avi-tag, PinPoint-tag, Bio-tag, T7-peptide, V5-peptide, hemagglutinin A and Glu-Glu. In general, small size protein tags have not so much influence on the ►**fusion protein** as larger tags. Small size protein tags consist of a maximum of 12 amino acids, medium size tags of a maximum of 60 amino acids and large tags of more than 60 amino acids.

Small Size Tags

Polypeptide tags with a maximum of 12 amino acids can be defined as small size tags.

The Arg-tag was first described in 1984 and usually consists of five or six arginines. Arginine is the most basic amino acid (1). Arg 5-tagged proteins can be purified on the cation exchange resin SP-Sephadex, to which most of the contaminating proteins do not bind. After binding, the tagged proteins are eluted with a linear NaCl gradient at alkaline pH. Polyarginine might affect the tertiary structure of proteins whose C-terminal region is hydrophobic. The C-terminal series of arginine residues can be removed by carboxypeptidase B treatment. This enzymatic process has been successfully used in several instances, but has often been limited by poor cleavage yields or by unwanted cleavage occurring within the desired protein sequence. The Arg-tag can be used to immobilize functional proteins on flat surfaces; this is important for studying interactions with ligands. While the Arg-tag is not used very often, in combination with a second tag it can be an interesting tool for protein purification.

The His-tag was first described in 1987 (1, 3) and usually consists of from five up to ten histidines. Histidine is the amino acid that exhibits the strongest interaction with immobilized metal ion matrices, as electron donor groups of the histidine imidazole ring readily form coordination bonds with the immobilized transition metal. ►**Immobilized metal-affinity chromatography (IMAC)** (1, 2) is based on the interaction between a transition metal ion (Co^{2+} , Ni^{2+} , Cu^{2+} , Zn^{2+}) immobilized on a ►**matrix** and specific amino acid side chains. Hochuli has developed a nitrilotriacetic acid (NTA) adsorbent for metal-chelate affinity chromatography. Other materials like iminodiacetic acid (IDA), carboxymethylaspartate (CMA) and a specific silica matrix (SiMAC) are also applicable. Peptides containing sequences of consecutive histidine residues are efficiently retained on IMAC. Following washing of the matrix material, peptides containing polyhistidine sequences can be easily eluted by either adjusting the pH of the column buffer or by adding free imidazole (Table 1). The His-tag is also a useful tool for purifying recombinant proteins under denaturing conditions with urea or guanidinium HCl.

The Strep-tag was first described in 1993 and consists of eight amino acids (for the sequence, see Table 2). The core of the Strep-tag is a histidine-proline-glutamine sequence. The tag binds to the biotin-binding pocket of streptavidin. A streptavidin mutant with a specific mutation at positions 44, 45 and 47 has a higher affinity for the Strep-tag than the native form. The streptavidin variant is called Strep-Tactin. In physiological buffer conditions, Strep-tagged proteins are bound to the biotin-binding pocket and can be eluted gently with biotin derivatives (1, 4). Elution is also possible by transiently reducing the pH. Elution with 2.5 mM desthiobiotin is recommended. The matrix can be regenerated with 4-hydroxy azobenzene-2-carboxylic

Protein Tags. Table 1 Matrices and elution conditions of affinity tags

Affinity Tag	Matrix	Elution condition
Arg-Tag	Cation-exchange resin	NaCl linear gradient from 0 to 400 mM at alkaline pH > 8.0
His-Tag	Ni ²⁺ -NTA, Co ²⁺ -CMA (Talon)	Imidazole 20 to 250 mM or low pH
FLAG-Tag	Anti-FLAG monoclonal antibody	pH 3.0 or 2-5 mM EDTA
Strep-Tag II	Strep-Tactin (modified Streptavidin)	2.5 mM desthiobiotin
c-myc-Tag	Monoclonal antibody	Low pH
S-Tag	S-fragment of RNaseA	3 M guanidine thiocyanate, 0.2 M citrate pH 2, 3 M magnesium chloride
HAT-Tag (natural histidine affinity Tag)	Co ²⁺ -CMA (Talon)	150 mM imidazole or low pH
Calmodulin-binding peptide	Calmodulin	EGTA or EGTA with 1 M NaCl
Cellulose-binding domain	Cellulose	Family I: Guanidine HCl or urea > 4 M
SBP-Tag	Streptavidin	Family II/III: Ethylene glycol 2 mM biotin
Chitin-binding domain	Chitin	Fused with Intein: 30-50 mM dithiothreitol, β -mercaptoethanol or cysteine
Glutathione S-transferase	Glutathione	5-10 mM reduced glutathione
Maltose-binding protein	Cross-linked amylose	10 mM maltose

acid, which is yellow in solution and red when bound on the matrix. The binding conditions are very specific and buffer conditions are variable. Purifications under anaerobic conditions, with chelating agents, with low or high salt concentration, with non-ionic, ionic and zwitter-ionic detergents are possible. The Strep-tag is a useful tool for [▶protein-protein interaction](#), purifying membrane proteins and [▶protein chips](#).

The FLAG-tag was first described in 1988 and consists of a short, hydrophilic 8-amino-acid peptide (for the sequence, see Table 2). The FLAG peptide binds to the antibody M1. Whether binding is calcium-dependent or -independent remains controversial. Additional targets are the monoclonal antibodies M2 and M5, each with different recognition and binding characteristics. The system is non-denaturing and thus allows active fusion proteins to be purified. The complex can be dissociated by chelating agents such as EDTA or by transiently reducing the pH (Table 1). The FLAG-tag can be removed by treatment with enterokinase, which is specific for the five C-terminal amino acids of the peptide sequence. A disadvantage of the system is that the monoclonal antibody purification matrix is not as stable as others, e.g. Ni-NTA or Strep-Tactin.

The murine anti-c-myc antibody 9E10 was developed in 1985 (5) and is used as an immunochemical reagent in cell biology and in protein engineering. The antibody epitope of eleven amino acids (for the sequence, see Table 2) can be expressed in a different protein context and still confer recognition by the 9E10 immunoglobulin. The c-myc-tag has been successfully used in Western blot technology, immunoprecipitation and flow cytometry. It is therefore useful for monitoring expression of recombinant proteins. The successful co-immunopurification of interacting proteins expressed in *Agrobacterium*-transformed *Arabidopsis* cells was also reported. C-myc-tagged proteins can be affinity-purified by coupling Mab 9E10 to divinyl sulphone-activated agarose. The washing conditions are physiological followed by elution at low pH, which could exert a negative effect on protein activity. It is a widely used detection system but is rarely applied for purification.

Medium Size Tags

In general, polypeptide and protein domain tags with a minimum of 13 and a maximum of 60 amino acids can be defined as medium size tags.

Protein Tags. Table 2 Sequence and size of affinity tags

Tag	Residues	Sequence	Size (kDa)
Polyarg-Tag	5-6 (most 5)	RRRRR	0.80
Polyhis-Tag	2-10 (most 6)	HHHHHH	0.84
FLAG-Tag	8	DYKDDDDK	1.01
Strep-Tag II	8	WSHPQFEK	1.06
c-myc-Tag	11	EQKLISEEDL	1.20
S-Tag	15	KETAAAKFERQHMS	1.75
HAT-Tag	19	KDHLIHNVHKEFHAAHNK	2.31
3xFLAG-Tag	22	DYKDHDGDYKDHDIDYKDDDDK	2.73
Calmodulin-binding peptide	26	KRRWKKNFIAVSAANRFKKISSSGAL	2.96
Cellulose-binding domains	27-189	Domains	3.00-20.00
SBP-Tag	38	MDEKTTGWRGGHVVEGLAGELEQLRARLEHHPQGQREP	4.03
Chitin-binding domain	51	TNPGVSAWQVNTAYTAGQLVTYNGKTYKCLQPHTSLAGWEPSNVPALWQLQ	5.59
Glutathione S-transferase	211	Protein	26.00
Maltose-binding protein	396	Protein	40.00

The S-tag was first described in 1990 and consists of 15 amino acids (for the sequence, see Table 2). The sequence of the fusion-peptide allows detection by a rapid, sensitive homogeneous assay or by colorimetric detection in Western blots (1). The system is based on the strong interaction between the 15-amino-acid S-tag (Table 2) and the 103-amino-acid S-protein, both of which are derived from RNaseA. The S-protein/S-tag complex has a k_d of $\sim 0.1 \mu\text{M}$ which depends on pH, temperature and ionic strength. The tag is composed of four cationic, three anionic, three uncharged polar and five non-polar residues. This composition makes the S-tag soluble. The S-tag rapid assay is based on the reconstitution of ribonucleolytic activity. Tagged proteins can be bound on S-protein matrices. The elution conditions are very harsh, e.g. buffer with pH 2 (Table 1); however, cleaving the tag with protease is recommended to get functional proteins.

The HAT-tag was first described in 1999 and consists of 19 amino acids. It is a natural poly-histidine affinity tag (see His-tag) and has been developed to purify fusion

proteins under mild conditions. Purification with Co^{2+} -CMA resulted in higher elution product purity. Adsorption of weakly bound unspecific proteins was eliminated by using 5 mM imidazole in the equilibration and loading buffers and 150 mM imidazole was used to elute the HAT-tagged proteins. Elution of tagged proteins was also possible by decreasing the pH to 5.0. Urea turned out to have a much stronger negative effect on the binding of HAT-tagged proteins than guanidinium HCl.

The 3xFLAG system has been developed to improve the detection of the FLAG-tag, which was first described in 1988. This three-tandem FLAG epitope consists of 22-amino-acids (for the sequence, see Table 2) and can detect up to 10 fmol of expressed fusion protein. The 3x FLAG-tag binds in the same manner as the FLAG peptide to the antibody M1. The system is non-denaturing and thus allows active fusion proteins to be purified. The complex can be dissociated by chelating agents such as EDTA or by transiently reducing the pH (Table 1). The 3x FLAG-tag can be

removed by treatment with enterokinase, which is specific for the five C-terminal amino acids of the peptide sequence.

The calmodulin-binding peptide was first described in 1992. The peptide has 26 residues (for the sequence, see Table 2) derived from the C-terminus of skeletal muscle myosin light-chain kinase, which binds calmodulin with nanomolar affinity in the presence of 0.2 mM CaCl_2 . The tight binding allows more stringent washing conditions, ensuring that few contaminating proteins will be co-purified with the fusion protein. A second elution step with EGTA and 1 M NaCl is useful if the protein does not elute completely in the first step. Reducing agents and detergents in amounts up to 0.1% are compatible with the system.

The SBP-tag was first described in 2001. It is a new streptavidin-binding peptide and has a length of 38 amino acids (for the sequence, see Table 2). The dissociation constant of the tag from streptavidin is 2.5 nM. SBP-tagged proteins can be purified with immobilized streptavidin. The elution conditions are very mild, using 2 mM biotin. Proteins with C-terminal SBP-tagged proteins were expressed in bacteria and successfully purified. Little is known regarding further applications, but the tag seems to be an interesting tool for immobilizing proteins on streptavidin-coated chips. The chitin-binding domain from *Bacillus circulans* was first described in 1996 and consists of 51 amino acids. The affinity tag is commonly available in combination with self-splicing inteins. The [▶intein](#) from the *Saccharomyces cerevisiae* VMA1 gene, which consists of 454 amino acids, is often used. Other, shorter inteins have also been employed (6). Self-cleavage of the thioester bond can be induced by thiol reagents, such as 1,4-dithiothreitol or β -mercaptoethanol (Table 2). The C- or N-terminal amino acid residue of the target protein has an effect on *in vivo* and *in vitro* cleavage (6). A high salt concentration or the use of non-ionic detergents can be employed to reduce non-specific binding, thus increasing purity. The uncleaved fusion precursor and the intein tag remain bound to the chitin resin during target protein elution and can be stripped from the resin by 1% SDS or 6 M guanidine HCl.

More than 13 different families of proteins with cellulose-binding domains (CBDs) were classified and described between 1997 and 2002. CBDs can vary in size from 4 to 20 kD and occur at different positions within polypeptides, N-terminal, C-terminal and internal. Some CBDs bind irreversibly to cellulose and can be used for immobilization of active enzymes; others bind reversibly and are more useful for separation and purification. CBDs of family I bind reversibly to crystalline cellulose and are useful tags for affinity chromatography. Hydrogen bond formation and van der Waals interaction are the main driving forces for binding.

The advantages of cellulose are that it is inert, has low non-specific affinity, is available in many different forms and has been approved for many pharmaceutical and human uses. CBDs bind to cellulose over a moderately wide pH range, from 3.5 to 9.5. The affinity of the tag is so strong that an immobilized fusion protein can only be released with buffers containing urea or guanidine hydrochloride. These denaturing elution conditions make refolding of the fused target protein necessary. Fused proteins with CBDs of families II and III can be eluted gently from cellulose with ethylene glycol. This low-polarity solvent presumably disrupts the hydrophobic interaction at the binding site. Ethylene glycol can be removed easily by dialysis.

Large Size Tags

In general, protein domain and protein tags with a minimum of 61 amino acids can be defined as large size tags.

The glutathione S-transferase tag ([▶GST](#)) was first described in 1988 (7). A 26-kD GST of *Schistosoma japonicum* was cloned in an *E. coli* expression vector. Fusion proteins could be purified from crude lysate by affinity chromatography on immobilized glutathione. Bound fusion proteins can be eluted with 10 mM reduced glutathione under non-denaturing conditions. In the majority of cases, fusion proteins are soluble in aqueous solutions and form dimers. The GST-tag can be easily detected using an enzyme assay or an immunoassay. The tag can help to protect against intracellular protease cleavage and stabilize the recombinant protein. In some cases, GST fusion proteins are totally or partly soluble. It remains unclear which factors are responsible for insolubility, but in several instances insolubility of GST fusion proteins was associated with the presence of hydrophobic regions. Other insoluble fusion proteins either contain many charged residues or are larger than 100 kD. In some cases insoluble fusion proteins can be purified by [▶affinity chromatography](#) if they are solubilized in 1% Triton X-100, 1% Tween, 10 mM dithiothreitol, 0.03% SDS or 1.5% sarcosyl buffer. Sarcosyl inhibits co-aggregation of proteins with bacterial outer membrane components. Purification of other insoluble proteins must be done by conventional methods. A [▶site-specific protease](#) such as thrombin or factor X_a is recommended for cleaving the GST-tag from fusion proteins. The PreScission protease contains the human rhinovirus 3C protease including the GST-tag; the GST carrier and the protease can be removed after proteolysis by affinity chromatography on glutathione-agarose. GST fusion proteins have become a basic tool for the molecular biologist. They are also commonly used in studies on protein-DNA interactions, protein-protein interactions and as antigens for immunology or vaccination studies.

The 40-kD maltose-binding protein (MBP) was first described in 1988 and is encoded by the *malE* gene of *E. coli* K12. Fused proteins can be purified by one-step affinity chromatography on cross-linked amylose. Bound fusion proteins can be eluted with 10 mM maltose in physiological buffer. Binding affinity is in the micro-molar range. Some fusion proteins do not bind efficiently in the presence of 0.2% Triton X-100 or 0.25% Tween 20, while other fusion proteins are unaffected. Buffer conditions can range from pH 7.0 to pH 8.5 with up to 1 M salt. Denaturing agents cannot be used. A spacer sequence coding for ten asparagine residues between the MBP and the protein of interest increases the chances that a particular fusion protein will bind tightly to the amylose resin. The MBP-tag can be easily detected using an immunoassay. It is necessary to cleave the tag with a site-specific protease. N-terminal location can reduce the efficiency of translation. The MBP system is widely used in combination with a small affinity tag.

One disadvantage when heterologous proteins are produced in *E. coli* is that proteins frequently aggregate as insoluble folding intermediates, known as **inclusion bodies**. In order to recover an active protein, these must be solubilized with denaturing agents such as 8 M urea or 6 M guanidine hydrochloride. One possibility for avoiding inclusion bodies is to use large affinity tags such as GST or MBP. Hydrophilic tags, such as transcription termination anti-termination factor (NusA), *E. coli* thioredoxin (TrxA) or protein disulfide isomerase I (DsbA) can increase solubility. A disadvantage is, however, that proteins with these tags cannot be purified with a specific affinity matrix. The fusion construct must be used in combination with a small affinity tag for purification. The NusA protein increases the solubility of fusion proteins. Usually, *E. coli* NusA protein promotes hairpin folding and termination. Some insoluble proteins expressed in *E. coli* remained soluble when tagged N-terminally with NusA. NusA has often been used in combination with the His-tag. Thioredoxin can be fused to the amino or carboxyl terminus of the protein of interest, but typically the *trxA* sequence is placed at the 5' end. DsbA increases the solubility of the target protein in the cytoplasm and periplasm of *E. coli*. A site-specific protease is recommended for the cleavage of fusion proteins with NusA, TrxA or DsbA; the cleavage site can be used as a linker peptide.

Other Tag Systems

There are other tag systems in use, which are not described in detail here. Staphylococcal protein A gene fusion vectors were developed to purify recombinant proteins by IgG affinity chromatography. This protein is well suited for affinity purification due to its specific binding to the Fc part of immunoglobulins of many

species including human. Analogously to protein A, protein G from *Streptococcus* strain G148 and the Z-domain can be used in the same manner because they bind the Fc portion of IgG. Biotinylation of proteins using small peptide tags is commonly used for detection, immobilization and purification. Different tags, such as the Avitag, PinPoint X a protein purification system and Bio-tag have been described. The bacteriophage T7 and V5 epitopes are interesting tags for sensitive detection. Other epitope tags for detection are ECS (enterokinase cleavage site), HA (hemagglutinin A), and Glu-Glu.

Clinical Relevance

Protein tags are an important tool for protein purification. Disease relevant proteins can be isolated and used as drug targets. Some tags allow the rapid purification of specific complexes and are a useful tool for proteome exploration. Many tags with high affinity to their binding partner are also useful tools for immobilizing peptides or proteins on surfaces. Furthermore, the importance of affinity tag technology for use in peptide/protein chip design, high-throughput purification, peptide/protein libraries, large-scale production systems and drug delivery strategies will increase.

References

1. Terpe K (2003) Overview of tag protein fusions: from molecular and biochemical fundamentals to commercial systems. *Appl Microbiol Biot* 60:523–533
2. Porath J, Carlsson J, Olsson I et al (1975) Metal chelate affinity chromatography, a new approach to protein fractionation. *Nature* 258:598–599
3. Hochuli E, Döbeli H, Schacher A (1987) New metal chelate adsorbent selective for proteins and peptide containing neighbouring histidine residues. *J Chromatogr* 411:177–184
4. Schmidt TGM, Koepke J, Frank R et al (1996) Molecular interaction between the Strep-tag affinity peptide and its cognate target, Streptavidin. *J Mol Biol* 255:753–766
5. Evan GI, Lewis GK, Ramsay G et al (1985) Isolation of monoclonal antibodies specific for human c-myc proto-oncogene product. *Mol Cell Biol* 5:3610–3616
6. Xu M-Q, Paulus H, Chong S (2000) Fusions to self-splicing inteins for protein purification. *Method Enzymol* 236:376–418
7. Smith DB, Johnson K (1988) Single-step purification of polypeptides expressed in *Escherichia coli* as fusions with glutathione S-transferase. *Gene* 67:31–40

Protein Transduction Domains

Definition

Protein transduction domains are small protein regions that cross biological membranes through a still poorly

understood mechanism. Various compounds and macromolecules including proteins can penetrate living cells when linked covalently to PTDs. The most commonly used PTDs are the third alpha helix of the *antennapedia* transcription factor, a peptide derived from the basic domain of HIV Tat, a peptide derived from the HSV VP22 transcription factor and poly-arginine peptides.

► **Peptide Aptamers**

Protein/DNA Interaction

PATRICE MORAND, CHRISTOPH W. MÜLLER
EMBL, Grenoble Outstation and University Hospital,
Grenoble, France
mueller@embl-grenoble.fr

Definition

Protein/DNA interactions allow DNA-binding proteins to recognize different DNA sites within the genome in order to fulfill central roles during DNA transcription, replication, packaging and repair. Atomic interactions between DNA-binding proteins and the DNA backbone and bases involve direct and water-mediated ► **hydrogen bonds** as well as ionic, van der Waals and ► **hydrophobic interactions** (1). Sequence specific DNA-binding proteins achieve specificity through a combination of contacts with DNA bases in the major or ► **minor grooves** of the DNA and with the DNA backbone (Fig. 1).

Characteristics

Double-stranded DNA is the central storage form of genetic information. The overall ► **conformation of double-stranded DNA** is rather uniform, although different DNA sequences influence the conformation and flexibility of DNA. To recognize specific DNA sequences, DNA-binding proteins mainly distinguish DNA bases in the major or minor groove of the DNA. To fulfill this function DNA-binding proteins have to be exactly positioned on the DNA through interactions with the DNA sugar-phosphate backbone. Protein residues contacting DNA bases and backbone jointly define a surface able to recognize specific DNA sequences. Biochemical and structural analyses have characterized a large number of DNA-binding proteins and their cognate DNA sequences. These analyses show a surprising variety in the architecture of DNA-binding proteins and in the way they interact with DNA (2, 3).

Interaction Characteristics

Amino Acid / DNA Base Interactions

Different DNA bases have different chemical properties resulting in a specific signature for a given DNA sequence. DNA-binding proteins preferably bind to those sequences that allow them to optimize the number of favorable interactions. Most amino acid-base interactions occur in the ► **major groove** of the DNA, where DNA bases adenine, guanine, cytosine and thymine are better distinguishable than in the minor groove. In addition, the major groove of B-form DNA is wider and therefore offers better access to DNA bases.

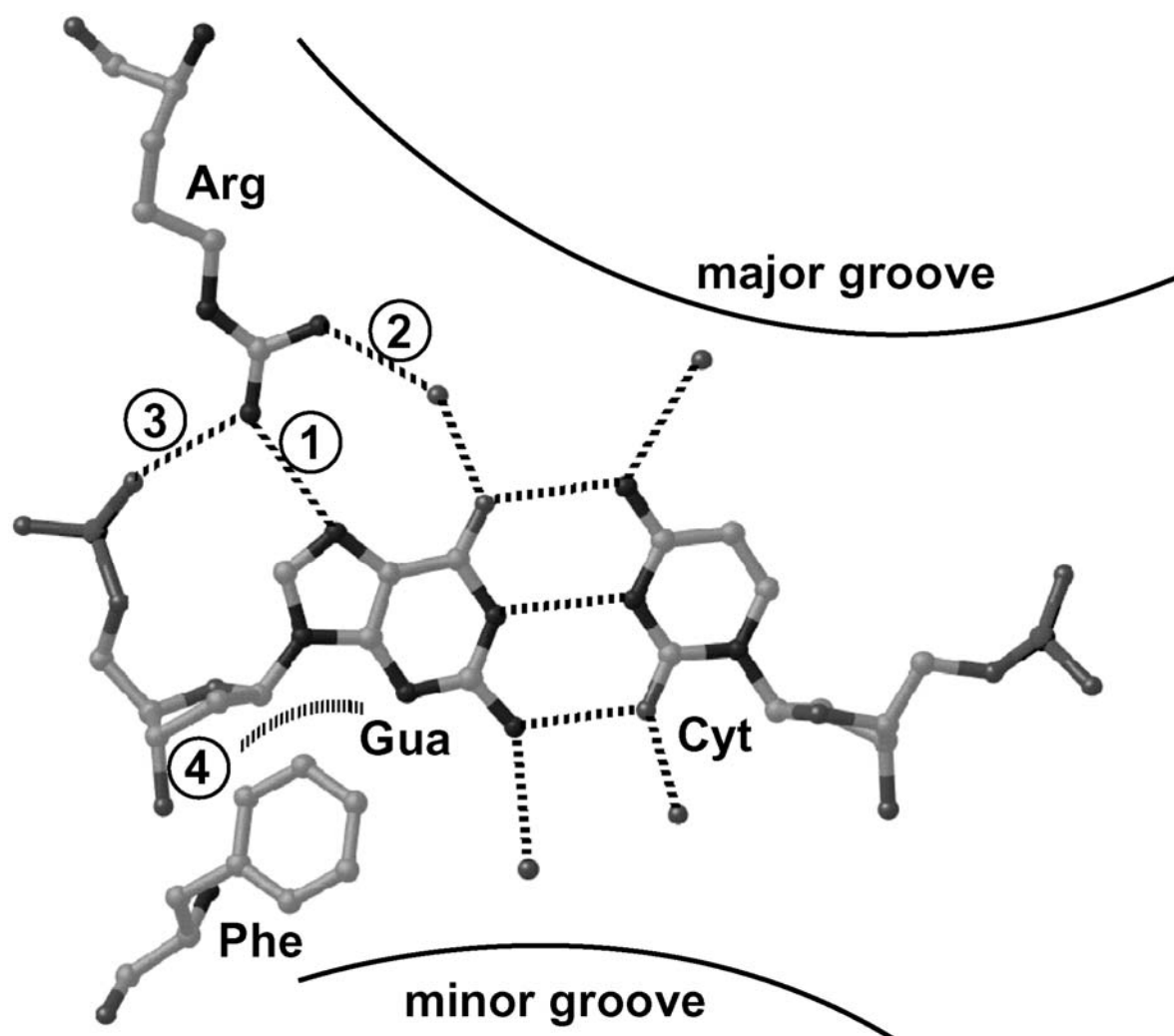
Hydrogen bonds between protein side chains and DNA bases account to a large extent for the sequence-specific recognition. About one third of these hydrogen bonds are single interactions between one residue and a DNA base; another third comprise bidentate interactions where at least two hydrogen bonds are formed between the protein residue and the DNA base or base pair. The remaining third comprise complex interactions where one residue contacts more than one base step (1).

► **Van der Waals interactions** also contribute to base specific recognition. For example, the exocyclic methyl group of thymine interacts with a number of hydrophobic side chains. Interactions of DNA bases with hydrophobic side chains like leucine, phenylalanine or proline are also particularly important for the binding of distorted DNA targets.

The increasing number of available protein/DNA complex structures allows the identification of certain preferences for interactions between protein side chains and DNA bases, like the bidentate interactions of arginine with guanine (Fig. 1) and of asparagine or glutamine with adenine bases. Despite such preferences, the overall conformation of the DNA-binding domain and its position on the DNA mainly determines the role of each residue in DNA recognition. No universal recognition code and no simple one-to-one relationship between amino acid residue and recognized DNA base has been established, and therefore the predictive power of such preferences remains limited.

Amino Acid / DNA Backbone Interactions

About two thirds of the total interactions between DNA-binding proteins and DNA comprise salt bridges, hydrogen bonds and hydrophobic interactions with the DNA sugar-phosphate backbone. Protein/DNA backbone contacts position DNA-binding domains onto the DNA to allow the formation of specific contacts with DNA bases. However, they also contribute to forming a complementary recognition surface able to recognize different DNA structures. Sequence-dependent variations in the DNA structure and flexibility can be “sensed” through protein/DNA backbone contacts. This mechanism allows the recognition of specific



Protein/DNA Interaction. Figure 1 Details of protein-DNA interactions. Sequence-specific recognition of a guanosine/cytosine base pair in the major groove through direct (1) and water-mediated (2) hydrogen bonds with an arginine residue. The arginine residue also contacts the DNA sugar-phosphate backbone (3). Hydrophobic interactions are observed between a phenylalanine residue, the guanine base and the backbone ribose (4). Note also additional water molecules interacting with the Gua:Cyt base pair.

DNA bases without directly contacting them and is often referred to as **indirect readout**. Protein/DNA backbone contacts are also particularly important for non-sequence specific DNA-binding proteins like the abundant **histone** proteins that form the protein core of the **nucleosome**.

Water-mediated Contacts

Water-mediated hydrogen bonds form a large part of all protein/DNA interactions. They are almost as abundant as direct hydrogen bonds and provide further complexity in protein/DNA interfaces. Most water molecules only appear to serve as “glue”, filling spaces between protein and DNA moieties. However, in some

cases precisely positioned water molecules partake directly in DNA sequence specific recognition.

Multimerization upon DNA-Binding

Many DNA-binding domains bind their target sites cooperatively as homo- or heterodimers. Homo- and heterodimerization allow sequence-specific DNA-binding proteins to increase their site specificity but also to extend the spectrum of potential binding sites by combining with various interacting partners. This principle is even taken one step further when different DNA-binding proteins bind cooperatively to DNA to form multimeric complexes.

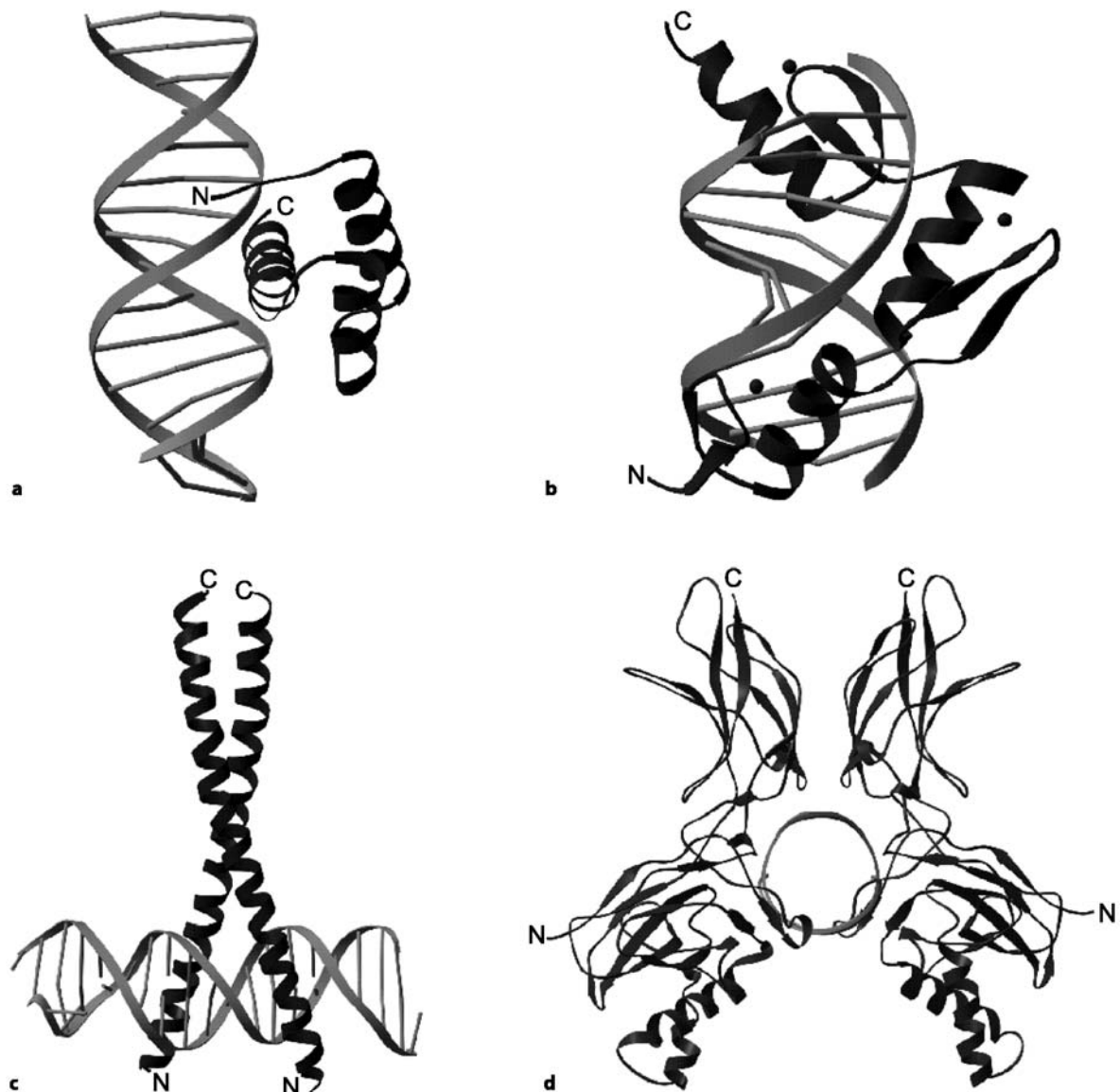
Classes of DNA-Binding Proteins

X-ray crystallography and ▶NMR spectroscopy have elucidated a large number of different structures of protein/DNA complexes with a surprising variety in the observed intermolecular interactions (Fig. 2). DNA-binding proteins often contain distinct DNA-binding ▶domains specialized in DNA-binding (2, 3). Many of these domains are rather small and consist of less than 100 amino acid residues with easily recognizable sequence signatures. More rarely, DNA-binding

requires large regions spanning several hundred amino acids. Different classification schemes have been proposed, which often use protein secondary-structure elements contacting the DNA as a criterion for classification.

DNA Recognition by α -Helices

Prominent examples of DNA-binding domains using α -helices inserted into the major groove for sequence specific recognition are proteins of the ▶helix-turn-helix



Protein/DNA Interaction. Figure 2 Examples of different protein/DNA complexes. (a) Engrailed homeo domain in complex with DNA. The helix-turn-helix motif is formed by the second and third helix. (b) Zn-finger protein Zif268 with three Zn-fingers arranged in tandem contacting the DNA. Zn-ions are depicted as spheres. (c) Basic-region-leucine zipper (bZip) DNA-binding domain of transcription factor GCN4. (d) NFκB P52 homodimer bound to DNA. DNA-binding is mediated by 5 loops per monomer.

transcription factor family as well as various Zn-coordinating or ▶leucine zipper containing DNA-binding domains (2, 3).

Helix-turn-helix DNA-binding domains are generally characterized by a stretch of 20 amino acid residues forming two α -helices connected by a short turn. The second of the two helices inserts into the major groove and contacts DNA bases and the backbone. N-terminal and/or C-terminal extensions are generally required to stabilize the helix-turn-helix motif and provide variability in site recognition. The helix-turn-helix motif forms the core of a large number of different prokaryotic and eukaryotic DNA-binding domains. Prominent examples are the bacterial repressors, eukaryotic homeodomain transcription factors that play crucial roles in development and winged helix-turn-helix proteins like the cell-cycle transcription factor E2F or the Ets domain transcription factors.

▶Zn-finger proteins of the Cys₂His₂ type form one group of Zn-coordinating DNA-binding domains where two anti-parallel β -strands are followed by an α -helix. Two cysteines and two histidines protrude from the two-stranded β -sheet and the α -helix to coordinate one Zn-ion. Each Zn-finger specifically recognizes three base pairs. DNA-binding domains generally contain several Zn-fingers arranged in tandem, which spiral around the DNA. Zn-fingers are a very abundant class of DNA-binding domains, with more than 300 representatives present in the human genome.

▶Basic-region-leucine zipper DNA-binding domains form dimers of uninterrupted α -helices that insert into the major groove. Prominent representatives are the yeast transcription factor GCN4, AP-1 (activator protein) and CREB (cAMP-response element-binding protein).

DNA Recognition by β -Sheets

DNA-binding proteins using β -sheets for sequence-specific recognition are less abundant. Examples include the bacterial repressors MetJ and Arc, where a two-stranded β -sheet runs through the major groove, the eukaryotic transcription factor GCM, which inserts a β -hairpin into the major groove and the general transcription factor TBP (TATA-box-binding protein), which binds in the minor groove and highly distorts the DNA.

DNA Recognition by Loops

▶Immunoglobulin-fold transcription factors share a common core related to immunoglobulin domains and use loops to contact DNA bases and backbone. Large regions of the protein and in some cases several domains are required for DNA-binding. So far, members of this family have only been observed in eukaryotes and include many medically important

transcription factors like P53, NF κ B, NFAT, STAT and T-box proteins.

Clinical Relevance

Normal development and physiology require the temporally and spatially coordinated expression of about 30,000 human genes. Regulation of gene expression is mainly provided by sequence-specific transcription factors that enhance or inhibit target gene expression by binding to specific promoter sequences. Up to 3000 human transcription factors are encoded in the human genome, and many diseases are entirely or partially caused by mutations affecting transcription factor activity.

In many cases, the mutations do not directly alter the transcription factor but rather change interacting partners upstream in the signaling cascade, like receptors, kinases or other effector molecules, leading to changes in the activation or repression of transcription factors. Nevertheless, a considerable number of disorders are directly caused by deletions, insertions and mutations of transcription factor genes. As a result the encoded proteins are either not properly expressed or deficient in DNA-binding, trans-activation or interaction with other partners.

For a number of human syndromes transcription factor ▶haploinsufficiency has been identified as the underlying cause (4 and references therein). Whereas one allele produces a functional transcription factor, the second allele is defective, consequently decreasing transcription factor levels to 50%. Haploinsufficiencies of T-box transcription factors TBX1, TBX5 and TBX3 have been identified as reasons for human DiGeorge, Holt-Oram and ulnar-mammary syndromes, respectively. Point mutations in TBX5, which directly affect residues contacting DNA in the major and minor groove, have been found in several Holt-Oram patients, providing a molecular explanation for this syndrome. In most human cancers a relatively small number of transcription factors, including c-JUN, STATs, NF κ B and AP-1, are over-activated and prevent apoptosis of cancer cells, whereas the activity of many tumor suppressors is down-regulated as is the case for P53. Most P53 residues that are frequently mutated in different cancers are near or at the protein/DNA interface. The two most frequently mutated residues in the DNA-binding domain of P53 are two arginine residues, of which one contacts the minor groove and the other forms a DNA backbone contact (5).

The pharmacological modulation of transcription factor activity represents an attractive therapeutic strategy. Mainly, two medical approaches have been explored to interfere with transcription factor activity. Currently, the most frequent strategies impact upon the signal transduction cascades that regulate transcription factor phosphorylation and dephosphorylation and/or nuclear

import. More specific strategies seek natural ligands or synthetic small molecules capable of binding to the transcription factor to modulate activation of the functional domains or inhibit their binding to DNA (6, 7). For example, tamoxifen or raloxifen used in breast cancer bind and induce a conformational change in the activation function domain 2 (AF2) of the estrogen receptor (a nuclear receptor belonging to the group of Zn-coordinating DNA-binding proteins), leading to a useful anti-estrogen effect. Some drugs may also inhibit the binding of transcriptional repressors to promoter elements and therefore allow the derepression of a gene of medical interest. Given the size and complexity of protein/DNA interfaces, the directed inhibition or stimulation of the interaction of DNA-binding proteins with their DNA target sites is difficult. Nevertheless, recent advances in drug discovery technologies may help to identify new molecules that act directly at the level of protein-DNA interaction.

References

1. Luscombe N, Laskowski RA, Thornton JM (2001) Amino acid-base interactions: a three-dimensional analysis of protein-DNA interactions at an atomic level. *Nucl Acids Res* 29:2860–2874
2. Harrison SC (1991) A structural taxonomy of DNA-binding domains. *Nature* 353:715–719
3. Garvie CW, Wolberger C (2001) Recognition of specific DNA sequences. *Mol Cell* 8:937–946
4. Seidman JG, Seidman C (2002) Transcription factor haploinsufficiency: when half a loaf is not enough. *J Clin Invest* 109:451–455
5. Cho Y, Gorina S, Jeffrey PD et al (1994) Crystal structure of a p53 tumor suppressor-DNA complex: Understanding tumorigenic mutations. *Science* 265:346–355
6. Darnell Jr (2002) Transcription factors as targets for cancer therapy. *Nat Rev Cancer* 2:740–749
7. Emery JG, Ohlstein EH, Jaye M (2001) Therapeutic modulation of transcription factor activity. *Trends Pharmacol Sci* 22:233–240

Protein-Protein Analysis, Suppressor Hunting

RANDY STRICH

Fox Chase Cancer Center, Philadelphia, USA

r_strich@fccc.edu

Definition

The biochemical analysis of a protein often leads in two directions, namely what does this protein do and how is

this activity regulated. Identifying another factor that directly interacts with the protein of interest can provide significant insight into these two questions. The use of genetic approaches, in particular ►**second site suppression**, provides a powerful *in vivo* tool for identifying protein:protein interactions that can help define a particular pathway. Second site suppressors are defined as a ►**mutation** in a second gene that restores the wild type ►**phenotype** to an individual already harboring a mutation in the gene of interest. For a suppressor ►**allele** to encode a candidate interactive protein, it should have two characteristics. First, the suppressor allele should be dominant, not recessive in nature. Second, the suppressor should be ►**allele-specific** meaning that it is not able to suppress a deletion mutation in the gene of interest.

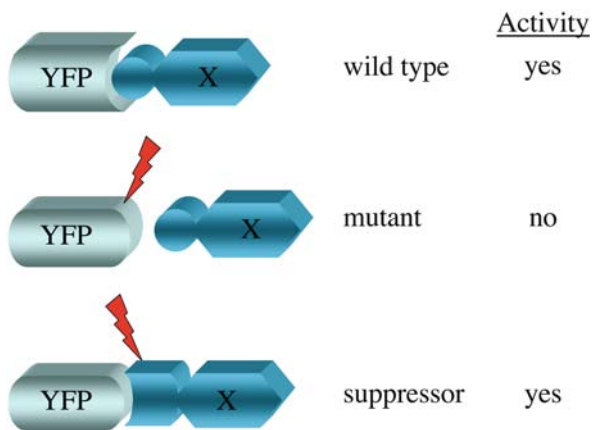
Characteristics

Classic Suppressor Analysis

A cartoon illustrating the second site suppressor analysis is shown in Fig. 1. Your favorite protein (YFP) is the protein of interest that has been defined in previous studies. For YFP to function properly, it requires the interaction of an unknown protein (X) (top line). A mutation in YFP that alters its conformation such that it no longer binds X (middle line) will no longer function thus producing a mutant phenotype. This phenotype makes the cell different under the particular experimental conditions employed, thus distinguishing it from the wild type. For example, this phenotype can result in cell death under certain growth conditions or be subtle such as a change in cell shape. A second round of mutagenesis generates a new mutation in X that restores its ability to interact with the mutant form of YFP thereby restoring wild type function (i.e. suppression, bottom of Fig. 1). This scenario makes two predictions for X if it is a direct binding partner of YFP. First, the mutated X should be able to suppress the YFP mutation even in the presence of the wild type X protein. This “dominant” activity over the wild type protein is the necessary gain of function required to recognize the mutant YFP protein. The second prediction is that the X-dependent suppression will only occur with the mutated YFP allele used in the suppressor hunt. For example, this suppression will not happen if the YFP gene has been deleted, as both proteins are necessary for normal function. Allele-specific suppression and genetic dominance are hallmarks of a directly interacting protein.

Dosage Suppressor Analysis

Like classical suppressor analysis described above, the rationale for dosage suppression is that a mutation in YFP reduces its ability to bind protein X causing a mutant phenotype (Fig. 2). The increased concentration

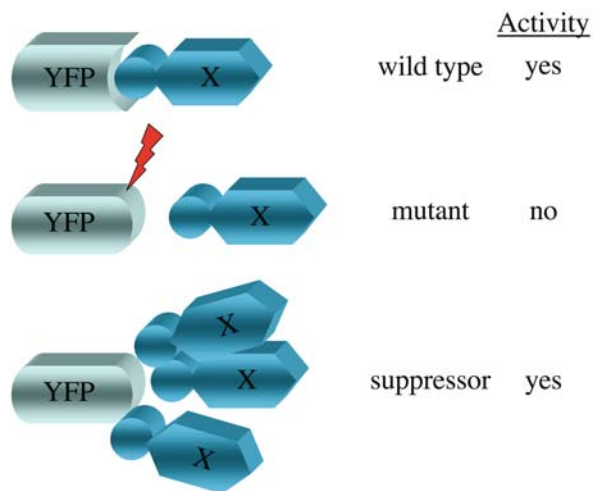


Protein-Protein Analysis, Suppressor Hunting. Figure 1 Classical suppressor analysis. The interaction between your favorite protein (YFP) and another factor (X) is required for wild type activity (top panel). The mutation in YFP (lightning bolt) disrupts this interaction causing a mutant phenotype (middle panel). A new mutation in X (second bolt) restores this interaction thereby reconstituting the pathway and generating a wild type phenotype (bottom panel).

of the X protein will compensate for this reduced binding affinity and therefore restore wild type activity. In this suppressor strategy, the increased “dosage” can be derived from increasing the copy number of X through use of a high-copy plasmid or placing X under the control of a strong promoter that causes X to be transcribed at a significantly higher frequency or both. Dosage suppression has three advantages over the classical suppressor analysis described in Fig. 1. First, elevated gene dosage is normally derived from a [▶library](#) of random genomic fragments or cDNAs inserted into a high-copy plasmid. Once a suppressor has been identified, the plasmid containing the activity can be rapidly isolated and the suppressor gene identified. Second, the source of the library can be derived from any organism thus allowing cross-species analysis. This option may be especially important if a heterologous system is being established (see below). Finally, a new mutagenesis cycle, which can introduce new and unwanted mutations, is avoided. The rules described for classical suppressor analysis (e.g. dominant activity and allele specificity) still hold for dosage suppression.

Selection or Screen, Phenotypic Choices

The use of suppressor analysis to identify interacting proteins must start with a phenotype that separates the individual harboring the mutant form of YFP from the rest of the population. Phenotypes can be sorted by two basic methods, selections or screens. A selectable phenotype usually relies on life and death to distinguish



Protein-Protein Analysis, Suppressor Hunting. Figure 2 Dosage suppression. As in panel A, wild type function requires the interaction between YFP and X (top panel). A mutation in YFP reduces the binding affinity to X resulting in a mutant phenotype. Increasing the effective concentration of X (lower panel) compensates for the reduced binding ability thus producing wild type activity.

between wild type and mutant individuals respectively. Many conditions can be used to provide the trigger that leads to lethality in the mutant strain. For example, growth at elevated temperatures is a commonly used condition that can render mutant proteins inactive even though they are functional at lower temperatures. In this example, a suppressor mutation would simply restore the ability of the individual containing a mutation in YFP to grow at high temperature. As only individuals that survive the elevated temperature are of interest, the use of a selection offers the power to examine a large number of individuals for suppressor activity. Conversely, a screen requires more effort as each individual must be examined for the restoration of the wild type phenotype. For example, the original mutation causes a fly eye to change color. A suppressor that reverts the eye back to the original color will only be found by examining each fly. Therefore, whenever possible, the use of a selective protocol is recommended both to save time and saturate the system.

All Mutations Are Not Created Equal

The selection of the target allele to initiate the suppressor hunt is critical for enhancing the chances of success. Two criteria must be satisfied for a target allele to be used in an interacting suppressor hunt. First, and foremost, the mutant protein must be present in the cell and be maintained at normal (or near normal) levels. Without the target protein present and in its normal cellular concentration and location, it is

unlikely that an interacting protein can be mutated in any fashion that will provide suppressor activity. Second, when at all possible, an allele in the target protein should be employed that harbors an inactivating ►[missense](#) mutation in an amino acid that is on the surface of the protein. If the goal is to identify interacting proteins, then mutations that cause large deletions of the coding region should be avoided.

Suppressor Analysis

The isolation of a second site suppressor requires several validation steps. The first step is to determine if the allele isolated is able to suppress a null mutation of YFP. If it can, then this allele is characterized as a bypass suppressor and is unlikely to interact directly with your protein of interest. These types of suppressor typically function downstream or independently of YFP activity and would be of interest if the entire pathway were under investigation. This characteristic of suppressor analysis has allowed investigators to cast a wide net for proteins that function within a particular pathway. For dosage suppressors, simply re-introduce the rescued plasmid into a strain harboring a null allele of the target gene. Suppressors that still function in the presence of a null allele again most probably represent bypass suppressors.

The second condition of allele specificity is the gold standard for directly interacting suppressors. Allele specificity suggests that the change in the suppressor protein is able to compensate specifically for the target mutation. This information is invaluable in delimiting interacting domains for both proteins as the mutant alleles can be sequenced to determine where the new mutations are located. For suppressor mutations, outcrossing to a different target allele is necessary. For dosage suppressors, the connection between allele specificity and direct interactors is less well documented. However, the dosage suppressor can be tested for allele specificity by simply re-introducing the suppressor-harboring plasmid into a mutant strain that harbors a different mutation in YFP.

Clinical Relevance

The genetic approach of second site suppression is most commonly conducted in model genetic organisms such as yeasts, fruit flies or nematodes. Similar approaches have proven difficult and/or too costly to recapitulate in mammalian cell culture systems or in whole animal studies. However, an alternative approach is to establish a genetic system in yeast based on a foreign target gene. The gene selected can have direct relevance to human disease or be a component of a basic biological process (e. g. cell-cycle control, signal transduction) that can have an indirect role in generating a disease state. This method takes advantage of the considerable homology observed between basic

biological processes in yeast and mammals. For example, the glucocorticoid receptor/transcription factor can activate reporter gene expression in yeast (1, 2). Moreover, G-coupled protein receptors have been modified to activate the mating pheromone mitogen activated protein (MAP) kinase cascade (3, 4). Establishing an assay with human YFP that provides a phenotypic readout in yeast will then allow suppressor analysis to be performed with a mutant YFP. An alternative approach is to find a homolog of the mammalian YFP in the model organism. This strategy relies on the conservation of pathways and activities of YFP in higher and lower eukaryotes. In either case, suppressor analysis may permit the isolation of interacting proteins with significance in human disease.

References

1. Schena M, Yamamoto KR (1988) Mammalian glucocorticoid receptor derivatives enhance transcription in yeast. *Science* 241:965–967
2. Garabedian MJ, Yamamoto KR (1992) Genetic dissection of the signaling domain of a mammalian steroid receptor in yeast. *Mol Biol Cell* 3:1245–1257
3. Klein C, Paul JI, Sauve K et al (1998) Identification of surrogate agonists for the human FPRL-1 receptor by autocrine selection in yeast. *Nat Biotechnol* 16:1334–1337
4. Reilander H, Weiss HM (1998) Production of G-protein-coupled receptors in yeast. *Curr Opin Biotechnol* 9:510–517

Proteinase 3

Definition

Proteinase 3 (also called myeloblastin) is a serine protease found in azurophilic granules of myeloid cells.

►[Morbus Wegener](#)

ProteinChip Array

Definition

ProteinChip Array is a specific form of a protein biochip (►[Protein Chips](#)) incorporating ►[SELDI](#) technology. ProteinChip Arrays are used in conjunction with a time-of-flight mass spectrometer detector thus enabling protocols of protein capture and detection from complex samples without additional detection labels. ProteinChip

Arrays can be designed with broad capture affinities based on chromatographic interactions or more specific affinities based on protein-protein, DNA-protein or receptor ligand interactions.

►Mass Spectrometry: SELDI

ProteinChip Reader

Definition

ProteinChip Reader refers to a time-of-flight mass spectrometer that is used to detect proteins desorbed from a ►ProteinChip Array by a laser in ►SELDI technology

►Mass Spectrometry: SELDI

ProteinChip Technology

►Mass Spectrometry: SELDI

Protein-Ligand Interaction Studied by NMR

CHRISTIAN FREUND

Freie Universität Berlin and Institute for Molecular Pharmacology, Berlin, Germany
freund@fmp-berlin.de

Definition

The molecular interactions of a ►protein with a small organic molecule, a peptide or another protein can be followed by the acquisition of nuclear magnetic resonance spectra. Depending on the type of ►NMR experiments used, qualitative information about the amino acid residues important for binding can either be obtained for the protein or for the ►ligand or for both. In addition, a full NMR-based structure determination of the protein-ligand complex allows the interaction to be viewed at atomic resolution.

Description

►NMR spectroscopy represents a powerful method of examining the structural and dynamic properties of

molecules in aqueous solution. The interaction of a macromolecule of interest, for example a protein, with another molecule (referred to as the ligand) results in changes in NMR observable parameters. Typical NMR methods for the investigation of protein-ligand complexes make use of the difference in ►chemical shifts between free and bound protein (or ligand) resonances or they exploit the different biophysical properties (►relaxation, diffusion) of a small molecule ligand as compared to the target protein. In addition, ►NOE- and RDC- (residual dipolar coupling) based NMR methods allow the determination of high-resolution three-dimensional structures of protein-ligand complexes.

Ligand-Exchange Rates

Two limiting cases for the ►chemical exchange of protein and ligand binding can be observed by NMR spectroscopy. If the exchange rate is large compared to the chemical shift difference between the free and bound ligand/protein resonances, the observed signal from either the protein or the ligand will gradually shift upon addition of the interaction partner (►fast exchange limit). Alternatively, if the exchange rates are small compared to the chemical shift difference between free and bound ligand/protein resonances, the signal intensity of the free form will decrease, whereas the signals of the bound form will increase (►slow exchange limit). If the time scale of chemical exchange and the time scale defined by the chemical shift difference between the free and the bound forms are similar, the lines become very broad and coalescence is observed. Interactions defined by this intermediate exchange regime are difficult to analyse and it is advantageous to try to separate the time scales by temperature shifts, acquiring the spectra at different magnetic field strengths or by chemically modifying the ligand (thereby either enhancing or decreasing the kinetic on- or off-rates). Since the resonance frequency difference between the free and bound form of the binding partners can be small or large for individual nuclei, different exchange regimes might be observed for the different atoms of the protein or the ligand. In many cases the kinetic on-rate for a typical protein-ligand interaction is diffusion limited and the different off-rates of an intermolecular interaction determine whether fast or slow exchange is observed. Interactions described by dissociation constants larger than 10 μM often fall into the fast exchange limit of modern high field spectrometers, while smaller dissociation constants result in intermediate or slow exchange compared to the NMR defined time scale.

Biological systems provide many examples of interactions that belong to either of the two exchange regimes. For example, many interactions that are part of the signal transduction process in living cells are characterized by large off-rates that reflect the transient

nature of the respective molecular complexes. In contrast, molecular recognition of antigens by affinity-matured antibodies is exclusively characterized by relatively small off-rates and tight binding. Protein-ligand interactions that are characterized by the fast exchange limit have the advantage that the course of the chemical shift change can be followed when increasing amounts of the binding partner are added. Under the assumption that the change in chemical shift is proportional to the fraction of bound molecules, a binding constant can be derived in favourable cases. For example, if the change in protein resonances is followed after addition of increasing amounts of ligand and if the binding is characterised by a simple two-state mechanism ($A + B \leftrightarrow AB$) the fraction bound can be derived from the solution of a quadratic equation that includes the dissociation constant as a fit parameter. While the qualitative mapping of the binding site is easily possible for protein ligand complexes that comply with the fast exchange limit, an **▶NOE**-based elucidation of the three-dimensional structure is often difficult to achieve. The fast exchange of the ligand might hamper the build-up and detection of NOEs between the protein and the ligand that are crucial for the process of structure determination (1). However, chemically altered ligands, intramolecular tethering of the ligand to the protein or acquisition of NOESY spectra at different protein-ligand ratios and with different NOESY mixing times might offer ways to detect intermolecular NOEs under the fast-exchange condition. These NOEs have to be interpreted cautiously within the context of structure calculation algorithms, since the isolated spin-pair approximation less adequately describes the correlation between distance and signal intensity. In contrast, slowly exchanging ligands can often be analysed within a 1:1 complex and the structures of many protein-ligand and protein-protein complexes with tightly bound ligands have been solved.

Isotope Labelling

In many cases the resonances of the protein and the ligand will fall into overlapping spectral regions. To distinguish the resonances of the individual binding partners, isotopic labelling of either the protein or the ligand allows the acquisition of NMR spectra that select for or against the incorporated isotope.

In most labelling protocols either the protein of interest or the ligand will be labelled by isotopes that contain the favourable magnetic properties of spin-1/2 nuclei (aside from the naturally abundant protons). Incorporation of ^{13}C and ^{15}N is most frequently used and typically obtained by recombinant expression of the protein of interest in bacteria or eukaryotic cells that are grown on isotopically enriched media. Since the natural abundance of these isotopes is low (1.1% for ^{13}C and

0.36% for ^{15}N), it is possible to detect only the signals from the isotopically labelled component of the molecular complex. Alternatively, the protons of either of the two binding partners can be “silenced” by substitution with deuterons, and only the proton signals of the non-deuterated binding partner will be detected. For full deuteration, the adaptability of *E.coli* to grow on > 98% deuterated components is most often exploited in conventional labelling schemes. In the case of small molecule ligands, a synthetic incorporation of the desired isotope can often be achieved. Since many modern NMR pulse schemes rely on the selection of coherence pathways based on the direct coupling of two spin-1/2 nuclei, there are many different ways to selectively detect the resonances of the labelled component of a complex.

Chemical Shift Changes

The chemical shift is a very sensitive indicator of changes in the chemical environment of atomic nuclei. Chemical shift changes can therefore be seen in cases of temperature shifts, solvent exchange, chemical modifications, conformational changes or non-covalent interaction with other molecules. Protein-ligand interactions are defined by non-covalent interactions such as van der Waals interactions, hydrogen bonds and electrostatic forces. The change in chemical environment upon molecular encounter can be due to neighbouring effects of the closely bound protein/ligand or be caused by a conformational change that accompanies the binding event. For rigid proteins, the chemical shift change of the detected nuclei is therefore a direct qualitative measure of the distance from the ligand atoms (2).

Detection of Protein NMR Signals-SAR by NMR

For the detection of protein resonances in ligand binding experiments, one of the most widely used methods is the ^{15}N - ^1H correlation spectrum, where, under acidic to neutral pH conditions, all amide groups give rise to a resonance in the corresponding two-dimensional NMR spectrum. The abundant use of experiments that exploit this correlation is explained by the low price of ^{15}N -labelled $\text{NH}_4\text{Cl}/(\text{NH}_4)_2\text{SO}_4$ and the good dispersion of the ^{15}N -resonances. For folded polypeptides, the resonance of proteins and protein domains of up to 300 amino acids can often be resolved to the level of individual NH groups. In cases where the individual resonances have been assigned in the ^{15}N - ^1H spectrum, the amino acids that are involved in binding can be identified. This so-called structure activity relationship (SAR) by NMR has found wide application, especially in target based screening procedures (3). If the three-dimensional structure of the protein is known, the binding site can be qualitatively mapped onto the surface of the protein. The limitation of this

method is the restricted size of the protein that can be investigated. Since T_2 relaxation increases monotonously with molecular weight, resonance assignments get severely hindered for proteins with a molecular weight greater than 50–100 kD. Although spectroscopic techniques such as TROSY offer ways to select the slowly relaxing components of NMR signals, the effort and costs incurred in obtaining spectral assignments and determining structures of protein complexes greater than 50 kD are enormous. The size limitation does not apply to the detection of ligand resonances that are measured in the presence of substoichiometric amounts of a given protein.

Detection of Ligand Resonances

In order to selectively detect ligand resonances, isotopic labelling can be used to enable the selective NMR experiments to be used in a manner analogous to that described for proteins. Alternatively, if the molecular size difference of the protein and the ligand is large (>10), the biophysical properties associated with differently sized binding partners offer a way to distinguish the much sharper resonances of the ligand from the broad signals of the high-molecular weight component. NMR transverse relaxation (T_2 relaxation) is the major pathway for the rapid decay of magnetization and addition of increasing, sub-stoichiometric amounts of the protein will lead to line-broadening of ligand signals that are now affected by the much slower tumbling of the protein-ligand complex. The finding that protein-ligand interactions cause measurable NMR relaxation enhancement of the ligand even under conditions of large ligand excess is also the basis for the measurement of transferred-NOEs (4). The nuclear Overhauser enhancement (NOE) is dependent on the longitudinal cross-relaxation between nuclear spins that are close in space (2–6 Å). Small molecules display weakly positive or zero NOEs. In contrast, macromolecules are characterized by large negative NOEs. Within the molecular complex, NOEs between bound ligand protons build up and, under conditions of fast ligand exchange, are transferred to the free ligand molecules whose NMR resonance signals are monitored. Since NOE intensities as a first order approximation are proportional to the inverse sixth power of the distance between protons, the conformation of the bound ligand may be derived from transferred NOE measurements.

A widely used transferred NOE-based method for the screening of small molecules is the saturation transfer difference technique (STD-NMR) (6). Here, regions of the protein that do not show overlap of protein resonances with ligand signals are saturated and magnetisation is subsequently transferred to the whole protein by spin diffusion. Spin diffusion is mediated by NOEs between spatial neighbours and increases with

the size of a protein. Eventually, magnetisation will also spread to bound ligand molecules, which in turn is transferred into solution by chemical exchange. Therefore, ligands interacting with the protein show attenuated signals and a comparison of the saturation transfer experiment with a spectrum acquired without saturation of the protein results in a saturation transfer difference (STD) spectrum that selectively displays the resonances of the interacting ligand.

Another method for distinguishing free and bound ligand resonances is based on the diffusion time of molecules (5). Pulsed field gradients that are able to create a spatially encoded dephasing of magnetization have a differential effect on slow and fast diffusing molecules. The amount of magnetization that can be regained by a second gradient pulse (rephasing) will be different for the more slowly diffusing bound ligand molecules than for the free ligand.

Structures of Protein-Ligand Complexes

The NMR structure determination of protein-ligand complexes is based on the NOEs observed between the protein and the ligand (1). Once resonance assignments have been obtained for both molecules, distance restraints can be derived from NOESY-spectra. Isotope edited NOESY experiments that use isotopically labelled protein or ligand allow the identification of important intermolecular NOEs. For example, half-filtering methods select for or against the magnetization of certain nuclei based on spin-echo difference modules that distinguish between directly coupled and uncoupled magnetisation (7). Another approach uses the asymmetric deuteration of protein and ligand and the recording of a ^{15}N -NOESY-HSQC spectrum. NOEs observed between NH and aliphatic protons of a small ligand will be intermolecular. Novel methods exploit the selective $^{13}\text{C}/^1\text{H}$ -labelling of isoleucine, leucine and valine in a deuterated background (8). Complementary, protonated aromatic side chains might be introduced into the deuterated protein and both labelling schemes allow the identification of intermolecular NOEs between aliphatic methyl groups and/or aromatic side-chains. NOE signals are then converted into distance restraints that are used in structure calculation programs as for example CNS or CYANA. In addition to NOE derived distance restraints, dipolar couplings between spin-1/2-pairs can, in principle, be used to estimate distances between atom pairs. To measure non-zero internuclear dipolar couplings accurately, the molecules of interest must be either weakly aligned with the magnetic field in a dilute liquid-crystalline medium (9) or be investigated by solid-state NMR. Especially in cases where NOEs are scarce, the measurement of dipolar couplings can add significantly to the precision and quality of a three-dimensional structure determined by NMR spectroscopy.

Clinical Applications

Several of the techniques mentioned in the previous sections (SAR by NMR, STD-NMR) have been efficiently used in the screening of libraries of small molecules. The NMR based screening methods often provide complementary information to ►HTS-methods, since they are able to detect very weak intermolecular interactions. Structural information on ligands identified in a screen can often be obtained directly or in combination with a few additional NMR experiments. NMR can also be used to test for false positives obtained with an HTS-approach since NMR should always detect a direct molecular interaction in aqueous solution. For these reasons, protein-ligand interactions investigated by NMR have become an important tool for the development of drugs with biological and clinical potential.

References

1. Neuhaus D, Williamson M (1989) The Nuclear Overhauser Effect in Structural and Conformational Analysis. VCH Publishers, New York, pp 1–522
2. Goldfarb V, Wittekind M, Jeffrey PD et al (1993) Production and characterization of an antibody Fv fragment ¹⁵N-labeled in the VL domain only. *J Mol Biol* 232:15–22
3. Shuker SB, Hajduk PJ, Meadows RP et al (1996) Discovering high-affinity ligands for proteins: SAR by NMR. *Science* 274:1531–1534
4. Ni F (1994) Recent developments in transferred NOE methods. *Progress NMR Spectroscopy* 26:517–606
5. Stejskal EO, Tanner JE (1965) Spin diffusion measurement: Spin echoes in the presence of a time-dependent field-gradient. *J Chem Phys* 42:288–292
6. Mayer M, Meyer B (1999) Characterization of ligand binding by saturation transfer difference NMR spectra. *Angew Chem Int Ed Engl* 35:1784–1788
7. Otting G, Senn H, Wagner G et al (1986) Editing of 2D H1-NMR spectra using X half-filters – combined use with residue-selective N-15 labeling of proteins. *J Magn Reson* 70:500–505
8. Gross J, Gelev VM, Wagner G (2003) A sensitive and robust method for obtaining intermolecular NOEs between side chains in large protein complexes. *J Biomol NMR* 25:235–242
9. Tjandra N, Bax A (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* 278:1111–1114

Proteinogenic Amino Acids

Definition

Proteinogenic amino acids comprise of the twenty standard amino acids completed by some special amino acids like seleno cystein and others, which are coded in

the mRNA and incorporated in the peptide chain during protein synthesis at the ribosome.

►Amino Acids: Physicochemical Properties

►Classification of Active Centers

Protein-Protein Interactions

PETER UETZ, CAROLINA S. VOLLERT

Institute for Genetics, Forschungszentrum Karlsruhe, Karlsruhe, Germany
peter.uetz@itg.fzk.de

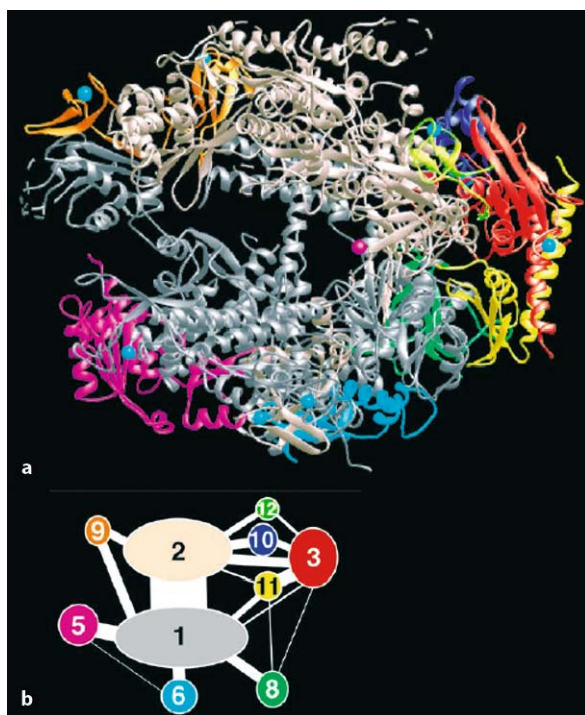
Definition

Specific interactions between two or more proteins, i.e. enzyme-inhibitor complex; antibody-antigen complex; receptor-ligand interactions; multiprotein complexes such as ribosomes or RNA polymerases (Fig. 1).

Characteristics

Protein-protein interactions affect all processes in a cell; proteins need to interact in order to shape the cell as well as organelles and molecular machines such as ribosomes and RNA polymerases. Multi-subunit channels or receptors in membranes are also held together by protein-protein interactions. Protein-protein interactions can be arbitrarily classified based on the proteins involved (structural or functional groups) or based on their physical properties (weak and transient, “non-obligate” vs. strong and permanent). Protein interactions are usually mediated by defined domains, hence interactions can also be classified based on the underlying domains. This domain specific interaction distinguishes such interactions from random collisions that happen by Brownian motion in aqueous solutions inside and outside cells. Note that many proteins are known to interact, although at present it remains unclear whether certain interactions have any physiological relevance. In a single-cell organism such as yeast each of the roughly 6000 proteins interacts with at least 3 other proteins, which adds up to a total of ~20,000 interactions or more. By extrapolation, there may be approximately ~100,000 interactions in the human body.

Most protein-protein interactions are detected as interacting pairs or as components of protein complexes. Simple eukaryotes such as yeast contain several hundred protein complexes, which may contain dozens or even hundreds of protein subunits each (e.g. ribosomes, spliceosomes). It has been proposed that all proteins in a given cell are connected in a complex



Protein-Protein Interactions. Figure 1 A large protein complex and its protein-protein interactions. (a) Ribbon representation of the RNA Polymerase II complex structure. (b) Schematic interaction diagram for the 10 subunits. The thickness of the connecting lines corresponds to the surface area buried in the corresponding subunit interface. Colors of subunits are identical in (a) and (b). From Cramer et al. (2001) Structural basis of transcription: RNA polymerase II at 2.8 Å resolution. *Science* 2001, 292:1863–1876.

network, in which distinct protein interactions are forming and dissociating constantly.

Structural Features of Protein-Interaction Sites

Hundreds of protein complexes have been analyzed by ▶X-ray crystallography and other methods. Data about the structures of proteins and complexes are available from the Protein Databank (PDB, ▶<http://www.rcsb.org/>). The following aspects about geometry and energetics of protein interactions have been drawn from comparative analyses of up to 100 protein pairs and complexes that have been crystallized.

The contact area between two proteins is usually larger than 1100 Å² with each of the interacting partners contributing at least 550 Å² of complementary surface. On average each partner loses about 800 Å² of solvent-accessible surface upon contact, contributed by some 20 amino acid residues of each partner, i.e. the average interface residue covers some 40 Å². On average, dimers contribute 12% of their accessible surface area

to the contact interface, trimers 17.4% and tetramers 20.9%. However, variations exist and total interface areas range from 6% in inorganic pyrophosphatase homodimers to 29% in tryptophan repressor homodimers. For data sets of 10 enzyme-inhibitor complexes and 6 antibody-antigen complexes, the mean interface area is about 780 Å².

Forces that mediate protein-protein interactions include electrostatic interactions, hydrogen bonds, the van der Waals attraction and hydrophobic effects. The average protein-protein interface is not less polar or more hydrophobic than the surface remaining in contact with the solvent. Water is usually excluded from the contact region. In contrast, non-obligate complexes tend to be more hydrophilic, as each component has to exist independently in the cell. It has been proposed that hydrophobic forces drive protein-protein interactions, whereas hydrogen bonds and salt bridges confer specificity. Van der Waals interactions occur between all neighboring atoms, but these interactions at the interface are no more energetically favorable than those made with the solvent. However, they are more numerous, as the tightly packed interfaces are denser than the solvent and hence they contribute to the binding energy of association. Hydrogen bonds between protein molecules are more favorable than those made with water. Interfaces in permanent associations tend to have fewer hydrogen bonds than interfaces in non-obligate associations. The number of hydrogen bonds is about 1/170 Å² buried surface. A standard size interface (~1600 Å²) buries about 900 Å² of the non-polar surface, 700 Å² of polar surface, and contains 10 (5) hydrogen bonds. In a set of reasonably stable dimers there are on average 0.9–1.4 hydrogen bonds per 100 Å² of contact area buried (interfaces usually covering >1000 Å²). The number of hydrogen bonds varies from 0 in some complexes (e.g. uteroglobin) to 46 in variant surface glycoprotein. Side-chain hydrogen bonds represent approximately 76–78% of the interactions.

Only 56% of homodimers were found to possess salt bridges (many having none, and at the most five).

Independent studies of the shape showed that 83–84% of interfaces are relatively flat. With few exceptions, the interfaces are approximately circular areas on the protein surface in both permanent and non-obligate complexes. Interfaces in permanent associations tend to be larger, less planar, more highly segmented (in terms of sequence) and closer packed than interfaces in non-obligate associations.

Complementarity can be measured in terms of “fitting surface shape”. Interfaces in homodimers, enzyme-inhibitor complexes and permanent heterocomplexes are the most complementary, while the antibody-antigen complexes and the non-obligate heterocomplexes are the least complementary.

Proteins consist of 3 kinds of secondary structures: α -helices, β -sheets and loops, which usually connect the first two types and lie on the surface of proteins. In one study (involving 28 homodimers), 53% of the interface residues were α -helical, 22% β -sheets and 12% $\alpha\beta$, with the rest being coils or unstructured loops. The amino acid composition of interfaces has been shown to be more hydrophobic than the exterior but less hydrophobic than the interior of a protein. In one study, 47% of interface residues were hydrophobic, 31% polar and 22% charged. Permanent complexes have interfaces that contain hydrophobic residues, while the interfaces in non-obligate complexes favor the more polar residues. Site-directed mutagenesis showed that in many cases a majority (i.e. >50%) of interface residues can be mutated to alanine with little effect on the K_d ; i.e. the functional epitope is a subset of the structural epitope.

Thermodynamics

Protein-protein interactions can be described as simple chemical reactions:



A = Protein A

B = Protein B

AB = protein complex, i.e. an interacting pair of proteins

Multiprotein complexes usually assemble subunits successively.

Qualitative Description of Stability

Protein interactions can be weak and extremely short-lived (“non-obligate”) or strong and permanent (“stable”). For example, an enzyme binds and phosphorylates a protein substrate and dissociates after less than a microsecond. Other protein complexes such as the triple-helix of collagen reside in bones and other tissues for weeks or even years without dissociation.

Quantitative Description of Protein-Protein Interactions

Quantitatively, the interaction of two proteins follows the mass action law:

$$\frac{[A][B]}{[AB]} = \frac{1}{K_a} = K_d = \frac{k_d}{k_a} \quad (2)$$

k_a = second-order rate constant for the bimolecular association reaction,

k_d = first-order rate constant for the uni-molecular dissociation reaction,

$K_d = k_d / k_a$ = equilibrium constant for dissociation.

K_d is related to the concentrations of A, B and AB at thermodynamic equilibrium. K_d has the dimension of a

concentration and is expressed in mole/liter (or “M”). The range of K_d values observed in biologically relevant processes that rely on protein-protein interactions is extremely wide and extends over at least 12 orders of magnitude from 10^{-4} to 10^{-16} M. K_d values in the micromolar range are considered to be weak, values in the nanomolar range or below to be strong interactions (e.g. trypsin – pancreatic trypsin inhibitor [PTI] has a dissociation constant of the order of 10^{-14} M). However, the biological strength may depend on other effects such as cooperativity. Several weak interactions between the subunits of a complex can result in a highly stable complex.

Energetics

With a K_d ranging from 10^{-4} to 10^{-14} M, the free ΔG_d of dissociation δG_d ranges from 6 to 19 kcal/mol, i.e. 19 kcal are required to separate 1 mol of trypsin-PTI. Actual K_d values can be retrieved from some of the databases (Table 1). A single pairwise interaction between amino acids may account for as much as 6 kcal/mol. Residue pairs that form salt bridges and charged hydrogen bonds yield the largest values. Pairs that make neutral hydrogen bonds or non-polar interactions are in the 0–3 kcal/mol range. This is much less energy than that of a hydrogen bond, which implies that an interaction between the two residues in a complex is only marginally stronger than an interaction with water, with which they interact as free proteins. In complexes of known 3D structure, the peptide group participates in at least half of the hydrogen bonds at protein-protein interfaces. It has been estimated from an empirical correlation that for non-polar surfaces (i.e. hydrophobic interactions), there is an energy gain of approximately $25\text{--}72 \text{ cal}/\text{\AA}^2$. Many strong dimeric interactions that have association constants greater than 10^{-16} M have to be denatured to separate the complex into monomers.

Protein-Protein Interactions. Table 1 Databases of interacting proteins

Database of Interacting Proteins (DIP)	► http://dip.doe-mbi.ucla.edu/
BIND	► http://www.bind.ca/
AMAZE	► http://www.amaze.ulb.ac.be/
MINT	► http://cbm.bio.uniroma2.it/mint/
PDB (structures)	► http://www.rcsb.org/

Methods for Studying Protein-Protein Interactions

Details on the methods for the analysis of protein-protein interactions can be found in other sections of this Encyclopedia. Important experimental techniques are listed in Table 2.

Regulation of Protein-Protein Interactions

Interactions between proteins are tightly regulated. The most important ways of regulation involve the following mechanisms.

Expression

Proteins can only interact when expressed in the same place. RNA and protein expression (i.e. transcription and translation) are therefore regulating such interactions in a time- and space-dependent manner. For example, various fibroblast growth factors (FGFs) are expressed in different tissues such as limbs, brain, kidney etc. while their cognate receptors are similarly expressed in particular tissues where their interaction activates signaling pathways that in turn lead to the expression of certain target genes. FGFs are also expressed in a time-dependent manner, e.g. FGF4 and 8 are only expressed in embryos while others are mostly expressed in adult tissues. The same is true for FGF receptors.

Modifications

Covalent modification of proteins can prevent or induce binding of others. Examples include proteins that contain the so-called SH2 domain that binds only

peptides which contain a phosphorylated tyrosine residue.

Subcellular Localization

Several transcription factors such as NF κ B are present in the cytoplasm or even in membranes (such as Notch). They can be activated by translocation to the nucleus where they interact with other transcription factors in order to activate gene transcription.

Stability

Proteins can only interact when they are present. Thus, regulated degradation of proteins prevents them from further interactions. Prominent examples are the cyclins, which are broken down at specific time points during the cell-cycle, so they can no longer interact with their partners, cyclin-dependent kinases (CDKs).

Ligands

Small molecules or other proteins can regulate protein interactions. For example, binding of guanosine triphosphate (GTP) to trimeric guanosine nucleotide binding proteins (G proteins) promotes their dissociation while binding of guanosine di-phosphate (GDP) favors their association.

Prediction of Protein-Protein Interactions

Proteins of Known Structure

Prediction of interaction sites in proteins of known structure usually focuses on the location of hydrophobic surface clusters on proteins. One such study

Protein-Protein Interactions. Table 2 Methods to study protein-protein interactions

Method	Notes
Purification of protein complexes and their separation by ►2-dimensional gel electrophoresis or ►liquid chromatography.	After separation the components of a protein complex can be identified by the staining of protein bands by various dyes such as Coomassie Blue. Furthermore, the protein bands can be analyzed by ►mass spectrometry which can also be used to sequence proteins.
In-vitro binding experiments	Involve purified proteins whose interaction is detected by retention of one partner by a second protein that has been immobilized on some sort of matrix (such as agarose or glass).
In vivo methods	E.g. ►two-hybrid systems, ►FRET
Equilibrium ultracentrifugation ►Microcalorimetry	Used to determine dissociation constants in the micromolar range. When K_d is in the nanomolar range, radioisotope labelling or ELISA techniques can be used.
Site directed mutagenesis	Used to identify the role of individual amino acids.
NMR X-ray crystallography	Physical methods to determine the 3D-structure of proteins and protein complexes

predicted the correct interaction site in 25 out of 29 cases. Other methods include ►[solvation potential](#), ►[residue interface propensity](#), ►[hydrophobicity](#), planarity, ►[protrusion](#) and ►[accessible surface area](#). Among a test set of 28 homodimers, the known interface site was found to be amongst the most planar, protruding and accessible patches and among the patches with highest interface propensity. Nevertheless, one of the algorithms (PATCH) that uses multiple parameters predicted the location of interface sites in known complexes only for 66% of the structures.

Protein Pairs from Genome Sequences

Several attempts have been made to predict protein-protein interactions from genome sequences. Two major methods employ ►[Rosetta stone proteins](#) and phylogenetic profiles.

Rosetta Stone Proteins

Some protein sequences are split into two independent proteins in some organisms, whereas they are maintained together in other organisms. From such sequences, it has been concluded that the two independent proteins form a complex, based on the association in the fusion protein. Such fusion proteins are called Rosetta stone proteins. Supposedly, they predict interactions among related proteins. For example, human succinyl CoA transferase is split in *E. coli* into acetate CoA transferases α and β subunits.

Phylogenetic Profiles

Some gene pairs (or groups) are maintained together in many different organisms during the course of evolution. It has been concluded that such “co-evolving” genes encode proteins that are associated either functionally or physically, i.e. by a protein-protein contact. While the latter is not always true, it has been found to be true in a number of cases, such as the yeast proteins Hog1 and Fus3 and their homologs in other eukaryotes.

Clinical Relevance

Biologically active proteins such as peptide hormones or antibodies act by interacting with other proteins such as receptors or antigens respectively. Knowing their interaction sites allows the modification of the activity of such proteins or the changing of their specificity. In addition, small molecules may be designed that block such interactions, e.g. molecules that inhibit the binding of virus coat proteins to their cellular receptors, thereby blocking infection. Proteins and their interactions are therefore potential drug targets.

Protein-protein interactions can also be disadvantageous. Insulin for example tends to form dimers and

hexamers which are less active than monomers. Genetically engineered insulin molecules do not dimerize and hence retain biological activity.

References

1. Alberts B, Johnson A, Lewis J et al (2002) Molecular Biology of the Cell, 4th edn. Garland Science, New York
2. Lodish H, Berk A, Zipursky L et al (2000) Molecular Cell Biology, 4th edn. WH Freeman, New York
3. Golemis E (2001) Protein-Protein Interactions – A Molecular Cloning Manual. Cold Spring Harbor Laboratory Press, New York
4. Kleanthous C (2000) Protein-Protein Recognition. Oxford University Press. [contains good reviews on structural aspects and thermodynamics of protein-protein interactions]
5. Thorner J (2000) Applications of Chimeric Genes and Hybrid Proteins, Part C: Protein-Protein Interactions and Genomics. Methods in Enzymology, vol 328. Academic Press

Protein-Tyrosine Sulfotransferase

►Tyrosine Sulfation of Proteins

Proteoglycan

Definition

Proteoglycan is any protein with one or more covalently attached glycosaminoglycans.

►Glycosylation of Proteins

Proteolytic Degradation

Definition

Proteolytic degradation designates the hydrolysis of one or more peptide bonds of a protein by the action of proteases. Proteolytic degradation is involved in many physiologic operations, e.g. apoptosis, cell signalling, protein maturation or turnover.

►Recombinant Protein Expression in Yeast

Proteome

Definition

The term ‘proteome’ refers to the total PROTEins expressed by a genome, and to the systematic analysis of the protein profile of a given cell or tissue.

- ▶ Automated High-Throughput Functional Characterization of Human Proteins
- ▶ *Drosophila* as a Model Organism for Functional Genomics
- ▶ Protein Databases
- ▶ Protein Microarrays as Tool for Protein Profiling and Identification of Novel Protein-Protein Interactions
- ▶ Proteomics – from Proteins to Disease
- ▶ *Xenopus* as a Model Organism for Functional Genomics

Proteome Profiling

Definition

Protein profiling denotes the study of protein constitution or protein characteristics of a particular tissue or organism.

- ▶ Proteomics in Cardiovascular Disease

Proteomic Signatures

Definition

Proteomic signatures are proteomic phenotypes that characterize a biological response. A given phenotype may only be observed in a single experiment. Proteomic signatures are reproducibly elicited whenever a given cellular process is activated or inhibited.

- ▶ Two-Dimensional Gel Electrophoresis

Proteomics

Definition

Proteomics is the discipline that studies and analyzes the ▶ proteome of a cell. The proteome comprises the

total amount and composition of proteins present in a cell in a particular state at a given time.

- ▶ High-Throughput Approaches to the Analysis of Gene Function in Mammalian Cells
- ▶ Large-Scale Homologous Recombination Approaches in Mice
- ▶ Mass Spectrometry: ESI
- ▶ Structure-Based Drug Design
- ▶ Proteomics in Ageing
- ▶ Proteomics – from Proteins to Disease
- ▶ Proteomics in Cancer
- ▶ Proteomics in Cardiovascular Disease

Proteomics in Ageing

OLIVIER TOUSSAINT¹, FLORENCE CHAINIAUX¹,
ALINE CHRÉTIEN¹, VÉRONIQUE FLAMENT¹,
EDOUARD DELAIVE¹, JEAN-FRANÇOIS DIERICK²

¹University of Namur, Namur, Belgium

²Biovallee, Gosselies, Belgium

olivier.toussaint@fundp.ac.be

jf.dierick@biovallee.be

Definitions

Proteomics and Complexity of Age-related Protein Modifications

Proteomics is the analysis of an entire protein set expressed by a biological system under given conditions. Classically proteome analysis is a two-step process, 2DGE separates proteins as spots that are quantified. Recent refinements of 2DGE include ▶ differential in-gel electrophoresis. Then, quantified spots of interest are identified by mass spectrometry (MS) methods. Peptide mapping determines the molecular weights of peptides obtained after proteolytic cleavage and peptide sequencing allows partial sequencing of isolated peptides.

Human ageing is characterized by a progressive loss of physiological functions, increased tissue damage and defects in various tissue renewal systems. A fundamental biogerontological question is how do cells slowly accumulate irreversible modifications in a normal physiological milieu or in response to abnormal stress, eventually reaching a critical threshold leading to irreversible degeneration, functional losses and eventually death?

Proteomics is excellently suited to analyze the complex age-related biochemical modifications of our multiple cell types and subcellular compartments. With ageing,

Proteomics – from Proteins to Disease Mechanisms and the Development of Novel Therapeutic Strategies

ERICH E. WANKER, BIRGIT KERSTEN

Department of Neuroproteomics, Max Delbrück
Center for Molecular Medicine Berlin-Buch, Berlin,
Germany
b.kersten@mdc-berlin.de

Definition

Proteomics is the large-scale systematic study of the proteins of an organism, a specific type of tissue or a cell, using biochemical, cell biological and genetic methods (2, 62). The word proteomics has traditionally been used to describe the separation of proteins prepared from tissues and cell lines by ►two-dimensional (2D) polyacrylamide gel electrophoresis (1). In this sense, the first proteomics studies were performed in the 1970s, when researchers started the systematic cataloguing of protein spots from 2D gels and the building of databases of all expressed proteins (39). However, although the separation of proteins on 2D gels could be performed reproducibly in many different laboratories, the identification of proteins was difficult because of a lack of analytical methods for protein characterisation. In the 1990s, this limitation was overcome by the development of highly sensitive ►mass spectrometry (MS), which has to be called the most significant breakthrough in proteomics research in the last 20 years (23). Thus, the combination of 2D gel electrophoresis for separation of complex protein mixtures with mass spectrometry techniques for protein identification marks the beginning of a new era of proteomics research. Since then, the area of inquiry covered by the term proteomics has widened considerably. Today, it includes the systematic analysis of ►protein-protein interactions using the two-hybrid system or protein arrays (46), as well as large-scale protein localisation studies by ►immunofluorescence (IF) microscopy or studies to determine protein structure by X-ray crystallography (43). However, strategies exclusively targeting genes or messenger RNA, such as large-scale mutagenesis screens (51) or systematic antisense experiments (18) are not usually considered to be part of proteomics.

Introduction

The human genome project has laid the foundation for proteomics by providing the sequence data essential

for systematic protein analysis. Now, our goal is to understand the complex biological processes in which proteins are involved, using high throughput biochemical and cell biological approaches. In this review, we will describe how proteomics in the future will contribute to our understanding of biology and medicine through the global analysis of gene products.

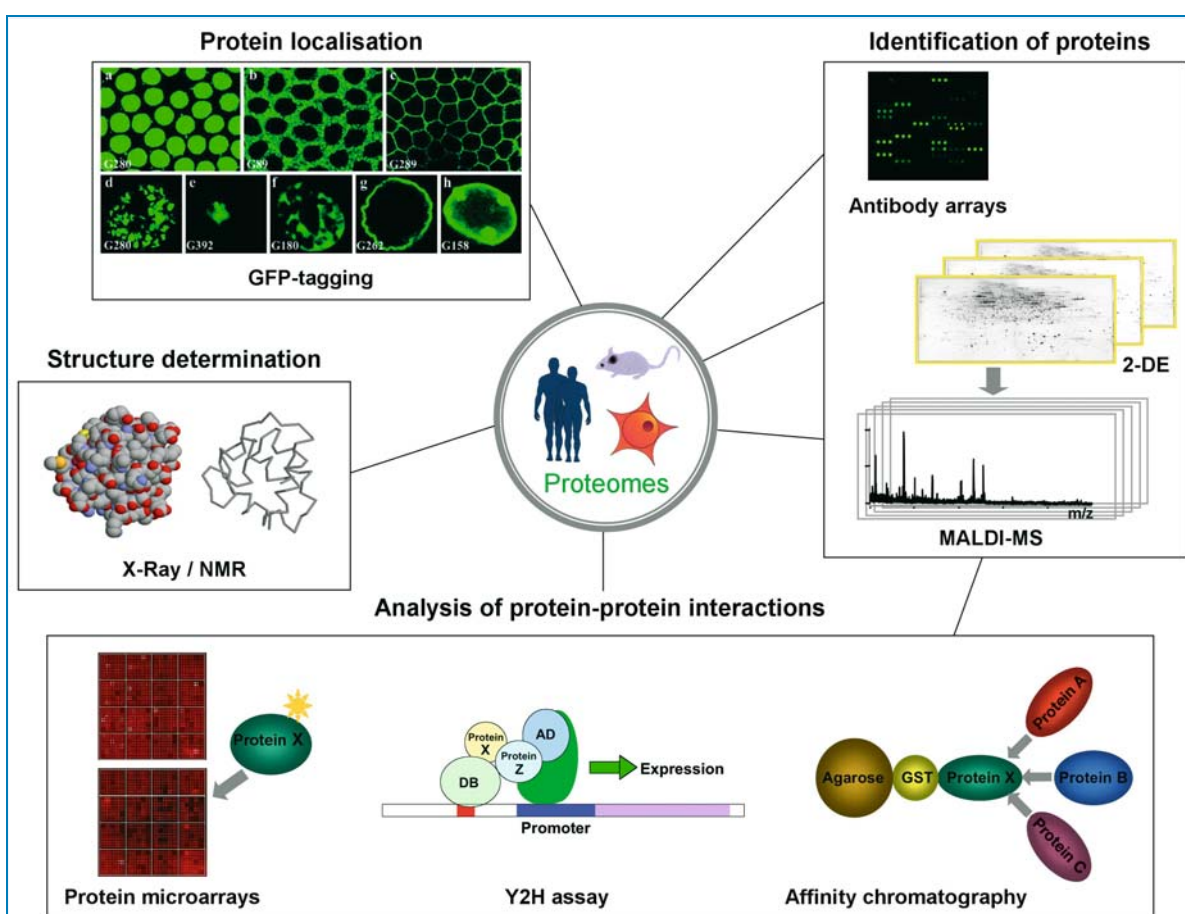
Why do we need proteomics research? During the past fifteen years, comprehensive DNA sequence information has been generated by the various genome projects. However, this information *per se* is not sufficient to understand the function of proteins in living organisms. A cell usually depends on a multitude of metabolic and regulatory pathways and there is no strict linear relationship between the genome and its protein complement, the “proteome”. Only a fraction of genes is activated in a specific cell type at a given time and protein synthesis and folding, rather than merely the genetic layout, determine the cell’s phenotype. Therefore, the systematic analysis of proteins and their interaction partners will significantly contribute to our understanding of cell function and the identification of novel drug targets for therapeutic intervention.

The analysis of ►open reading frames (ORFs) in genomic data sets for example, does not necessarily allow the prediction of functional proteins. This is particularly true for small genes and genes with low homology (9, 42). For reliable human gene prediction, it is therefore critical to verify the respective gene product by proteomics methods.

Modifications such as ►phosphorylation or ►methylation of proteins are also not apparent from the DNA sequence. Proteomics studies on tissue specific protein extracts are needed to identify active isoforms in protein complexes as well as post-translational modifications. As mRNA levels do not mostly correlate with protein expression levels (20), it is also necessary to quantify the amounts of proteins in cell extracts for the systematic analysis of gene regulatory networks. Moreover, the subcellular localisation of proteins as well as their interaction partners cannot be predicted from the DNA sequence. Co-localisation as well as protein-protein interaction studies are important to determine the molecular composition of cellular structures at the protein level. Together, these areas of proteomics research (summarised in Fig. 1) will increase our understanding of protein function and disease processes. They will allow the identification of new biomarkers for diagnosis as well as of new targets for therapy and accelerate drug development for complex diseases.

Identification of Proteins

A critical step in proteomics is the obtaining and handling of protein samples. Protein samples



Proteomics – from Proteins to Disease Mechanisms and the Development of Novel Therapeutic Strategies.

Figure 1 Different approaches of functional protein research. The image of the antibody array was obtained from Zhou et al. (2004); <http://genomebiology.com/2004/5/4/R289>, licensee BioMed Central Ltd. The GFP-tagging image was taken from Morin et al. (2001); Copyright 2006 National Academy of Sciences, U.S.A.; The GFP-image gives the subcellular distribution of trapped *Drosophila* proteins. Targeting of trapped *Drosophila* proteins to the nucleus (a), cytoplasm (b), or membrane (c). Different distribution of GFP-tagged *Drosophila* proteins in the giant nuclei of larval salivary glands (d-h).

prepared from tissues or cell lines contain >10,000 different proteins, which can be partly separated by 1D and 2D gel electrophoresis (2-DE). Although the best 2D gels are capable of resolving up to 10,000 protein spots (25), only the most abundant proteins of crude protein extracts can be visualised by 2D gel electrophoresis because the amounts of several proteins lie below the detection limit of the protein stains used. Subcellular fractionation techniques or affinity based purification strategies are applied to overcome this difficulty and to reduce sample complexity prior to electrophoresis. After silver or Coomassie staining, spots/bands of interest are excised and proteins are identified by mass spectrometry, which has replaced the classical technique of Edman degradation, because it is much more sensitive and offers higher throughput.

There are two major approaches for mass spectrometric protein identification, the “peptide mass mapping” approach and the “peptide fragmentation” approach. In the peptide mass mapping approach, initially described by (23), the mass spectrum of the eluted peptide mixture is acquired by ▶**matrix-assisted laser desorption/ionisation** (MALDI) measurement. The method has been improved by the automation of the MALDI identification procedure, whereby hundreds of protein spots can be excised, enzymatically digested and automatically analysed (6). In the “peptide fragmentation” approach, individual peptides in a mixture are analysed to gain sequence information (12). The peptides are ionised by ▶**Electrospray Ionization Mass Spectrometry** (ESI) directly from the liquid phase. They are then sprayed into a “tandem mass spectrometer”, which has

the ability to resolve peptides in a mixture, isolate them by species and dissociate them into N- or C-terminal fragments. This tandem method (ESI-MS/MS) is technically more complex than MALDI fingerprinting, but has the advantage of yielding the sequences of peptides, rather than listing only peptide masses. The fragmentation data can be used to search protein sequence databases and also nucleotide databases such as ► **expressed sequence tag** (EST) resources, which increases the probability of identifying an uncharacterised protein.

The development of mass spectrometry techniques for the identification of proteins is a rapidly growing field. A mass spectrometer unit that combines a MALDI ion source with an efficient tandem mass spectrometer for fragmentation of peptides is now available (53). It combines the high throughput capacity of the peptide mapping method with the specificity of the peptide sequencing method and allows an automated one step analysis strategy. Ongoing efforts are directed at speeding up sample preparation and separation in highly integrated appliances.

A main future goal is to analyse complex protein mixtures by mass spectrometry without prior gel separation in a single automated experiment and thereby to overcome problems of the 2-DE based protein identification, such as missing the identification of hydrophobic and low abundance proteins. The introduction of a liquid chromatography based peptide separation strategy named “► **multidimensional protein identification technology**” (MudPIT) was an important step forward in this direction (33). The basis of this technique is the digestion of a complex mixture of proteins before they are separated by two independent liquid chromatographic separation systems. From the second column the peptides are eluted directly into an ion trap mass spectrometer, where they can be measured and identified in a fully automated fashion.

Quantification of Proteins and Differential Display Proteomics

2D Gel Electrophoresis Approach

In most proteomics laboratories, 2D gel electrophoresis is used to identify and quantify proteins that are up- or down-regulated in specific cells or tissues. Biological materials derived from two different conditions are separated by 2D gel electrophoresis and proteins are visualised by Coomassie or silver staining. Using image analysis, proteins that are only present in one condition are selected and identified by mass spectrometric analysis. Using this approach, several diagnostic markers as well as potential therapeutic targets were identified in the past (see

Clinical Proteomics section). A recent study compared the protein complements from different mouse strains by 2D gel electrophoresis and, surprisingly, over 1,000 genetically variant protein spots were identified (26), indicating that 2D gel proteomics can be used to study genetic variability in mice in high throughput. However, the 2D approach has the disadvantage that the reproducible preparation of protein samples is difficult and requires a lot of experimental experience. An important recently made improvement in the 2D gel approach is the use of difference gel electrophoresis (► **DIGE**), in which two pools of proteins are labelled with different fluorescent dyes and separated in the same 2D gel (58). The advantages of this system lie in the increased reproducibility of the protein separation because the samples to be compared are separated in the same 2D gel, which decreases the number of 2D gels to be run. A further advantage lies in the high sensitivity and the broad and linear dynamic ranges of the fluorescent dyes, which should help to quantify even low abundance proteins accurately.

Analytical Protein Microarrays

► **Protein microarrays** permit the analysis of hundreds or even thousands of addressable proteins in parallel, printed in a systematic order at high density on coated glass slides (69). To profile proteins or antibodies in crude protein samples of interest, analytical protein microarrays consisting of antibodies or protein antigens respectively, are increasingly used. In diagnostics, analytical microarrays are applied to discover new disease markers (see Clinical Proteomics).

To profile antibodies from biological samples by antigen arrays, the bound antibodies are often detected by fluorescently labelled secondary antibodies (49). Antibody arrays can be probed with protein extracts from two different cell states labelled with two different fluorescent dyes (55) or with unlabeled extracts followed by incubation with a fluorescently labelled antibody binding to a different epitope from that recognized by the capture antibody (sandwich) (37). To detect proteins present in low amounts, such as IgE or cytokines, by analytical microarrays, rolling circle amplification offers a very sensitive detection (52). Protein profiling experiments with antibody arrays depend very much on well-characterised antibodies and many technical problems have yet to be overcome for routine application (3). However, some promising results have already been obtained, such as the profiling of receptor kinases (37) or cytokines (52). We predict that antibody arrays for quantification of proteins in cell extracts will be a highly important and widely used proteomics tool in the future.

Protein Quantification by Mass Spectrometry

A combination of mass spectrometry and radioactive labelling can also be used to quantify proteins in cell extracts (38). Bacteria, for example, are grown in two different media, with the natural abundance of nitrogen isotopes or enriched in N^{15} . Protein preparations are then mixed, separated and analysed by mass spectrometry. Two versions of any peptide in the protein mixture can be detected by their number of nitrogen atoms leading to a pair of peaks from each peptide. The ratio of peak heights allows the accurate quantification of the corresponding proteins. As an alternative, the isotope-coded affinity tag (▶ICAT) method, where cysteine-containing peptides are labelled *in vitro* prior to mass spectrometry, was developed (19).

Protein Modifications

Post-translational modifications, such as phosphorylation, glycosylation and sulphation, while not apparent from genomic sequence or mRNA data, modulate the activity of many proteins and are enormously important for protein function.

The 2-DE protein separation technique often has sufficient resolution to separate the modification states of a protein directly. Some of these protein modifications can be recognized on the 2D gels, as “trains of spots”, namely groups of protein spots showing a regular spacing in their molecular mass or in their isoelectric point (pI) distributions. For example, phosphorylation changes the protein charge and is often indicated by a horizontal train of spots. Protein modifications can also be determined with the aid of mass spectrometry. The combination of function or structure based purification of modified “subproteomes”, such as phosphorylated proteins, with mass spectrometry is suitable. Stable isotope labelling strategies in combination with mass spectrometry have been applied to study the dynamics of protein modification (34).

Antibody arrays are other promising tools in this respect. Nielsen et al. (2003) used antibody microarrays to monitor the amounts and modification states of signal transduction proteins from crude cell lysates. They captured proteins of interest by specific antibodies. Phospho-specific antibodies were used to detect the phosphorylation state of the proteins.

To study protein phosphorylation *in vitro*, protein and peptide microarrays are increasingly used (11). Zhu et al. (2000) developed protein microarrays bearing microwells in which 17 different protein substrates were covalently immobilized. Nearly all of the protein kinases from *S. cerevisiae* were analyzed and many novel activities were found as well as 27 protein kinases with an unexpected tyrosine kinase activity.

To screen for potential protein substrates of barley CK2 α kinase, Kramer et al. (2004) generated protein microarrays including 800 different recombinant barley proteins and identified 21 substrates, including known as well as novel substrates of CK2 α .

Protein Localisation

A proteomics strategy of increasing importance is the systematic analysis of protein localisation in cells. A proteome scale localisation study has been performed in *S. cerevisiae* by immunolocalisation of epitope-tagged proteins (29). These experiments established the subcellular localisation of 2,744 proteins, 955 of which were of unknown function. The results obtained by this systematic analysis were integrated with previously published data and supported the thesis that there is a high correlation between protein function and localisation in yeast. Using GFP-tagged proteins, large-scale localisation studies were also performed for *S. pombe* (10) and *Drosophila* (36). Moreover, this approach was used to study the subcellular localisation of proteins in mammalian cells (54). In the future, protein localisation studies of cells will be fully automated. They will yield important insights into the subcellular organisation of cells and the interplay of proteins on a molecular level.

Protein-Protein Interactions

Besides quantity and subcellular localisation, the interactions of any given protein with other proteins provide useful insights into its function. Systematic analysis of protein-protein interactions is a key proteomics approach that will lead to a better understanding of how organisms develop, function, reproduce and age. The generation of interaction networks will also lead to the identification of new drug targets and therapeutic strategies.

Isolation and Characterisation of Protein Complexes

One attractive way of studying protein-protein interactions is to purify whole protein complexes from cell extracts by affinity chromatography (48). This can be achieved with affinity chromatography tags such as glutathione S-transferase (GST), antibodies, DNA, RNA or small molecules that specifically bind to a cellular target protein.

The human spliceosome has been purified in this way using biotinylated RNA as a “bait” (5). Its components were displayed by 2D gel electrophoresis and 19 new factors could be identified. Co-localisation studies with IF microscopy confirmed that many of the proteins were indeed associated with the spliceosome in cells. Another approach uses GST fusion proteins for the isolation of protein complexes from cell extracts. The

multi-protein complex associates with the bait, which is immobilised on beads *via* the GST tag. After washing, the complex is eluted, separated by gel electrophoresis and analysed by mass spectrometry. In a single experiment, many different components of a protein complex can be identified. Using this approach, a combination of proteins containing huntingtin, clathrin and α -adaptin was affinity purified using GST-HIP1 protein and functionally characterised by cell biological methods (59). Furthermore, epitope tagged protein, can be over-expressed in a cell. The protein is then immunoprecipitated with its interaction partners attached, by an antibody that recognises the epitope (60). This approach requires an expression clone, but no time needs to be spent on the generation of a specific antibody against the bait protein. Because full-length cDNAs for many human genes are already available, this approach will allow the systematic analysis of many unknown human protein complexes in the near future.

The analysis of protein complexes opens new lines of functional investigation. For example, a proteomics study of profilin I and II affinities using chromatography/mass spectrometry resulted in the identification of unknown signalling molecules that regulate actin polymerisation and participate in endocytosis (64). The affinity chromatography/mass spectrometry procedure was also used for systematic analysis of the entire yeast proteome (15). Tandem affinity purification (**▶TAP**) was employed to purify 589 protein complexes and determine their compositions. Comparison of yeast and human protein complexes showed conservation across species and allowed the generation of elaborate protein-protein interaction networks. Similarly, a physical and functional map of 221 molecular associations of the TNF- α /NF- κ B signalling transduction pathway was generated very recently (7).

Two-hybrid System

The **▶yeast two-hybrid** (Y2H) system has been shown to be a powerful tool for the study of protein-protein interactions (13). It is based on the molecular structure of transcription factors. Close proximity of the DNA binding domain (DBD) to the activation domain (AD) induces transcription of target genes. Proteins fused to a DBD and an AD are coexpressed in yeast. If they interact, growth (HIS3) and colour (LacZ) reporters are activated. By monitoring the growth of yeast clones on selective plates and assaying β -galactosidase activity, protein-protein interactions can be identified on a large scale. Currently, the Y2H system is used in an array based format and a library format. In the array method, yeast

clones expressing AD and DBD fusions are spotted onto selective plates and the proteins to be tested for potential interactions are screened against the entire grid by interaction mating (17). Once established, this procedure is cheap, fast and results in highly reproducible results. In the library method, one set of proteins (e.g. the AD fusions) is pooled to generate a library and the library is then screened against the DBD fusions of interest to detect protein-protein interactions. This method is relatively time consuming and requires the repeated sequencing of prey clones after each library screening experiment. Using the Y2H system, genome scale protein-protein interaction studies for *S. cerevisiae* (57), *C. elegans* (32) and *Drosophila* (16) have been performed. For the mapping of yeast interactions, both the library and array approaches were used and approximately 5,000 interactions were detected (57). The vast majority of the interactions found by Y2H screening were new. Bioinformatic analysis confirmed a certain number of the Y2H interactions on the basis of data from previous genetic and biochemical studies. However, the relevance of many others cannot be easily determined. Thus, Y2H studies allow the identification of potential interactions that have to be verified by other methods and functional assays. The main advantage of the Y2H system is that it can be performed with high throughput and in an automated manner. In the future, it will allow the generation of comprehensive interaction networks for many prokaryotic and eukaryotic organisms.

Protein Arrays and Phage Display

Protein arrays are not only a powerful tool to quantify the amounts of specific proteins in cell extracts or body fluids (see Protein Quantification section), they can also be used for global analysis of protein-protein interactions (47, 67). For interaction studies for example, 5,600 purified yeast proteins were arrayed on glass chips and screened for their ability to interact with proteins and phospholipids. Using this approach, many unknown calmodulin and phospholipid-interacting proteins were identified (67). Similarly, using a protein chip technology, pair-wise interactions among 29 human DNA replication initiation proteins were detected and characterised (47). The advantage of the protein chip approach is that a comprehensive set of individual proteins can be directly screened *in vitro* for a large variety of activities, including protein-protein, protein-DNA, protein-RNA and protein-lipid interactions, under a wide range of different conditions (69).

▶Phage display is a method where bacteriophage particles are engineered to express a peptide or protein of interest fused to a capsid or coat protein (63). It is

very efficient in screening peptide-protein as well as protein-protein interactions. Like the Y2H system, it is simple and can be performed with high throughput. Using this approach, new molecules in the epidermal growth factor (EGF)-receptor signalling pathway (70) and antigen-antibody interactions have been identified (24).

Structural Proteomics

A major challenge for post-genomic research is to assign functions to all proteins from higher eukaryotes. One of the crucial factors in the understanding of the function of uncharacterised proteins is the analysis of their experimental or modelled 3D structures. In recent years, different structural proteomics initiatives were initiated for high throughput investigation of protein structures (50). Currently, the techniques mainly used for large-scale structure determination are ►X-ray crystallography and ►nuclear magnetic resonance (NMR). Both techniques can reveal the three-dimensional structure of proteins and other biomolecules in atomic detail. X-ray crystallography uses a beam of X-rays striking a protein crystal. The diffracted (scattered) rays are detected by X-ray film. NMR spectroscopy makes use of changes in spin states of atomic nuclei that are exposed to a magnetic field. Upon absorption of electromagnetic radiation of an appropriate frequency, a transition from the lower to the upper spin state occurs. The resonance frequency of a given nucleus depends on the local chemical environment. X-ray methods give the highest resolution, but crystals are required. NMR methods, in contrast, are effective with proteins in solution, but highly concentrated solutions (~1 mM) are needed. Furthermore, NMR provides a wealth of information about dynamics. Thus NMR and X-ray techniques nicely complement each other in structure analysis. Moreover, cryo electron crystallography techniques are applied to study membrane structures. These techniques use electrons for diffraction. Since electrons are diffracted much more strongly than X-rays, very thin crystals can be studied. Accordingly, cryo electron crystallography has been applied to extremely thin, so-called two-dimensional crystals. Using this technique mainly membrane bound proteins as well as ribosomal proteins have been analysed (61).

Currently, about 16,500 structures have been deposited in public databases; however, only 300 of them are of human origin. This is due to the difficulty of human protein generation *in vitro*. Structure determination relies heavily on the high throughput production of soluble, folded proteins. Systematic expression studies have shown that only a fraction of all human proteins (20%) can be produced in soluble form in

E. coli (8). Thus, the challenge for all structural proteomics initiatives lies in the efficient production of biologically active proteins that can be used for 3D structure determination.

The knowledge of the molecule structure is a starting point for several applications, such as “►structure-based drug design”. In this iterative design process, small molecules are optimised to fit precisely into the binding site of the protein that is critical for the protein function.

Clinical Proteomics

Despite considerable advances in our understanding of the molecular basis of diseases such as cancer, substantial gaps remain in our knowledge of disease pathogenesis and the development of effective strategies for diagnosis and treatment. Proteomics approaches are used to investigate disease mechanisms and help to close these gaps. Particularly promising areas of research are altered protein expression (not only in whole cells, but also in subcellular structures), the study of protein complexes in biological fluids, the development of biomarkers for early diagnosis and the identification of new targets for therapeutic intervention.

Disease Expression Profiling

In many studies in past years, 2D gel electrophoresis was used to study protein expression in disease states. Using this approach, it has been demonstrated e.g. that leukaemias can be classified into different subtypes (22). Comparison of the protein profiles of normal human and myoepithelial breast cells allowed the detection of 170 protein spots that are differentially expressed (41). Moreover, 2D gel electrophoresis and mass spectrometry allowed the identification of proteins that are critical for the acute onset of heart disease (4).

An important recently made improvement in the 2D gel approach is the use of DIGE (see Protein Quantification section). In one study, DIGE was applied to quantify the differences in protein expression between oesophageal carcinoma and epithelial cells and a large number of proteins were found to be either up- or down-regulated in cancer cells (65). So far, mainly the most abundant proteins can be separated and analysed with 2D gels. The major challenge for future application is to reduce sample complexity prior to protein analysis. Subcellular fractionation, for example, might allow the proteome analysis of cell organelles, while protein complexes of low abundance might be enriched significantly by affinity chromatography techniques. An elegant demonstration of subproteome analysis is

the 2D gel based analysis of phagosomes, which led to the identification of over 250 proteins from this organelle (14).

One major drawback of 2D gel disease expression profiling is that it requires a relatively large amount of protein for analysis. Clinical samples especially, are generally obtainable only in limited amounts. Laser capture micro-dissection for instance, which allows defined cell types to be isolated from tissues, yields protein amounts too small to be analysed by 2D gels. Such samples need to be screened for protein expression changes using analytical protein microarrays (see Protein Quantification section).

Currently, microarray profiling studies of disease tissues are beginning to emerge. Knezevic et al. (2001) for example, analysed protein expression changes in carcinoma tissue by an antibody array approach. Using laser capture micro-dissection, specific cellular populations were studied. Differences in the expression patterns of various proteins that correlate with tumour progression were identified in epithelial cells. Most proteins were found to be involved in signal transduction pathways (27). Pawletz et al. (2001) used a reverse-phase protein array approach to quantify tissue proteins. In this, protein extracts are spotted onto microarrays and proteins of interest are quantified with antibodies. It was found that cancer progression is associated with phosphorylation of serine/threonine kinase Akt and decreased phosphorylation of extracellular signal-regulated kinase (ERK).

Another clinically relevant application of microarrays is the identification of proteins that induce an antibody response in autoimmune disorders (49). Protein arrays were incubated with patient serum and fluorescently labelled secondary antibodies were used to detect autoantibody binding to specific protein antigens in rheumatoid arthritis. Such protein arrays are powerful tools to study immune responses in a variety of diseases, including cancer.

Disease Biomarkers

There is substantial interest in applying proteomics methods to the identification of disease markers. Approaches include comparison of protein expression in healthy and diseased tissue, analysis of secreted proteins and direct serum profiling. Ostergaard et al. (1999) for example, have found the putative urinary marker psoriasin, which can be used to follow patients with bladder squamous cell carcinomas by 2D gel electrophoresis. A technology much used at the moment for serum analysis is surface enhanced laser desorption/ionisation mass spectroscopy (56). In this method, microlitre quantities of serum are applied to a protein binding surface and bound proteins/peptides

are analysed by MALDI-MS. Peptide patterns that differ between cancer patients and healthy individuals have been reported for several cancers using this method (45). Another very productive approach for the identification of cancer markers has been the analysis of serum for autoantibodies against tumour proteins. A number of antigens have been detected by screening of expression libraries with patient sera (21) or by using a random peptide-library approach (Mintz et al., 2003). Multiple proteins that induce autoantibodies have also been identified by using 2D gels to separate tumour proteins (30).

Various strategies are currently in use to identify protein complexes and protein-protein interactions (see Protein-protein Interaction section). However, these approaches have not yet been applied extensively to identify disease modulators and biomarkers. Using the two-hybrid system, a network of more than 180 interactions has recently been generated for Huntington's disease. Based on the resulting information, a potential modulator of disease pathogenesis has been identified (17). This suggests that interaction studies of diseases may lead to the identification of potential drug targets and novel biomarkers.

Proteomics and Drug Development

At present, the pharmaceutical industry is very interested in integrating modern proteomics technologies into their research programs to increase output in drug target research and lead compound development. Several studies illustrate the application of functional proteomics for identification of potential drug targets in specific signalling pathways. Lewis et al. (2000) for example, detected proteins regulated by the mitogen-activated protein kinase kinase (MKK)/ERK cascade using functional proteomics methods. Twenty-five targets were identified, only five of which had been previously characterised as MKK/ERK effectors. In another project, proteases potentially suitable for drug development have been identified in a microtitre plate-screening assay. Colorectal carcinoma biopsies were searched for protease activity and higher levels of metalloprotease expression were detected in tumours as compared to control tissue (35).

Outlook

Proteomics provides a powerful set of tools for the large-scale study of protein function. In particular, high throughput analysis of gel separated proteins by mass spectroscopy will lead to a better understanding of functional complexes and their modifications in a physiological context. We predict that proteomics in the future will move away from expression analysis

by 2D gel electrophoresis to more sensitive techniques such as protein arrays, which will allow fast and reliable quantification of small amounts of proteins in diseased and control tissues. We also suggest that systematic and hypothesis-driven protein-protein interaction, co-immunoprecipitation and co-localisation studies will be central to proteomics research in the next ten years. These studies will provide a wealth of new information concerning signalling cascades and functional protein complexes, which will have a very high impact on biological science. Because full-length cDNAs for all human proteins will be available soon, systematic proteomics studies in the coming years will focus on the human proteome. The understanding of complex physiological processes in a cell will feed into systems biology, where new hypotheses can be tested by traditional proteomics and genomics methods as well as by computational approaches.

References

- Anderson NG, Anderson NL (1996) Twenty years of two-dimensional electrophoresis: past, present and future. *Electrophoresis* 17:443–453
- Anderson NG, Matheson A, Anderson NL (2001) Back to the future: the human protein index (HPI) and the agenda for post-proteomic biology. *Proteomics* 1:3–12
- Angenendt P, Glokler J (2004) Evaluation of antibodies and microarray coatings as a prerequisite for the generation of optimized antibody microarrays. *Methods Mol Biol* 264:123–134
- Arnott D, O'Connell KL, King KL et al (1998) An integrated approach to proteome analysis: identification of proteins associated with cardiac hypertrophy. *Anal Biochem* 258:1–18
- Barabino SM, Sproat BS, Ryder U et al (1989) Mapping U2 snRNP-pre-mRNA interactions using biotinylated oligonucleotides made of 2'-OMe RNA. *EMBO J* 8:4171–4178
- Berndt P, Hobohm U, Langen H (1999) Reliable automatic protein identification from matrix-assisted laser desorption/ionization mass spectrometric peptide fingerprints. *Electrophoresis* 20:3521–3526
- Bouwmeester T, Bauch A, Ruffner H et al (2004) A physical and functional map of the human TNF- α /NF- κ B signal transduction pathway. *Nat Cell Biol* 6:97–105
- Bussow K, Quedenau C, Sievert V et al (2004) A catalog of human cDNA expression clones and its application to structural genomics. *Genome Biol* 5:R71
- Claverie JM (1997) Computational methods for the identification of genes in vertebrate genomic sequences. *Human Mol Genet* 6:1735–1744
- Ding DQ, Tomita Y, Yamamoto A et al (2000) Large-scale screening of intracellular protein localization in living fission yeast cells by the use of a GFP-fusion genomic DNA library. *Genes Cells* 5:169–190
- Feilner T, Kreutzberger J, Niemann B et al (2004) Proteomic studies using microarrays. *Curr Proteomics* 1:283–295
- Fenn JB, Mann M, Meng CK et al (1989) Electrospray ionization for mass spectrometry of large biomolecules. *Science* 246:64–71
- Fields S, Song O (1989) A novel genetic system to detect protein-protein interactions. *Nature* 340:245–246
- Gagnon E, Duclos S, Rondeau C et al (2002) Endoplasmic reticulum-mediated phagocytosis is a mechanism of entry into macrophages. *Cell* 110:119–131
- Gavin AC, Bosche M, Krause R et al (2002) Functional organization of the yeast proteome by systematic analysis of protein complexes. *Nature* 415:141–147
- Giot L, Bader JS, Brouwer C et al (2003) A protein interaction map of *Drosophila melanogaster*. *Science* 302:1727–1736
- Goehler H, Lalowski M, Stelzl U et al (2004) A Protein Interaction Network Links GIT1, an Enhancer of Huntingtin Aggregation, to Huntington's Disease. *Mol Cell* 15:853–865
- Gonczy P, Echeverri C, Oegema K et al (2000) Functional genomic analysis of cell division in *C. elegans* using RNAi of genes on chromosome III. *Nature* 408:331–336
- Gygi SP, Rist B, Gerber SA et al (1999a) Quantitative analysis of complex protein mixtures using isotope-coded affinity tags. *Nat Biotechnol* 17:994–999
- Gygi SP, Rochon Y, Franz BR et al (1999b) Correlation between protein and mRNA abundance in yeast. *Mol Cell Biol* 19:1720–1730
- Hanash S (2003) Disease proteomics. *Nature* 422:226–232
- Hanash SM, Madoz-Gurpide J, Misek DE (2002) Identification of novel targets for cancer therapy using expression proteomics. *Leukemia* 16:478–485
- Henzel WJ, Billeci TM, Stults JT et al (1993) Identifying proteins from two-dimensional gels by molecular mass searching of peptide fragments in protein sequence databases. *Proc Natl Acad Sci USA* 90:5011–5015
- Huften SE, Moerkert PT, Meulemans EV et al (1999) Phage display of cDNA repertoires: the pVI display system and its applications for the selection of immunogenic ligands. *J Immunol Methods* 231:39–51
- Klose J, Kobalz U (1995) Two-dimensional electrophoresis of proteins: an updated protocol and implications for a functional analysis of the genome. *Electrophoresis* 16:1034–1059
- Klose J, Nock C, Herrmann M et al (2002) Genetic analysis of the mouse brain proteome. *Nat Genet* 30:385–393
- Knezevic V, Leethanakul C, Bichsel VE et al (2001) Proteomic profiling of the cancer microenvironment by antibody arrays. *Proteomics* 1:1271–1278
- Kramer A, Feilner T, Possling A et al (2004) Identification of barley CK2 α targets by using the protein microarray technology. *Phytochemistry* 65:1777–1784
- Kumar A, Agarwal S, Heyman JA et al (2002) Subcellular localization of the yeast proteome. *Genes Dev* 16:707–719
- Le Naour F (2001) Contribution of proteomics to tumor immunology. *Proteomics* 1:1295–1302

31. Lewis TS, Hunt JB, Aveline LD et al (2000) Identification of novel MAP kinase pathway signaling targets by functional proteomics and mass spectrometry. *Mol Cell* 6:1343–1354
32. Li S, Armstrong CM, Bertin N et al (2004) A map of the interactome network of the metazoan *C. elegans*. *Science* 303:540–543
33. Link AJ, Eng J, Schieltz DM et al (1999) Direct analysis of protein complexes using mass spectrometry. *Nat Biotechnol* 17:676–682
34. Mann M, Jensen ON (2003) Proteomic analysis of post-translational modifications. *Nat Biotechnol* 21:255–261
35. McKerrow JH, Bhargava V, Hansell E et al (2000) A functional proteomics screen of proteases in colorectal carcinoma. *Mol Med* 6:450–460
36. Morin X, Daneman R, Zavortink M et al (2001) A protein trap strategy to detect GFP-tagged proteins expressed from their endogenous loci in *Drosophila*. *Proc Natl Acad Sci USA* 98:15050–15055
37. Nielsen UB, Cardone MH, Sinskey AJ et al (2003) Profiling receptor tyrosine kinase activation by using Ab microarrays. *Proc Natl Acad Sci USA* 100:9330–9335
38. Oda Y, Huang K, Cross FR et al (1999) Accurate quantitation of protein expression and site-specific phosphorylation. *Proc Natl Acad Sci USA* 96:6591–6596
39. O'Farrell PH (1975) High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* 250:4007–4021
40. Ostergaard M, Wolf H, Orntoft TF et al (1999) Psoriasin (S100A7): a putative urinary marker for the follow-up of patients with bladder squamous cell carcinomas. *Electrophoresis* 20:349–354
41. Page MJ, Amess B, Townsend RR et al (1999) Proteomic definition of normal human luminal and myoepithelial breast cells purified from reduction mammoplasties. *Proc Natl Acad Sci USA* 96:12589–12594
42. Pandey A, Liewitter F (1999) Nucleotide sequence databases: a gold mine for biologists. *Trends Biochem Sci* 24:276–280
43. Pandey A, Mann M (2000) Proteomics to study genes and genomes. *Nature* 405:837–846
44. Paweletz CP, Charboneau L, Bichsel VE et al (2001) Reverse phase protein microarrays which capture disease progression show activation of pro-survival pathways at the cancer invasion front. *Oncogene* 20:1981–1989
45. Petricoin EF, Liotta LA (2004) SELDI-TOF-based serum proteomic pattern diagnostics for early detection of cancer. *Curr Opin Biotechnol* 15:24–30
46. Phizicky E, Bastiaens PI, Zhu H et al (2003) Protein analysis on a proteomic scale. *Nature* 422:208–215
47. Ramachandran N, Hainsworth E, Bhullar B et al (2004) Self-assembling protein microarrays. *Science* 305:86–90
48. Rigaut G, Shevchenko A, Rutz B et al (1999) A generic protein purification method for protein complex characterization and proteome exploration. *Nat Biotechnol* 17:1030–1032
49. Robinson WH, DiGennaro C, Hueber W et al (2002) Autoantigen microarrays for multiplex characterization of autoantibody responses. *Nature Medicine* 8:295–301
50. Sali A, Glaeser R, Earnest T et al (2003) From words to literature in structural proteomics. *Nature* 422:216–225
51. Schindewolf C, Lobenwein K, Trinczek K et al (2000) Comet assay as a tool to screen for mouse models with inherited radiation sensitivity. *Mamm Genome* 11:552–554
52. Schweitzer B, Roberts S, Grimwade B et al (2002) Multiplexed protein profiling on microarrays by rolling-circle amplification. *Nature Biotechnology* 20:359–365
53. Shevchenko A, Loboda A, Ens W et al (2000) MALDI quadrupole time-of-flight mass spectrometry: a powerful tool for proteomic research. *Anal Chem* 72:2132–2141
54. Simpson JC, Wellenreuther R, Poustka A et al (2000) Systematic subcellular localization of novel proteins identified by large-scale cDNA sequencing. *EMBO Rep* 1:287–292
55. Sreekumar A, Nyati MK, Varambally S et al (2001) Profiling of cancer cells using protein microarrays: discovery of novel radiation-regulated proteins. *Cancer Res* 61:7585–7593
56. Tang N, Tornatore P, Weinberger SR (2004) Current developments in SELDI affinity technology. *Mass Spectrom Rev* 23:34–44
57. Uetz P, Giot L, Cagney G et al (2000) A comprehensive analysis of protein-protein interactions in *Saccharomyces cerevisiae* [comment]. *Nature* 403:623–627
58. Unlu M, Morgan ME, Minden JS (1997) Difference gel electrophoresis: a single gel method for detecting changes in protein extracts. *Electrophoresis* 18:2071–2077
59. Waelter S, Scherzinger E, Hasenbank R et al (2001) The huntingtin interacting protein HIP1 is a clathrin and alpha-adaptin-binding protein involved in receptor-mediated endocytosis. *Hum Mol Genet* 10:1807–1817
60. Wen YD, Cress WD, Roy AL et al (2003) Histone deacetylase 3 binds to and regulates the multifunctional transcription factor TFII-I. *J Biol Chem* 278:1841–1847
61. Werten PJJ, Rémy H-W, de Groot BL et al (2002) Progress in the analysis of membrane protein structure and function. *FEBS Lett* 529:65–72
62. Wilkins MR, Sanchez JC, Gooley AA et al (1995) Progress with proteome projects: why all proteins expressed by a genome should be identified and how to do it [review]. *Biotechnol Genet Eng* 13:19–50
63. Winter G, Griffiths AD, Hawkins RE et al (1994) Making antibodies by phage display technology. *Annu Rev Immunol* 12:433–455
64. Witke W, Podtelejnikov AV, Di Nardo A et al (1998) In mouse brain profilin I and profilin II associate with regulators of the endocytic pathway and actin assembly. *EMBO J* 17:967–976
65. Zhou G, Li H, DeCamp D et al (2002) 2D differential in-gel electrophoresis for the identification of esophageal scans cell cancer-specific protein markers. *Mol Cell Proteomics* 1:117–124
66. Zhou H, Bouwman K, Schotanus M et al (2004) Two-color, rolling-circle amplification on antibody microarrays for sensitive, multiplexed serum-protein measurements. *Genome Biol* 5:R28
67. Zhu H, Bilgin M, Bangham R et al (2001) Global analysis of protein activities using proteome chips. *Science* 293:2101–2105

68. Zhu H, Klemic JF, Chang S et al (2000) Analysis of yeast protein kinases using protein chips. *Nat Genet* 26:283–289
69. Zhu H, Snyder M (2003) Protein chip technology. *Curr Opin Chem Biol* 7:55–63
70. Zozulya S, Lioubin M, Hill RJ et al (1999) Mapping signal transduction pathways by phage display. *Nat Biotechnol* 17:1193–1198

the ratio between the rates of translation and degradation of specific proteins changes, modifying the half-life and/or the total amount of these proteins. Proteins with increased half-life become prone to post-translational modifications such as glycation, oxidation or phosphorylation; cell-signaling cascades are regulated through phosphorylation/dephosphorylation. Age-induced modifications will change the ratios between the various isoforms of a protein and reshape the network of interactions between signaling proteins, affecting gene expression in turn. Proteomics in ageing allows the detection and quantification of various proteins isoforms, based on discrimination by 2DGE and/or MS.

Characteristics

Proteomics and *In Vitro* ▶ Replicative Senescence

Human diploid fibroblasts (HDFs) divide a finite number of times in culture, counted in terms of cumulative population doublings (CPDs). Replicative senescence (RS) is due to telomere attrition, which can be accelerated by mild ▶oxidative stress. The occurrence of RS has been demonstrated for many cell types excepting embryonic germ cells and the majority of tumor-derived cells. The transcription level of many genes is altered in RS. However there is generally poor correlation between the transcription level and the abundance of the corresponding protein. Several proteomics studies compared extracts of proteins of HDFs at early, intermediate and late CPDs. Extracts of total soluble cytoplasmic, nuclear, membrane-bound and secreted proteins of relatively pure clones of morphotypes were obtained. Spots on 2DGE were found which were specific to each of these clones of morphotypes. The different age-related HDFs have been classified morphologically into 7 types called “morphotypes”. 2DGE gels were also used successfully to compare protein expression of HDFs at increasing CPDs taken from patients with ▶Werner syndrome, an autosomal recessive disease resembling accelerated ageing and from age-matched control patients (1).

Recent proteomics identified many new biomarkers in WI-38 fetal lung HDFs and rat embryo fibroblasts. For instance, the glucose-6 phosphate dehydrogenase level is decreased by 40% while the pyruvate kinase (PK) level is increased by 50% in HDFs, suggesting decreased activity of the pentose phosphate pathway and increased glycolysis. This impinges upon various

cellular functions like ATP production, DNA duplication and antioxidant defense by the glutathione reductase/glutathione oxidase system (2).

Proteomics and Stress-Induced Premature Senescence

Many normal human proliferative cell types (fibroblasts, ▶melanocytes, endothelial cells, retinal pigment epithelial cells, etc.) exposed to stress at subcytotoxic level (UVB, H₂O₂, ethanol, hyperoxia, γ irradiations, homocysteine, hydroxyurea, *tert*-butylhydroperoxide (*t*-BHP), etc.) undergo stress-induced premature senescence (SIPS) (2). SIPS covers the effects observed in proliferative cell types from 48–72 h after subcytotoxic exposure to stress, including irreversible growth arrest. An acceleration of the transition through the sequence of the morphotypes has been observed after exposure of HDFs to repeated subcytotoxic stress with UV, mitomycin C or electromagnetic fields. 2DGE run after isolation of relatively pure clones of HDFs of a given morphotype after repeated stress displayed expression changes similar to those found in RS. However further proteomic comparison indicated that, at the level of protein expression, SIPS and RS are different phenotypes sharing similarities. Thirty proteins with changes of expression level specific or common to SIPS induced by *t*-BHP or ethanol and/or RS were identified. These proteins are involved in energy metabolism, defense systems, maintenance of the redox potential, cell morphology and transduction pathways. In particular, PK was also over-expressed in SIPS. Several isoforms of the antioxidant enzyme peroxiredoxin VI displayed differential expression levels in RS and *t*-BHP-induced SIPS when compared to cells at early CPDs. There were also proteins involved in nucleus structure (1, 2). 5-bromodeoxyuridine (BrdU) and ▶5-chlorodeoxyuridine (CldU) are DNA intercalating agents that induce a senescence-like phenotype, probably due to a change in nuclear matrix structure. Nuclear matrix proteins of HeLa cells cultured with CldU were separated by 2DGE and identified by MS. Since these alterations were observed within 24 h after addition of CldU, these proteins may be involved in an early step of the senescence-like phenotype.

Proteomics and Other Biogerontological Systems

Protein profiling of the human epidermis from the elderly reveals up-regulation of a signature of Interferon-gamma-induced polypeptides that include

manganese superoxide dismutase and the p85 beta subunit of phosphatidylinositol 3-kinase.

Mouse brain proteins were isolated from five regions (cerebellum, cerebral cortex, hippocampus, striatum, cervical spinal cord) at five ages from week 10 to month 24 and separated by 2DGE. Seventeen proteins were differentially expressed in the course of ageing. A mouse brain proteome database was constructed. Another study identified modified crystallins associated with ageing of lens and produced 2DGE proteome maps of crystallins in mouse lens.

Post-translational Modifications of Proteins and Ageing

2DGE can reveal post-translational modifications inducing obvious changes in the molecular weight and/or the isoelectric point of the proteins. Specific modifications can be searched for on 2DGE with specific antibodies, staining or labeling methods. MS allows identifications of amino acids with any modification changing the molecular weight of a peptide such as phosphorylation, glycosylation or oxidation. Phosphoproteins can be enriched by immunoaffinity methods utilizing antibodies recognizing phosphoserine, phosphothreonine or phosphotyrosine (1, 3).

To investigate the occurrence of proteasome structural and functional age-related alterations, 26S proteasome was purified from peripheral blood lymphocytes of 20–63 year-old donors. An age-related decline of 26S proteasome-specific activity was found to be associated with an increased yield of post-translational modifications of proteasome subunits, while proteasome content and subunit composition were unchanged. In particular, some catalytic and assembly subunits of the 20S proteasome were preferentially modified with age by glycosylation, conjugation with a lipid peroxidation product and/or ubiquitination (4).

► ‘Drusen’ are extracellular deposits that accumulate below the retinal pigment epithelium on Bruch’s membrane. Drusen are risk factors for developing age-related macular degeneration (AMD). Liquid chromatography tandem MS analyses of drusen preparations from 18 normal donors and five AMD donors identified 129 proteins. Immunocytochemical studies localized approximately 16% of these proteins in drusen. Oxidative protein modifications were observed in drusen. This supported the hypothesis that oxidative injury contributes to the pathogenesis of AMD and suggested that oxidative protein modifications may have a critical role in drusen formation (reviewed in 3).

Clinical Relevance

Proteomics on *Ex Vivo* Cultivated HDFs

A recent study compared the proteome of *in vitro* cultured dermal fibroblasts from healthy subjects of different ages. The study identified 38 proteins

exhibiting more than three-fold reproducible variations with ageing. This investigation highlighted the role of synthetic and degradative pathways in modulating the whole cell machinery and emphasized that metabolic impairment with age could depend partly on different expression of a number of genes leading to an imbalance among functional proteins.

Molecular Scars and SIPS *In Vivo*

Molecular scars are stress-induced changes in the expression level of specific proteins. They appear several days after stress exposure to stimuli such as chemicals, UV, inflammation, ischaemia-reperfusion or ethanol abuse. The analysis of molecular scars in such conditions could help understand their pathophysiological outcome.

Cells displaying cellular characteristics of senescence are found in atherosclerotic lesions of the coronary arteries from patients who died from ischemic heart diseases, in legs of patients with venous insufficiencies, in the inflammatory context of osteoarthritis, in intestinal metaplasia from the stomach, in arteries subjected to angioplasty, in atrophic nephrons of kidneys of aged rats, in tissue surrounding liver carcinomas and in prostatic hyperplasia (2).

Future of Proteomics in Experimental Gerontology

Proteomics could be extremely valuable in gerontopharmacology and geronto-toxicology especially since the therapeutic efficiency and the possible side effects of drugs might differ greatly with age. Chemical agents can also have long-term effects on gene expression (molecular scars). Proteomics can help to evaluate how the tissues react to a particular drug with increasing age and how to match treatments with individual age classes (1).

Alteration of the mitochondrial proteome and altered mitochondrial function are involved in a variety of degenerative diseases, in heart disease, in ageing and in cancer. Based upon the human genome there are approximately 1,000–2,000 proteins constituting the mitochondrial proteome. Only about 600 mitochondrial proteins have been identified and characterized at the molecular level.

Alternative methods to 2DGE are now being employed, including different MS-based approaches following both one-dimensional SDS-PAGE and gel-free approaches, blue native gel electrophoresis (BN-PAGE), proteome simplification by subcellular fractionation and affinity chromatography (1, 3).

The findings arising from such studies will be stored and processed in gerontology-related bioinformatic databases such as the human genome and proteome resources (5).

References

1. Dierick JF, Dieu M, Remacle J et al (2002) Proteomics in experimental gerontology. *Exp Gerontol* 37:721–734
2. Toussaint O et al (2003) Role of subcytotoxic stress in tissue ageing. Biomarkers of senescence, signal transduction and proteomics in stress-induced premature senescence. In: Nystrom T, Osiewicz HD (eds) *Topics in Current Genetics*, vol 3, Springer Verlag, Berlin Heidelberg New York, pp 269–294
3. Schöneich C (2003) Proteomics in gerontological research. *Exp Gerontol* 38:473–481
4. Carrard G, Dieu M, Raes M et al (2003) Impact of aging on proteasome structure and function in human lymphocytes. *Int J Biochem Cell Biol* 35:728–739
5. de Magalhaes JP, Costa J, Toussaint O (2005) HAGR: the Human Ageing Genomic Resources. *Nucleic Acids Res* 33:537–543

Proteomics in Cancer

FRANÇOIS LE NAOUR
INSERM U602, Hospital Paul Brousse, Villejuif,
France
flenaour@vjf.inserm.fr

Definition

The ►**proteome** designates all the proteins expressed by a single cell at a particular moment of its life. Therefore, the proteome is more complex and dynamic than the genome. Indeed, the proteome is characterized by the expression level of the proteins as well as by their partitioning, their post-translational modifications, their molecular associations and by the regulation of their activity. The study of the proteome is called proteomics, which involves a variety of sophisticated techniques and approaches. Proteomics allows the global analysis of the proteome during physiological or pathological processes.

Cancer is a somatic genetic disease. Tumorigenesis is a multistep process involving mutations and other genetic events leading to gain of function of dominantly acting ►**proto-oncogenes** and to the loss of function of ►**tumor suppressor genes**. The modified genes have an impact in the expression of genes and proteins, thus leading to disorders in terms of cell growth and differentiation.

Characteristics

The recent intense interest in proteomics has been made possible by the sequencing of the human genome, giving access to the sequences of all the genes, as well as by the recent advances in ►**mass spectrometry** for the analysis of macromolecules (techniques for ionization and increased resolution and sensitivity). In cancer

research, proteomics provides novel opportunities to understand the tumoral process as well as to define novel markers and therapeutic targets.

Proteomics: the State of the Art

The proteome can be analyzed by ►**two-dimensional gel electrophoresis** (2DE) allowing the separation of the proteins in two dimensions. The proteins are separated according to their charge in a first dimension where they stop their migration after reaching their isoelectric point and then in a second dimension according to their molecular mass. The proteins can be visualized by different techniques of staining such as Coomassie blue or silver, and then digitalized. 2D electrophoresis is a highly resolutive technique leading to the visualization of 1000–2000 protein spots in one experiment. Therefore, 2DE allows comparative analysis of protein expression levels as well as of protein isoforms. A recent development in 2D electrophoresis has been introduced by differential labeling of proteins using dyes also called ►**DIGE** (difference gel electrophoresis). This system allows comparative and quantitative studies after mixing of two protein extracts in the same gel, which improves the sensitivity and the reproducibility of the 2D electrophoresis (1). The identification of the proteins can be performed after enzymatic digestion of the proteins in the gel followed by analysis of the resulting peptides using mass spectrometry. The mass spectrometry analysis can be performed by ►**MALDI-T** of providing a ►**peptide mass map**, which is used for database searching to identify the protein. Another type of mass spectrometry, called ►**MS/MS**, allows the identification of a protein from a single peptide after fragmenting that peptide and using the second spectrum (MS/MS or MS²) for determining its sequence or a sequence tag. This information is used for database searching. MS/MS is therefore a powerful tool for analyzing the composition of complex mixtures as well as for characterizing post-translational modifications. The most recent and spectacular advances in proteomics have been achieved by coupling ►**liquid chromatography** with mass spectrometry (LC-MS/MS). A major advantage of LC-MS/MS is that it allows the identification of basic or highly hydrophobic proteins as well as protein having extreme molecular weights, which are difficult to resolve in 2D electrophoresis. The sequential use of several types of liquid chromatography (ion exchange, reverse phase...), also called ►**multidimensional chromatography**, allows one to separate and identify complex mixtures containing thousands of peptides in experiment. In addition, quantitative analysis can be performed using LC-MS/MS after differential isotopic labeling of the proteins from two cell extracts. The advantage of this approach is that it simultaneously quantifies and identifies the regulated proteins. One of

these methods is the ►ICAT system (isotope-coded affinity tag) (2). Furthermore, the chemical derivatization of some post-translational modifications (i.e. phosphorylation, glycosylation) combined with isotopic labeling has led to the identification of peptides exhibiting these modifications in complex mixtures and their relative quantification.

Proteomics and Cancer

The first level of proteomic analysis of cancer is global protein profiling and proteome mining. Since 2DE has appeared, the technology has been applied to cancer research. This has led to comparative studies of tumors with healthy tissues. Many cancer types were investigated by 2DE leading to the characterization of novel oncogenes or proteins with clinical interest (3, 4). For example, it was demonstrated using 2DE that leukemias could be classified into their different subtypes (1). Proteomic analysis has also been combined with laser capture microdissection prior to 2DE analysis. Indeed, clinical tissues are often heterogeneous and tumor cells are often contaminated with normal stromal cells, infiltrating lymphocytes, fibroblasts and blood vessel cells. Laser capture microdissection was developed for separating tumor cells from their microenvironment. In the search for novel tumor markers, a proteomic-based approach, using 2D Western blots of sera from cancer patients, has enabled the identification of tumor-associated antigens that elicit a humoral response (1, 3, 4, 5). This revealed a high frequency of autoantibodies in sera from several types of cancer. Some of these antigenic proteins may be novel cancer markers and may have clinical applications in diagnosis as well as in determining prognosis.

Proteomics provides the opportunity to focus on cell compartments or subproteomes such as the phosphorylated proteins or the cell surface proteome. Indeed, the proteomic analysis of cell compartments has modified our knowledge of cell compartments by identifying many proteins that were not expected to be there (e.g. phagosome, mitochondria) (1). A relevant example of proteomic analysis of a subproteome is the recent study of the cell surface proteome. Indeed, there is no method so far for performing a systematic characterization of cell surface proteins, whereas they play a major role in interactions and communication of cells with their environment. In addition, cell surface proteins are low in abundance making their identification very challenging. Recent advances in proteomics combined selective chemical labeling of the cell surface proteins using biotin with specific affinity purification (based on the very high affinity of avidin for biotin) allowing the profiling of the cell surface proteome and characterization using mass spectrometry. Applied to cancer cells, this method led to the identification of many chaperone proteins expressed at the cell surface (1). This

intriguing observation raises the possibility that many proteins expressed at the cell surface are not characterized yet and that novel tumor markers might be discovered in this way.

Quantitative proteomic analysis using LC-MS/MS is still in its initial stages. However, the approaches that avoid gel separation open novel possibilities in proteomics. A recent study has focused on the modulated protein in terms of expression after transfection of the oncogene c-myc using the ICAT system followed by LC-MS/MS analysis. The approach allowed the identification, from a whole proteome, of many proteins up- and down-regulated which might be relevant to cancer, such as proteins involved in adhesion or cell growth (6). The approaches combining isotope labeling of proteins with LC-MS/MS are very promising in cancer research because they allow the identification and quantification of low abundance proteins in complex mixtures.

Future Directions

Proteomics is now moving on to the analysis of whole proteins in complex mixtures. ►Protein chips (or microarrays) are emerging as an alternative to 2DE for differential expression analysis at the protein level in oncology (7). For example, profiling of mixtures is now possible using ►SELDI-ToF mass spectrometry. SELDI-ToF (Surface-enhanced laser desorption/ionization time-of-flight) analysis enables the high throughput characterization of lysates from a very few tumor cells and may be best suited for clinical biomarker studies. Another emerging device in proteomics is ►FT-ICR mass spectrometry (ultra high resolution). FT-ICR mass spectrometry leads to obtaining precise mass measurements of proteins and information on their modifications. This approach, also called 'top down', is complementary to the analysis of peptides, called 'bottom up'. This technology is also suitable for characterizing post-translational modification by a method of dissociation (►ECD, electron capture dissociation) specific to this type of mass spectrometry. Finally, proteomics offers an unprecedented opportunity to understand protein regulatory networks. The proteins are interacting with partners and form a complicated network in the cell. A change in one node of the network can lead to changes in proteins that may not be physically associated with the initiating protein. The understanding of protein networks will help to understand pathological processes such as cancer and to determine the best therapeutic targets.

Clinical Relevance

Cancer induces disorders in term of expression of genes and proteins. In cells, the mRNA molecules are

translated into proteins that carry out the basic biological functions. The tumorigenesis process can be addressed at the mRNA level and at the protein level. Because there is no strictly linear relationship between the mRNA and the proteins in terms of the expression level, both approaches may give complementary information. Proteomics is suitable for the characterization of cell compartments and subproteomes (i.e. cell surface proteome) that might enable us to define novel markers. The characterization of post-translational modifications of proteins induced by cancer might be a promising field. The understanding of molecular networks will help to define novel targets. Finally, proteomic approaches will play an important role in the discovery of new drugs and maybe in the development of personalized medicines.

References

1. Hanash S (2003) Disease proteomics. *Nature* 422:226–232
2. Patterson SD, Aebersold RH (2003) Proteomics: the first decade and beyond. *Nat Genet* 33:311–23
3. Cancer Proteomics (2001) Proteomics, vol 1, N°10. Special issue.
4. Oncoproteomics (2002) *Technol Cancer Res Treat* 1:219–316. Special issue.
5. Wulfkühle J, Liotta LA, Petricoin EF (2003) Proteomics applications for early detection of cancer. *Nat Rev Cancer* 3:267–275
6. Shiio Y, Donohoe S, Yi EC et al (2002) Quantitative proteomic analysis of Myc oncoprotein function. *EMBO J* 21:5088–96
7. Protein microarrays (2003) Proteomics, vol 3, N°11. Special issue.

Proteomics in Cardiovascular Disease

RODNEY LUI, SEAN LAL,
CRISTOBAL G. DOS REMEDIOS
Muscle Research Unit, Institute for Biomedical
Research, University of Sydney, Sydney, Australia
crisdos@anatomy.usyd.edu.au

Definitions

► **Cardiovascular disease** (CVD) encompasses any dysfunction of the heart and its own or other blood vessels. This may progress to damage life-sustaining organs like the brain and the heart itself. The aetiology of many CVDs is not known. A number of reports have focused on the development of molecular screening tools that detect changes in gene expression or gene sequences (genomics). Only within the past decade has

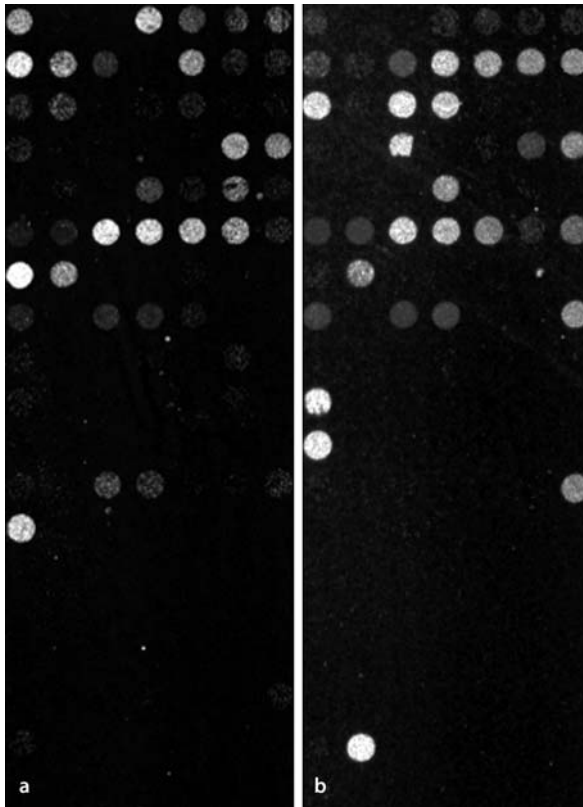
► **proteomics** emerged as a practical alternative to genomics. Proteomics is defined as the study of proteins expressed in a particular condition at a particular time by a particular cell or tissue type. Proteins are involved in virtually every cellular and regulatory function and changes in either the amount or specific type of protein can lead to the development of CVD.

Characteristics

Organs express only a limited subset ($\leq 10^4$) of total possible proteins ($\geq 10^6$) encoded in the approximately 30,000 genes in the genome. ► **Proteome profiling** and ► **functional proteomics** are designed to provide molecular snapshots of the proteins that determine the structure and function of particular organs. The expectation is that this information will identify protein differences between diseased and non-diseased tissues. Proteomic studies have so far focused on the two common types of CVD, ischaemic heart disease (IHD) and dilated cardiomyopathy (DCM). These two cause more than 70% of cases of human heart failure. IHD underlies most of them.

Ischaemic Heart Disease

The primary cause is atherosclerosis, a disease of the cells that line the inner wall of arteries (endothelial cells), particularly the coronary arteries of the heart. Recently, IHD has been regarded as an inflammatory disorder (► **inflammation**) mediated by a dysfunctional immune system (1). Accordingly, proteomic research has focussed on proteins and smaller peptides associated with the immune system. This began by demonstrating elevated levels of small numbers of membrane pro-inflammatory receptors and adhesion molecules found in the membranes of ► **endothelial cells** and ► **leukocytes**. These receptors orchestrate the ► **inflammatory response** and the development of atherosclerotic plaques. Traditionally, these surface proteins were detected by immunohistochemistry or flow cytometry (FACS) using monoclonal antibodies directed against ► **cluster of differentiation** (CD) antigens on leukocytes. Expression levels of certain CD antigens may change in a way that indicates the progress of IHD. For example, there are changing levels of circulating CD11b (an adhesion molecule found on monocytes and neutrophils) and CD36 (a scavenger receptor with a high specificity for oxidised ► **LDLs**, which are the major lipid component of atherosclerotic plaques) expressed on leukocytes in IHD (2). This may be significant because leukocytes are the mediators of inflammatory responses. Pathophysiology is further complicated by the release into the circulation of cytokines (plasma proteins/peptides secreted by inflammatory cells) that activate other inflammatory cascade proteins. All of this has large effects on the cardiac tissue



Proteomics in Cardiovascular Disease. Figure 1 Dark-field images of antibody arrays that captures proteins that are expressed on the cell surface of peripheral blood leukocytes. Images illustrate the differences in both the pattern and levels of proteins expressed between (a) a non-diseased blood donor and (b) an IHD patient.

surrounding the microvasculature (3). ►Cytokines also have systemic effects by recruiting more inflammatory cells and pro-inflammatory circulating proteins from lymphatic tissues.

Dilated Cardiomyopathy

Several proteomic investigations of human heart failure have focused on DCM. In most cases the ►etiology of DCM is unknown (idiopathic DCM). Proteome analyses have largely been achieved by ►two-dimensional polyacrylamide gel electrophoresis (2-DE). More recently 2-DE has been combined with ►mass spectrometry (MS) to identify proteins based on peptide ‘fingerprints’ (resulting from the exhaustive proteolysis of isolated 2-DE protein spots), or by limited sequence analyses of proteins/peptides (MS-MS). These data have been incorporated into CVD proteome databases of which the major ones are: ►www.harefield.nthaus.nhs.uk/nhli/protein; ►userpage.chemie.fu-berlin.de/~pleiss/dhzb.html and ►www.mdc-berlin.de/~emu/heart. These

have identified as many as 100 cardiac-specific proteins from whole cardiac tissue lysates containing up to 10,000 protein spots. Significant numbers of these proteins are now known to change in the terminal stages of the disease process. However the number of mRNA changes reported from gene arrays (see below) exceeds the number detected on the protein level (4). Most of these have been categorized into three groups:

1. Defective cytoskeletal and myofibrillar proteins that are consistent with contractile dysfunctions in ►congestive heart failure (5).
2. Mitochondria-associated and energy-producing proteins consistent with impaired energetics of the myocardium.
3. Proteins associated with cellular stress responses consistent with differences in expression levels of proteins such the 27 kD ►heat shock protein (hsp27).

Until recently, most cardiac proteomic studies have been based on 2-DE separations. While this technique continues to improve, so far it has not fulfilled its promise to identify and quantify differences in protein expression between diseased and control hearts. This can be attributed to several factors including difficulties in solubilizing proteins prior to separation, failure to reproducibly detect low-abundance proteins, difficulties in separating proteins with high or low isoelectric points (pIs), and generally a limited ability of 2-DE to quantify protein expression levels. While some of these shortcomings can be addressed by the addition of MS, 2-DE is rapidly being overtaken by protein microarrays currently under development.

Protein microarrays superficially resemble the better-known gene arrays carrying oligonucleotides or cDNA fragments, except that they are based on proteins, but there are major differences. Nucleotide hybridisation relies on precise base pairing and unique probes for known genes are relatively easy to synthesize. By contrast, there are more than 20 amino acids and they lack the ability to pair precisely with synthesized probes. Protein arrays rely on the formation of high-affinity and highly specific complexes with immobilized ligands. These may – but need not – be antibodies. Furthermore, proteins are much less stable than nucleotides. In principle, proteins can be complexed to non-protein ligands including lipids or nucleic acids such as aptamers (►<http://pubs.acs.org/subscribe/journals/mdd/v05/i05/html/05ttb.html>).

Clinical Relevance

In most Western communities, the incidence of CVD continues to increase. Although the application of proteomics is still in its infancy, it is highly likely to lead to clinical therapies (6). We predict that proteomics

will not only accurately and specifically identify cardiovascular disease, but will also help to find early markers of disease and therefore identify CVD at a stage when it is eminently treatable. In addition to its diagnostic and prognostic potential, it may also identify possible drug targets and thus play a role in the development of new therapies. Given the huge range of responses to drugs in the population, we envisage that protein arrays will be used to identify and predict individual responses to drug therapies. Companies are currently developing protein arrays that will use complex diagnostic information to produce treatments that are capable of being personalized for each patient.

References

1. Libby P, Ridker PM, Maseri A (2002) Inflammation and atherosclerosis. *Circulation* 105:1135–1143
2. Febbraio M, Hajjar DP, Silverstein RL (2001) CD36: a class B scavenger receptor involved in angiogenesis, atherosclerosis, inflammation, and lipid metabolism. *J Clin Invest* 108:785–791
3. Niebauer J (2000) Inflammatory mediators in heart failure. *Int J Cardiol* 72:209–213
4. dos Remedios CG, Liew CC, Allen PD et al (2003) Genomics, proteomics and bioinformatics of human heart failure. *J Muscle Res Cell Motil* 24:251–260
5. Alpert NR, Mulieri LA, Warshaw D (2002) The failing human heart. *Cardiovasc Res* 54:1–10
6. Loscalzo J (2003) Proteomics in cardiovascular biology and medicine. *Circulation* 108:380–383

Proteomics in Human-Pathogen Interactions

HENGLI TANG
Department of Biological Science, Florida State
University, Tallahassee, FL, USA
tang@bio.fsu.edu

Definition

Proteomics of human-pathogen interactions investigates alterations in protein content and characteristics that happen during or after incursion of host cells by a pathogen, either in humans or in the pathogen. Changes that are revealed by these studies can aid in a wide range of research and development efforts to combat the pathogen, such as the development of ►**biomarkers** for diagnostic purposes and the discovery of novel drug targets or prophylactic vaccines.

Characteristics

Proteomics of human pathogens examines the protein expression program of the pathogen. Typically, protein contents and properties of wild type and mutant pathogens are compared to identify the underlying molecular determinants of the phenotypic differences. In addition, ►**protein expression profiling** that identifies spatial and temporal patterns of expression may correlate the potential pathogenic factors with disease progression.

Applying a proteomics approach to study changes in host cells follows the footsteps of ►**DNA microarray**-based studies of host-virus interactions. As with the ►**RNA profiling**, global changes in gene expression in response to bacterial or viral infections are measured, albeit at the protein level. Again, changes induced by wild type and mutant pathogens as well as temporal regulation of host protein expression are often analyzed.

Infectious Diseases Proteomics Technologies

Profiling of total protein contents with ►**two-dimensional gel electrophoresis** followed by identification of proteins of interest using ►**mass spectrometry** is a common theme for a proteomics project. This process can be greatly facilitated when complete DNA sequence information is available for the particular organism. Microbes, having much smaller genomes than their hosts, are especially amenable to genomics and proteomics techniques. For example, according to the National Center for Biotechnology (NCBI), there are 140 completed microbial genomes and another 180 bacterial genome projects in progress, in addition to 1506 listed viral genomes as of February 2004. The human genome, despite having approximately 3 million base pairs and being one of the largest genomes, has also been fully sequenced with only small gaps to be filled. These technological advances in the genomics field paved the way for efficient proteomics analysis of both pathogen and human host during an infection.

Proteomics databases also play an important role in studies of host-pathogen interaction. In addition to the various 2-D gel databases created by individual researchers for specific pathogens, there are also companies that specialize in establishing large, comprehensive proteomics databases. For example, the protein interaction maps (PIM) produced by Hybrigenics contain protein-protein interaction data based on whole genome information on microbes and pathogen-host cell interplay. For example, the ►**polypeptides** of hepatitis C virus (HCV) are known to interact with one another during the life cycle of the virus, thus elucidating these interactions and those between HCV and hepatic factors will probably facilitate anti-HCV drug and vaccine research. Other human pathogens

included in this PIM database are *Helicobacter pylori* as well as human immunodeficiency virus. Several 2-D PAGE databases have also been generated for Epstein-Barr virus-transformed B-lymphoblastoid cell lines (6). Most of the data in the current databases has been generated using high throughput methods. As a result, no single set of data is to be completely trusted and relied upon alone. Rather, a definitive conclusion is only reached after assimilation of all the information from different sources and validation of results with alternative methods.

Proteomics Applied to Human-Pathogen Interactions

Proteomics studies aimed at host-pathogen interactions can be divided into two main categories:

1. Proteomics of the pathogen and its adaptation to the host
2. Proteomics of the host and its response to the infection by the pathogen.

Proteomics Studies of the Pathogen

The proteomes of many bacterial pathogens have been determined by 2-D electrophoresis and mass spectrometry. These include *Staphylococcus aureus*, *Bacillus subtilis*, *Pseudomonas aeruginosa*, *Helicobacter pylori*, *Salmonella typhimurium*, *Mycoplasma pneumoniae*, *Mycobacterium tuberculosis*, *Chlamydia pneumoniae* and *Plasmodium falciparum* (1, 3). Protein profiling has been used to identify pathogenic determinants, comparing non-virulent with ►virulent strains/forms of the pathogen. For example, *Mycobacterium tuberculosis* is the pathogen that causes tuberculosis (TB) in humans, while the vaccine strain *Mycobacterium bovis* BCG is non-virulent. When the proteomes from these two strains were compared by 2-D gel electrophoresis, more than two dozen proteins were found to be differentially expressed (3). Subtle differences between isogenic mutant strains H37Rv and H37Ra of *Mycobacterium tuberculosis* revealed by proteomics analysis may also help explain the mechanism of attenuation of *M. tuberculosis* H37Ra (2). Infection of a mammalian host by *Yersinia pestis* is accompanied by a temperature change from 26°C (the temperature of the flea) to 37°C. When bacterial proteins for *Y. pestis* grown at these different temperatures were compared using a protein array approach, factors implicated in virulence were identified from the proteins that were up-regulated at the higher temperature.

Another area of microbial research where proteomics tools are useful is the molecular physiology study of pathogens. Comparative proteomics of the malaria parasite *Plasmodium falciparum* revealed distinct protein expression and localization profiles in four different stages of the life cycle, defining underlying mechanisms at the protein level for the metabolism,

►stress response and other physiological processes (1). These proteomics studies contribute to our overall understanding of the pathogens and add to our abilities to fight them by identifying potential new drug targets. Emerging antibiotic resistance in human microbial pathogens is a big concern and an area of intense research by medical microbiologists. Comparative proteomics have been used to examine the differences in protein content and characteristics exhibited by antibiotic-sensitive and resistant strains of *Streptococcus pneumoniae*, *Helicobacter pylori* and *Staphylococcus aureus*. Proteins that are up-regulated in the resistant strains were identified by mass spectrometry and their possible roles in establishing the resistance examined. Alternatively, antibiotic-binding proteins can be analyzed using a proteomics approach to study the mechanism of resistance.

Systematic identification and characterization of cell surface proteins of bacterial pathogens can lead to identification of new targets for vaccines. Based on available genome sequences, potential cell surface or secreted proteins can be identified *in silico* by a combination of computer prediction and ►homology search. These candidate proteins are expressed and purified as recombinant proteins that are in turn used as immunogens to generate ►antisera in mice. FACS staining of un-permeabilized cells confirms the cell surface exposure of these proteins. Finally, the validated antisera are tested for their ability to neutralize the pathogen in an ►in vitro infection assay. A cell surface or secreted protein that can elicit antibodies capable of suppressing infection meets the criteria for a candidate vaccine. This genomics/proteomics-based vaccine development approach, known as “reverse vaccinology” has been successfully used to identify vaccine candidates for *Neisseria meningitidis* B and *Chlamydia pneumoniae* (2, 3).

Direct analysis of ►immunogenic factors that react specifically with infected patient sera can also result in a selection of vaccine candidates for the particular pathogen. For example, 2-D PAGE was coupled with Western blotting to identify *Helicobacter pylori* proteins that were reactive to pooled patient sera; the positive spots were identified by in-gel digestion and mass spectrometry. One of the proteins identified in this study was proposed as a vaccine candidate.

An important aspect of protein function pertains to its interaction with other proteins inside a cell or a macromolecular complex. Identification of cellular factors interacting with viral proteins has been and continues to be a popular research project for molecular virologists and has produced a vast amount of useful information on virus-host cell interaction. Equally important are the interactions between microbial proteins. Traditional techniques for studying protein-protein interactions such as ►co-immunoprecipitation

and ►yeast two-hybrid can be adapted to high throughput, systematic studies. A comprehensive yeast two-hybrid analysis of all potential pair-wise combinations of the 266 vaccinia virus proteins (which total over 70,000) revealed 28 previously unidentified interactions (4). Similar studies to discover interactions between bacterial proteins have also been carried out for *Helicobacter pylori* and *Rickettsia sibirica*.

Proteomics Studies of the Host

RNA profiling (►differential display) of gene expression in a human cell pre- and post-infection by a pathogen has been done for a large number of pathogens including HIV-1. In addition to generating considerable numbers of mRNA profiles that can guide and validate the corresponding proteomics experiments, this work has also been a driving force behind the development of sophisticated ►bioinformatics tools. Comprehensive protein profiling of a human cell however, is technically more challenging, due largely to the fact that the current resolution of proteomics techniques prevents simultaneous and exhaustive analysis of the very large number of proteins. Several approaches have been undertaken to circumvent this problem and identify differentially expressed/modified proteins in infected cells or in cells manifesting a disease phenotype.

Firstly, alternative detection methods and sophisticated imaging software help increase the resolution of the 2D gels. Using computer-aided imaging analysis, Toda et al. (6) compared protein spots on silver- as well as Cypro Ruby- stained 2-D gel images from three pairs of pre- and post-immortalized Epstein-Barr virus-transformed B-lymphoblasts (LCLs). Even though no spot was found to be commonly regulated in the three cell lines pre- and post-immortalization, the expression of one protein, later identified to be stathmin, decreased dramatically in two of the three cell lines after immortalization. Importantly, the expression level of this protein in the remaining cell line is extremely low both pre- and post-immortalization, consistent with the authors' hypothesis that this change represents an early event in the immortalization process which may have already begun in the third cell line (6).

Secondly, conducting focused investigations of subpopulations of protein can cut down the number of proteins that need to be analyzed in a single study. For example, the ►"sub-proteome" of cell-surface proteins should contain most of the proteins involved in parasite receptor binding and an ►affinity purification step with phospho-specific antibodies can significantly narrow the field for an experiment aimed at detecting changes in host protein ►phosphorylation in response to infection. Alternatively, total protein samples can be interrogated with a collection of well-characterized monoclonal antibodies to discover potential changes in a group of pre-defined genes. Such a multiplex Western blotting technology was used in an

effort to define proteome changes in dermal microvascular endothelial cells (DMVEC) upon infection by Kaposi's sarcoma associated herpesvirus (KSHV) (5). Approximately 850 monoclonal antibodies recognizing ►transcription factors and cell signaling molecules (BD Biosciences) were used in Western blots to detect differential protein expression between age- and passage-matched DMVEC cell pairs with or without KSHV infection. A total of 52 proteins were found to be differentially expressed with 29 proteins up-regulated in infected cells. One of the up-regulated proteins, heme oxygenase-1 (HO-1), was characterized in detail.

Finally, using body fluids or supernatants of cell cultures as starting material can also significantly reduce the number of proteins to be analyzed. Protein profiling of body fluids from infected patients is an area of proteomics that holds great promise for biomarker and vaccine development. Coupled with high-accuracy mass spectrometry, the Proteinchip® technology offered by Ciphergen is well suited for clinical diagnostic research. Since protein separation and mass identification are done sequentially on the same surface without any gel running steps, this technology allows rapid processing of large numbers of clinical samples in parallel. It was also used to identify anti-viral factors that can inhibit HIV infection (7). Although the cell source for the factors turned out to be identified incorrectly, it nevertheless demonstrated the ability of the technology to detect and identify active anti-viral components out of a plethora of protein factors secreted by a mixture of different types of primary cells.

Clinical Relevance

Tremendous progress has been made in the proteomics study of pathogens and their interactions with the human host, fueled by advances in protein chemistry technology and the rapid expansion of genomic sequence databases. Although the majority of the current proteomics studies of pathogen-host interaction have been done with *in vitro* cell culture models of the diseases, the scientific information generated by these studies should nevertheless lead to better understanding of the complex biological processes involved in pathogenesis, host-defense and mechanisms of drug resistance of both viral and bacterial infections ►in vivo. This will greatly facilitate the development of early and accurate diagnostic markers as well as the discovery of pertinent drug targets and prophylactic vaccine candidates that promise a better outcome of disease management.

References

1. Carucci DJ, Yates 3rd, Florens L (2002) Exploring the proteome of *Plasmodium*. Int J Parasitol 32:1539–42
2. Fraser CM (2004) A genomics-based approach to bio defence preparedness. Nat Rev Genet 5:23–33

3. Kellam P (2001) Post-genomic virology: the impact of bioinformatics, microarrays and proteomics on investigating host and pathogen interactions. *Rev Med Virol* 11:313–329
4. McCraith S, Holtzman T, Moss B et al (2000) Genome-wide analysis of vaccinia virus protein-protein interactions. *Proc Natl Acad Sci USA* 97:4879–4884
5. McAllister SC, Hansen SG, Ruhl RA et al (2004) Kaposi's Kaposi sarcoma-associated herpesvirus (KSHV) induces heme oxygenase-1 expression and activity in KSHV-infected endothelial cells. *Blood* 103:3465–3473
6. Toda T, Sugimoto M (2003) Proteome analysis of Epstein-Barr virus-transformed B-lymphoblasts and the proteome database. *J Chromatogr B* 787:197–206
7. Zhang L, Yu W, He T et al (2002) Contribution of human alpha-defensin 1, 2, and 3 to the anti-HIV-1 activity of CD8 antiviral factor. *Science* 298:995–1000

Proteomics in Microfluidic Systems

JAISREE MOORTHY, DAVID J. BEEBE
Biomedical Engineering, University of Wisconsin,
Madison, WI, USA
djbeebe@wisc.edu

Definition

► **Microfluidics** deals with transporting and manipulating small volumes (nl– μ l) of fluids within confined regions (► **microchannels**) and allows for characterizing samples in a high-throughput manner. Over the last decade, microfluidic platforms have been explored for chemical and biological studies. In molecular biology, the picture that has emerged as the central dogma is that knowledge of genes alone cannot provide the critical clues to understanding the functioning of the cell; there is a dynamic and intricate network of interactions between DNA, protein and other molecules within the cellular environment (1). Moreover, the presence, quantity and reactivity of proteins and other biomolecules inside the cell vary with time and external conditions. Characterizing the expression, structure, function and interaction profiles of proteins within the cell during different cell-cycle stages and external conditions is commonly referred to as proteomics. In order to carry out these studies in a high-throughput manner, proteomics requires improved technology to handle small volumes, to perform multiple tests and to detect molecules at low concentrations. As microfluidic systems offer these capabilities, they provide an attractive platform for proteomics.

Characteristics

Proteomics primarily involves separation, sensing (or detecting) and interaction studies. Performing one or

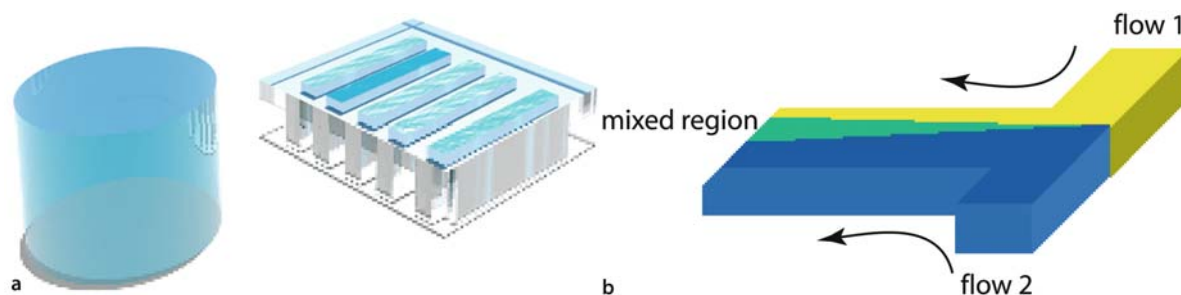
more of these studies is the first step towards understanding the molecular orchestration that occurs within cells. In the following sections, the basic features of microfluidic devices and their ability to perform these studies will be discussed.

Basics of Microfluidics

A typical microfluidic device consists of a myriad channels and other components such as pumps and ► **valves** to transport and manipulate fluid. The volume of fluid in a single microchannel is typically a few micro liters. Figure 1a compares the volume in one well of a microtiter plate (384-well) with that in a microfluidic channel. It is possible to carry out at least 10 assays in the microchannel for every assay done in the microtiter well. The surface area to volume ratio is very high within the channel, allowing a higher fraction of molecules in the sample volume to effectively “see” the surface. High surface area to volume ratio becomes useful when the sensor (detector) is at the surface and also for effective heat dissipation (e.g. during electrophoretic separation). ► **Laminar flow** occurs inside micron-sized channels, where the viscous properties of the fluid dominate; the flow of water through a microchannel is analogous to flow of a viscous fluid such as honey through a significantly larger channel (2). Laminar flow is smooth, steady, devoid of fluctuations (in velocity) in space and time and is predictable (Fig. 1b). Moreover, in laminar flow, two or more streams flowing adjacently mix at their interface mainly by ► **diffusion**. Slow mixing (► **Mixing Time**) can be used to advantage for separation and detection. For example, particles can be separated based on size, as larger molecules diffuse more slowly than smaller ones. A key advantage of the microfluidic platform is that the environment within the channels can be controlled in both spatial and temporal manners. This ability allows for dynamic studies that are not possible in conventional methods. For example, one can analyze cell signaling under varying external conditions or study the kinetics of biomolecular interaction.

The advantages of microfluidics include (3):

- a) ability to handle small volumes by virtue of their dimensions
- b) high-throughput, many assays can be performed in parallel
- c) quicker assays due to the larger sensing surface and better heat dissipation
- d) low evaporation and less exposure to toxins as the microsystems are enclosed
- e) occupation of less space making the devices portable for use in point of care diagnostics
- f) higher reproducibility because the sample volumes can be kept constant, as devices can be manufactured with precise tolerances



Proteomics in Microfluidic Systems. Figure 1 (a) Comparison of volumes in a microtiter well (384-well) to those in microfluidic channels. With the same amount of sample, at least 10 more assays can be performed in the microfluidic platform. (b) The viscous properties of water flowing through the channel are similar to those of honey; the flow is smooth, steady and devoid of fluctuations in velocity. Two streams flowing adjacently to each other (directions shown by arrows) mix only at their interface by diffusion, which can be slow for large molecules (e.g. proteins). However, slow mixing can be used to advantage for separation or detection as well as to define spatial regions within the channels.

- g) single cell analysis and cell based assays because the physics of transportation within micron-sized channels allow for the study of biological phenomena occurring at the same scale
- h) ability to perform experiments that require precise spatial and temporal control, which are difficult to perform in regular micro-titer plates

Fabrication

A microfluidic device consists of channels (or cavities) of a pre-defined size and is connected to other cavities. The channels can be created by either etching a surface or by forming walls. Fabrication methods have evolved from traditional silicon-based protocols to ►micro-molding and ►liquid phase photo-polymerization (LP3) (Fig. 2) using organic or biopolymers (e.g. gelatin) (4). In micromolding, a “master” containing raised features (for channels) is first fabricated, over which elastomeric polymers are cured by pouring the monomer mixture and baking for few hours. The cured elastomer is released from the “master” and sealed to form enclosed channels. In LP3, one starts with an enclosed cavity and fabricates various components *in situ* by flowing monomer solution and polymerizing channel walls and other components (eg. valve, filter) by irradiation through a mask. The mask defines the location of polymerized structures within the cavity. Using this method, the time to make a device has been reduced from days to minutes (Fig. 2). LP3 also allows the user to arrange the components in any desired location based on the assay requirements (5).

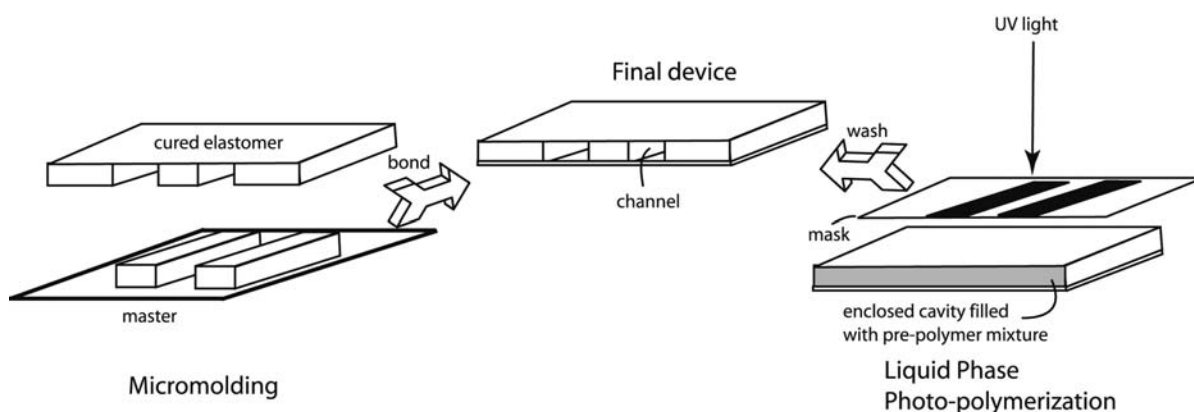
Separation

One of the most common ways to separate proteins is *via* electrophoresis. The biomolecules separate based

on their charge and mass; the presence of a protein is indicated by a band (of dyed protein) and the expression level is reflected in the intensity of the signal. By virtue of their size, separation in microchannels requires small sample volumes. Separation can be done quickly with a higher electric field because of better heat dissipation (increased surface area to volume ratio) mitigating joule heating. The microfluidic platform is well developed for capillary electrophoresis and is widely used in DNA studies. Gel electrophoresis (GE), both 1D and 2D has also been demonstrated in this platform. After separation, the proteins are usually digested into smaller peptides, further separated and detected using other techniques such as mass spectroscopy (MS). Alternatively, if the functionality of the protein is known beforehand, affinity based separations can be used to both separate and identify the proteins. Both protein digestion and affinity separations have been performed in microchannels. Moreover, separation of proteins from single cells has been performed inside microfluidic systems (6).

Sensing

Identifying a specific protein or a class of proteins is important for characterizing its functionality and to develop appropriate drugs. Increased exposure to surfaces inside a microchannel has been used to perform effective sensing on cantilever beams and surfaces. The physics of fluid flow in microchannels (slow mixing in laminar flow) has been used to develop a T-sensor, wherein biomolecules are detected by flowing antibodies in an adjacent stream (Fig. 1b); specific binding between the biomolecules was identified by tracing the diffusion of molecules at the interface of the streams (7). Analogous to DNA microarrays, protein arrays have been developed (8), wherein antibodies or fragments are immobilized on a



Proteomics in Microfluidic Systems. Figure 2 Fabrication of a microfluidic device using micromolding and liquid phase photo-polymerization (LP3). In micro-molding, elastomeric polymers are cured over a “master” and sealed to form enclosed channels (final device). In LP3 various structures (channel walls, pumps, valves, etc.) are photo-polymerized *in situ* in an enclosed cavity. A pre-polymer mixture containing monomer, cross-linker and photo-initiator is introduced into the cavity and polymerized by irradiating with UV light through the mask. Upon removing the unpolymerized mixture and washing, the final device is obtained.

surface. By using a channel network (along with protein arrays) extra dimensions in space and time can be utilized. However, unlike DNA fragments, proteins must be hydrated to maintain their functionality. A significant challenge is to keep the immobilized proteins hydrated and functioning naturally. One way to keep proteins hydrated is to position them within ►hydrogels. Another problem with protein studies is that of cross-reactions and false positives. Using channel networks, multiplexed detection (e.g. fluorescence and cantilever) can be performed to minimize false positives and non-specific interactions.

Applications

The biochemical (and biological) function of a protein is characterized by studying interactions with other proteins under varying conditions. Such studies allow for the development of an interaction map, which can help in understanding the involvement of a specific protein in various pathways and how its excess or absence can affect the functioning of the cell as a whole. Because many tests can be carried out in parallel, microfluidic devices can be used to study the parameter space (buffer type, conc. etc.) influencing the interactions in a systematic manner. Multiple interactions can potentially be studied in a microfluidic platform with the connectivity in the microchannel mimicking the signaling or metabolic pathways within the cell. The size of microfluidic systems allows for manipulating and studying a single cell and the ability to precisely control the physics of transportation inside microfluidic channels allows for manipulating the spatial and temporal environment around the cell. Takayama et al. have shown that spatial control of environment around the cell is possible by exposing a

single cell to different solutions (9). This approach can be extended to stimulate the cell at specific regions while monitoring changes (e.g. calcium waves) *in vivo* so as to elucidate signal transfer and protein interaction networks. Another area where microfluidic platforms can be useful is in crystallization studies. Finding the right conditions for crystal formation is still an art. By utilizing the capability of the platform to manipulate small volumes of fluid and to perform large number of tests in parallel, Hansen et al. demonstrated that the parameter space for crystallization could be explored quickly to find the optimal set of conditions (10).

The capabilities of traditional technologies such as 2D GE for separation and mass spectroscopy for detection of biomolecules are being improved for proteomic studies and there is growing interest in using the microfluidic platform to link multiple tools. Although microfluidic devices may quickly find applications as links between existing technologies, the strengths of these devices in handling small volumes, multiplexing and performing studies at the cellular level will drive the growth of this platform as an independent integrated system.

Clinical Relevance

Several diseases are caused by the inappropriate presence of a protein at a certain stage of cell growth, mis-folded or improper post-translation modifications of proteins, and excess (or scarcity) of a specific protein. One way to address the problem is to pinpoint the protein (►marker) that can be assayed for diagnosis and/or prognosis. As microfluidic platforms for sensing proteins are being realized, detecting a specific marker can be carried out efficiently. Multiple functionalities such as sensing, storing reagents and fluid manipulation

(pumps and valves) can be incorporated to develop portable diagnosis systems for use in remote areas (5). However, as the molecular basis of the cell is being understood, it is becoming evident that a single marker alone may not suffice; multiple markers and/or their interactions may be the cause of many diseases. In the absence of a specific marker, a fingerprint of the protein expression and interaction can be obtained. By using a microfluidic platform, the proteins expressed in the diseased cell can be quickly obtained using separation techniques and the interaction of the proteins can be analyzed. These results can then be compared with “healthy” profiles for use in prognosis and diagnosis. Other uses of the platform include high-throughput drug screening and testing for side effects of a specific drug by mapping its interaction with a set of proteins.

References

1. Tyers M, Mann M (2003) From genomics to proteomics. *Nature* 422:193–197
2. Beebe DJ, Mensing GA, Walker GM (2002) Physics and applications of microfluidics in biology. *Annu Rev Biomed Eng* 4:261–286
3. Mitchell P (2001) Microfluidics: downsizing large-scale biology. *Nat Biotechnol* 19:717–721
4. Moorthy J, Beebe DJ (2003) Organic and biomimetic designs for microfluidic systems. *Anal Chem* 75:292–301
5. Moorthy J, Mensing G A, Kim D et al (2004) Microfluidics tectonics platform: A colorimetric, disposable botulinum toxin ELISA system. *Electrophoresis* 25:1705–1713
6. Cannon DM, Winograd N, Ewing AG (2000) Quantitative chemical analysis of single cells. *Annu Rev Biophys Biomol Struct* 29:239–263
7. Weigl BH, Yager P (1999) Microfluidics – Microfluidic diffusion-based separation and detection. *Science* 283:346–347
8. Mirzabekov A, Kolchinsky A (2002) Emerging array-based technologies in proteomics. *Curr Opin Chem Biol* 6:70–75
9. Takayama S, McDonald JC, Ostuni E (1999) Patterning cells and their environments using multiple laminar fluid flows in capillary network. *Proc Natl Acad Sci USA* 96:5545–5548
10. Hansen CL, Skordalakes E, Berger JM et al. (2002) A robust and scalable microfluidic metering method that allows protein crystal growth by free interface diffusion. *Proc Natl Acad Sci USA* 99:16531–16536

Proteopathies

► Defective Protein Folding Disorders

Prothrombin

Definition

Prothrombin, also referred to as coagulation factor II, is involved in the clotting cascade. The inactive precursor prothrombin is converted into the active thrombin (factor IIa) by proteolytic cleavage which is catalysed by factor Xa, the product of the preceding step of the clotting cascade.

► Polyadenylation

Protocol Mutagenesis

Definition

Protocol mutagenesis refers to the process where a practical laboratory protocol is continually modified in an uncontrolled and often unrecorded way, often rendering it useless.

► Large-Scale Homologous Recombination Approaches in Mice

Protonation States

Definition

Protonation states of the amino acids Asp, Glu, His, Tyr, Cys, Lys, and phosphorylated residue side chains can vary depending on the pH and local environment of the residue. Protonation states are derived from titration curves of proteins, and show a continuous change of total protein charge over pH ranges. However, crystal structures are commonly solved with simple assumptions regarding charges on individual amino acid residues. As hydrogen atom electrons are not usually resolved in crystal structures, assignment of coordinates to hydrogen atoms is a modeling task.

► 3D Structure Determination by X-Ray

► Amino Acids, Physicochemical Properties

Proto-Oncogene

Definition

Proto-Oncogene describes a normal gene that is usually concerned with the regulation of cell proliferation. By mutation, it can be converted into a cancer-promoting

► oncogene.

- ▶ Mutagenesis Approaches in Yeast
- ▶ Proteomics in Cancer
- ▶ Ras Signalling
- ▶ RNA Polymerase III
- ▶ Tumor Suppressor Genes

Protoplast Transformation

Definition

In general the term transforming comprises the transfer of genetic information between bacteria or RNA and viral DNA in eukaryotic cells by uptake of DNA sequences. In yeast genetics, the yeast cell wall has to be removed (remainder called protoplast) to make the cell susceptible to the uptake of foreign DNA. The uptake is usually further enhanced by chemical perturbation of the cell membrane (e.g. by the addition of polyethyleneglycol).

- ▶ [Recombinant Protein Expression in Yeast](#)

Protrusion

Definition

In general the term protrusion refers to anything that protrudes. In molecular biology the term describes the state that a residue protrudes the surface of a protein. It is quantified by calculating a protrusion index; this gives an absolute value for the extent to which a residue protrudes from the surface of a protein. In cell biology it is used as a general term to describe highly dynamic projections of the cell surface, like filopodia or lamellipodia, that help cells to sense their external environment.

- ▶ [Protein-Protein Interaction](#)
- ▶ [Two Hybrid System](#)

Provirus

Definition

Provirus is the integrated DNA form of a retrovirus.

- ▶ [Repetitive DNA](#)
- ▶ [Retroviruses](#)

Proximal Spinal Muscular Atrophy

- ▶ [Spinal Muscular Atrophy](#)

PRP

- ▶ [Pattern Recognition Receptors](#)

PrP^C

Definition

PrP^C is the native, cellular isoform of the prion protein. PrP^C is a glycosylphosphatidylinositol (GPI)-anchored protein on the surface of many cell types of mammals. Its physiological functions are poorly understood. Changes in the conformation of PrP^C have been related to certain transmissible degenerative diseases.

- ▶ [Prion Diseases](#)

PrP^{Sc}

Definition

PrP^{Sc} is the Scrapie isoform of the prion protein. This protein shows increased resistance to digestion with proteinase K. It is closely associated with infectivity, and in terms of the prion hypothesis, PrP^{Sc} is part of the infectious agent, the prion.

- ▶ [Prion Diseases](#)

Prx

- ▶ [Peroxiredoxin](#)

PS

- ▶ [Phosphatidyl Serine](#)

PS1/PS2

Definition

PS1/PS2 stands for presenilin2/presenilin 2. PS1 protein is an essential component of the γ -secretase complex. Presenilin 2 (PS2) protein is a homologue to PS1 and may replace PS1 as essential component of the γ -secretase complex. PS1 and PS2 are integral membrane proteins that contain 6-8 transmembrane domains, and are predominantly localized within the endoplasmic reticulum and Golgi of neurons within the brain. Mutations in PS1 and PS2 have been linked to a rapid increase in plaque accumulation in ►[Alzheimer's Disease](#).

PSD95

►[PDZ Domain](#)

PSEN1/PSEN2

Definition

PSEN1/PSEN2 refers to the human genes for ►[PS1/PS2](#) proteins.
►[Alzheimer's Disease](#)

Pseudoachondroplasia

Definition

Pseudoachondroplasia (PSACH) is a dominantly inherited chondrodysplasia characterized by disproportionate short stature, abnormal joints and early-onset osteoarthritis requiring joint replacement. PSACH is caused by mutations in the cartilage oligomeric matrix protein (COMP).

►[Repeat Expansion Diseases](#)

Pseudoautosomal Region

Definition

Pseudoautosomal region in the human refers to a segment of the human DNA found in the distal short

arms of the X and Y chromosomes. These segments contain homologous sequences promoting X-Y cross-over during meiosis.

►[SRY – Sex Reversal](#)

►[X-Chromosome Inactivation](#)

Pseudogene

Definition

A pseudogene is a non-functional sequence of DNA in the genome, which is almost identical to a functional gene sequence. Pseudogenes arise by ►[gene duplication](#) (or insertion) followed by accumulation of mutations which render it nonfunctional, but in the presence of the functional original gene such mutations constitute no selective disadvantage and can persist in the gene pool. Processed pseudogenes have no introns, and are therefore thought to have arisen by the action of viral reverse transcriptase. In order to be propagated in the gene pool, the pseudogenes must have arisen early in development or in germ cells, and are therefore often associated with embryonic genes. Processed pseudogenes are abundant for embryonic simple epithelial keratins k8 and K18.

►[Intermediate Filaments](#)

►[Neurofibromatosis Type 1 \(NF1\), Genetics](#)

Pseudohypoaldosteronisms

Definition

Pseudohypoaldosteronisms (PHA) are characterized by end-organ resistance to aldosterone inducing hyperkalemia and hyperaldosteronism. There are two forms of PHA classified according to the level of blood pressure with either hypotension (►[Type 1 Pseudohypoaldosteronism](#) or ►[PHA 1](#)) or hypertension (►[Type 2 Pseudohypoaldosteronism](#) or ►[PHA 2](#)).

Pseudohypoparathyroidism

Definition

Pseudohypoparathyroidism is an endocrine disease resembling true hypoparathyroidism, in that serum levels of calcium and phosphate are reduced. However,

in contrast to hypoparathyroidism, circulating levels of parathyroid hormone (PTH) are not reduced, but increased in most instances.

- ▶ [G-Proteins](#)
- ▶ [Hyper- and Hypoparathyroidism](#)

Pseudoknot

Definition

Pseudoknot refers to an RNA secondary structure involving a stem loop, which forms a second base pairing interaction with an adjacent, usually downstream, sequence of the same mRNA. The structure consists of two base paired regions (stem1 and stem2) joined by two loops (loop1 and loop2); the two stems usually form a coaxial stacked structure.

- ▶ [Translational Frameshifting, Non-standard Reading of the Genetic Code](#)

Pseudomonomolecular Process

Definition

Pseudomonomolecular process describes a process that involves the association or dissociation of two or more molecules displaying diminished or absent concentration dependence.

- ▶ [Thermodynamic Properties of DNA](#)

Pseudopregnant Mouse

Definition

Pseudopregnant mouse defines a female mouse after copulation with a vasectomized male. The mouse behaves hormonally pregnant allowing its use as a recipient for embryos; e.g. for chimera production.

- ▶ [Large-Scale Homologous Recombination Approaches in Mice](#)
- ▶ [Transgenic and Knock-out Animals](#)

Pseudotyped Virus

Definition

Pseudotyped virus refers to a retrovirus, the host range of which has been widened. They were initially developed for human gene therapy.

- ▶ [Medaka as a Model Organism for Functional Genomics](#)

PSF

- ▶ [Point Spread Function](#)

PSI

Definition

The Proteomics Standards Initiative (PSI) is aimed at defining community standards for data representation in proteomics to facilitate data comparison, exchange and verification. The Proteomics Standards Initiative launched the HUPO meeting in Washington D.C., April 28-29, 2002. As a first step, the PSI will develop standards for two key areas of proteomics: ▶ [mass spectrometry](#) and ▶ [protein-protein interaction](#) data.

- ▶ [Protein Databases](#)

PSI

Definition

The aim of the protein structure initiative (PSI-NIGMS) is to determine a large number of protein structures not redundant with those in the PDB (protein data bank), which can be collected as an inventory of all protein structure families in a library that is widely available for basic biological studies. This project involves: 1) organizing known protein sequences into families; 2) selecting family representatives as targets; 3) solving

the 3D structure of targets by X-ray crystallography or NMR spectroscopy; and 4) building models for other proteins by homology to solved 3D structures. This project was initially developed and funded by National Institute of General Medical Sciences (NIGMS).

► [Protein Databases](#)

PSI-BLAST

Definition

PSI-BLAST stands for Position-Specific Iterated ► [BLAST](#). Following a standard BLAST (searching tool for nucleotide and protein homologs to a query sequence in large databases) search, found hits are combined to a multiple sequence alignment, which is converted into a position specific score matrix. In iterated searches, this profile is used as a query.

► [Protein Domains](#)

Psoriasis, Molecular Basis

KHUSRU ASADULLAH¹, WOLFRAM STERRY²

¹Corporate Research Business Area Dermatology, Schering AG, Berlin, Germany

²Department for Dermatology and Allergology, Medical School Charité, Humboldt University Berlin, Berlin, Germany
khusru.asadullah@schering.de

Definition and Characteristics

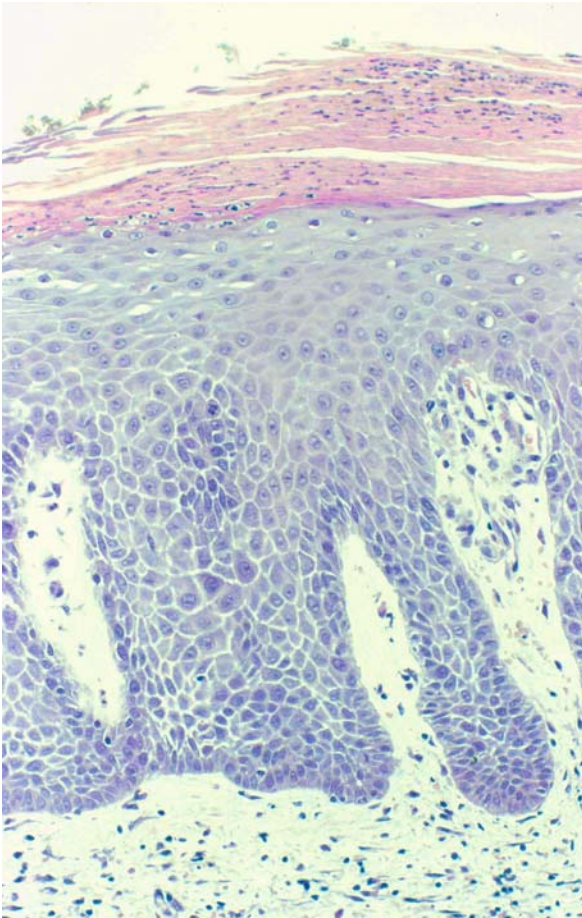
Psoriasis is a multigenic, cutaneous disorder characterized by inflammation and abnormal epidermal proliferation with a prevalence of 2–3% in the general population (1). It is defined as a clinical entity affecting skin, nails and, in approximately 5% of the patients, joints. Typical findings are ► [keratinocyte](#) hyperproliferation with hyper- and para-keratotic differentiation, epidermal influx of polymorphonuclear leukocytes and the presence of a mononuclear infiltrate in the papillary dermis and the epidermis. Based on the current pathophysiological understanding, psoriasis is considered to be a ► [T lymphocyte](#) mediated dermatosis (see below). In contrast to type 2 psoriasis, type 1 psoriasis is characterized by early onset (before age 40) and a positive family history.

Clinical Relevance Diagnostics

The observation of typical clinical features is sufficient for making the diagnosis in the majority of patients. The usual presentation of the skin lesions is that of sharply demarcated erythematous (red), hyperkeratotic (scaling) plaques of varying extent, distributed symmetrically over the skin (1, 2; Fig. 1). Histological analyses of skin biopsies show akathosis (elongation of the epidermis), abnormal differentiation and inflammatory infiltrates (Fig. 2). Molecular analyses are not necessary for the diagnosis. However, some hints of an expansion of certain T lymphocyte families expressing specific subsets of antigen recognition molecules in skin lesions has been reported, as well as an association of early onset psoriasis and a positive family history for psoriasis with the genotype ► [human leukocyte antigen](#) (HLA)-Cw6, B57 (see also below), which may contribute to a better understanding of the disease rather than have diagnostic impact.



Psoriasis, Molecular Basis. Figure 1 Typical symmetric skin lesions in a psoriatic patient.



Psoriasis, Molecular Basis. Figure 2 Typical histological findings in a psoriatic patient.

Therapy

The current options for the treatment of psoriasis are limited. Topical therapies include the use of dithranol, calcipotriol, and steroids. These, however, are usually not very successful in severe psoriasis. UV-radiation therapy is commonly used but it is associated with development of side effects like tumor induction after application over several years. The options for systemic antipsoriatic therapies include cyclosporine A, FK506, and fumaric acid esters. Yet, these treatments may also fail in certain patients and, in particular, may be of limited value due to fairly serious side effects. Consequently, these medications are only used in severe psoriasis. This situation indicates the demand for novel therapeutic strategies. New understanding of the pathophysiology of psoriasis might offer opportunities for well-targeted therapeutic interventions. Indeed several immunomodulatory approaches including intervention in the ►cytokine network or inhibition of T lymphocytes or their costimulation by antigen presenting cells as well as inhibition of T-cell infiltration are currently in clinical development (3).

Genetics and Other Pathophysiological Aspects

Psoriasis is a multigenic disorder. Genetic factors are likely to be of fundamental importance in the expression of the disease. This is supported by numerous family, twin and HLA allotype studies (reviewed in 4). Although these observations suggested that patients are genetically predisposed to psoriasis, they imply that environmental factors are also involved.

Three factors are well recognized as triggering the onset, causing new lesions or inducing a flare in the disease, stress, skin injury and infection (5). Infections seem to be of particular importance. More than 40% of psoriatic patients reported a provoking infection. Moreover, 25% of psoriatic children had the onset of their disease after an infection, commonly of the upper respiratory tract. In children already suffering from psoriasis, over 50% had a psoriasis flare within 2–3 weeks of an upper respiratory infection (reviewed in 5). The impact of bacterial infections is clinically most obvious in the frequent association of acute guttate psoriasis with streptococcal throat infection. Recent investigations suggest that this seems to result from a ►superantigen-driven process. However, the pathophysiological importance of viral infections for psoriasis has been recognized, too. Beside the association of virally induced upper respiratory tract infection with guttate flares, hepatitis C, human immunodeficiency and herpes virus infections are considered to trigger, or even to induce, the onset of psoriasis. It is likely that proinflammatory cytokines produced by infected cells or by other cells in response to the infection are involved in these processes.

The association between certain major histocompatibility complex (MHC), i.e. HLA-antigens and psoriasis and the persistence of the disease throughout life once it has appeared suggest the existence of a “memory”, and the spontaneous exacerbation and remission of disease activity is characteristic of diseases involving a chronic immune response (4). Genome-wide scans for disease susceptibility have generated evidence for a major locus situated in the MHC termed PSORS1 (for psoriasis susceptibility 1), on the short arm of chromosome 6. Subsequent studies have sought to refine PSORS1 boundaries by means of linkage disequilibrium fine mapping. Comparison between the different minimal intervals found, points to a 100 kb segment telomeric to HLA-C, and alpha-helix coiled-coil rod homolog genes. Due to methodical limitations, however, the 100 kb interval can only be considered a provisional assignment, and candidate genes lying in its proximity cannot be excluded only on the basis of their position. Studies of positional candidate genes have also been undertaken, focusing on HLA-C, corneodesmosin, and alpha-helix coiled-coil rod homolog genes. Several reports have repeatedly indicated the

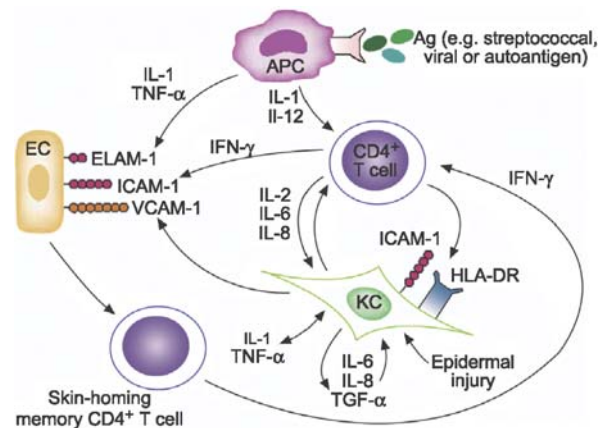
HLA-Cw6*0602 allele as the most significant marker for disease prediction (6).

Several observations indicate that T lymphocytes and cytokines are of major importance in the pathogenesis of this chronic skin disease. These observations are supported by the beneficial effects of systemic administration of immunosuppressive drugs like cyclosporine A, FK506, DAB389IL-2 and fumaric acid esters, known to act on T-cells and to influence the cytokine pattern. Finally, recent investigations using SCID mice demonstrated a central role of immunocytes, in particular T cells, in the pathophysiology of psoriasis. The arguments for considering psoriasis a T-cell mediated dermatosis can be summarized as follows: (3).

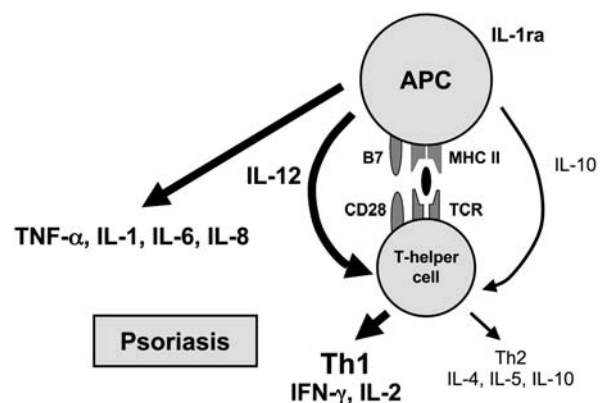
- Presence of activated T-cells in the skin lesions
- Cure of the disease by bone marrow transplantation from healthy persons and transfer of the disease by transplantation of bone marrow from psoriatic patients
- Demonstration of the impact of immunocytes by SCID mice experiments
- Therapeutic effects of immunosuppressants targeting T lymphocytes (e.g. cyclosporine-A, anti-T cell antibodies)

Altogether, based on current knowledge, psoriasis is considered to be a T-cell-dependent (auto)immune disease, probably initiated by presentation of so far unknown “psoriasis-related antigens” by specialized cutaneous APC. Presentation of these antigens and/or superantigens activates T-effector lymphocytes leading to enhanced cytokine formation. These activated T-cells are considered to change epidermal growth homeostasis, leading to increased keratinocyte proliferation and abnormal differentiation. The current pathophysiological concept is shown in Fig. 3.

The cutaneous and systemic over-expression of several proinflammatory cytokines, in particular type 1 cytokines, has been demonstrated in psoriasis. This includes ►interleukin (IL)-2, IL-6, IL-8, IL-12, ►interferon (IFN)- γ and ►tumor necrosis factor (TNF)- α . IL-8 was considered to be of particular importance, since it acts as a chemoattractant in particular for granulocytes, cumulating in psoriatic lesions. Limited efficacy of anti-IL-8 therapy, however, challenged this hypothesis. In contrast, impressive effects of anti-TNF α therapy demonstrate the major impact of this proinflammatory cytokine in the disease. Considering all the data and the evidence that “classical” type-1 and type-2 patterns only mark extremes in a wide variety of motifs, the cytokine profile in psoriasis can be considered as “proinflammatory, type 1-like”. The resulting ►immunological imbalance is shown in Fig. 4. This seems to be of particular importance, since it is characteristic for an



Psoriasis, Molecular Basis. Figure 3 Model for the immunopathogenesis of psoriasis. Presentation of antigen by the antigen-presenting cells (APC), e.g. Langerhans cells, to CD4⁺ T lymphocytes leads to the synthesis of cytokines which stimulate keratinocyte (KC) proliferation and the expression of adhesion molecules by endothelial cells (EC) and keratinocytes. Keratinocytes, in turn, are stimulated to secrete their own cytokines which can act in an autocrine and/or paracrine manner to maintain the psoriatic process. In addition, expression of adhesion molecules by endothelial cells allows the extravasation of leukocytes into the lesion including skin-homing CD4⁺ memory T cells expressing the HECA-452 (CLA) antigen. These cells can also be activated in an antigen-specific manner leading to the secretion of more cytokines to perpetuate the process. Direct activation of keratinocytes by epidermal injury (Koebner reaction) could initiate a psoriatic lesion. (Adapted from a model initially developed by Baker and Fry (3) reproduced with permission from Trends Immunol).



Psoriasis, Molecular Basis. Figure 4 Immune imbalance in psoriasis. The interaction of antigen presenting cells (APC) with T lymphocytes, leading to their activation and differentiation into T-helper (Th)–1 cells is a crucial, early step in the immune activation in psoriasis. In addition, type-1 and several other pro-inflammatory cytokines are over-expressed.

ongoing cellular immune response. Remarkably, a similar cytokine pattern is found in other immune diseases, such as inflammatory bowel diseases e.g. Crohn's disease, transplant rejection and rheumatoid arthritis. Beside the regulatory effects of cytokines on inflammation, which seem to be crucial, several cytokines might directly enhance keratinocyte proliferation. It has been shown that IL-1 and IL-6 are mitogenic for keratinocytes. However, several other factors are involved in the control of keratinocyte growth as well. These include epidermal growth factors (GF), insulin-GF-1, keratinocyte-GF and hepatocyte-GF. The growth factors responsible for the increased keratinocyte proliferation might come from a wide variety of cells, including autocrine and paracrine growth factors from keratinocytes themselves. It is important to notice that there might be a special mix of factors responsible for the keratinocyte hyperproliferation and that other factors not identified so far might be involved, too.

References

1. Christophers E, Sterry W (1993) Psoriasis. In: Fitzpatrick TB, Eisen AZ, Wolff K (eds) et al *Dermatology in General Medicine*. McGraw-Hill, New York, p 489–515
2. Bos JD, De Rie M (1999) The pathogenesis of psoriasis: immunological facts and speculations. *Immunol Today* 20:40–46
3. Asadullah K, Volk HD, Sterry W (2002) Novel immunotherapies for psoriasis. *Trends Immunol* 23:47–53
4. Baker BS, Fry L (1992) The immunology of psoriasis. *Br J Dermatol* 126:1–9
5. Valdimarsson H, Baker BS, Johnsdottir I et al (1986) Psoriasis: a disease of abnormal keratinocyte proliferation induced by T lymphocytes. *Immunol Today* 7:256–259
6. Capon F, Munro M, Barker J et al (2002) Searching for the major histocompatibility complex psoriasis susceptibility gene. *J Invest Dermatol* 118:745–751

PSP

- Progressive Supranuclear Palsy

PTC

- Premature Termination Codon

PTDs

- Protein Transduction Domains

PTH

- Parathyroid Hormone

PTHrP

- Parathyroid Hormone-Related Protein

PTM

- Post-Translational Modification

Ptosis

Definition

Ptosis is also called “drooping eyelid”. It is caused by weakness of the muscle responsible for raising the eyelid, damage to the nerves that control those muscles, or laxity of the skin of the upper eyelids.

- Mitochondrial Myopathies

PTS

Definition

Peroxisomal matrix proteins have to be imported into their target organelle post-translationally. The major translocation pathway depends on a peroxisomal targeting signal (PTS). Two different PTS signals have been identified and are called PTS1 and PTS2.

- Peroxisomal Disorders

P-Type ATPase

Definition

P-type ATPase comprise of a large protein family of transmembrane proteins, which use the energy released by hydrolysis of ATP to transport typically mono-or divalent cations across cellular membranes. However, members of a subfamily of P-type ATPases have been implicated in the transbilayer movement of lipids and bile acids. The enzymatic mechanism of P-type ATPases involves a phosphorylated intermediate, hence the name P-type.

► [Biological Membranes](#)

Pulse Sequence

Definition

In NMR spectroscopy, a series of programmed radio frequency pulses are generated to irradiate NMR samples in a high magnetic field, in order to produce response signals to allow detection of correlations between different nuclei. A pulse sequence always starts with the relaxation period and ends with the detection of an FID. In-between the pulse sequence can be a varying complexity.

► [3D Structure by NMR](#)

► [Multidimensional NMR Spectroscopy](#)

Pulsed-Field Gel Electrophoresis

Definition

Pulsed-field gel electrophoresis refers to an agarose gel electrophoresis method for separating large DNA fragments. Conventional electrophoresis, which separates DNA fragments by sieving (DNA has a constant charge/mass ratio), is difficult to use for fragments above 20 kbp, and fails altogether at about 100 kbp because the DNA is vastly longer than the pore size of even the softest gels, and is drawn through end-on. Shorter fragments can tumble, and thus do present a diameter to the gel that depends on their size. Pulsed-field electrophoresis techniques allows separation of DNA fragments into the tens of megabases by periodically switching the direction of electrophoresis

(various geometries), differentially retarding molecules based on their relaxation (reorientation) time, a function of length.

► [YAC and PAC Maps](#)

Purifying Selection

Definition

If the rate of silent substitutions in the coding portion of a gene exceeds the rate of non-silent substitutions, clearly the gene can be regarded as optimised in a way that specific constraints prevent further stochastic base substitutions in non-silent sites.

► [Sequence Annotation in Evolution](#)

Purine Nucleotide Biosynthesis

► [Nucleotide Biosynthesis](#)

Purines

Definition

Purines are one of the two types of nitrogenous organic bases found in nucleic acids, where adenine and guanine are the most common. They consist of 6-member and 5-member fused heterocyclic rings, with nitrogen atoms in the 1-, 3-, 7-, and 9-positions, and there may be oxygen substituents at the 2- and 6-positions of the 6-member ring as for pyrimidines.

► [Nucleotide Biosynthesis](#)

► [tRNA](#)

pVHL

► [HIF-Prolyl-Hydroxylase](#)

PWS

► Prader-Willi Syndrome

Pyg

► Pygopus

Pygopus

Definition

Pygopus (Pyg) is a PhD finger domain containing protein which is required for activation of transcription by β -catenin.

► Wnt/Beta-Catenin Signaling Pathway

PYP

Definition

PYP stands for photoactive yellow protein, and is thought to be a blue light photoreceptor in certain bacteria like *Halorhodospira halophila*, from which it was first isolated. It has high sequence and structural homology to ► PAS domains.

► Photoreceptors

Pyrimidine Nucleotide Biosynthesis

► Nucleotide Biosynthesis

Pyrimidines

Definition

Pyrimidines are one of the two types of nitrogenous organic bases found in nucleic acids, where cytosine, thymine and uracil (for RNA) are the most common. They consist of a 6-member conjugated heterocyclic ring, with nitrogen atoms in the 1- and 5-positions, and oxygen substituents at the 2- and 6-positions.

► Nucleotide Biosynthesis

► tRNA

Pyruvate Kinase

Definition

Pyruvate kinase is a glycolytic enzyme that converts phospho(enol)pyruvate to pyruvate generating ATP from ADP. A deficiency of this enzyme in red blood cells leads to hemolytic anemia.

► Hemochromatosis

Q10

Definition

Q₁₀ (also known as CoQ₁₀, coenzyme Q₁₀, ubiquinone) is a fat-soluble coenzyme involved in the electrontransport of the respiratory chain. Q₁₀ also serves as an antioxidant.

QC

► Quality Control

QSAR

► Quantitative Structure-Activity Relationship

Q-T Interval

Definition

Q-T interval defines the time from the electrocardiogram, the Q-TC interval indicates the time from the Q wave to the end of the T wave corresponding to electrical systole.

► Hyper- and Hypoparathyroidism

Q-TOF

► Quadrupole Time of Flight Mass Spectrometer

Quadrupole

Definition

Quadrupole is used as a mass filter consisting of four parallel electrodes. By applying combined alternate and direct voltage to the electrodes, only ions with a certain mass to charge ratio oscillate on a stable trajectory before they finally reach the detector. Stable trajectories can be calculated using Mathieu's equations. In a measurement, different voltages are used to scan a mass spectrum over a broad range of mass to charge ratios.

► [SNP Detection and Mass Spectrometry](#)

Quadrupole Mass Analyzer

Definition

A quadrupole mass analyser (QMA) is a device for the selection of ions of a particular mass/charge ratio (m/z) in mass spectrometers. The QMA consists of four parallel rods that have fixed direct current (DC) and alternating radio frequency (RF) potentials applied to them. Ions produced in the ion source of the mass spectrometer are passed through the QMA and selected according to their m/z by variation of the RF potentials. The mass range and resolution depends on the length and diameter of the rods. QMAs are nowadays mainly used in mass spectrometers coupled to gas- or liquid-chromatography. They were developed by Nobel Prize winner Wolfgang Paul.

► [Mass Spectrometry: ESI](#)

Quadrupole Time of Flight Mass Spectrometer

Definition

A Q-TOF is a hybrid mass spectrometer with ► [MS/MS](#) capabilities consisting of a quadrupole and a time-of-

flight mass analyzer. A collision cell is located between the quadrupole and the TOF analyzer to induce fragmentation in MS/MS experiments. Q-Tof instruments have very high sensitivity, resolution and mass accuracy.

► [Mass Spectrometry: ESI](#)

Quality Control

Definition

Quality control can be generally defined as a control system ascertaining that the product of a process fulfills its specifications and stated requirements. In cell biology, quality control refers to a biosynthetic process in the endoplasmic reticulum (ER) involving chaperones and other factors, to ensure that misfolded proteins are not transported to distal secretory compartments beyond the ER.

► [Glycosylation of Proteins](#)

► [Large-Scale Homologous Recombination Approaches in Mice](#)

► [Protein Folding](#)

Quantitative Structure-Activity Relationship

MICHAEL WIESE

Rheinische Friedrich-Wilhelms University, Bonn, Germany

mwiese@uni-bonn.de

Definition

QSAR stands for quantitative structure-activity relationship. This methodology aims to relate differences in structures of a series of molecules to differences in their biological activity through a mathematical equation, called the QSAR model. To develop a QSAR model one needs: (i) activity data, (ii) structural descriptors and (iii) an appropriate mathematical function. Activity data are biological activities of a series of related derivatives. They are usually numbers characterizing a standardized effect (binding and inhibition constants or constants related to effects like EC_{50} , LD_{50}).

Structural descriptors represent molecular properties of the investigated compounds and can be either measured or calculated. They are also expressed as numbers derived from data believed to be related to the molecular property of interest. The conversion of the chemical information into a mathematical description has to conserve the information content inherent in the molecules and should represent the changes in biological activity as a function of the changes in molecular structure. A mathematical function is required that connects changes in biological activity to changes in molecular structure. Regression techniques have been proven to be effective, but nowadays additional techniques such as artificial neural networks and genetic algorithms are also applied.

Description

A Brief History of QSAR

The earliest report on a relationship between molecular structure and biological properties comes from Crois in 1863. He observed that the toxicity of alcohols to mammals increased as the water solubility of the alcohols decreased. In 1868 Crum Brown and Fraser concluded, based on their studies of the biological effects of certain alkaloids, that the physiological (biological) activity Φ is a function of the chemical constitution C (Eq. 1).

$$\Phi = f(C) \quad (1)$$

But even today Eq. 1 cannot be applied as it was formulated; only differences in biological activities $\Delta\Phi$ can be correlated with differences in molecular structure and properties ΔC (Eq. 2).

$$\Delta\Phi = f(\Delta C) \quad (2)$$

In the end of the 19th century Meyer and Overton noted that the toxicity of organic compounds depended on their [lipophilicity](#). As a model system to determine the partitioning they proposed the olive oil/water system.

Little progress in QSAR occurred until the middle of the 1930s when Hammett observed that substituents have a similar effect on the equilibrium constants and reaction rates of aromatic compounds. He showed that equilibrium constants K and reaction rate constants k of substituted benzene derivatives can be described by Eqs. 3, 4.

$$\log \frac{K}{K_0} = \rho\sigma \quad (3)$$

$$\log \frac{k}{k_0} = \rho\sigma \quad (4)$$

K_0 and k_0 denote the equilibrium constant and reaction rate of the unsubstituted compound, respectively, while K and k refer to the substituted derivative. The parameter ρ is specific for a certain reaction. The dissociation of substituted benzoic acids was used as a reference system with ρ equal to one. The parameter σ depends only on the electronic properties of the substituent. Electron-withdrawing substituents have positive σ -values and electron-donating substituents negative σ -values. The relationships developed by Hammett are termed the linear free energy relationship (LFER) as the free energy is related to an equilibrium constant according to Eq. 5.

$$\Delta G = -RT \ln K \quad (5)$$

Such relationships are also termed “extrathermodynamic” as they are stated in terms of thermodynamic parameters, but no thermodynamic principle states that these relationships should hold.

Classical or 2D QSAR

It took another 30 years to start the modern QSAR methodology, still widely used and termed classical or 2D QSAR to differentiate it from 3D QSAR methods. Hansch is recognized as a founder of the modern QSAR. In 1964 he derived a Hammett-type equation for assigning lipophilicity values to aromatic substituents, noting that it is an additive physicochemical property (Eq. 6).

$$\log \frac{P_{R-X}}{P_{R-H}} = \pi_X \quad (6)$$

Further Hansch and Fujita combined σ and π in one equation and correlated these parameters with the biological response using multiple linear regression (Eq. 7):

$$\log \frac{1}{C} = a\sigma + b\pi + \text{const.} \quad (7)$$

where C is the molar concentration or dose that leads to a defined biological effect (for example EC_{50} , LD_{50}). This formalism, known nowadays as a Hansch analysis, provides a flexible model that is able to describe different kinds of biological activities, especially from *in vitro* experiments. Later it was extended to more than two parameters and also included nonlinear terms to describe partitioning in more complex systems like organisms (Hansch and Fujita). The first extension, also derived by Hansch, used a parabolic dependence on $\log P$ to model partitioning. However it proved to be inappropriate for describing all kinds of lipophilicity-activity relationships. Using computer simulations of drug transport, Kubinyi developed the so-called “bilinear model” that described

nonlinear relationships in a much more flexible way (Eq. 8).

$$\log \frac{1}{C} = a \log P - b \log(\beta P + 1) + \text{const.} \quad (8)$$

In the same year, 1964, Free and Wilson independently developed a model in which individual substituents X_i were additively contributing to the biological activity. It can be formulated according to Eq. 9, where a_i ($i = 1, 2, \dots, n$) are the individual contributions of all n substituents attached to the molecules at various positions and is the calculated activity of a reference compound. The presence of a given substituent is indicated by 1 and its absence by 0. Any compound can be taken as a reference and often the unsubstituted derivative is used.

$$\log \frac{1}{C} = \sum a_i + \mu \quad (9)$$

The Free-Wilson analysis can be easily applied, as no physicochemical parameters are needed. However it has several restrictions compared to the Hansch analysis. First, no information can be gained about the effects of substituents not present at a given position. Second, if a large number of parameters are involved, the regression analysis can lead to statistically insignificant contributions from many substituents and loss of useful information. This necessitates a thorough selection of the relevant substituents, which can be done by stepwise regression analysis or, more efficiently, by evolution strategies or genetic algorithms (see below). It is also possible to derive the underlying physicochemical properties from an analysis of the contributions of the substituents by a Free-Wilson analysis deriving in this way a Hansch analysis (1). Both types of analysis can be combined; positions with larger structural variation can be described by physicochemical descriptors, those with small variation by Free-Wilson indicator variables. A comprehensive overview on classical QSAR is given in (2).

Molecular Descriptors Used in Classical QSAR

Three kinds of molecular descriptors can be classified:

- based on physicochemical properties, either measured or calculated,
- based on topological properties,
- derived from quantum mechanical calculations.

Descriptors based on physicochemical properties can be largely divided into three groups: (i) electronic, (ii) lipophilic and (iii) steric (bulk). Those that characterize electronic properties are σ -Hammett values and related scales, \mathcal{F} -values that represent the inductive (field) effect of a substituent and \mathcal{R} -values that characterize the resonance effect of a substituent, spectroscopic data

like IR-frequencies of a common functional group and NMR chemical shifts. The latter descriptors are derived from the actual series of derivatives and usually describe the electronic effects of the substituents better than those derived from model systems. The drawback of such descriptors is that they are available only for derivatives already synthesized. Parameters describing hydrogen bonding could also be related to this group. Interestingly, they have been mostly neglected in classical QSAR, with the exception of the linear salvation energy relationship (LSER) developed by Kamlet and Taft and independently by Abraham. Another approach to describing hydrogen bond acceptor and donor properties has been developed by Raevsky. In the second group of parameters the logarithm of the partitioning coefficient (P) is mostly used. For the organic phase, 1-octanol is now the established standard organic phase, though it was shown that other organic phases and especially liposomes can provide better model systems. Another parameter to characterize the lipophilicity of a substituent is the π -value introduced by Hansch. Fragment values of atoms and/or functional groups are also used. There are several computational methods that aim to calculate $\log P$ solely from the molecular structure (3). None of them has been proven to be accurate in all cases, so they should be used with care, establishing that they are able to calculate correct $\log P$ differences for the series of compounds investigated by comparison of measured with calculated values for a few compounds.

Lipophilicity parameters refer to the neutral, unionized form of the compound. However, most drugs are partly ionized at physiological pH. As long as the pK_a values of the compounds under investigation do not differ too much this poses no problem in the QSAR analysis. However if the pK_a values differ significantly, the different degree of ionization has to be taken into account. Partitioning of acids and bases is pH-dependent. For bases the partition coefficient is constant at high pH and decreases with lower pH when ionization starts. When pH is further decreased the partition coefficient decreases by about one log-unit per pH unit until the $\log P$ of the ionized form is reached. The difference between the partitioning of the neutral and the ionized form has been found to be 3–4 orders of magnitude in many cases. To account for the different degree of ionization, the distribution value $\log D$, or $\log P_{app}$, can be measured at a certain pH or calculated from $\log P$. Neglecting the partitioning of the ionized form, Eq. 10 is used for acids and Eq. 11 for bases.

$$\log D = \log P - \log(1 + 10^{pH - pK_a}) \quad (10)$$

$$\log D = \log P - \log(1 + 10^{pK_a - pH}) \quad (11)$$

It is nowadays more and more recognized that membranes are an integral part of biological systems and that even passive transport across membranes can be a very selective process so that no single model partitioning system can account for transport across all kinds of membranes (4).

The third group of parameters relates to the steric (bulk) properties of the substituents. The molar refractivity MR, a measure of compound volume and polarizability is very popular. Another steric descriptor is the van-der-Waals volume V_w derived from Bondi. Both measures assume that the substituent is a sphere and neglect differences in spatial requirement in 3D space of nonspherical substituents. The STERIMOL parameters, developed by Verloop are an attempt to incorporate the shape of substituents. The substituent is described by its length (L), minimal width (B_1) and maximal size (B_4 , B_5). The major problem in this approach is that it requires up to four descriptors per substituent position, in this way necessitating a larger data set.

Soon after the start of the modern QSAR methodology many molecular descriptors based on topological properties of the molecular graphs or information indices were calculated and investigated for their use in QSAR analysis. They can easily be computed from the 2D representations of the chemical structures; however, even if a significant relationship is found, it is hard to interpret their meaning.

Quantum chemical, molecular orbital-based parameters are also used. Yet computed structural parameters may only be applicable to the set of compounds under investigation, while physicochemical characteristics of molecules can, in principle, be extended to other groups of molecules. A comprehensive summary of QSAR parameters is given in (5).

Variable Selection

The aim of a QSAR analysis is the development of a model that can fit the data, can predict the property of interest and can be interpreted to guide further synthesis and optimization of the desired properties. The availability of several hundred molecular descriptors poses the problem of selecting the “right” ones. Another problem relates to the increased risk of chance correlations by testing a large number of potential descriptors. Several statistical techniques like forward and backward elimination of variables or stepwise regression are used to cope with large numbers of potential descriptors. However, experience has shown that they often end up in a local optimum, especially if large numbers of variables are considered. Evolutionary and genetic algorithms have been proven to be superior, as they generate optimal or nearly optimal models very fast and efficiently.

To decide on the best QSAR models is not straightforward. The best fit can easily be checked by statistical parameters of the regression and cross-validation is often used to check predictivity. Here one compound at a time is removed from the data, the model is developed and the activity of the missing compound is predicted from the model. This process is repeated till all compounds have been left out once (leave-one-out, LOO-cross-validation). A more stringent way is to remove a larger part of the compounds temporarily from the analysis and to predict their activities from the model generated from the reduced data set (leave-group-out). Again this is repeated until all compounds have been left out and their activities have been predicted. As this procedure is not deterministic, it is necessary to repeat it several times, selecting the compounds to be left out at random, to get a valid statistical estimate. The predictive power of the model is measured by Q^2 as calculated according to Eq. 12, where y_{obs} , and y_{pred} are the observed and predicted activities and y_{mean} is the mean activity of the training set.

$$Q^2 = 1 - \frac{\sum_y (y_{\text{pred}} - y_{\text{obs}})^2}{\sum_y (y_{\text{obs}} - y_{\text{mean}})^2} \quad (12)$$

The real criterion for the validity of a model can only be its predictive ability for an external test set that has never been used in the process of model development. To select a proper test set from the available compounds is a hard problem. The test set should be representative and cover the chemical space as much as possible. On the other hand the chemical spaces of the training set used to build the model and the test set must not be too different. Often cluster or principal component analysis of the available compounds is performed in the descriptor space and from the resulting clusters one or more compounds are selected for the test set, while the others are used to form the training set. This ensures that training and test sets cover nearly the same chemical space but it can result in overoptimistic estimation because of similarity between the two sets. Another approach is to randomly select the test set from the available compounds, taking into account that the test set should span a similar range of activities as the training set. Ideally, in either case the procedure should be repeated several times, which is rarely done. The predictivity for the test set can be estimated by linear regression of the actual and predicted activities. This measure is biased as the slope of the regression line and the intercept are fitted. Q^2 can also be applied to the test set. However, Q^2 values can

be biased if the means of the activities of the training and the test set differ. Silverman has developed a new figure of merit P^2 that gives an unbiased estimate (Eq. 13), where p_i and p_j are the predicted activities, m_i and m_j are the measured activity values of the external test set and the summation runs over all pairs of compounds i and j in the test set (6).

$$P^2 = 1 - \frac{\sum_{i,j} [(p_i - p_j) - (m_i - m_j)]^2}{\sum_{i,j} (m_i - m_j)^2} \quad (13)$$

Though cross-validation is accepted as a measure of internal predictivity, it is not a reliable measure of external predictivity (7). A more reliable procedure to test the validity of the model and to avoid chance correlations is the so-called Y-scrambling. Here one attempts to correlate randomly ordered activities (Y-values) with the original descriptors. If no models of comparable fit are obtained after several scrambling runs, there is some confidence in the validity of the tested model.

Three-Dimensional QSAR

A limitation of classical QSAR is that it does not explicitly consider the 3D properties of the compounds such as shape and [conformation](#). It is therefore limited to homologous, or congeneric series of derivatives that share a common skeleton. In 1988 a method called comparative molecular field analysis (CoMFA) was published. This described molecular properties by fields around the molecules calculated as interaction energies on a regular grid (see below). The field values are correlated with biological activity values using partial least squares (PLS) analysis. A set of molecules that are assumed to act *via* the same mechanism is selected, but they need not have the same molecular skeleton. For all molecules, low energy 3D structures are generated. If rigid analogs or analogs having different conformational constraints are available this helps to define the bioactive conformations of the flexible molecules. Then an alignment is performed by superimposing all molecules according to the [pharmacophore](#) suggested. The alignment is the most difficult and important step in the analysis. However, it is not necessary to find and use the real bioactive conformations for the alignment, providing the essential pharmacophore pattern is correctly matched. For example, considering a set of molecules than can exist in an extended or folded conformation, very similar results will be obtained, as long as all molecules are modeled either in the extended or folded conformation. This follows from the way the field values are

correlated with the biological activities by PLS (see below). The overlaid molecules are put into a box that is sufficiently large and the interaction energies with a probe atom are calculated. Usually a C_{sp3} carbon with plus 1 charge is used as a probe atom. The box is divided into many cubes of a reasonable size (usually a default value of 2 Å is used), the probe atom is put on every grid point and the steric and electrostatic interaction energies between the probe and every molecule are calculated. Besides the steric (Lennard-Jones potential) and electrostatic (coulombic potential) interactions other fields like hydrogen bond donor and acceptor or hydrophobic fields can be considered. The interaction energies calculated at the grid points are unfolded to yield a table where each column corresponds to a given grid point and type of interaction and each row represents one molecule under investigation. As the interaction energies increase steeply near the molecular surface due to the form of the Lennard-Jones and the coulombic potentials, high values are truncated to a predefined cutoff (usually ± 30 kcal/mol) in order to decrease their influence on the result of the PLS analysis. PLS is a multivariate statistical technique that, like principal component analysis, extracts the maximum variance from a block of variables as fewer uncorrelated variables called principal components or latent variables. Unlike principal component analysis, in PLS the latent variables are extracted so that they cover a maximum of the variance of the X-block (structural parameters) and simultaneously are maximally correlated to the Y-block (the activity parameters). Cross-validation is used to determine the number of latent variables to be extracted. The results can be visualized by contour maps showing favorable and unfavorable regions of the different fields. The visual display can guide the decision as to in which parts of the molecules structural changes should be made to enhance biological activity. However, the contour maps are often fragmented and not easily interpreted. Additionally, the standard settings for the display of the contour maps can lead to wrong impressions. Over the years CoMFA has become the most popular 3D QSAR method. A critical review of published CoMFA results together with an overview on new trends can be found in (8).

Another popular 3D QSAR approach developed by Klebe and coworkers is the comparative molecular similarity indices analysis (CoMSIA). In CoMSIA, similarity fields, also calculated on a regular grid, are used instead of interaction fields. Gaussian functions centered at the atoms are used to describe different atomic properties and are compared to give a measure of similarity. The use of Gaussian functions eliminates the need for a cut-off and additionally, tolerates small

deviations in the molecular superposition without a significant impact on the results due to their smoother characteristics. CoMSIA contour maps are also usually smoother than CoMFA maps. In contrast to CoMFA maps that surround the regions with favorable and unfavorable contributions to activity, CoMSIA maps are directly located at the favorable and unfavorable groups and therefore easier to interpret. For an overview of CoMFA and other 3D QSAR methods see (9).

Other types of 3D approaches have also been developed that do not depend on the alignment of the molecules. The grid-independent descriptors (GRIND) are autocorrelation transforms of 3D field values that are independent of the relative orientation of the molecules (10). Another approach to describe molecular properties independent of alignment is proposed by Baumann (11).

“Multidimensional” QSAR

The limitation of 3D QSAR methods like CoMFA and CoMSIA to a single conformation is overcome by the recently proposed multidimensional QSAR. Instead of one conformation, an ensemble of conformations and alignments can be considered, inserting the “fourth” dimension in the QSAR methods (12, 13). Taking additionally into account the induced fit of the biological target, an introduction of the “fifth” dimension has also been proposed (14).

References

- 1) Kubinyi H (1990) The Free-Wilson method and its relationship to the extrathermodynamic approach. In: Hansch C, Sammes P, Taylor J (eds) *Comprehensive Medicinal Chemistry. The Rational Design, Mechanistic Study & Therapeutic Application of Chemical Compounds*, vol 4. Pergamon Press, Oxford, pp 589–645
- 2) Kubinyi H (1993) QSAR: Hansch Analysis and Related approaches. In: Mannhold R, Krogsgaard-Larsen P, Timmerman H (eds) *Methods and Principles in Medicinal Chemistry*, vol 1. VCH, Weinheim
- 3) Mannhold R, van de Waterbeemd H (2001) Substructure and whole molecule approaches for calculating log P. *J Comput-Aided Mol Des* 15:337–354
- 4) Seydel JK, Wiese M (2002) Drug-Membrane Interactions. In: Mannhold R, Krogsgaard-Larsen P, Timmerman H (eds) *Methods and Principles in Medicinal Chemistry*, vol 15. VCH, Weinheim
- 5) Todeschini R, Consonni V (2000) Handbook of Molecular Descriptors. In: Mannhold R, Krogsgaard-Larsen P, Timmerman H (eds) *Methods and Principles in Medicinal Chemistry*, vol 11. VCH, Weinheim
- 6) Silverman BD (2000) The Thirty-one Benchmark Steroids Revisited: Comparative Molecular Moment Analysis (CoMMA) with Principal Component Regression. *Quant Struct Act Relat* 19:237–246

- 7) Golbraikh A, Tropsha A (2002) Beware of q². *J Mol Graphics Modelling* 20:269–276
- 8) Kubinyi H, Folkers G, Martin YC (1998) 3D QSAR in Drug Design, vol 3. Recent Advances, Kluwer/ESCOM, Dordrecht
- 9) Kubinyi H (1993) 3D QSAR in Drug Design. Theory Methods and Applications, ESCOM Science Publishers, Leiden
- 10) Pastor M, Cruciani G, McLay I et al (2000) Grid-Independent Descriptor (GRIND): A Novel Class of Alignment-Independent Three-Dimensional Molecular Descriptors. *J Med Chem* 43:3233–3243
- 11) Baumann K (2002) Distance Profiles (DiP): A translationally and rotationally invariant 3D structure descriptor capturing steric properties of molecules. *Quant Struct Act Relat* 21:507–519
- 12) Santos-Filho OA, Hopfinger AJ (2002) The 4D-QSAR Paradigm: Application to a Novel Set of Non-peptidic HIV Protease Inhibitors. *Quant Struct Act Relat* 21:369–381
- 13) Vedani A, McMasters DR, Dobler M (2000) Multi-Conformational Ligand Representation in 4D-QSAR: Reducing the Bias Associated with Ligand Alignment. *Quant Struct Act Relat* 19:149–161
- 14) Vedani A, Dobler M (2002) Multidimensional QSAR: Moving from three- to five-dimensional concepts. *Quant Struct Act Relat* 21:382–396

Quantum Yield

Definition

Quantum yield is defined as the yield of fluorescence photons per absorbed photons. It denotes a parameter that generally quantifies the efficiency of dye molecules to fluorescence.

► [Fluorescence Correlation Spectroscopy](#)

Quasispecies

Definition

Viruses with RNA genomes such as HIV–1 display a high level of genetic variation, and therefore exist as distributions of closely-related genomic variants, referred to as quasispecies, which evolve in response to selection pressure. The best known example is haemoglobin, consisting of 4 non-covalently associated folded chains (two alpha-chains and two beta-chains).

► [Reverse Transcriptase](#)

Quaternary Structure

Definition

The quaternary structure of a protein describes the interconnection and arrangement of polypeptide chains within a protein. Only proteins consisting of more than one polypeptide chain have a quaternary structure.

► [Classification of Active Centers](#)

► [Protein Databases](#)

Quiescence

Definition

Quiescence describes a reversible cell-cycle arrest which is caused by the absence of mitogens.

► [Senescence](#)

RA

- Rheumatoid Arthritis

Rac

- Rho, Rac, Cdc42

RACE

Definition

RACE stands for ‘rapid amplification of cDNA ends’ (also known as ‘anchored PCR’). It is a technique that allows for the amplification of mRNA for which partial sequence information is available. RACE is commonly used to identify the insertion site of the vector in gene-trapping experiments by amplifying the 5’ or 3’ end of the fusion transcript (a fusion of the endogenous exonic sequence of the locus and the reporter gene).

- [Mouse Genomics](#)

RAD51

Definition

Radiation sensitivity gene 51 was originally identified in a yeast mutant as the eukaryotic homologue of the bacterial RecA protein. Such genes are involved in repair of DNA double strand breaks by homologous recombination.

- [Chromosomal Instability Syndromes](#)

Radiation Hybrid Map

Definition

Radiation hybrid map describes a physical analogue of linkage mapping in which recombinations are replaced by radiation-induced breaks. Phenotypic markers can only be used in rare cases in such mapping, but radiation hybrid panels have been useful in the mapping of DNA sequence polymorphisms.

- [YAC and PAC Maps](#)

Radiation Sensitivity Gene 51

- [RAD51](#)

Radioactive Isotope

Definition

Radioactive isotope (radioisotopes) consists of unstable atoms with an unusual number of neutrons. This causes them to decay radioactively, during which they give off energy and subatomic particles. Radioisotopes do occur naturally, but are also artificially produced by bombarding particular elements with neutrons. Radioisotopes are used due to their chemical properties, for instance as tracers, and as sources of radiation.

- [Molecular Imaging Mod.](#)

Radioactivity/Radioactive Decay

Definition

The spontaneous disintegration of certain atomic nuclei (isotopes) is named radioactive decay. This process is accompanied by the emission of high energy radiation or particles and the energy produced during decay is the

basis for detection. The isotopes ^3H , ^{14}C , ^{15}N , ^{32}P , ^{33}P and ^{35}S are frequently used in biological experiments.

► [Protein Microarrays as Tool for Protein Profiling and Identification of Novel Protein-Protein Interactions](#)

Radixin

► [ERM Protein](#)

Rafts

Definition

Lipid rafts are small (<100 nm diameter) cell membrane domains enriched in sphingolipids and cholesterol that accommodate a selective set of proteins. These domains are biochemically defined by their resistance to detergent solubilisation at 4°C, owing to the tight packing of their lipid acyl side chains, which display an unusually high melting temperature and are destabilised by cholesterol- and sphingolipid-depleting agents.

► [Biological Membranes](#)

Ragged Red Fibres

► [RRF](#)

Ran

Definition

Ran is a small G protein of the Ras family that regulates the directionality of nucleocytoplasmic transport by cycling between its GDP- and GTP-bound form. However, since the intrinsic GTP hydrolyzing activity of Ran is very low, so-called effector proteins such as

► [RanGAP](#) and ► [RanGEF](#) are needed to accomplish

GDP/GTP switching.

► [Nuclear Import and Export](#)

► [Nuclear Pore Complex](#)

Ran GTPase Activating Protein

► [RanGAP](#)

Ran Guanine Nucleotide Exchange Factor

► [RanGEF](#)

Random Amplification of Polymorphic DNA Markers

Definition

Random amplification of polymorphic DNA markers is a genetic marker technique using PCR amplification from short (about 10 bp) segments of arbitrary sequences to look for polymorphisms.

► [Mutagenesis Approaches in Medaka](#)

RanGAP

Definition

RanGAP is an effector protein of Ran. RanGAP is localized in the cytoplasm and induces the hydrolysis of Ran-bound GTP to GDP. As a result, cytoplasmic RanGTP levels are low.

► [Nuclear Import and Export](#)

RanGEF

Definition

RanGEF besides RanGAP is another protein effector of Ran. RanGEF is found in the cell nucleus where it exchanges Ran-bound GDP for GTP. Therefore, in the nucleus, RanGTP levels are high.

► [Nuclear Import and Export](#)

Ran-GTP

Definition

Ran-GTP is a small GTPase that serves as a cofactor for karyopherin/exportin receptors.

►RNA Export

RANKL

Definition

Receptor activator of NfκB ligand (RANKL) (also named ODF, OPGL, or TRANCE) is a tumor necrosis factor (TNF)-like family cytokine and, like other family members, signals to the activation of the transcription factors NFκB and, through the kinase Jnk, c-Jun. RANKL induces differentiation of bone marrow hematopoietic precursor cells into bone-resorbing osteoclasts.

Rapamycin (Sirolimus)

Definition

Rapamycin is a 31-member macrocyclic lactam of the FK506 type, isolated from *Streptomyces Hygroscopicus*. It has antiproliferative properties and antifungal effects. It shows strong affinity for PPIases of the FKBP family.

►Peptidyl Prolyl Cis/Trans Isomerases

Rapid-Eye-Movement Sleep

►REM (Rapid-Eye-Movement) Sleep

RAPDs

►Random Amplification of Polymorphic DNA Markers

Rare Mutation

Definition

A rare mutation is a permanent, heritable change in the genetic information that occurs with an allele frequency lower than 1% within the population.

►Pharmacogenomics

Ras

Definition

Ras stands for rat sarcoma. The Ras proto-oncogenes are normal cellular components. These genes encode 21 kDa proteins that bind to the inner surface of plasma membranes. The ►Ras proteins are thought to be important for the transduction of signals required for proliferation and differentiation. Mutation or overexpression of Ras results in activated p21. Abnormalities of p21 have been associated with a variety of human cancers, including a majority of colorectal cancers.

►Signal Transduction: Integrin-Mediated Pathways

►Neurofibromatosis Type 1 (NF1), Genetics

►Protein Prenylation

Ras Guanyl Nucleotide Exchange Factor

Definition

Ras guanyl nucleotide exchange factor refers to a protein that binds to Ras or a Ras-related GTP (guanosine 5'-triphosphate)-binding protein and activates it by stimulating it to release its bound GDP and bind GTP instead.

►Two-Hybrid System

Ras Proteins

Definition

Ras proteins are small GTPases that relay signals from receptor tyrosine kinases to the nucleus regulating cell differentiation, proliferation and apoptosis.

►Neurofibromatosis Type 1 (NF1), Genetics

Ras Signalling

REINHOLD SCHÄFER

Laboratories of Molecular Tumour Pathology and of Functional Genomics, Charité, University Medicine Berlin, Berlin, Germany
reinhold.schaefer@charite.de

Definition

Ras proteins are critical nodal points of highly conserved signal transduction pathways, which integrate diverse extracellular ligand-mediated stimuli via a wide spectrum of cell surface receptors with the intrinsic transcriptional machinery. Ras signalling pathways regulate cytoplasmic processes as well as gene expression and cell-cycle progression. Depending on the cellular context, the phenotypic end-points of Ras signalling are proliferation, survival, growth arrest or differentiation. Mutated Ras proteins are crucially involved in tumorigenesis.

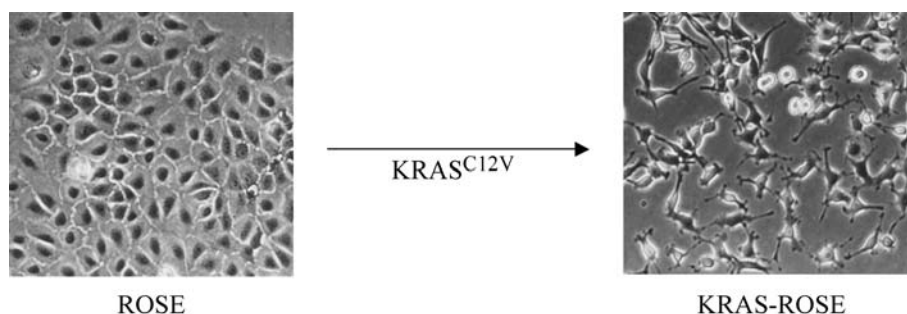
Characteristics

Identification of Oncogenic Ras Proteins by Reverse Genetic Approaches

Ras proteins are the founding members of a superfamily of small GTPases. Research on small GTPases and the signalling events associated with them was largely fuelled by the finding that the Ras proteins were encoded by viral and cellular **▶oncogenes**. Mutated and transforming versions of Ras proteins were identified in human and experimental cancers in the early 1980's. Ras proteins had already earlier been identified as **▶transforming proteins** in the genomes of RNA tumour viruses and had obtained their names from tumours in rodent cells infected by retroviruses. For example, Harvey sarcoma virus was found to encode the H-ras protein, Kirsten sarcoma virus the K-ras protein. In mammalian cells, the

transforming activity of Ras proteins was recognised in gene transfer experiments aiming at identifying cancer genes of cellular rather than of viral origin. For this purpose, genomic DNA prepared from cancer cells was transfected into normal murine recipient cells and the populations of transfectants were examined for morphologically altered cell colonies ("foci"). After cloning and further cultivation, the tumour DNA transfectants were injected into nude mice and gave rise to progressively growing, undifferentiated tumours. Comparing the transferred oncogene causing the tumorigenic conversion with the **▶proto-oncogene** present in normal cells revealed **▶point mutations** preferentially affecting amino acids 12, 13, and 61. The typical effect of Ras oncogenes introduced into immortalised, phenotypically normal rat recipient cells is depicted in Fig. 1. However, depending on the cell type, Ras proteins may exert other effects as well (Table 1). The prerequisites for an efficient oncogenic conversion of normal cells triggered by Ras oncogenes have been worked out in detail during recent years. Human cells require expression of telomerase for **▶immortalisation** and the inactivation of the TP53 and RB **▶tumour suppressor gene** pathways for Ras-induced oncogenesis (1).

Initially, Ras signalling *via* the non-mutated form of the protein was studied in genetically tractable model organisms. In *Drosophila melanogaster* the Ras signal transduction pathway controls eye development, in *Caenorhabditis elegans* the development of sex organs. In *Saccharomyces cerevisiae* the plasma membrane proteins Ras1 and Ras2 sense the nutritional status of the environment, e.g., the abundance and quality of available carbon sources. In particular, the genetic dissection of eye development in *Drosophila* allowed the elucidation of the order of protein interactions between the receptor level, receptor adapters, nucleotide exchanging factors and Ras. Protein interaction studies permitted identification of



Ras Signalling. Figure 1 The transfer of a recombinant expression vector harbouring the *KRAS* oncogene into immortalized rat ovarian surface epithelial cells results in epithelial-mesenchymal transition. Cells lose their epithelial shape and become motile. *KRAS* expression is associated with the loss of E-cadherin, a protein controlling adhesion of epithelial cells (Ref. 2).

Ras Signalling. Table 1 Selection of RAS oncogene effects mediated by forced expression in mammalian cells

Recipient cells	Phenotypic changes	Properties of cells
Mouse 3T3, rat 208F fibroblasts	Focus formation, loss of density-dependent growth control, anchorage independence, tumorigenicity	Immortal, spontaneously transformed
Rat ovarian surface epithelial cells	Epithelial-mesenchymal transition, anchorage dependence, tumorigenicity	Immortal, spontaneously transformed
Diploid rat fibroblasts	Focus formation, loss of density-dependent growth control, anchorage independence, tumorigenicity	Immortalisation by complementary oncogene required (e.g. Myc), additional genetic alterations (loss of tumour suppressor gene function)
Rat PC12 pheochromocytoma cells	Growth arrest, neurite formation	Chromaffin cell precursors
Human diploid fibroblasts	Premature senescence	Limited life span
Immortalized human fibroblasts and epithelial cells	Loss of density-dependent growth control, anchorage independence, tumorigenicity	Immortalisation by hTERT (catalytic subunit of telomerase), inactivation of tumour suppressors TP53 and RB by SV40

cytoplasmic kinases downstream of Ras in the signalling pathway. Initially it was believed that activated Ras would activate a simple linear cascade of cytoplasmic kinases in order to complete the link between the cell surface and the nucleus. However, Ras can affect multiple effector-mediated signalling pathways and is a component of a complex cellular signalling network showing mutual regulation and ►[cross-talk](#) (Fig. 2).

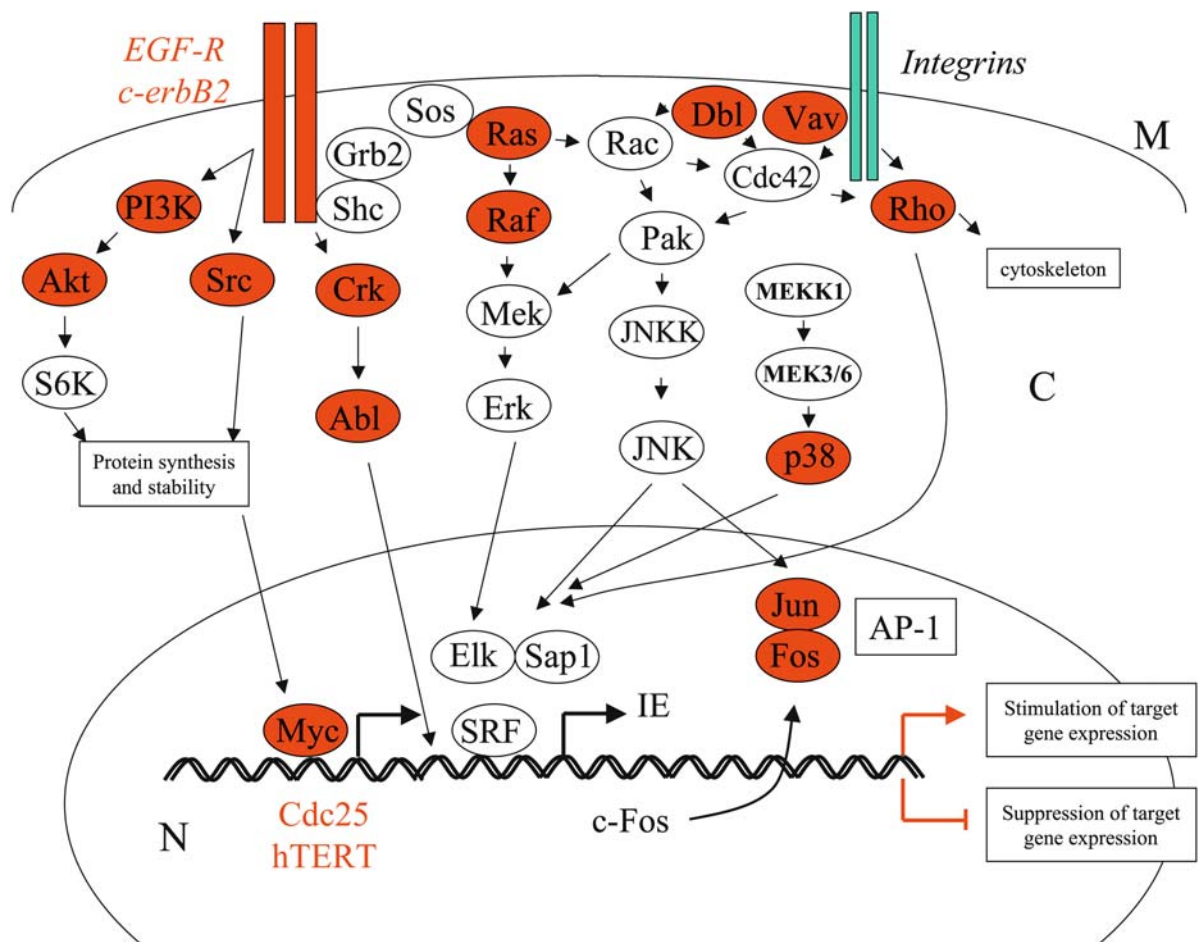
Biochemical Properties of Ras Proteins – Molecular Switches

The three mammalian *ras* genes encode four highly related 21 kD GDP/GTP-binding and hydrolysing proteins, which share approx. 90% amino acid identity, H-ras, K-Ras4A and K-ras4B (two alternatively spliced forms differing in their COOH-termini) and N-ras. The four Ras proteins show complete amino acid identity in regions important for interaction with several binding proteins (effectors). Ras proteins are located at the inner face of the plasma membrane. Their association with the membrane is mediated by a series of post-translational modifications, which target Ras proteins to functionally relevant micro-domains of the plasma membrane and are essential for their biological activity. This modification occurs in two consecutive steps. In the first step, the carboxyl terminus (CAAX-motif of the protein sequence) is farnesylated. Trafficking to the membrane is completed by ►[palmitoylation](#) at cysteine residues amino-terminal to the CAAX-motif of H-Ras, K-Ras4A and N-Ras. In the K-Ras4B protein, a lysine-rich sequence is responsible for membrane association.

Ras proteins exist in two functional conformations. The GDP-bound proteins are inactive; the GTP-bound forms are active. GDP can be exchanged for GTP (nucleotide dissociation) and GTP can be hydrolysed to GDP and P_i (hydrolytic GTP cleavage). Upon GTP/GDP cycling, the conformation in protein regions called switch I (corresponding to amino acids 30–38) and switch II (amino acids 59–76) changes tertiary structure. Together with the flanking region ranging from amino acids 25–45, the switch I and II regions make up the effector domain of Ras, which is responsible for binding to interacting proteins. In the GTP-bound state, a high affinity binding site is exposed. Nucleotide exchange and GTP hydrolysis on Ras proteins are facilitated by guanine nucleotide exchange factors (GEFs) and GTPase activating proteins (GAPs), respectively. Multiple GEFs and GAPs specific for Ras proteins have been identified. GEFs enhance the release of bound nucleotides and promote loading with GTP because of the higher intracellular concentration of GTP relative to GDP. GAPs enhance the rate of GTP hydrolysis and thus promote inactivation of Ras proteins. Oncogenic forms of Ras are locked in their active GTP-bound state and are insensitive to GAP activity (for review see refs 4 and 5).

Signalling Upstream of Ras Proteins

Extracellular signals that activate Ras proteins can be mediated through different types of receptors, tyrosine kinase receptors (e.g. platelet-derived growth factor receptor, PDGFR or epidermal growth factor receptor,



Ras Signalling. Figure 2 Schematic and simplified view of the cellular signalling network involving Ras proteins and some of their effectors (modified from ref. 3). Proteins marked in red have been shown to exhibit transforming activity in mammalian cells. M, plasma membrane; C, cytoplasm; N, nucleus; IE, immediate early genes.

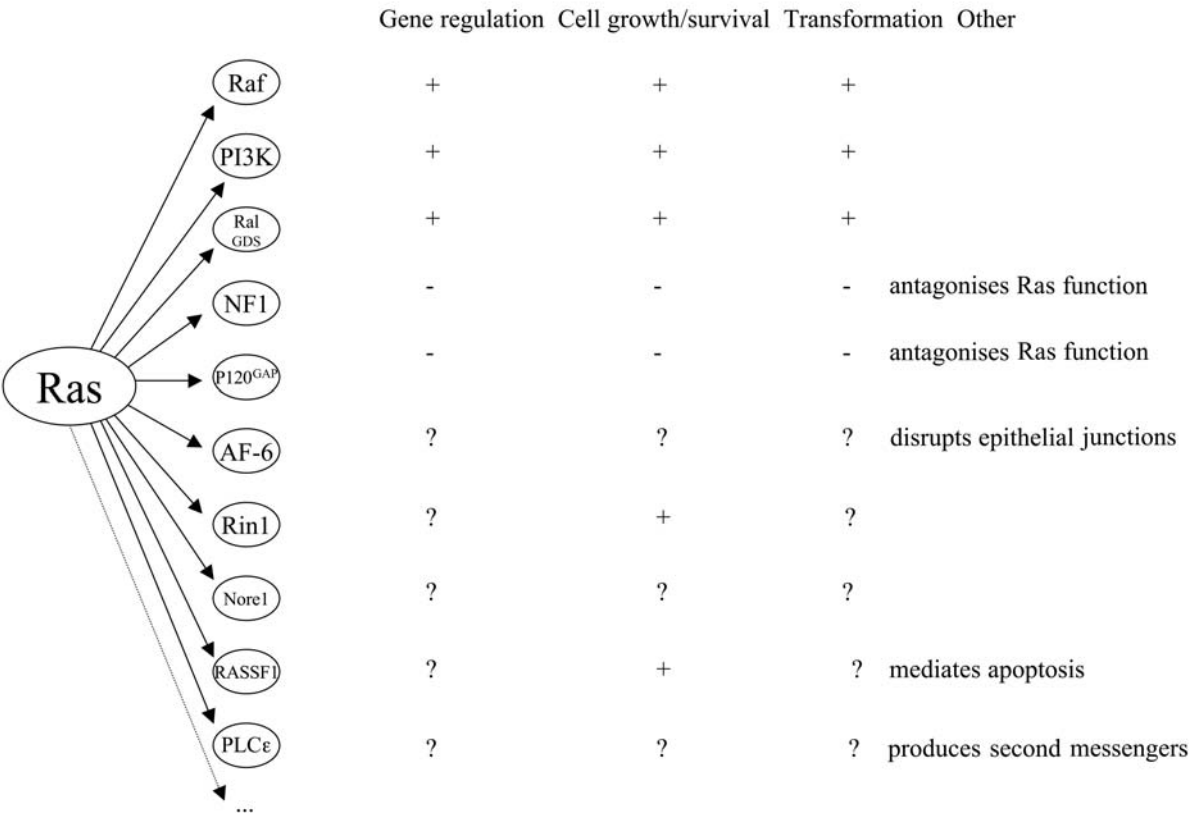
EGFR), cytokine receptors (e.g. the interleukin-2-receptor), T-cell receptors and subunits of heterotrimeric G-proteins. One of the most studied signalling processes related to Ras function is Erk activation, mediated by growth factors and tyrosine kinase receptors. Following ligand binding, Src homology domain (SH2)-binding proteins such as Grb2 or Shc bind to the phosphotyrosine residues of the activated receptor. Grb2 binds to and activates Sos, a GEF for Ras. Alternative upstream activation mechanisms involve the subunits of G-proteins or protein kinase C. The nucleotide exchange factor Sos forms a complex with Ras resulting in the release of GDP. The high intracellular GTP/GDP ratio causes replacement of GDP by GTP. The GEF then dissociates from Ras-GTP. Besides Sos, the Ras exchange factors RasGRF (cdc25Mm), RasGRP, CNRasGEF and phospholipase C- ϵ , may promote GTP/GDP exchange.

Since the intrinsic dissociation of Ras-GTP is slow and the intracellular GTP concentration is high, GAPs are required for stimulating GTP hydrolysis. Several GAPs are known in mammalian cells including p120-GAP and the neurofibromatosis type I protein (NF1).

Signalling Downstream of Ras Proteins

GTP-bound Ras proteins transmit signals down the pathway by interacting with effector proteins. A wide spectrum of proteins that interact specifically with active Ras has been isolated (Fig. 3) and it is very likely that various signals can be transduced via distinct effectors. Complex phenotypes can be achieved by the action of several signalling pathways downstream of Ras. In view of the clinical importance of permanent Ras activation, the pathways involved in tumorigenic conversion have been investigated in detail; however, the precise contribution of each effector pathway is still

Functional roles of Ras effector pathways



Ras Signalling. Figure 3 Effector proteins interacting with Ras-GTP. (+) a functional role in regulating gene expression, cellular proliferation, survival or malignant transformation has been established. (–) antagonistic function to Ras-GTP, stimulation of GTP-GDP hydrolysis and inactivation, GAPs may exert transformation suppressing function. (?) functional contribution yet unknown.

unclear. These pathways include the Raf serine/threonine kinases, Ral GEFs and phosphatidylinositol 3-kinase (PI3K). The pathways downstream of Ras have, in part, been functionally dissected by using Ras effector domain mutants. The H-Ras (12V/37G) and H-Ras (12V/40C) mutants fail to bind to Raf and activate the Ral-GDS and PI3K-pathways respectively, while the H-Ras (12V/35S) mutant preferentially activates the Raf pathway. In addition to Raf, PI3K and Ral, Ras utilises related small GTPases such as members of the Rho family to exert effects on growth and transformation (5).

Ras-GTP mediates the translocation of the serine/threonine kinases Raf (A-Raf, B-Raf and c-Raf-1) to the membrane, where other accessory factors activate the kinase function. Raf, in turn, activates the dual specificity kinases Mek 1/2. Activated MEK phosphorylates and activates ERK 1/2 mitogen-activated kinases (▶MAPKs). Phosphorylated Erk activates the cytoplasmic target p90^{RSK} serine/threonine kinase to

regulate protein synthesis and translocates into the nucleus to activate transcription factor targets such as Elk 1 and Ets family members (Fig. 2). In this way, mutated Ras proteins achieve a permanent activation of the down-stream signalling cascade and the reprogramming of gene expression. The Raf/MAPK pathway has been found to play a critical role in the control of cell-cycle progression. However, it can also protect against certain apoptotic stimuli.

In addition, Ras interacts with the p110 catalytic subunit of PI3K and stimulates its lipid kinase activity. This results in the conversion of phosphoinositides that bind to the serine/threonine kinase Akt/PkB and Rac GDP-GTP exchange factors. Akt/PkB can phosphorylate multiple targets that prevent apoptosis in various ways. One of the Akt-mediated effects is the inactivation of the AFX forkhead transcription factor. This inhibits AFX from entering the nucleus and prevents expression of the death-activating Fas ligand. Another effect is activation of IKK, which causes inactivation of



the NF κ B-inhibitor I κ B. NF κ B can translocate into the nucleus and stimulate expression of anti-apoptotic genes. PI3K also causes activation of the small GTPase Rac, which can also activate NF κ B *via* the production of reactive oxygen species. Although the PI3K/PkB/Akt pathway was initially implicated in mediating anti-apoptotic signals generated by growth factors, extracellular matrix and Ras, it can also contribute to cell-cycle progression.

Several lines of evidence suggest that RalGDS and Ral GTPase contribute to Ras-mediated transformation. Ras-mediated activation of RalGDS, like Raf activation, involves recruitment to the membrane and subsequent activation of RalA and B. Downstream of Ral, the Ral-binding protein RBP1, a GTPase activating protein, negatively regulates Rac and cdc42 GTPases. Another Ral interaction protein is phospholipase D1. The RalGDS effector pathway may affect the activity of transcription factors regulated by Rho GTPases (4, 5).

Studies on Ras-induced transformation of fibroblasts have revealed that the activity of Rho family GTPases is needed. Signalling crosstalk involves several members of the Rho family such as RhoA and B, Rac1,

cdc42 and others. A functional link between RhoA-GTP and cell-cycle regulation is established by inhibition of p21^{Waf/Cip1}. Ras and Rho pathways regulate the nuclear targets Elk-1 and SRF, respectively. These factors in turn act synergistically to activate c-Fos transcription.

Nuclear Targets of Ras Signalling Pathways

Transcription factors activated as a consequence of Ras signalling comprise Ets 1/2, Elk-1, NF κ B, SRF, c-Fos, c-Jun, c-Myc and E2F. The regulatory sequences of genes up-regulated on Ras activation harbour binding sites for these transcriptional regulators. The functional roles of these binding sites readily explain the impact of Ras signalling on the stimulation of gene activity. The up-regulation of genes exhibiting a critical function in the tumorigenic process may explain why the molecular Ras switch is of central importance for the malignant properties of cancer cells (6). More recently, the number of candidate Ras-responsive genes has increased dramatically due to comparative transcriptome analysis of normal and Ras-transformed cells. Detailed functional studies are required to find out which of these candidate

Ras Signalling. Table 2 Selection of Ras-responsive genes and their functional roles in transformation and tumorigenesis

Gene product	Function
Up-regulated upon RAS-mediated signalling	
Cyclin D1	Cell-cycle progression
COX-2 (inducible cyclooxygenase-2)	Promotion of cell survival
MMPs (matrix metalloproteases)	Destruction of extracellular matrix, invasion, metastasis
VEGF (vascular endothelial growth factor)	angiogenesis
RhoC	metastasis
Fra-1	Transformation, motility, invasion
Down-regulated upon RAS-mediated signalling	
TMs (high molecular weight tropomyosins)	Stabilisation of actin filaments, maintenance of cell shape
Lox (lysyl oxidase)	Stabilisation of extracellular matrix
Tsp-1 (thrombospondin)	Inhibition of angiogenesis
Timp-1 (tissue inhibitor of metalloproteases)	Inhibition of invasive properties
E-cadherin	Cell-cell adhesion, inhibition of invasive properties
Several transformation suppressors: WT1 (Wilms tumour suppressor) HTS1 (HeLa tumour suppressor) Lot1 ("lost in ovarian tumours", putative suppressor of ovarian carcinoma)	Tumour suppression

genes are essential for the transformed phenotype or are secondary to transformation. Gene expression profiling studies have also suggested that the number of genes down-regulated equals the number of up-regulated genes (2, 7). The mechanisms for target gene repression upon expression of Ras oncogenes are currently unknown. Some Ras-responsive genes exhibiting a critical role in proliferation and transformation are listed in Table 2.

Clinical Relevance

Ras mutations that lock the molecular switch in its active state have been found overall in about one third of all human cancers. The prevalence of mutations differs in distinct tumour types, e.g. approx. 90% in pancreatic cancer, about 50% in carcinomas of the colon, lung and thyroid. Mutations are infrequent in carcinomas of the breast, ovary and cervix (8). However, there are alternative mechanisms leading to persistent Ras activation in cancer. Over-expression of Her2/neu/ErbB2, a receptor tyrosine kinase, is prevalent in mammary and ovarian cancer. Similarly, inactivating mutations of NF1 encoding a GAP in malignant schwannomas produce high Ras-GTP levels. Moreover, the Ras-mediated signalling pathway may be activated by mutation of the downstream kinase Raf in melanoma and ovarian carcinoma. In view of the crucial role of Ras-mediated signalling in the control of proliferation and the tumorigenic phenotype, therapies targeting Ras proteins as well as active components of the signalling network such as membrane-bound receptors and cytoplasmic kinases have been developed (9) and are currently being tested in clinical trials.

References

1. Hahn WC, Counter CM, Lundberg AS et al (1999) Creation of human tumour cells with defined genetic elements. *Nature* 400:464–468
2. Tchernitsa OI, Sers C, Zuber J et al (2004) Transcriptional basis of KRAS oncogene-mediated cellular transformation in ovarian epithelial cells. *Oncogene* 23:4536–4555
3. Hunter T (1997) Oncoprotein networks. *Cell* 88:333–346
4. Malumbres M, Pellicer A (1998) Ras pathways to cell cycle control and cell transformation. *Frontiers in Bioscience* 3:887–912
5. Shields JM, Pruitt K, McFall A et al (2000) Understanding Ras: 'it ain't over 'til it's over'. *Trends Cell Biol* 10:147–154
6. Bernards R, Weinberg RA (2002) A progression puzzle. *Nature* 418:823
7. Zuber J, Tchernitsa OI, Hinzmann B et al (2000) A genome-wide survey of Ras transformation targets. *Nature Genet.* 24:144–152
8. Bos JL (1989) Ras oncogenes in human cancer: a review. *Canc Res* 49:4682–4689
9. Sebolt-Leopold JS (2000) Development of anticancer drugs targeting the MAP kinase pathway. *Oncogene* 19:6594–6599

Ras Superfamily of Small GTP Binding Proteins

Definition

Like all other GTPases, members of the Ras superfamily of small GTPases cycle between GDP and GTP-bound conformations. Generally, both binding and hydrolysis of GTP require assistance of accessory proteins. These include guanine nucleotide exchange factors (GEFs) which catalyze exchange of GDP for GTP, and GTPase-activating proteins (GAPs) which catalyze the hydrolysis of GTP to form GDP. In many cases, the GTP-bound conformation is membrane associated via an exposed isoprenoyl or fatty acyl anchor covalently attached to the GTPase, and interacts with specific effector proteins. Among the effector proteins of ARF-like GTPases are coat proteins, but also lipid-modifying enzymes.

► [Protein and Membrane Transport in Eukaryotic Cells](#)

Ras-Like GTPases

Definition

Ras-like GTPases belong to the large superfamily of monomeric 20–40 kDa GTP-binding proteins, which oscillate between GTP-bound (active) and GDP-bound (inactive) states. They serve as molecular switches in many fundamental cellular processes like proliferation and differentiation.

► [G-Proteins](#)

RasMol

Definition

RasMol is a program for molecular graphics visualization originally developed by Roger Sayle. (<http://www.openrasmol.org/>)

► [Protein Databases](#)

Rb

Definition

Rb is the Retinoblastoma susceptibility gene product. It functions as a negative regulator of the G1–S transition of

the cell-cycle, mainly through the repression of the E2F transcription factors. Its activity is regulated by phosphorylation by specific cyclin/cdk complexes. In the cancer syndrome retinoblastoma, the gene is mutated.

- ▶ [Cell Cycle – Overview](#)
- ▶ [Growth Factors](#)
- ▶ [Retinoblastoma \(Rb\) Protein](#)
- ▶ [Senescence](#)
- ▶ [Tumor Suppressor Genes](#)

RB1

- ▶ [Retinoblastoma Gene](#)

RBS

- ▶ [Ribosomal Binding Site](#)

RCA

Definition

RCA stands for rolling circle amplification (RCA). RCA displays an enzymatic method that is based on the polymerase chain reaction employed for signal amplification.

- ▶ [Protein Microarrays as Tool for Protein Profiling and Identification of Novel Protein-Protein Interactions](#)

RCDP

- ▶ [Rhizomelic Chondrodysplasia Punctata](#)

rDNA

- ▶ [Ribosomal DNA](#)

Reactive Oxygen Species

Definition

Reactive oxygen species are short-lived, highly reactive reduced products of oxygen such as the superoxide anion ($O_2^{\cdot-}$), hydroxyl radical ($\cdot OH$), nitric oxide ($NO\cdot$), peroxynitrite ($ONOO^-$) and ozone. They can react with most molecules in their vicinity inducing inflammatory and growth-promoting actions.

- ▶ [Free Radicals](#)
- ▶ [Morbus Wegener](#)

Reading Frame

Definition

During translation of a sequence of nucleotides into a protein, reading frame means reading progressively through the nucleotide sequence by grouping of non-overlapping triplets (codons). Since the codons are made up of three nucleotides, there are three possible frames; only one of them is the reading ("sense") frame, also called frame zero. If genes are not read in a multiple of three, or translation begins with a different starting point, they are considered out of the reading frame, spoiling regular translation.

- ▶ [Duchenne Muscular Dystrophy](#)
- ▶ [Genetic Code](#)
- ▶ [Open Reading Frame](#)

Readlength

Definition

Readlength refers to the total number of nucleotide bases that are generated from an individual sequencing reaction. Each nucleotide sequence is termed a read, so readlength refers to the total number of nucleotide bases in an individual read.

- ▶ [Shotgun Libraries](#)

Readthrough Suppression

Definition

Readthrough suppression in some retroviruses refers to the process by which ribosome can, at low frequency, bypass a terminator at the end of ►gag and continue translation of ►pro and ►pol in the same reading frame.

►Retroviruses

RecA

Definition

RecA stands for an essential recombination protein, which is highly conserved in prokaryotes and functions to anneal a single strand of DNA to a complementary DNA strand of a duplex DNA molecule.

►DNA Helicases

Real-time PCR

►Real-Time Polymerase Chain Reaction

Real-Time Polymerase Chain Reaction

Definition

Real-time PCR quantitates the initial amount of a template by monitoring the emission of a fluorescent reporter that is specific for a certain gene during each PCR cycle, as opposed to the endpoint detection by conventional quantitative PCR methods. The amount of fluorescence emitted during the reaction is an indicator of the amplicon production during each PCR cycle.

►Rheumatism Related Genes

Receptor for Egg Jelly Protein

Definition

Fertilization of the sea urchin egg depends upon the sperm acrosome reaction that is responsible for binding and fusion of the sperm to the egg. The acrosome reaction is triggered by an acute increase in inward cation flux, stimulated by interaction between fucose sulfate polymer in the egg jelly and a 210kD sperm membrane glycoprotein on the sperm head called the receptor for egg jelly (REJ).

►Polycystic Kidney Disease Autosomal Dominant

Receptor Serine/Threonine Kinases

ARISTIDIS MOUSTAKAS, SERHIY SOUCHELNYTSKYI, CARL-HENRIK HELDIN

Ludwig Institute for Cancer Research, Uppsala, Sweden

c-h.heldin@licr.uu.se

Definition

Receptor serine/threonine kinases (►RSKs) are transmembrane proteins of the plasma membrane and are characterized by extracellular ligand-binding domains and cytoplasmic kinase domains. All RSKs of the human genome act as signaling receptors for members of the transforming growth factor β (►TGF- β) superfamily of secreted polypeptides (5). Accordingly, RSKs are known to play major roles in multiple physiological and pathological processes of metazoan life, including early embryogenesis, adult tissue homeostasis, atherosclerosis, tissue fibrosis and cancer (1, 3).

Rearrangement of Immunoglobulin Genes

Definition

Immunoglobulin molecules are coded by hybrid genes that are created randomly by the genomic rearrangement of DNA sequences.

►Chromosomal Instability Syndromes

Characteristics

Type I and Type II RSKs

RSKs can be divided into two subgroups, the type I and type II RSKs, based on primary amino acid sequence comparison (5). Upon ligand binding, a heteromeric complex of type I and type II receptors are formed, in which the type II RSK phosphorylates and activates the type I RSK. Type I RSKs therefore play the major role in activating downstream effectors of the signaling pathway. The type I receptors appear to have diverged more in relation to a hypothetical common ancestral precursor of RSKs. The vertebrate family includes seven type I and five type II receptors which bear multiple names in the literature. For a complete nomenclature we refer to the LocusLink database (<http://www.ncbi.nlm.nih.gov/LocusLink/>).

RSKs During Evolution

The human RSK family of 12 members is considerably smaller than the 58-membered receptor tyrosine kinase (▶RTK) family. RSK members are abundant in all metazoan species and appear even in the lowest phyla of Porifera (e.g. the sponge *Ephydatia*) and Cnidaria (e.g. the coral *Acropora*). They have been studied to various degrees in all metazoan model organisms, including *A. californica*, *C. elegans*, *D. melanogaster*, *X. laevis*, *D. rerio*, *M. musculus*, *R. norvegicus* and *H. sapiens*. Despite more than 300 receptor-like kinases identified in plants such as *A. thaliana*, counterparts of the vertebrate RSKs seem to be absent from plants and also from fungi or protists. Thus, the evolutionary emergence of the RSKs in metazoans correlates with their functional roles in mediating the physiological effects of TGF- β superfamily ligands and thus regulating cell-cell interactions and cellular movements important for the development of multilayer tissue patterns.

Invertebrates have fewer RSKs than vertebrates. Among invertebrates with fully sequenced genomes, *C. elegans* has three RSKs, one type II receptor (Daf-4) and two type I receptors (Daf-1, Sma-6) and *D. melanogaster* has four RSKs, two type II receptors (Punt, Wit) and two type I receptors (Tkv, Sax). The twelve human or mouse RSKs are commonly classified on the basis of the specific ligand they recognize or on the basis of the effector pathway they stimulate (Fig. 1A). Thus, we distinguish the TGF- β type II (▶T β R-II), the ▶activin types IIA and IIB (▶ActR-IIA, ▶ActR-IIB), the ▶BMP type II (▶BMPr-II) and the AMH type II (▶AMHR-II) receptors. On the other hand we recognize the TGF- β type I (▶T β R-I, also known as ▶ALK-5, activin receptor-like kinase 5), the activin/nodal types IB and IC (▶ActR-IB/ALK-4, ▶ActR-IC/ALK-7), the BMP types IA and IB (▶BMPr-IA/ALK-3, ▶BMPr-IB/ALK-6) and the dual specificity TGF- β /BMP type I receptors (ALK-1, ALK-2).

Structure of RSKs

All RSKs are single-span transmembrane proteins with their N-terminal ligand-binding domains extracellularly and their C-terminal kinase domains intracellularly (Fig. 1B) (5,6). The ectodomains include a primary cysteine-rich fold with a characteristic cluster of three cysteines proximal to the transmembrane domain. In addition, the extracellular domains of RSKs are glycosylated. The α helical and amphipathic transmembrane domains are important to orient the kinase domain properly in the receptor complex thereby allowing signaling. The cytoplasmic juxtamembrane domains are generally short and of divergent sequence content that may play roles in signal transduction. This is well documented for the type I RSKs whose characteristic TTSGSGSG (GS) motif immediately preceding the kinase domain plays an important role in regulation of the type I receptor kinase activity. The main part of the RSK cytoplasmic domain consists of a kinase domain that shows specificity towards phosphorylation of serines and threonines. Finally, RSKs have cytoplasmic tails which are mostly short but which in certain receptors can be longer. These tails are very divergent in sequence composition, and like the juxtamembrane domains, may contribute to propagation of signal transduction.

Molecular Interactions

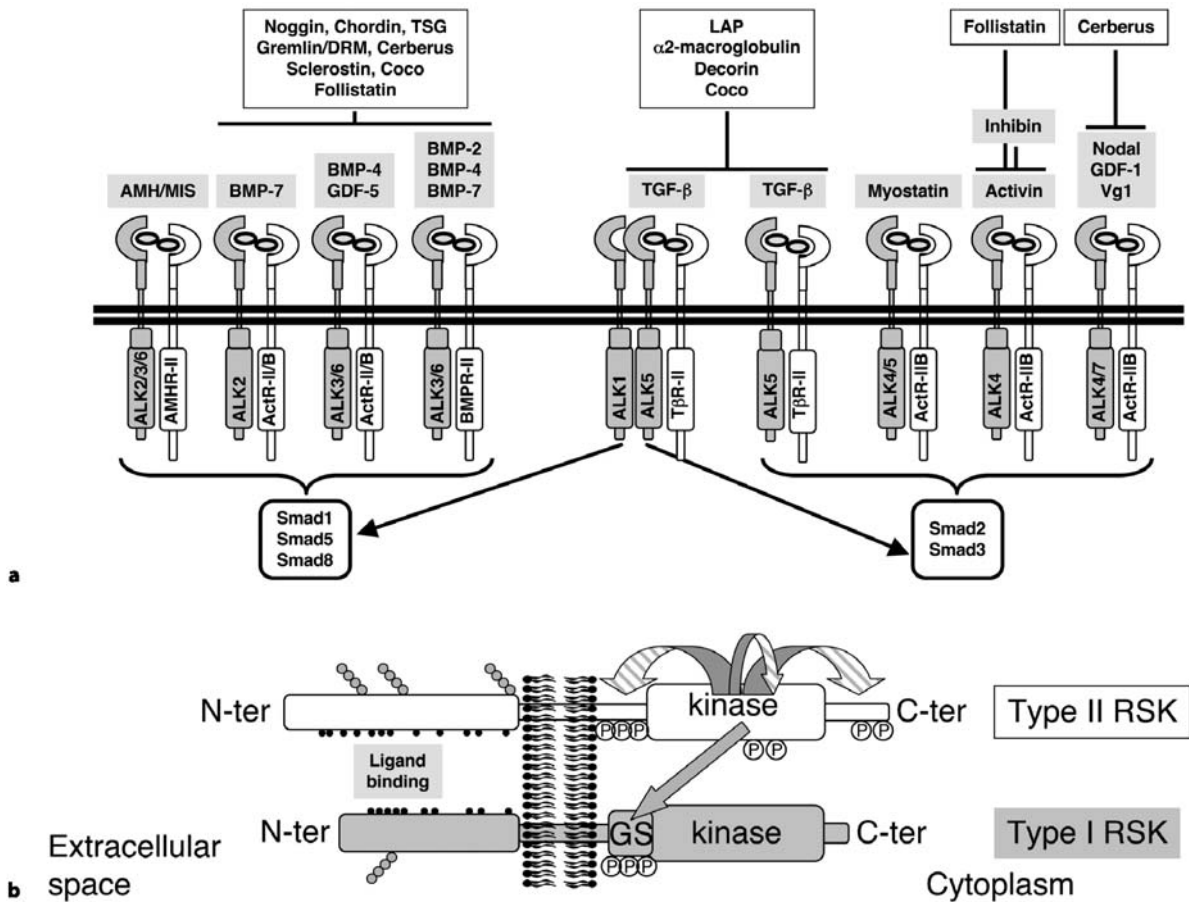
Extracellular Interactions

RSKs participate in signal transduction cascades initiated by TGF- β superfamily members. Access of these extracellular ligands to the receptor complex is regulated by several ligand-binding proteins and by accessory receptors.

Ligand-Receptor Interactions

After synthesis at ribosomes of the rough endoplasmic reticulum, RSKs traffic *via* Golgi and exocytic vesicles and eventually reside on the plasma membrane as monomers or homodimers. After ligand binding, a heteromer of two type II and two type I receptors is formed (6). Whereas both type I and type II receptors are needed for high affinity ligand binding, the relative contribution differs between TGF- β and BMP receptors. Thus, the extracellular domains of T β R-II and ActR-IIA/ActR-IIB have high affinity for TGF- β and activin, respectively. In contrast, ectodomains of T β R-I or ActR-IB have weak affinity for the same ligands. Inversely, BMPr-II has weak affinity for BMP ligands, whereas BMPr-IA or BMPr-IB show stronger affinity for the same ligands.

The ligand-receptor model has recently been understood at the atomic level and provides a unifying mechanism of initiation of signal transduction by TGF- β members (7). Yet, the model cannot provide unique



Receptor Serine/Threonine Kinases. Figure 1 RSK family of TGF- β superfamily receptors. (a) Ligand-RSK combinatorial specificity. Heteromeric RSKs embedded on a lipid bilayer (double line) are shown, whereby type I RSKs (grey) pair with type II RSKs (white) and bind dimeric TGF- β superfamily ligands (light grey ovals). The receptor cartoons highlight their modular structure. The name of each RSK is shown and when multiple type I receptor combinations have been described with a single type II receptor or vice versa, the multiple receptors are listed in a consecutive acronym. The known corresponding ligands for each RSK combination are listed (light grey background). Ligand-binding proteins (white boxes) that sequester and block (inhibitory arrows) ligand release and binding to RSKs are listed on top. Nomenclature of abbreviated ligand-binding inhibitors: TSG, twisted gastrulation, DRM, drumstick, LAP, latency associated peptide. The specific Smads activated by each RSK complex are listed below the RSK cartoon. (b) Structure-function relationships of RSKs. Type I (grey) and type II (white) RSKs are drawn to highlight their modular structure. Their N- and C-termini, ligand-binding domains (light grey background), kinase domains and the glycine/serine-rich motif (GS) of type I RSKs are marked. Small black circles indicate conserved cysteine residues. Chains of grey circles indicate N-linked carbohydrates. Circled P indicates phosphorylated serine or threonine residues and curved striped arrows point to auto-phosphorylation events, while a flat grey arrow points to trans-phosphorylation events.

heteromeric pairs of RSKs for each TGF- β ligand. This is obvious since, for example, the human genome has approximately 40 TGF- β superfamily ligands, yet only five type II and seven type I RSKs occur. Accordingly, multiple ligands of the superfamily form functional complexes with the same type I/type II RSK complex. The exact mechanisms underlying signaling specificity in various cellular contexts remain to be elucidated. Thus, Fig. 1a illustrates the current knowledge of

ligand-receptor associations, which may be refined in the future. A good example of the combinatorial complexity created by the hetero-tetrameric model of the RSKs is the known use of the type I receptor T β R-I. T β R-I together with T β R-II makes the classic TGF- β receptor complex expressed in most cell types. However in endothelial cells, TGF- β signals *via* two type I receptors, T β R-I and ALK-1, and thereby transmits a dual set of signals (Fig. 1a), leading to distinct

downstream signaling cascades and physiological responses. Finally, T β R-I serves as a component of the myostatin receptor together with ActR-IIB. Thus, a single RSK can serve as a receptor subunit for multiple ligands, it can pair with multiple type II receptors and can also combine with alternative type I receptors.

Ligand Antagonists and Coreceptors

The recognition of TGF- β superfamily ligands by RSK complexes is tightly regulated in space and time. Usually, TGF- β ligands are released in a well-controlled manner leading to graded responses that can be regulated by multiple feedback loops and alternative signaling pathways that operate in the same or a neighboring cell simultaneously (5). For these reasons, a growing number of extracellular or membrane-bound factors (some of which are listed in Fig. 1a) cooperate with or antagonize the ligand-receptor interaction. Some of these proteins act negatively by sequestering ligands in the extracellular environment (6). They can be polypeptides derived from the same propeptide as the mature ligand, as is the well-documented case of TGF- β 1, whose N-terminal latency associated peptide (\blacktriangleright LAP) binds non-covalently to the mature dimeric TGF- β 1. The complex is also bound by additional proteins called latent TGF- β binding proteins (\blacktriangleright LTBPs), which constitute integral components of the extracellular matrix. Other examples are the protein follistatin that binds with high affinity activins and thus prohibits their binding to their cognate receptor complex and noggin, a BMP-specific antagonist that also traps extracellular BMPs and precludes them from binding to their receptors. Some other proteins act in a positive manner and these are usually transmembrane receptors themselves or proteins that link to the plasma membrane *via* glyco-phosphoinositide tails (5, 6). For example, the transmembrane proteoglycan receptor betaglycan and its structurally related transmembrane receptor endoglin bind TGF- β with high affinity and mediate its presentation to the signaling RSK complex, thus acting as accessory receptors. In contrast, the truncated receptor BAMBI pairs with BMP RSKs and inhibits proper signal transduction by BMPs.

Intracellular Interactions

Intracellular interacting proteins serve as direct substrates of RSK, as regulators of the kinase activity and as modulators of signaling *via* affecting accessibility of receptors to down-stream targets and *via* compartmentalization of receptors in cells. More than 30 different intracellular proteins have been found to interact with

RSKs (2, 4, 6). RSKs themselves form stable complexes with each other due to an interaction of their intracellular parts. Formation of such complexes, initiated by ligand binding, leads to activation of kinases and is the key event in initiation of the intracellular signaling.

Smads

\blacktriangleright Smad proteins are essential intracellular signal transducers of RSKs (4, 6). Smad proteins are divided into three classes; receptor-regulated Smad1, 2, 3, 5, and 8, and inhibitory Smad6 and 7 interact with RSKs, whereas common-mediator Smad4 has not been found in complex with RSKs. Receptor-dependent phosphorylation of receptor-regulated Smads is the triggering event in Smad-induced signaling. After phosphorylation by RSKs, Smads form complexes with various proteins, including common-mediator Smad4, transcription factors and co-activators; these Smad complexes then regulate gene expression. Despite structural similarity with other Smads, Smad6 and Smad7 inhibit the signaling by interacting with RSKs.

Modulatory Proteins

RSKs form complexes with modulatory proteins (2, 4, 6). As for RTKs, scaffolding, adaptor and kinase-regulating proteins regulate RSK signaling. The interaction of T β R-I with \blacktriangleright FKBP12 prevents the receptor kinase from spurious activation by stabilizing the kinase in an inactive conformation.

Proteins interacting both with Smads and with receptors serve as adaptors for Smad-dependent signaling. For example, an interaction of T β R-I with \blacktriangleright SARA enhances receptor-dependent phosphorylation of Smad2. Another protein, Dab-2, also forms a link between T β R-I/T β R-II and Smad2, thus increasing signaling efficiency. TRAP1 that forms a complex with Smad4 and with RSKs facilitates the interaction between Smad4 and receptor-regulated Smads. On the other hand, interaction of Smad3 with ARIP1 in complex with the ActR-IIA receptor results in inhibition of Smad3-dependent transcription, suggesting both enhancement and inhibition of Smad signaling by interacting proteins.

Two proteins interacting with BMPR-IA receptor, BRAM1 and XIAP, were claimed to form a link from RSKs to the TAB1 protein, thus enhancing the interaction with the TAK1 kinase (2). This link provides an example of a mechanism whereby RSKs activate the MAP kinase pathways or NF κ B signaling. Links of TGF β and activin signaling to induction of apoptosis can be provided by Daxx, which forms a complex with T β R-II and Fas receptor. Another adaptor protein, Dok-1, acts as a link

between activin receptors and the Smads; Dok-1 augments activin-induced apoptosis in B cells. An interaction of T β R-I with the B α subunit of PP2A phosphatase was shown to mediate an inhibition of p70 S6 kinase and induce cell-cycle arrest. Formation of a complex between T β R-I, T β R-II and the p85 subunit of PI3 kinase may provide a link to regulation of cytoskeletal rearrangements, whereas interaction of LIM kinase 1 with BMPR-II suggests a more direct link to regulation of the cytoskeleton.

Interactions of RSKs with caveolin-1, STRAP, Smad7, ▶Smurf1 and Smurf2 suggest biochemical mechanisms for endocytosis of RSKs into caveolin-positive vesicles and further degradation or recycling of receptors. Interaction of RSKs with proteins, such as SARA, Hgs/Hrs and Dab-2 is associated with localization of RSKs to early endosomes, where Smads can be activated in addition to their signaling from the plasma membrane. ARIP2, TLP-1, sortin nexins and dynein light chain proteins are other examples of interacting proteins which affect endocytosis and trafficking of RSKs.

A number of interacting proteins still await functional studies of their roles in RSK signaling *in vivo*, e.g., apolipoprotein J/clusterin, LXR β , farnesyl transferase α and cyclin B (2). Recent application of proteomic technologies showed that the number of RSK-associated proteins is significantly higher; for example, 33 proteins were found to associate with BMPR-II. This suggests that RSKs form distinct complexes with different signaling potentials.

Regulatory Mechanisms

Regulation of RSK by Post-Translational Modifications

Three types of post-translational modifications have been described for RSKs, phosphorylation, N-glycosylation and ubiquitination.

Glycosylation

RSKs are N-glycosylated in their extracellular domains (5). For example, the T β R-I has one and the T β R-II has three glycosylation sites (Fig. 1b). Other RSKs show rather heterogeneous electrophoretic migration patterns that suggest glycosylation to varying degrees, which depends on the specific receptor and the cell type where the receptor is expressed. The glycosylated ectodomain of T β R-II was shown to have higher affinity for the ligand, as compared to the non-glycosylated receptor. This suggests that glycosylation may enhance recognition of the ligand by receptors.

Phosphorylation

A well-characterized mechanism for regulation of RSK activity is phosphorylation (6, 7). All studied RSKs are phosphorylated on multiple sites in their intracellular domains; phosphopeptide maps showed at least 16 phosphorylated peptides in T β R-I, 17 phosphopeptides

in T β R-II and 18 phosphopeptides in BMPR-II. Many of the phosphorylation sites are due to receptor autophosphorylation or to phosphorylation by another RSK. However, phosphorylation of kinase inactive receptors has also been observed, suggesting that RSKs may also be phosphorylated by other kinases. The T β R-I kinase is activated by phosphorylation in the juxtamembrane GS-rich region located immediately N-terminally of the kinase domain (Fig. 1b). Phosphorylation of Thr185, Thr186, Ser187, Ser189 and Ser191 in the GS-region of T β R-I changes receptor conformation to the active state. A similar model of kinase activation may be valid for T β R-II, as phosphorylatable Ser213, Ser225, Ser228 and Ser229 residues are located in similar proximity to the kinase domain as the GS-region in T β R-I (Fig. 1b). The patterns of total phosphorylation of other RSKs in response to ligand stimulation are similar to those of T β R-I and T β R-II, but a detailed characterization of phosphorylation sites remains to be performed. Such studies will unveil whether the model of activation of the T β R-I kinase is valid for other RSKs.

In addition to phosphorylation sites in the juxtamembrane region, T β R-II phosphorylation was described at Ser409 and Ser416 in the kinase domain with positive and negative effects on the kinase activity, respectively. Ser551 and Ser553 in the C-terminal tail of T β R-II are also phosphorylated, the functional importance of which remains to be elucidated. There is one example of a phosphorylation that did not affect the kinase activity but rather modulated signaling; phosphorylation of Ser165 in T β R-I enhanced TGF β -dependent growth inhibition but decreased pro-apoptotic signaling.

Structural studies showed that the T β R-I kinase has certain features which are more similar to RTKs than to non-receptor serine/threonine kinases (6, 7). This observation suggested that not only serine and threonine, but also tyrosine residues might be phosphorylated by RSKs. In agreement with this notion, T β R-II was found to autophosphorylate Tyr259, Tyr336 and Tyr424 residues; mutation of these tyrosine residues inhibited the kinase activity.

Ubiquitination

Poly-ubiquitination is known to trigger protein degradation. T β R-I interaction with Smad7 in complex with the E3 ubiquitin ligases Smurf1 or Smurf2 was found to result in ubiquitination of the receptor, followed by targeting of the ubiquitinated T β R-I to degradation (4, 6). The sites of ubiquitination in the T β R-I remain to be identified, as well as whether ubiquitination occurs also in other RSKs.

Function of RSKs in Development and Human Disease

Being mediators of TGF- β superfamily signaling, RSKs are critical regulators of embryonic patterning, cell

proliferation, apoptosis, differentiation and pathogenic processes that range from tissue fibrosis and cancer to atherosclerosis and autoimmune disease.

Knockout of RSKs in Mice

The developmental role of many of the RSKs has been determined *via* genetic ablations in the mouse (1). Complete knockout of the TGF- β receptors T β R-II and ALK-5 and the endothelial TGF- β receptor ALK-1, all lead to embryonic lethality due to severe defects in vascular development. In particular, the T β R-II knockout phenocopies the complete loss of the TGF- β 1 ligand in mice. Loss of activin receptor ActR-IIA leads to partial (25%) perinatal lethality due to defects in craniofacial development and the survivors exhibit defects in fertility, whereas loss of ActR-IIB leads to perinatal lethality due to defects in left-right and antero-posterior embryonic deregulation. Similarly, loss of the activin type I receptor ALK-4 blocks early development at the primitive streak stage and causes lethality, a phenotype largely shared by three BMP RSKs, BMPR-II whose loss arrests gastrulation, ALK-2 whose loss affects gastrulation and causes defects in endoderm and mesoderm formation and ALK-3 whose loss inhibits formation of mesoderm. In contrast, knockout of the BMP RSK ALK-6 does not affect viability, but leads to defects in chondrogenesis of the limb and female-specific defects in ovarian development. Finally, knocking-out AMHR-II phenocopies loss of the corresponding ligand, [▶AMH/MIS](#), by inducing uterine development and infertility in males. The knockout of the type I receptor ALK-7 is still awaited. Based on the predominant embryonic defects that lead to early lethality by genetic ablation of most RSKs, recent attempts focus on the generation of tissue-specific (conditional) knockouts that will allow analysis of the role of these receptors in late stages of development and in adult organs. One such example is the recent hair follicle-specific knockout of the BMP type I receptor ALK-3, which supports a role of this RSK in the progenitor cell formation of the follicle. Such technologies will most certainly bring new biological links to the normal functions of the RSKs.

RSK Mutations in Human Disease

A number of RSKs have been shown to accumulate genetic mutations in humans that lead to their malfunction and subsequent development of human genetic disorders (3,). Thus, hereditary hemorrhagic telangiectasia, a severe angiogenic defect, is linked to germline mutations in the genes encoding ALK-1 or the accessory receptor endoglin. Germline mutations in ALK-1 but also in BMPR-II are linked to the development of familial primary pulmonary hypertension. ALK-6 mutations have

been uncovered in association with brachydactyly type A2, a rare disorder of the hand. The most tissue-specific of the RSKs, AMHR-II, accumulates germline mutations that lead to persistent Müllerian duct syndrome. In addition, most RSKs seem to be mutated in human cancers. Germline mutations in the ALK-3 and ALK-5 genes are known in human cancers, such as juvenile polyposis, breast, pancreatic, biliary, cervical, head and neck tumors and chronic lymphocytic leukemia, whereas somatic mutations that primarily affect epigenetic control of ALK-5 and T β R-II are abundant in the same large spectrum of human cancers, but at least in the case of T β R-II, also in atherosclerosis.

The current rapid pace of research in the RSK/TGF- β field promises to reveal additional genetic associations between RSKs and various human disorders, reflecting the function of the RSKs as mediators of signals from TGF- β superfamily members.

▶RSK (Receptor Serine/Threonine Kinase)

References

1. Chang H, Brown CW, Matzuk MM (2002) Genetic analysis of the mammalian transforming growth factor- β superfamily. *Endocr Rev* 23:787–823
2. Derynck R, Zhang YE (2003) Smad-dependent and Smad-independent pathways in TGF- β family signalling. *Nature* 425:577–584
3. Dijke P ten, Goumans M-J, Itoh F et al (2002) Regulation of cell proliferation by Smad proteins. *J Cell Physiol* 191:1–16
4. Moustakas A, Souchelnytskyi S, Heldin C-H (2001) Smad regulation in TGF- β signal transduction. *J Cell Sci* 114:4359–4369
5. Piek E, Heldin C-H, ten Dijke P (1999) Specificity, diversity, and regulation in TGF- β superfamily signaling. *FASEB J* 13:2105–2124
6. Shi Y, Massagué J (2003) Mechanisms of TGF- β signaling from cell membrane to the nucleus. *Cell* 113:685–7006
7. Souchelnytskyi S, Moustakas A, Heldin C-H (2002) TGF- β signalling from a 3-dimensional perspective: insight into selection of partners. *Trends Cell Biol* 12:304–307

Receptor Tyrosine Kinases

Definition

Receptor tyrosine kinases (RTKs) are a class of transmembrane receptors for growth factors and cytokines. They contain an extracellular ligand binding domain and an intracellular catalytic domain, which

mediates ATP-dependent autophosphorylation and phosphorylation of other proteins on tyrosine residues. These residues then serve as docking sites for SH2 domain (src-homology domain-2) containing downstream effectors, such as PI3 kinase.

- ▶ Focal Complexes/Focal Contacts
- ▶ Hemidesmosomes
- ▶ Neurotrophic Factors
- ▶ Transcription Factors and Regulation of Gene Expression
- ▶ Tyrosine Kinases

Receptors

Definition

Receptors are molecules that bind to a certain extra-cellular signal, initiating specific signal transduction pathways, so leading to the expression of a certain phenotype. Receptors may be located inside the cell, e.g. for steroid hormones, or at the plasma membrane, e.g. for peptide hormones. In pharmacology, a receptor is, more generically, any molecule that is a target for a drug.

- ▶ Cancer Invasion and Metastasis: Cellular, Molecular and Clinical Aspects
- ▶ Familial Hypercholesterolemia
- ▶ Receptor Serine/Threonine Kinase
- ▶ Receptor Tyrosine Kinases

Recessive Allele

Definition

Recessive allele is one of a series of terms applied to the phenotypic effect of a particular allele, in reference to another allele (usually the standard wild-type allele), with respect to a given trait. An allele “a” is said to be recessive with respect to the dominant allele “A”, if the A/A homozygote and the A/a heterozygote are phenotypically identical and different from the a/a homozygote. A recessive character, therefore, only manifests in the homozygote.

- ▶ Large-Scale ENU Mutagenesis in Mice

Recessive Mutation

Definition

Recessive mutation denotes a mutation that results in a loss of gene function mutation that does not produce

the mutant phenotype in the presence of the wild-type gene.

- ▶ Protein Interaction Analysis: Suppressor Hunting

Recklingshausen Disease

- ▶ Neurofibromatosis NF1
- ▶ Osteitis Fibrosa Cystica

Recoding

- ▶ Translational Frameshifting, Non-standard Reading of the Genetic Code

Recombinant Protein

Definition

Recombinant protein originates from recombinant DNA technology, i.e. the linkage of DNA molecules from different sources. Recombinant DNA is for instance the combination of vector DNA and an inserted DNA fragment. Recombinant proteins are expressed from cloned DNA templates.

- ▶ Recombinant Protein Expression in Bacteria

Recombinant Protein Expression in Bacteria

CLAUDIA LANGLAIS, BERNHARD KORN
RZPD, Deutsches Ressourcenzentrum für
Genomforschung GmbH, Heidelberg, Germany
korn@rzpd.de

Definition

- ▶ Recombinant protein expression in bacteria requires the insertion of a DNA fragment (▶ open reading frame,

►ORF) into an expression vector, routinely a ►plasmid vector and the transferral of this vector into bacterial cells (►transformation). The cells are then cultured and induced to express the desired protein. The cells are harvested by centrifugation, samples prepared and proteins detected by ►polyacrylamide gel electrophoresis and subsequent staining of the gel with Coomassie Brilliant Blue or silver stain or by ►immunoblotting. Protein expression in bacteria is highly scalable and can be adjusted from solid growth (colony expression), to conical flasks for liquid cultures (up to 3 l), to fermentation reaction chambers (often more than 100 l volume).

Characteristics

The growing field of proteomics is resulting in an ever-increasing demand for the production and purification of recombinant proteins for structural and functional studies, as it can be very time-consuming to isolate low abundance proteins from their native source in tissue or cells. Thus, high-throughput methods for the expression and characterisation of recombinant proteins are essential to obtain sufficient amounts of protein to be used for research. An advantage of recombinant protein expression is that the selected protein can be expressed with a fusion tag such as ►His-Tag/6×His, ►GST, ►MBP etc., which can considerably facilitate the identification and purification of the protein, as well as improve both the level of expression and the solubility. The expression of recombinant proteins in bacterial cells has become the most common method for expressing large amounts of protein and *Escherichia coli* (*E. coli*) is the most frequently used prokaryotic expression system for the high-level production of recombinant proteins (1, 2).

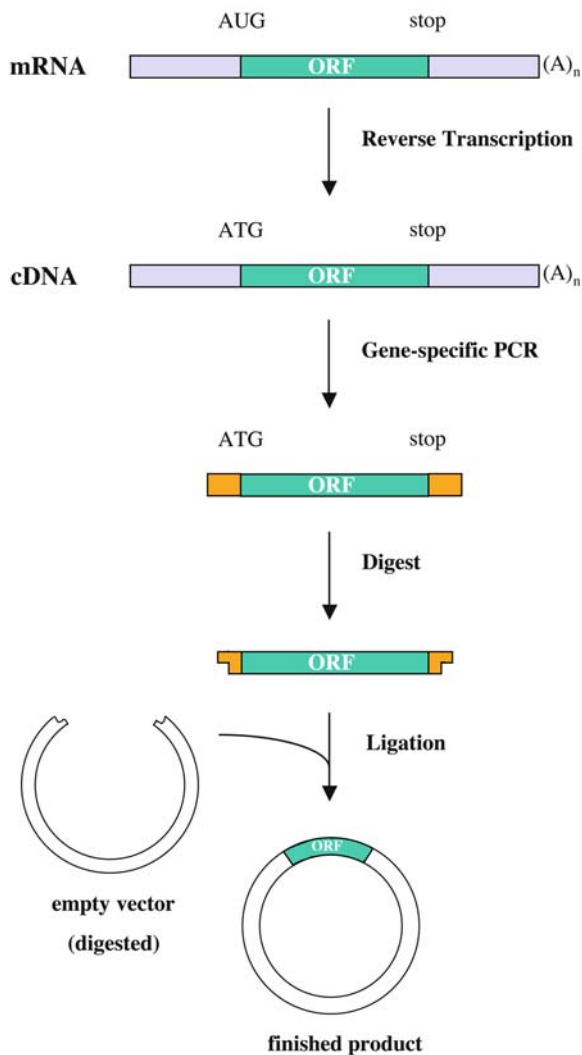
The use of *E. coli* has many advantages, including the ease of growth and manipulation of the organism, the availability of many different vectors and host strains that have been developed over the years and the amount of information concerning the genetics and physiology of *E. coli*. However, the *E. coli* expression system also has limitations such as the inability to carry out complex posttranslational modifications, for example glycosylation, myristoylation, phosphorylation, signal sequence cleavage and extensive disulfide bond formation. Another problem can be the production of misfolded or aggregated protein in insoluble inclusion bodies or problems with the expression of “toxic” gene products. Some gene products can be toxic to the bacterial host, disrupt normal metabolism and cause bacteria to die or reduce their growth rate, making it almost impossible to generate sufficient amounts of protein for further analyses. Low levels of expression can also be due to protein degradation (3), inefficient translation (due to mRNA secondary structure in the initiation region) or eukaryotic codon usage that is not

optimal for *E. coli*. Some eukaryotic codons are rarely found in *E. coli* and can cause translation to be delayed or stopped and therefore result in severely diminished expression (2). In order to minimise some of these problems, many specialised protein expression hosts have been designed to provide for instance enhanced protein yield in T7 promoter-based systems (BL21 StarTM (DE3), Invitrogen), the tight regulation and strong expression of toxic proteins (BL21-AITM, Invitrogen) or the reduction of the basal expression of target genes via expression of T7 lysozyme (BL21 (DE3) pLysE or pLysS). Other expression hosts contain rare tRNA genes (RosettaTM, Novagen) or promote disulfide bond formation (OrigamiTM, Novagen). Thus, the decision as to which particular strain of bacteria is to be used depends on the individual application and the protein of interest.

The Cloning of DNA into a Vector

In order to express a complete protein, the exact full-length coding sequence (CDS) of the protein is required. This information can usually be obtained from databases available on the Internet such as SwissProt (►www.expasy.org/sprot/sprot-top.html), NCBI (►www.ncbi.nlm.nih.gov/) and others. The tissue, which contains the desired protein in the highest abundance, is then used to extract mRNA and in a reaction called ►reverse transcription, cDNA is generated from the many different mRNAs. The cDNA which codes for the protein of interest (open reading frame, ORF) is amplified by ►PCR using gene-specific primers and at the same time additional sequences are added, which contain restriction sites that are used to clone the DNA into a vector. Both the PCR fragment and the vector are digested using ►restriction enzymes and then joined together in a ►ligation reaction (Fig. 1). Instead of restriction enzymes, other methods can be used to clone DNA into a vector and these include the Gateway[®] Cloning System (Invitrogen) or the Creator[®] System (Clontech), both of which use ►homologous recombination and are highly efficient. Another alternative is, of course, to obtain the appropriate cDNA clone from one of the companies or institutions that distribute them; the largest of them, the RZPD (Resource Centre for Genome Research) has more than 35 mio. clones available online (►www.rzpd.de). In some cases, subcloning of the full or partial ORF is required.

There are many different plasmid vectors available for protein expression and several points need to be considered before choosing a vector, such as the type of bacterial strain which will be used, if and what kind of tag will be added to the protein, the desired level of protein expression etc. All expression vectors contain a selectable antibiotic-resistance gene, a promoter, a ►replication origin (ori), and a region into which

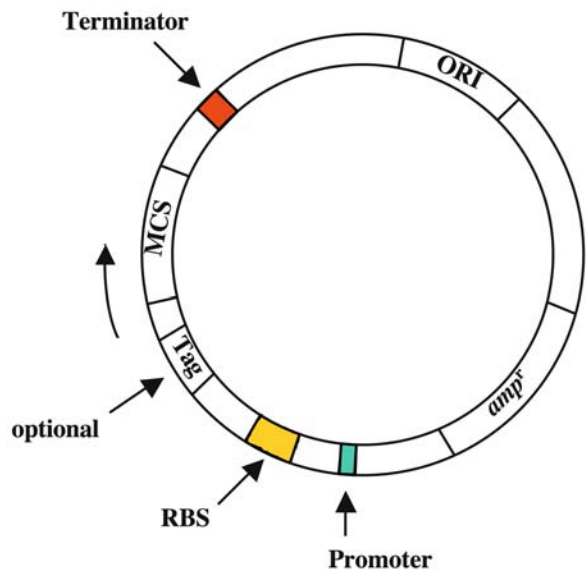


Recombinant Protein Expression in Bacteria.

Figure 1 Schematic representation of the steps required to obtain an open reading frame from mRNA. A mixture of many different cDNAs is produced from mRNA by using reverse transcription. With the use of gene-specific PCR, the cDNA of interest is amplified and at the same time overlaps are added to the PCR product. The overlaps contain cleavage sites recognised by certain restriction enzymes. The PCR product with the ORF is digested with these restriction enzymes, as well as the vector into which the ORF is to be cloned. Thus, compatible ends are generated in the vector and the PCR product. The ends are ligated and the ORF cloned into the vector.

foreign DNA can be inserted (known as the multiple cloning site, MCS) (Fig. 2).

The antibiotic-resistance gene facilitates the selection of positive transformants (see Transformation below). Medium is supplemented with antibiotics to kill

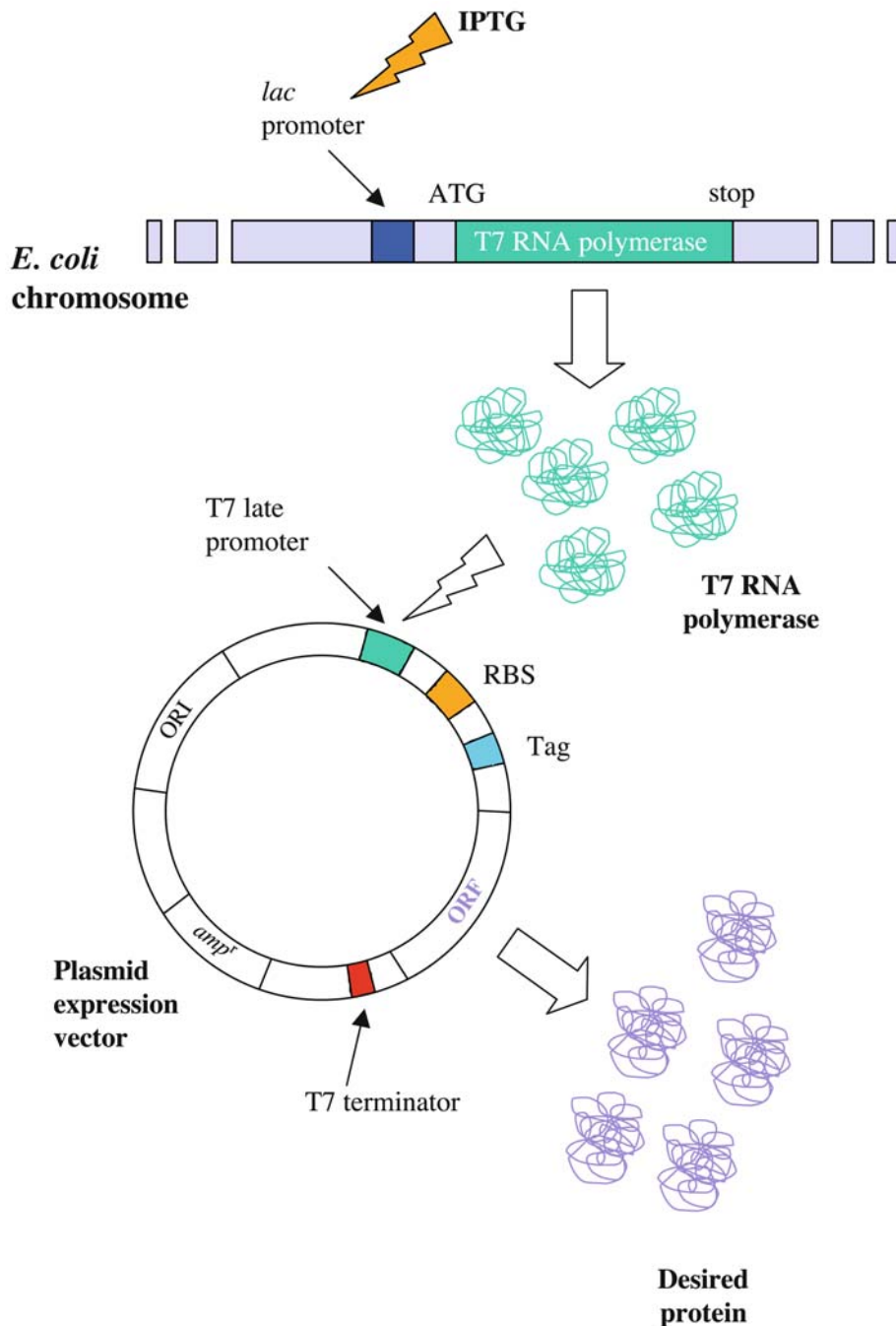


Recombinant Protein Expression in Bacteria.

Figure 2 Diagram of a simple expression vector. The vector contains an origin of replication (ORI), an antibiotic resistance gene (here the gene is *amp^r*, which encodes β -lactamase, an enzyme which inactivates ampicillin), a promoter, ribosome binding site (RBS), N-terminal tag, multiple cloning site (MCS) and terminator.

plasmid-free cells and maintain those that contain the plasmid with the antibiotic-resistance gene and the ORF.

The promoter is the section of DNA where RNA polymerase initiates transcription. An example is the *lac* promoter, which initiates transcription only when lactose, or the non-hydrolysable lactose analogue IPTG (isopropyl- β -D-1-thiogalactopyranoside), is added to the culture. Following induction with IPTG, the gene under control of the *lac* promoter is transcribed into mRNA, which is then translated into protein. A very popular promoter is the bacteriophage T7 late promoter (1). The T7 promoter is part of a more complicated expression system, as two levels of amplification are involved (Fig. 3). Firstly, genetically modified bacterial strains are used, which in their genome carry the gene encoding T7 RNA polymerase from the bacteriophage T7 under control of the *lac* promoter. The T7 RNA polymerase transcribes the gene under control of the T7 late promoter on the plasmid expression vector. Thus, when IPTG is added to the culture medium, the *lac* promoter initiates transcription of T7 RNA polymerase, which is translated into protein and then initiates transcription at the T7 late promoter on the expression plasmid. Thus the desired protein is produced (Fig. 3). This system leads to the synthesis of large amounts of



Recombinant Protein Expression in Bacteria. Figure 3 The two-step mechanism of protein expression with the T7 late promoter. The figure shows a section of an engineered *E. coli* chromosome, which contains the gene for T7 RNA polymerase under control of the *lac* promoter. The host also contains an expression plasmid vector with an ORF cloned into the vector under control of the T7 late promoter. When IPTG is added to the culture medium, the *lac* promoter initiates transcription of the T7 RNA polymerase gene and the mRNA is translated into protein. T7 RNA polymerase then initiates transcription at the T7 late promoter on the expression vector and the ORF is translated into the desired protein.

mRNA and results in the accumulation of the desired protein in high concentrations. The recombinant protein can represent up to 50% of total cell protein.

The origin of replication determines the approximate number of copies of the plasmid in a single bacterial cell. High copy number plasmids will yield more DNA

following a DNA preparation than medium or low copy number plasmids. High copy number plasmids also cause an increase in expression level, since the number of plasmids in a single cell correlates with the amount of protein that is expressed (when using the same promoter). However, if the desired protein is harmful to the bacterial metabolism, then a low copy vector is more appropriate, as the basal expression will be lower and thus cause less damage. Some promoters, and in particular strong promoters like the T7 late promoter, can be “leaky”, which means the protein of interest is expressed in low amounts before protein expression is induced with IPTG. This basal expression is caused by a) the high activity of T7 RNA polymerase and b) lactose in the medium. The problem can be minimised by using low copy plasmids and/or a tightly regulated promoter construct such as the arabinose p_{bad} promoter. For cloning purposes, several bacterial hosts offer advantages in particular applications. The DH10B and DH5 α strains for instance include many useful genetic markers, such as improved DNA cloning capabilities due to the *endA1* mutation (a mutation of the non-specific endonuclease endonuclease I results in reduced levels of intra- and intermolecular recombination events), which increases plasmid quality.

Transformation

Once the open reading frame (ORF) cDNA has been cloned into the appropriate vector and the strain of bacteria to be used for protein expression has been chosen, the DNA is ready for insertion into the bacterial cells. This process is called transformation.

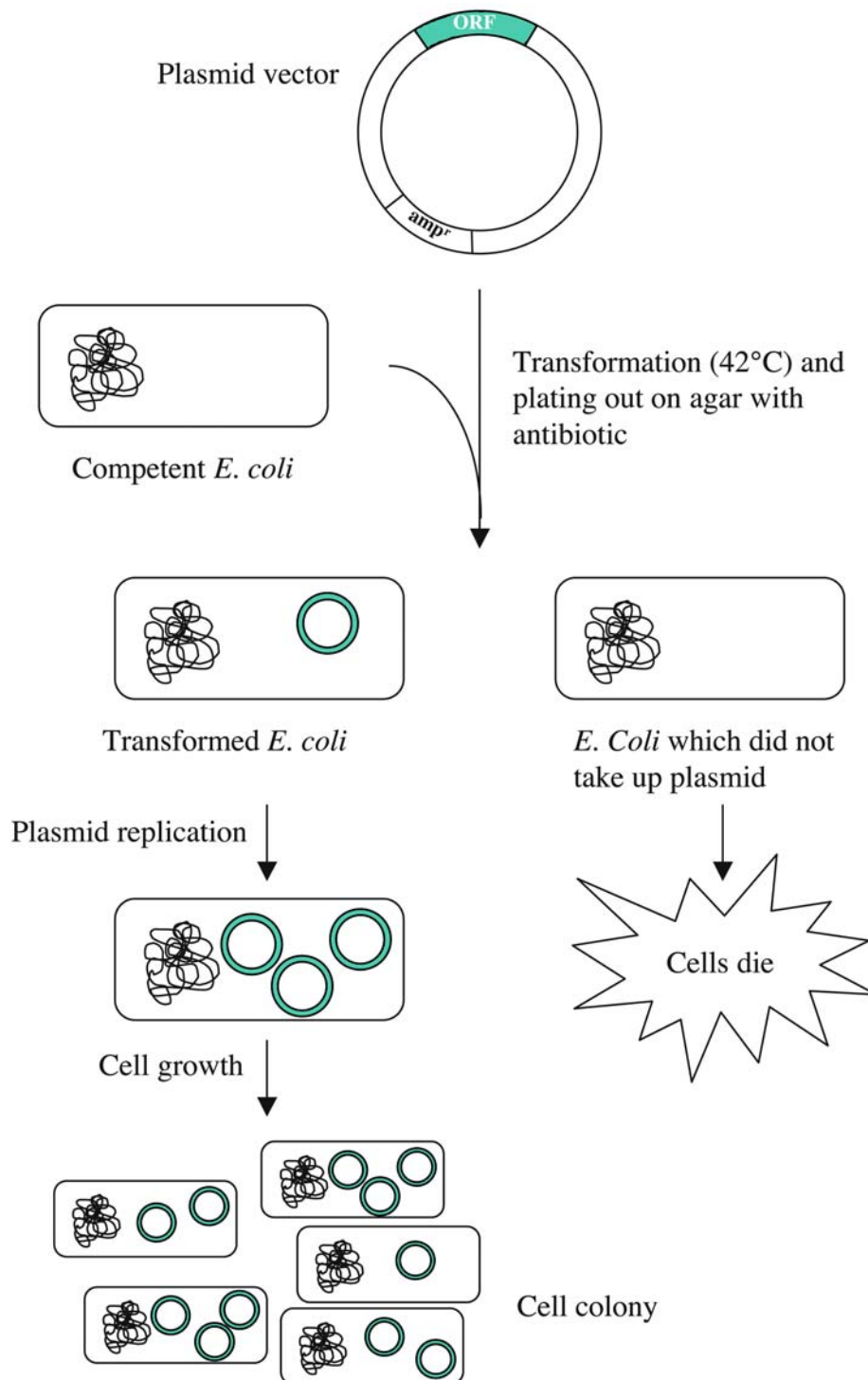
In order for a transformation to be successful, *E. coli* cells first need to be made competent i.e. permeable to foreign DNA, since normal *E. coli* cells rarely take up foreign DNA. A classical method of making bacterial cells competent is to expose them to high concentrations of calcium chloride, though the mechanism of how this is accomplished is not fully understood. In order to transform bacteria, competent cells are mixed with plasmid expression vector, briefly incubated at 42°C, grown in medium at 37°C and then plated out on a dish of agar containing an antibiotic (Fig. 4). The antibiotic is very important, as it selects for those cells, which have taken up the plasmid that contains the antibiotic-resistance gene and the ORF. All cells, which do not take up the plasmid, either die or undergo growth arrest in the presence of the antibiotic. For instance, when employing a plasmid vector that confers resistance to ampicillin, bacteria that do not take up the plasmid lack the resistance gene *amp*^R and die in the presence of this particular antibiotic. In the case of a plasmid vector conferring resistance to kanamycin, however, cells that do not contain the resistance gene *km*^R will not die, but undergo growth arrest. Thus, in order to be used in DNA cloning, a plasmid vector

should contain a resistance gene that allows for positive selection of transformants. The agar plate with the transformants is incubated overnight and colonies are picked the following day. Each colony originates from a single cell and consequently is made up of cells which all contain identical copies of the same recombinant plasmid, the expression vector with the ORF. The picked cells are incubated in medium overnight and an aliquot of the overnight culture is used to create a glycerol stock (stored frozen at –80°C) as a reference for future experiments.

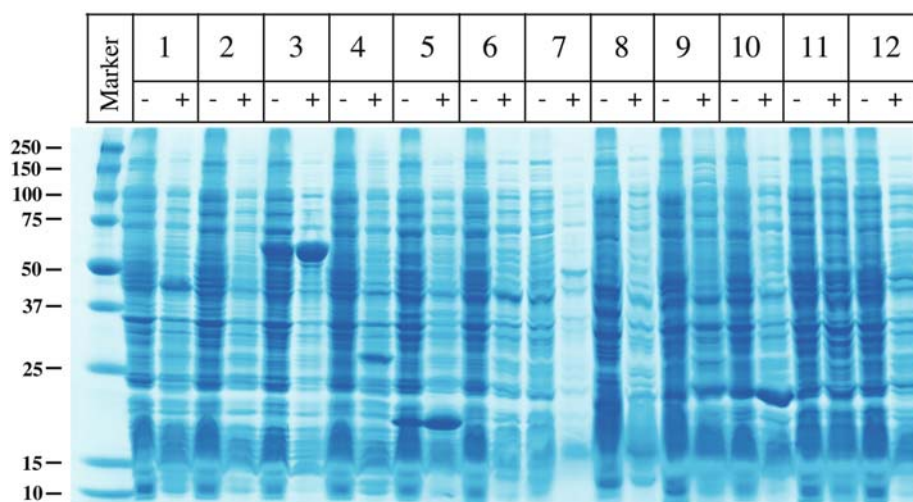
Protein Expression and Analysis

For protein expression, an overnight culture is diluted in fresh medium to allow the cells to quickly return to logarithmic growth and reach the optimal cell density for induction of protein expression. The cells should be in mid-log phase when inducing protein expression, as they grow rapidly in this phase. The method of induction depends on the plasmid vector and is usually done by addition of IPTG (in the case of the T7 late promoter and many other promoters) or by temperature shift-up (in the case of the P_L promoter in combination with a specific host strain). There are different media available to culture *E. coli*, such as LB, 2YT or SOC and bacteria display different growth rates depending on the medium used. The temperature at which the cells are induced to express protein is also of importance. *E. coli* can be induced to express protein in a range of ~12°C to 42°C. The lower the temperature, the more soluble the expressed protein is likely to be, because expression proceeds at a slower rate and the protein can fold correctly and thus stay in the soluble fraction. In general, the cells are cultured for a longer time period when using a low induction temperature in order to build up sufficient amounts of protein. Thus, varying the time period and/or temperature of induction of expression is useful to elucidate the optimal conditions of protein accumulation.

Protein expression is analysed by pelleting the cells, resuspension of the pellet in ►SDS sample buffer and fractionation of the sample by polyacrylamide gel electrophoresis. The protein bands on the gel are visualised by subsequent staining of the gel with Coomassie Brilliant Blue stain (Fig. 5) or silver stain. Alternatively, if the protein has been expressed with a tag, such as a 6×His-tag or there is a protein-specific antibody available, the protein can be blotted from the electrophoresis gel onto a membrane (►Western Blot) and detected with the appropriate antibody. The solubility of the expressed proteins can be easily determined by lysing the cells, fast centrifugation and analysis of the supernatant and pellet by immunoblotting. Expressed proteins can also be analysed for function (e.g. kinase assay), but the protein may need to be purified for this purpose.



Recombinant Protein Expression in Bacteria. Figure 4 The transformation of competent *E. coli* cells. Competent *E. coli* cells are mixed with plasmid DNA, briefly exposed to 42°C (heatshock), incubated and then plated out on agar containing an antibiotic. The plasmid vector confers antibiotic resistance to those cells that have taken up the plasmid. Thus, the transformed cells survive, while the others die. The plasmid replicates independently within the cell and the transformed cells multiply to form a colony of cells all containing copies of the same plasmid.



Recombinant Protein Expression in Bacteria. Figure 5 Coomassie stained gel of protein expressions in *E. coli*. Samples were taken prior to (–) and following (+) induction of protein expression with IPTG, fractionated by gel electrophoresis and the gel stained in Coomassie Brilliant Blue stain. The molecular weight marker is indicated at the left of the figure. The gel demonstrates, how some proteins can be expressed in large amounts (lane 3 and 5), whereas others are expressed in much smaller quantities (lane 1 and 4). In lane 7, the expressed protein seems to inhibit bacterial growth, as the background staining is weaker compared to the other lanes.

Purification of Recombinant Proteins

Several vectors are available which allow expression of recombinant proteins that are fused at their N- or C-terminus to polypeptides or longer amino acid sequences. These tags can greatly facilitate purification of the expressed protein, as they bind to affinity resins. Some of the more popular tags include the 6×His tag

(which binds to ►Ni-NTA resin), GST (which binds to glutathione-Sepharose), maltose binding protein (MBP, which binds to amylose resin) or staphylococcal protein A (which binds to immunoglobulin G-Sepharose) (2, 4) (Table 1). Simple purification protocols have been devised for these and other tags, making the use of a tag one of the most common methods for cleaning up

Recombinant Protein Expression in Bacteria. Table 1 A selection of popular fusion partners and their ligands

Fusion partner	Ligand/matrix
6xHis	Ni ²⁺ -NTA
Glutathione-S-transferase (GST)	Glutathione-Sepharose
Maltose-binding protein (MBP)	Amylose resin
Galactose-binding protein	Galactose-Sepharose
Avidin	Biotin
Streptavidin	Biotin
Strep-tag	Streptavidin
c-Myc	Anti-Myc antibody
Staphylococcal protein A	IgG-Sepharose
Staphylococcal protein G	Albumin
FLAG peptide	Anti-FLAG monoclonal antibody
Dihydrofolate reductase	Methotrexate-agarose

a protein. Specific cleavage sites can be included in the sequence of the protein to allow the tag to be removed following purification (5). This is important for functional studies, since the tag can interfere with the correct folding and/or function of the protein. Proteins expressed without a tag can be purified using techniques such as gel filtration, ion exchange or hydroxyapatite, which separate samples according to protein size or charge.

Recombinant Proteins in Biomedical Research

Bacterially expressed proteins are used by researchers in order to elucidate their function and crystal structure, or to find novel interaction partners of the protein relevant for the development and progression of human diseases. Recombinant proteins can be used for a large number of different applications such as sub-cellular localisation studies using tagged proteins or antibody production. Other applications include functional assays of novel proteins or the development of protein chips in order to analyse or find interacting proteins. Protein interactions can also be studied using recombinant proteins in two-hybrid systems or bead-based assays and protein complexes can be assembled *in vitro* using recombinant proteins and thus be studied in a controlled environment. The ultimate goal of research is to learn more about the complex mechanisms involved in diseases with the purpose of eventually being in a position to treat and even cure them.

Clinical Relevance

Recombinant proteins are used in many different fields of study from apoptosis and cancer research to Parkinson, Alzheimer's and many other human diseases and some recombinant proteins are already being utilised in medicine as therapeutic agents. For instance insulin, which is essential for patients suffering from diabetes mellitus, is commonly obtained from the pancreas of pigs, but it is also produced using genetically modified *E. coli* (e.g. Huminsulin Basal[®], Humalog[®], Lilly). Drugs that modulate the immune system such as interferon alpha (treatment of chronic hepatitis B or C, leukaemia and other diseases) and interferon gamma (used in patients with chronic ►granulomatosis) are also generated using *E. coli* as well as interferon-beta 1b (Betaferon[®], Schering), which is used to treat multiple sclerosis. A relatively new compound is pegfilgrastim (Neulasta[®], Amgen), a growth factor that increases the number of neutrophils in the blood of cancer patients undergoing chemotherapy and a therapeutic agent that is still undergoing clinical trials is rViscumin (VISCUM), a genetically engineered plant protein from mistletoe. The activity of the compound is equivalent to that of the natural protein from mistletoe, which is traditionally used against many diseases including cancer.

References

1. Baneyx F (1999) Recombinant protein expression in *Escherichia coli*. Current Opinion in Biotechnology 10:411–421
2. Hannig G, Makrides SC (1998) Strategies for optimising heterologous protein expression in *Escherichia coli*. Trends in Biotechnology 16:54–60
3. Murby M, Uhlén M, Ståhl S (1996) Upstream Strategies to Minimize Proteolytic Degradation upon Recombinant Production in *Escherichia coli*. Protein Expression and Purification 7:129–136
4. Nilsson J, Ståhl S, Lundeberg J et al (1997) Affinity Fusion Strategies for Detection, Purification, and Immobilization of Recombinant Proteins. Protein Expression and Purification 11:1–16
5. Jenny RJ, Mann KG, Lundblad RL (2003) A critical review of the methods for cleavage of fusion proteins with thrombin and factor Xa. Protein Expression and Purification 31:1–11

Recombinant Protein Production in Mammalian Cell Culture

ANJA DRÖGE

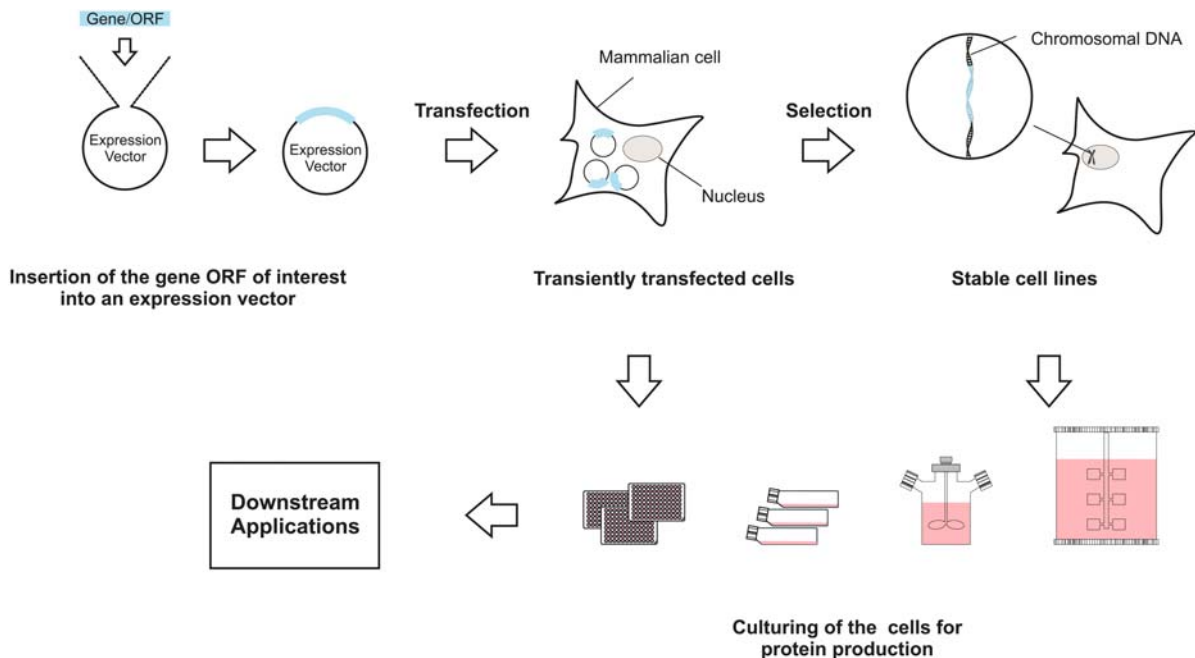
Max Delbrück Center for Molecular Medicine (MDC),
Berlin, Germany
a.droege@mdc-berlin.de

Definition

The cloning of genes in plasmid or viral vehicles and their subsequent introduction and expression in cultured mammalian cells has become an important tool for functional studies of eukaryotic proteins as well as for the industrial production of eukaryotic proteins for therapeutic use. This essay will cover the basic principles of protein expression in mammalian cell culture and summarize some of the most important applications of medical interest.

Characteristics

The production of eukaryotic proteins in mammalian cell culture offers many advantages *versus* protein production by the more economic bacterial cell cultures. One major advantage is that, in contrast to bacterial cells, mammalian cells are able to carry out proper ►protein folding, authentic ►glycosylation and a broad spectrum of ►posttranslational modifications (1). This assures the production of biologically active protein, which is required for many applications ranging from functional assays to therapeutic use. Furthermore mammalian cell cultures are ideal for studying eukaryotic proteins in



Recombinant Protein Production in Mammalian Cell Culture. Figure 1 Strategy for the production of a protein in mammalian cell culture. A DNA fragment coding for the protein of interest is inserted into an expression vector. The recombinant vector is introduced into cultured mammalian cells by transfection. For transient expression of the desired protein, the transfected cells are cultured for 24 to 96 hours. Alternatively, after transfection stable cell lines that integrated the desired coding sequence into their genome can be selected and propagated. Both transiently and stably transfected cells, can be used for various downstream applications.

convenient systems that can be experimentally manipulated but still resemble their natural environments. Using immunofluorescence microscopy and/or specific biochemical assays, essential information about the localization, the function and the interaction partners of proteins can be obtained. Such studies are of growing importance, given that after deciphering the human and other genomes, one major task is to understand the function of many so far uncharacterized proteins forming the complex proteomes of higher organisms.

Mammalian Expression Systems

The general strategy for the expression of a specific protein is outlined in Fig. 1. First, a DNA fragment that carries the **ORF** coding for the protein of interest (e.g. a **cDNA** fragment), is inserted into a **vector**, assuring its maintenance and expression. Subsequently the recombinant vector is introduced into a mammalian cell line of choice by **transfection**. Usually expression of the introduced coding sequences is seen between 12 and 96 h after transfection, depending on the vector used. Then expression decreases due to cell death or loss of the vector. This strategy often results in high levels of protein production but is transient i.e. persists only for a few days following transfection. However, transient expression systems are extremely useful for the rapid production of small quantities of protein.

For the continuous production of a specific protein e.g. for the production of therapeutics it is often advantageous to generate cell lines that permanently carry the ORF of interest. A commonly used strategy to accomplish this is to combine the relevant coding sequence with a selectable marker such as a gene conferring resistance to an antibiotic. After transfection cells are cultivated in the presence of the antibiotic to select for cells that integrated the resistance marker and the DNA carrying the ORF of interest into their chromosomal DNA. Once cells have been successfully transfected and cultured they can be used for diverse downstream applications including immunofluorescence microscopy, various biochemical assays or simply for the purification of the protein of interest.

In the following sections a survey of the most important tools and methods for mammalian expression will be given including expression vectors, transfection techniques, cell lines and cell culture types.

Vectors for Mammalian Expression

To deliver the coding sequence of interest into a mammalian cell it has to be inserted into an appropriate vector. The vector assures replication and maintenance as well as the controlled expression of the insert, which would otherwise be rapidly lost. In addition, many vectors offer the opportunity to combine the coding

sequence of interest with “tags” which code for short proteins or peptides that are fused to the target protein. For example, the hemagglutinin-tag codes for a short ▶**epitope** that is recognized by commercial antibodies allowing the immunological detection or affinity purification of the fusion protein. This is especially useful for the characterization of novel proteins for which antibodies are not yet available. Other tags like the His-tag that consists of 6 histidine residues additionally permit the convenient one-step purification of the tagged fusion protein on commercially available chromatography resins.

Numerous vectors are available for the delivery and expression of genes in mammalian cells (reviewed in 2, 3). The existing vectors can be divided into two categories, plasmid and viral vectors.

Plasmid Vectors

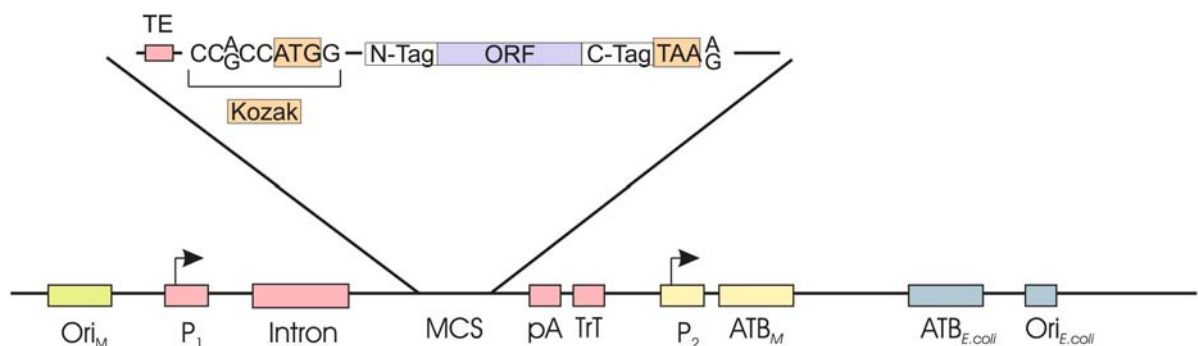
A broad variety of plasmid vectors is used for gene expression in mammalian cells. These plasmids contain functional elements for the replication and maintenance as well as for the efficient expression of the inserted sequence. The most important components of mammalian expression plasmids are illustrated in Fig. 2.

The sequence coding for the protein of interest is inserted into the ▶**multicloning site** (MCS) of the plasmid, which is located downstream of a mammalian ▶**promoter**, the site of transcription initiation. According to the specific requirements, one can choose among promoters for ▶**constitutive expression** or ▶**inducible expression**. Many plasmids include ▶**intron** sequences downstream of the promoter at the 5' end of the MCS, since their presence has been shown to improve expression of the usually intron-less coding sequences

(e.g. cDNA fragments). At the 3' end of the coding sequence a ▶**polyadenylation** signal is included. Polyadenylation increases the stability and the translatability of the ▶**mRNA**. To ensure correct transcriptional termination, a termination signal is placed at the 3' end of the coding sequence. This is required since eukaryotic expression plasmids usually consist of more than one transcriptional unit and continued transcription from one promoter inhibits the function of downstream promoters.

Efficient translation of the mRNA transcript in mammalian cells requires specific translational signals. The Kozak sequence CC(A/G)CCAUGG has been defined as the optimal translation initiation sequence in higher eukaryotes, with AUG being the start codon. Within this sequence the purines A/G three nucleotides upstream of the start codon and the G immediately following the start codon are most influential for optimal translation initiation. Likewise the efficiency of translational termination is modulated by the sequence context of the stop codon and is more effective if the base immediately following the stop codon is a purine rather than a pyrimidine. To allow the generation of stable cell lines as described above, most mammalian expression vectors carry marker genes suited for selection in mammalian cells such as genes conferring resistances to hygromycin or neomycin.

Last but not least, for amplification and replication of the recombinant plasmids in *Escherichia coli*, plasmids carry a bacterial origin of replication and a marker gene conferring resistance to an antibiotic. This is important because classical cloning procedures usually involve amplification of the recombinant plasmids in *E. coli*. Furthermore the recombinant plasmid-DNA can be



Recombinant Protein Production in Mammalian Cell Culture. Figure 2 Components of a mammalian expression plasmid. The functional elements depicted include signals required for bacterial (blue) and mammalian (green) replication of the plasmid, signals for correct transcription of the inserted sequence and modification of the transcript (pink) and signals assuring optimal translation of the protein of interest (orange). Abbreviations: ATB_M – antibiotic resistance gene for mammalian selection, ATB_{E.coli} – antibiotic resistance gene for bacterial selection, MCS – multiple cloning site; N-Tag/C-Tag – sequences coding for N- or C-terminal tags, respectively; Ori_M – mammalian origin of replication; Ori_{E.coli} – bacterial origin of replication; P – Promoter; pA – polyadenylation signal, TE – Transcriptional Enhancer, TrT – Transcription Terminator.

economically purified from *E. coli* cultures and used for the transfection into mammalian cells.

Viral Vectors

Viruses by their nature are efficient gene delivery vehicles and therefore excellent tools for introducing and expressing genes in mammalian cell lines. Most of the numerous different viral expression systems have in common that the structural genes that code for the proteins forming the virus capsids are deleted from the viral genome, while the gene of interest is inserted. To allow generation of infectious virus particles, these vectors are then either cotransfected with helper plasmids expressing the missing structural genes or are transfected into helper cells containing integrated copies of the required structural genes. These strategies result in the generation of viral particles that can be directly used to infect cultured cells with high efficiency but are replication-incompetent. This important feature assures the biosafety of these expression systems that are in many cases based on pathogenic viruses.

The most important viral expression systems for mammalian cells are based on Semliki Forest virus, adenovirus, lentivirus and vaccinia virus.

DNA Delivery

Unlike DNA embedded into virus particles, “naked” DNA, like other charged molecules, cannot freely pass the lipid bilayer of the eukaryotic cell membrane. Therefore tools for efficient DNA delivery have been developed (4). The choice of method depends on the cell type, the size of the culture and the required transfection efficiency. Ideally transfection should be fast and easy to perform, reproducibly yield high transfection efficiencies and cause minimal cytotoxicity. For the transfection of cultures on a large scale economic considerations can play an additional role. The methods used for transfection can be broadly classified into chemical and physical methods.

The principle of chemical methods is to complex the DNA with carrier molecules that overcome the cell-membrane barrier. These methods make use of the fact that the negatively charged nucleic acids adsorb to positively charged molecules or macromolecular complexes (the carriers) by electrostatic interaction. The carrier DNA-complexes are efficiently taken up by the cells and the nucleic acid is released within the cytoplasm or the nucleus. The first method that successfully applied this principle is the calcium phosphate method, in which DNA is co-precipitated with calcium chloride by addition of a phosphate buffer. The resulting calcium phosphate-DNA complexes adhere to the cell membrane and enter the cytoplasm. Other DNA carriers that are widely used for transfection include cationic polymers such as

DEAE-dextran, protamine, activated dendrimers and polyethylenimine polymers, liposomes and non-liposome lipids.

Several methods have been developed to deliver DNA into the cytoplasm of the target cells physically. These include the microinjection of DNA into cultured cells, the bombardment of cells with DNA immobilized onto microscopic heavy-metal particles (biolistic particle delivery) or the perturbation of the cell membranes by electrical pulses, which allows DNA entry (electroporation). The latter methods are usually not used for the routine transfection of cultured cells. Rather they are valuable tools for cells that are difficult to transfect by chemical methods or for specific applications such as DNA transfer into embryonic stem cells (microinjection), the transfection of plant protoplasts (electroporation) or the delivery of genes to cells *in vivo* (biolistic particle delivery).

Cell Hosts

Among the hosts that can be used for protein expression three major types, primary cultures, cell strains and continuous cell lines, can be distinguished. For primary cultures, cells are derived directly from excised tissue, which is typically disaggregated into single cells by enzymatic digestion with proteolytic enzymes such as trypsin. Cells are then cultured in flasks in appropriate media. The major advantage of primary cultures is that the cells have not been “modified” in any way and usually retain the distinctive characteristics they display *in vivo*. Therefore primary cells offer the opportunity to study proteins in an almost natural environment. Among the disadvantages of primary cultures on the other hand are the labor-intensive preparation of the cells and their limited life span. Moreover primary cells are much more difficult to transfect than other culture types.

Cell strains are derived from dividing primary cells by subculturing i.e. serially propagating them into cell culture dishes with fresh medium. The resulting cell strains have a finite lifetime in culture and usually die after 50–100 divisions.

Continuous cell lines have undergone genetic changes that allow them to divide indefinitely – they have “transformed” into the immortal state. Immortal cell lines can be derived from tumor cells that have undergone transformation in the organism. One example of a continuous cell line derived from a tumor is the widely used HeLa cell line that originates from a cervical carcinoma (4) and was called after the patient (Henrietta Lacks) from whom they were isolated. Continuous cell lines can also be derived from normal cell strains. Although as mentioned above these cells usually die after 50–100 divisions, variant cells that have undergone transformation arise at low frequency. Finally, transformation of cells can be experimentally

induced by treatment of normal cells with tumor viruses, chemicals or radiation. Due to the genetic modifications associated with immortalization, cell lines usually lose some of their original *in vivo* characteristics, which is a disadvantage for functional studies. On the other hand continuous cell lines have many advantages including almost limitless availability, fast growth, easy handling and consistency. Moreover continuous cell lines can be further modified by genetic manipulation and can therefore be adapted to the specific requirements of a protein production process. As a consequence most large-scale productions of mammalian recombinant proteins use continuous cell lines (1).

Culturing Cells

Cell cultures can either grow in suspension (as single cells or small free-floating clumps) or in monolayers adherent to the surface of a tissue culture flask. Both types can be grown at various scales. Adherent cells can be grown in the wells of 384-well microtitre plates with a surface of approximately 0.05 cm² as well as in large stackable chambers with each stack providing 2.5 m² of growth area. Likewise, the volumes of suspension cultures can range from a few ml in a small flask to up to more than 10,000 l in large stirred tanks for industrial production. Independently of the culture morphology the cell density has to be carefully controlled and contaminations have to be avoided by sterile working techniques and, if possible, the inclusion of antibiotics in the culture medium.

Clinical Relevance

Recombinant protein expression in mammalian cell culture is widely used for manufacturing of recombinant human proteins for therapeutic and diagnostic use (5, 6). Among the recombinant proteins that have already been approved as therapeutics are blood factors (e.g. Factor VIII, Factor VII), anticoagulants (e.g. tissue plasminogen activator), hormones (e.g. insulin, human growth hormone, erythropoietin), interferons and interleukins. Although most of these proteins can be, and earlier have been, extracted from their natural source, only large-scale recombinant protein production made them available in amounts sufficient for their common medical use. One good example of this is erythropoietin, which in former times had been isolated in small amounts from serum and urine. The increased availability of erythropoietin due to recombinant production allowed its widespread application for the treatment of anemia in chronic renal failure, which has led to a reduced morbidity and considerable improvements in the quality of life of patients. Recombinant protein production not only overcomes problems with the availability of therapeutic proteins, but also helps to avoid safety risks associated with natural sources such as human blood, urine or tissue. Another major advantage of recombinant protein production *versus* traditional

methods for the isolation of biopharmaceuticals is the possibility of modifying and thereby improving the naturally occurring proteins e.g. by generating faster or longer acting drugs or drugs with a higher activity.

Another important group of recombinant therapeutics produced in mammalian cell culture are monoclonal antibodies (6, 7). Naturally, antibodies are produced by B-lymphocytes. Each of these cells produces antibodies of unique specificity. By fusion of mouse B-cells to tumor cells derived from a mouse myeloma, immortal 'hybridomas' can be produced, which make the same antibody as the original B-cell but have the advantage that they can easily be cultured and used for the large-scale production of antibodies. Due to their murine origin these "first generation" monoclonal antibodies often provoked severe immune reaction in patients, especially after repeated treatment. To overcome these problems, chimeric and humanized antibodies were subsequently engineered and several of these antibodies are already approved and used for the treatment of a variety of diseases.

References

1. Chu L, Robinson DK (2001) Industrial choices for protein production by large-scale cell culture. *Curr Opin Biotech* 12:180–187
2. Makrides SC (1999) Components of vectors for gene transfer and expression in mammalian cells. *Protein Expres Purif* 17:183–202
3. Colosimo A, Goncz KK, Holmes AR et al (2000) Transfer and expression of foreign genes in mammalian cells. *BioTechniques* 20:314–331
4. Gey GO, Coffman WD, Kubicek MT (1952) Tissue culture studies of the proliferative capacity of cervical carcinoma and normal epithelium. *Cancer Res* 12:264
5. Walsh G (2000) Biopharmaceutical benchmarks. *Nat Biotech* 18:831–833
6. Walsh G (2003) Biopharmaceutical benchmarks – 2003. *Nat Biotech* 8:865–870
7. Dijk MA van Winkel JGJ van de (2001) Human antibodies as next generation therapeutics. *Curr Opin Chem Biol* 5:368–374

Recombinant Protein Production in Yeast

ROLAND WEIS, FRANZ HARTNER, ANTON GLIEDER
Research Centre Applied Biocatalysis, Institute of
Molecular Biotechnology, Graz University of
Technology, Austria
glieder@glieder.com

Definition

Proteins of industrial or medical interest are often produced in very low amounts by their natural hosts. However, for almost every application they are needed in both high yield and purity in order to meet safety and feasibility demands. This implies a need for high quality recombinant protein expression in heterologous production hosts. Prokaryotic systems such as *Escherichia coli* are cheap, fast and well established, but often fail to produce proteins in the required qualities, while on the other hand, production by cell cultures is time-consuming and cost-intensive.

As an alternative yeasts combine ease of genetic manipulation with high productivity and the ability to perform eukaryotic protein processing. Moreover, the scale-up to large-scale production is well established.

Characteristics

Standard procedures for expression of a foreign gene in yeast require some basic steps: (1) cloning of the gene into an appropriate expression vector and amplification of the DNA in the prokaryotic host *Escherichia coli*, (2) introduction of the expression cassette into the yeast of choice and (3) induced or constitutive expression of the foreign gene product.

Step (1) cloning and preparation of the DNA. A generalised scheme of an expression vector for yeast is shown in Fig. 1. It has to be supplied with an origin of replication and a selection marker for *E. coli*. Essential yeast-specific features include an appropriate selection marker (either resistance to antibiotics or ►[complementation](#) of metabolism-deficiencies), a eukaryotic promoter-region driving the expression of the gene of interest and a terminator-region ensuring proper termination of the transcript. The vector is either integrated into the genome of the yeast host by homologous recombination, or kept in the cell as extrachromosomal DNA. In the latter case, an additional DNA-sequence on the plasmid called the ►[autonomously replicating sequence](#) (ARS) is responsible for correct replication and segregation.

Step (2) transformation of yeast cells. The foreign DNA is introduced into the yeast cells by ►[electroporation](#) or ►[protoplast transformation](#). For gene integration the vector is linearized by cutting it with ►[restriction endonucleases](#) prior to transformation. Free DNA ends stimulate homologous recombination at a higher rate. The advantages of integrative transformation include e.g. expression strain stability, generation of multicopy strains and control of integration site.

Step (3) expression of foreign genes. Transformed yeast cells contain at least the gene of the ►[selection marker](#) and usually also the gene which should be expressed, as long as no recombination event has taken place that separated this gene from the selection

marker. Usually several transformants have to be analysed for expression of the desired gene product.

In an ideal expression strategy a compromise must be found combining high-level expression of recombinant protein with optimal cell growth. The choice of the promoter driving the production of the desired protein can be regarded as the committing step; the first option is to take a strong constitutive promoter leading to continuous protein expression during all stages of cell cultivation. However, constitutive expression cannot be employed for cytotoxic proteins (5).

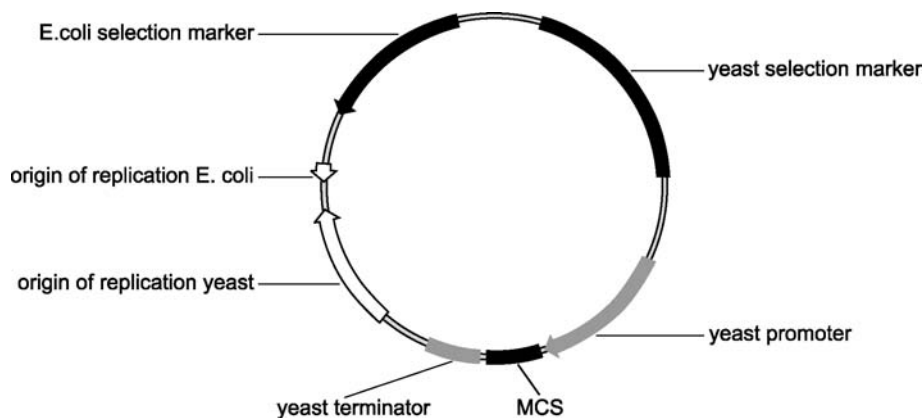
Alternatively, inducible promoters are available that are characterised by strong regulation; in the absence of inducer and/or presence of a repressing force, the gene is not expressed. Repressor depletion and/or supplementation of inducer initiate gene expression. In most cases the addition of inducer is strictly required to obtain considerable amounts of heterologous protein. Inducible promoter systems often provide highly efficient expression and the possibility of producing toxic proteins.

Employing yeasts for the expression of eukaryotic proteins usually helps to avoid difficulties with protein folding, insolubility or formation of protein oligomers due to their fully functional eukaryotic protein synthesis machinery. However, the success of recombinant protein expression depends on many experimental parameters. The decision for a specific host/protein-combination and the optimal ►[fermentation](#) process is a case-to-case choice that is dictated by an economic rationale and product properties. In both regards yeasts seem to be preferable to many other systems.

Saccharomyces cerevisiae

The yeast *Saccharomyces cerevisiae* is the workhorse for many industrial and scientific applications including heterologous protein production. Several reasons drive the extensive use of this yeast: (a) well-established, cost-effective fermentation and process technology; (b) efficient and highly developed methods for molecular manipulation and (c) its ►[GRAS](#)-status (generally regarded as safe).

Readily available expression vectors for this host organism fulfil two main requirements: high-copy-number for dosage-dependent high-level expression and mitotic stability under non-selective conditions (2). Yeast episomal vectors (YEp) can replicate independently from the yeast genome due to the presence of an ARS. Usually 30 or more copies of these vectors are present inside the cells. Vectors of the integrative-type (YIp) have no ARS sequence and are stably integrated into the genome in low copy numbers (<20). Several strategies for multiple-integration events have been developed, as e.g. targeting the recombinant DNA to the ►[LTRs](#) (long terminal repeats) of the retrotransposon Ty or the ►[ribosomal DNA](#) (rDNA) cluster.



Recombinant Protein Production in Yeast. Figure 1 Standard yeast plasmid for genomic integration or episomal replication using a yeast ARS. The “gene of interest” is inserted into the multiple cloning site (MCS); all necessary features for an *E. coli*-yeast species-shuttle vector are included (origins of replication, selection markers, promoter, terminator and MCS).

A variety of well-characterised constitutive, inducible and engineered (hybrid) promoter elements are in use to drive the expression of foreign genes. They are mainly derived from yeast genes of central metabolic pathways. Strong constitutive promoters include *ADH1* (alcohol dehydrogenase 1), *PGK1* (phosphoglycerate kinase) and *GAPDH* (glyceraldehyde-3-phosphate dehydrogenase) promoters. For inducible expression of foreign genes, the promoters of the *GAL*-system are employed. They are repressed in the presence of glucose and can be induced approximately 1000-fold by galactose (4).

Secretion of heterologous proteins is often an advantage due to simpler purification steps and in order to overcome problems with intracellular [▶proteolytic degradation](#). For several heterologous proteins effective secretion has been shown under the direction of their native [▶leader signals](#) but generally host-derived signal sequences are much more reliable. In *S. cerevisiae* the most widely used signal sequences derive from acid phosphatase (*PHO5*), invertase (*SUC2*) or mating factor α (MF α 1). The last is extensively used in a wide variety of yeasts (4).

Protein glycosylation is a [▶posttranslational modification](#) step common to all eukaryotes. While in mammalian cells highly complex N- and O-linked oligosaccharide-structures composed of a variety of sugars are produced, most yeasts contain mannose type structures lacking some typical mammalian sugar entities like galactose or sialic acid. These non-mammalian structures can result in severe problems for therapeutic applications, mainly due to their high potential for immunogenicity. To overcome these problems, efforts are on the way to humanise yeast-derived oligosaccharide structures. All eukaryotes

share the early steps of the glycosylation pathway, which produce a core structure in the endoplasmic reticulum (ER). In mammals this structure is usually trimmed by mannosidases before building the mature complex-type oligosaccharides. Yeasts do not trim the core structure to the same extent before adding further sugars, which are almost solely mannose moieties. *S. cerevisiae* especially even tends to hypermannosylate glycosylation targets. Co-expression of recombinant *Aspergillus saitoi* α -1,2-mannosidase was shown to produce human-like (Man₅GlcNAc) core structures. As an alternative the *mm9* mutant of *S. cerevisiae* can be used to avoid hyperglycosylation (1).

Non-Conventional Yeasts

During the last ten years, an increasing number of scientific studies have described the successful employment of non-conventional yeasts, especially of the methylotrophic genera, for recombinant protein production. Often higher protein yields were achieved than with any other host system.

Methylotrophic Yeasts

A limited number of yeast species is able to grow on methanol as sole energy and carbon source. They belong to the genera of *Pichia*, *Hansenula*, *Candida* and *Torulopsis*. When the cells are grown on methanol, the enzymes of the general methanol utilization pathway can account for up to 30% of total intracellular protein content.

Their genes provide strong promoters for high-level protein expression and methylotrophic yeasts can be grown to extremely high cell densities. The low level of secreted endogenous proteins allows efficient production of almost pure secreted heterologous proteins in

the culture supernatant. The massive abundance of alcohol oxidase upon induction by methanol makes this gene a preferred target for site-directed insertion of expression cassettes by gene replacement (2).

Hansenula polymorpha (*Pichia angusta*)

Standard expression vectors for this methylotrophic yeast carry the strong *H. polymorpha* *MOX* (methanol oxidase)- or *FMD* (formate dehydrogenase)- promoters. The *MOX* gene product introducing methanol into the catabolic pathway is present in high abundance in cells grown on methanol (induction). Since methanol is not strictly required for protein expression considerable amounts of protein are produced in cells grown in media supplemented with low levels of glycerol (derepression). However, upon induction by methanol, product levels are tremendously elevated. Several ▶auxotrophic *H. polymorpha* host strain/marker gene combinations as well as dominant markers are available, as described above for *S. cerevisiae*.

One observes a high frequency of plasmid integration, despite the presence of an ARS on a vector. Transformation results in mitotically stable strains containing varying copy numbers (up to 100) of expression cassettes integrated by non-homologous recombination. However, a high copy number does not necessarily confer high expression levels, since this depends also on the relative strength of the promoters involved, the 5'-untranslated region, the surrounding of the translation initiation site, the integration site and of course the nature of the heterologous gene and protein to be expressed. As for *S. cerevisiae*, Mfa1 is the preferred leader sequence for protein secretion. Glycosylation by methylotrophic yeasts does not differ from that of *S. cerevisiae* except that hypermannosylation is found to be less common. Unpredictable random integration of the expression cassette into the genome of *H. polymorpha* and the genetic instability of high copy number production strains are drawbacks for some applications (2).

Pichia pastoris

Unlike *H. polymorpha*, this species contains two alcohol oxidase genes, *AOX1* and *AOX2*, which offer different strategies for gene insertion. *AOX1* is the major enzyme for methanol-oxidation as compared to the weakly expressed *AOX2*. The lack of stable ARS-elements for *P. pastoris* raises the importance of stable gene integration for reliable usage in commercial high-level protein expression. For selection, complementation of metabolic deficiencies in the corresponding host as well as resistance to antibiotics is routinely used. The strong methanol-inducible *AOX1*-promoter (alcohol oxidase 1) and the strong constitutive *GAPDH*-promoter (both derived from *P. pastoris*) are frequently

used for protein expression. Less frequently employed are the weaker *FLD1*- (glutathione-dependent formaldehyde dehydrogenase, constitutive), *YPT1*- (gene encodes a GTPase involved in secretion, constitutive) or *PEX8*- (gene encodes a peroxisomal matrix protein, constitutive and inducible) promoters (5). The *AOX1*-promoter is tightly repressed by most carbon sources such as glucose, ethanol and glycerol and highly induced by methanol in the absence of these C-sources. Up to 20 g of heterologous protein are produced per litre of culture volume. For secreted proteins the Mfa1 signal sequence is the most frequently used, although many signal sequences from mammals and plants are functional as well. Recent publications demonstrated the feasibility of humanised protein glycosylation by *P. pastoris*. Therefore, *Pichia pastoris* can be regarded as a powerful system for the economical, recombinant large-scale production of mammalian polypeptides (1).

Fission Yeasts

The ▶fission yeast *S. pombe* diverges in some respects from the “classical” budding yeasts (e.g. *S. cerevisiae*). Apart from its asexual proliferation by cell division (fission), it shows a variety of typical properties of mammalian cells. For example, its mitochondrial genome structure, type of RNA-splicing, transcription-initiation mechanism, sugar chains of glycoproteins and posttranslational modifications resemble mammals more than fungi.

Interestingly, even though this species has been extensively studied in terms of molecular genetics and cell biology, it has not gained much importance in producing heterologous proteins yet. However, several of the features outlined above imply extraordinary beneficial capabilities of *S. pombe*, which might be exploited in the near future in the course of production of authentic heterologous polypeptides for therapeutic applications (3).

Clinical Relevance

The production of recombinant proteins in mammalian cell cultures is often difficult, error-prone and expensive, while still low yields are obtained. In addition, the assurance of virus-free products is a serious task. As an alternative, yeasts are employed efficiently for the production of therapeutically interesting recombinant proteins (Table 1), where *E. coli* cannot be used, usually due to difficulties in folding, posttranslational modification or quality demands (2, 5, 7, 8).

In order to avoid rejection or other immunological reactions, some proteins need an exactly matching mammalian glycosylation pattern. This cannot be done by native yeasts. Up to now no fully humanised, glycosylated recombinant protein has been produced in

Recombinant Protein Production in Yeast. Table 1 Recombinant therapeutic proteins produced in yeasts (all products are secreted into the culture media); HBsAg = Hepatitis B surface antigen; HSA = Human serum albumin

Host	Biopharmaceutical	Application
<i>S. cerevisiae</i>	hirudin	anticoagulant
<i>S. cerevisiae</i>	insulin	hormone - diabetes
<i>S. cerevisiae</i>	glucagon	hormone - diabetes
<i>S. cerevisiae</i>	HBsAg	vaccine
<i>S. cerevisiae</i>	urate oxidase	hyperuricaemia
<i>H. polymorpha</i>	HBsAg	vaccine
<i>H. polymorpha</i>	hirudin	anticoagulant
<i>H. polymorpha</i>	α 1-antitrypsin	anti-emphysemic
<i>P. pastoris</i>	HSA	blood plasma surrogate
<i>P. pastoris</i>	α 1-antitrypsin	anti-emphysemic
<i>P. pastoris</i>	hirudin	anticoagulant

yeasts, but the routine generation of metabolically engineered yeast strains is getting closer. It should be mentioned however, that mammalian cell lines, which are often derived from cancer cells, also usually show some deviations from human standard glycosylation patterns. Prior to administration, many heterologously expressed and purified proteins have to be treated further to become fully active. For example, insulin is derived from proinsulin *via* cleavage by carboxypeptidase Y. Often these “secondary” modifying enzymes that process precursors of therapeutically interesting targets are also produced in yeast systems (J.M. Cregg, personal communication).

As an example for the expression of a therapeutic protein in different hosts, the production of hirudin has been chosen. Hirudin is a low molecular weight anticoagulant peptide (~7 kD) naturally excreted from *Hirudo medicinalis*. Hirudin prevents fibrinogen from clotting, in addition to inhibiting several other thrombin-catalysed reactions such as the activation of factors V, VIII and XIII, as well as platelet activation. The most important application of hirudin is the treatment of the arterial thrombosis that causes cardiovascular diseases, e.g. heart attack and stroke, in cases of heparin-induced thrombocytopenia. Originally production of hirudin in *E. coli* yielded 1 mg/L intracellular protein, with the major drawbacks that a mixture of two hirudin forms with different activities were produced and proteolysis occurred. Successful attempts to secrete the target into the periplasma increased the protein level to 200–300 mg/L.

To circumvent the folding problems and limited secretion by bacterial hosts, yeasts have been employed as an alternative. In *S. cerevisiae*, hirudin was efficiently secreted into the culture medium at up to 500 mg/L; but it turned out that in some strains two derivatives lacking C-terminal residues were produced due to C-terminal processing. In order to improve the production yield in the yeast system, *H. polymorpha* and *P. pastoris* were investigated for the expression of hirudin. The yields of secreted recombinant hirudin in these species were in the range of 1.5 g/L. In addition, C-terminal proteolysis was much less pronounced (7).

To date, the vast majority of commercial processes rely on the well-known yeast *S. cerevisiae*. Due to increasing knowledge about non-conventional yeasts, together with the apparent advantages in terms of large-scale fermentation and glycosylation issues, alternative yeasts will be employed in many cases in the near future. Some important therapeutic proteins produced by *P. pastoris*, e.g. human serum albumin (HSA), endostatin and angiostatin, are currently in clinical trials, demonstrating that this reorientation of therapeutic protein production by yeasts is in progress (5, 6).

References

1. Brettthauer RK (2003) Genetic engineering of *Pichia pastoris* to humanize N-glycosylation of proteins. Trends Biotechnol 21:459–462

2. Gellissen G, Hollenberg CP (1997) Application of yeasts in gene expression studies: a comparison of *Saccharomyces cerevisiae*, *Hansenula polymorpha* and *Kluyveromyces lactis* – a review. *Gene* 190:87–97
3. Giga-Hama Y, Kumagai H (1999) Expression system for foreign genes using the fission yeast *Schizosaccharomyces pombe*. *Biotechnol Appl Biochem* 30:235–244
4. Hensing MCM, Rouwenhorst RJ, Heijnen JJ et al (1995) Physiological and technological aspects of large-scale heterologous-protein production with yeasts. *Antonie van Leeuwenhoek* 67:261–279
5. Lin Cereghino J, Cregg JM (2000) Heterologous protein expression in the methylotrophic yeast *Pichia pastoris*. *FEMS Microbiol Rev* 24:45–66
6. Mac Donald NJ, Shivers WY, Narum DL et al (2001) Endostatin Binds Tropomyosin. A potential modulator of the antitumor activity of endostatin. *J Biol Chem* 276:25190–25196
7. Sohn JH, Kang HA, Rao KJ et al (2001) Current status of the anticoagulant hirudin: its biotechnological production and clinical practise. *Appl Microbiol Biotechnol* 57:606–613
8. Walsh G (2003) Biopharmaceutical benchmarks – 2003. *Nat Biotechnol* 21:865–878

Recombination

Definition

Recombination refers to the exchange of genetic material between different chromosomes or chromatids resulting in changes of the phenotype.

- Genetic Epidemiology
- Huntington's Disease
- Repeat Expansion Diseases

RecQ Helicase Family

Definition

RecQ helicases belong to a gene family of helicases that is highly conserved in evolution. The protoType Is RecQ of *E. coli*, a 610 amino acid protein.

- Bloom Syndrome

Recurrence Risk

Definition

Recurrence risk is the probability that a genetic disease occurs in siblings or other relatives of an affected individual.

- Cleft Lip Palate
- Heritable Skin Disorders

Redefinition of Codons

Definition

Redefinition of codons describes a recoding event in which a different meaning is temporarily assigned to a codon. All cases known so far involve stop codons in special contexts. The typical example is UGA, which is recoded to specify the twenty-first amino acid selenocystein.

- Genetic Code

Reductive Unfolding

Definition

Reductive unfolding characterizes a process by which the disulfide bond(s) of a native protein is/are reduced under strongly reducing conditions, thereby denaturing its native structure.

- Protein Disulfide Bonds

Redundancy

Definition

Redundancy defines the state when two or more genes share similar functions, and upon single gene inactivation do not display a mutant phenotype. The loss of both/all genes is required in order to reveal the genes' functions.

- Morpholino Oligonucleotides, Functional Genomics by Gene 'Knock-down'

REE, RxRE

Definition

REE, RxRE describes the region on the HIV (human immunodeficiency virus) or HTLV (human T-cell leukaemia virus) genome to which Rev or Rex bind.

- Retroviruses

Reference State

Definition

Reference state defines a thermodynamic state relative to which all state functions are measured and to which binding is determined. Usually it is the unliganded native protein, or biopolymer. In DNA studies, for instance, it could be the double helix at specified pH, temperature and salt conditions.

- ▶ [Differential Scanning Calorimetry](#)
- ▶ [Isothermal Titration Calorimetry](#)
- ▶ [Thermodynamic Properties of DNA](#)

RefSeq

Definition

The Reference Sequence (RefSeq) collection is aimed at providing a comprehensive, integrated, non-redundant set of sequences, including genomic DNA, transcript (RNA), and protein products, for major research organisms.

- ▶ [Protein Databases](#)

Refsum Syndrome

Definition

Refsum syndrome or “Heredopathia atactica polyneuritiformis” is an autosomal recessively inherited peroxisomal metabolic disease, which is caused by a deficiency of phytanic-acid-oxidase and results in storage of phytanic acid in blood and tissues. The disease is characterized by ichthyosis of the skin, anomalies of bones, retinitis pigmentosa, polyneuropathy and cerebellar symptoms such as ataxia.

Region of Interest

Definition

Region of Interest describes an area of an image that will be photobleached or monitored during the time course of an imaging experiment.

- ▶ [FRAP](#)

Regulators of G Protein Signaling

Definition

Regulators of G protein signaling (RGS) consist of a large family of structurally related multifunctional signaling proteins, which bind directly to activated G α subunits and accelerate the rate of GTP hydrolysis.

- ▶ [G-Proteins](#)

REJ

- ▶ [Receptor for Egg Jelly Protein](#)

Relaxation Techniques/Relaxation

Definition

Relaxation techniques are primarily applied for quantitative determination of reaction kinetics in extremely fast chemical or biochemical reactions. The idea is to physically disturb an equilibrated system (e.g. by pressure or temperature jumps), and temporally resolve the dynamics of relaxation back to equilibrium. In NMR experiments, relaxation describes the disappearance of magnetization after excitation. The relaxation of nuclear magnetization depends on external (spectrometric) as well as internal (sample specific) parameters. Homogeneity of the magnetic field and temperature are important examples of external parameters that influence relaxation, while internal parameters as, for example, the shape, size and composition of a molecule determine the intrinsic relaxation properties of an NMR sample.

- ▶ [FCS](#)
- ▶ [Protein-Ligand-Interaction by NMR](#)

Release Factors

Definition

Release factors are protein factors that bind to the ribosome, and trigger the release of the finished polypeptide.

- ▶ [Translational Control in Eukaryotes](#)

Release Probability

Definition

Release probability describes the probability P_r with which a synaptic vesicle releases its content. The release probability is critically dependent on the influx of extracellular Ca^{2+} into the presynaptic terminal.

► [Neurons](#)

REM (Rapid-Eye-Movement) Sleep

Definition

REM sleep is one of two distinct brain states that occur during sleep. The name is derived from the fact that during REM sleep the eyes move rapidly behind the closed lids. REM sleep, also called paradoxical sleep, is a very active brain state as indicated by a desynchronized EEG similar to wakefulness, high variability of respiration and heart rate, and other signs of autonomic activation. Dreams occur during REM sleep but are not confined to this state.

► [Narcolepsy](#)

► [NonREM Sleep](#)

REM Sleep

► [REM \(Rapid-Eye-Movement\) Sleep](#)

Rendering

Definition

Rendering is a method to convert data into a visible format at the heart of all data visualization, and includes a large variety of graphic approaches. The most commonly used rendering methods are volume and surface rendering.

► [Electron Tomography](#)

Repeats

► [Repetitive Elements \(Repetitive DNA, Repeats\)](#)

Repeat Expansion Diseases

Definition

Repeat expansion diseases are defined as being a heterogenous group of generally rare, inherited disorders based on (dynamic) mutations/elongations of tandemly repeated DNA blocks of motifs spanning three to several base pairs.

► [Repeat Expansion Diseases, Dynamic Mutations Cause \(Neurological\) Model Disorder](#)

Repeat Expansion Diseases, Dynamic Mutations Cause (Neurological) Model Disorders

JÖRG T. EPPLN

Human Genetics, Ruhr-University Bochum, Bochum, Germany

joerg.t.epplen@rub.de

Definition

Repeat expansion diseases are caused by crucial extensions of naturally occurring polymorphic simple tandem DNA sequences within or in the vicinity of different human genes. A characteristic of these expanded alleles is their excessive instability in successive generations (► [dynamic mutations](#)) (4). Similarly, these alleles differ in lengths in cell populations of the same organism - mosaicism. A common clinical phenomenon in the so-called trinucleotide repeat disorders – the most abundant forms of repeat expansion diseases – is ► [anticipation](#), a clinical correlate of intergenerational repeat instability with a tendency towards further repeat elongation. Increased repeat lengths in successive generations lead to an earlier age of onset in affected individuals (and sometimes to accentuation of the symptomatology). So far, nearly two dozen human diseases with ► [Mendelian](#)

inheritance patterns have been shown to be caused by expansions of intrinsically polymorphic tandem repeats, mainly involving different trinucleotide motifs but also longer repetitive sequences up to 12mers. Interestingly, repeat expansion diseases have not been encountered thus far in animals or plants, but transgenic animal strains will be instrumental in understanding many facets of these diseases.

Characteristics

Two main categories of repeat expansion diseases are discernible on the basis of the expression of the repeats:

- **Polyglutamine disorders** are caused by moderate (CAG)_n expansions in coding regions (Table 1). The expanded trinucleotide blocks are translated into uninterrupted polyglutamine tracts.
- An even more heterogeneous group of disorders is due to large or very large expansions of (GAA)_n, (CTG)_n, (CGG)_n, (CCTG)_n, (ATTCT)_n and (CCCGCCCCGCG)_n blocks (Table 1). These repeats are not translated into protein, but whenever expanded in mutation carriers they interfere critically with the regulation of **gene expression** and/or **protein expression**.

Thus the pathophysiology of the two disease categories differs in principle.

Polyglutamine Disorders

Polyglutamine diseases are due to enlarged (CAG)_n blocks in the coding regions of a number of different genes. The (CAG)_n trinucleotide tracts are translated into polyglutamine domains within proteins. Depending on individual thresholds for the critical repeats, the resulting proteins form peptides containing β -sheet structures, accumulate in cells and are deposited as **inclusion bodies** in the cytoplasm and/or nucleus. Apparently, the **misfolded proteins** may interfere with the **proteasome** and metabolic pathways of the cell and finally lead to neuronal **cell death** (5). Polyglutamine diseases may share similar pathophysiological aspects with respect to repeat elongation, although the expansion thresholds are disease specific (Table 1). In addition, polyglutamine diseases usually present as adult onset, neurological traits with progressive neuronal dysfunction leading to **neurodegeneration** in specific brain regions. Up to now, this group comprises ten diseases, among which **Huntington's disease** (HD) is the most frequent.

The course of HD is characterized by neurological and/or psychiatric symptoms with onset usually in the 4th to 5th decade of life. On average, HD patients die after 10–20 years of disease. In contrast to the homogenous causation of HD, the genetically heterogeneous **spinocerebellar ataxias** (SCAs) are a growing group of dominantly

inherited neurodegenerative diseases with cerebellar atrophy. SCAs have a broad range of overlapping clinical symptoms with typical cerebellar signs. The disease entities SCA1–3, SCA6, SCA7 and SCA17 are based on critical (CAG)_n expansions (for further details on additional forms of SCA's see below). **Machado-Joseph disease** is the Japanese variant of SCA3 with a slightly different clinical symptomatology. **Dentatorubral-pallidoluysian atrophy** (DRPLA) is also more common in Japan and shares some symptoms of HD and the SCAs. **Kennedy disease** (**spinobulbar muscular atrophy**) is a X-chromosomal recessive trait presenting with neuromuscular symptoms and endocrine deficiency due to **androgen insensitivity**. The causative (CAG)_n expansion is localized within the **androgen receptor gene** encoded on the **X chromosome**.

Like polyglutamine disorders, (GCG)_n and (GAC)_n trinucleotide blocks result, albeit rarely, in elongated polyalanine as well as polyaspartate stretches. Yet, compared to the (CAG)_n expansions these motifs are only marginally increased in length and they do not result in neurological diseases, instead affecting bone differentiation. As a non-neurological exception among trinucleotide diseases, the autosomal dominantly inherited **synpolydactyly** has been shown to be caused by expansions of a polyalanine tract in the HOXD13 protein. Usually 15 alanine residues are expressed, whereas affected individuals harbor 7–9 extra alanine residues in the mutated allele products. In contrast to 'classical' trinucleotide diseases, the polyalanine tract in HOXD13 protein is more stable – like the polyaspartate block in **pseudoachondroplasia**. Thus these mutations appear to be products of non-reciprocal **recombination** or misalignment during recombination in contrast to the elusive mechanism(s) of the other **repeat expansion diseases** (see below).

Diseases with Repeat Expansions that Are Not Expressed on the Protein Level

Large-scale expansions of 'non-protein' coding, repetitive DNA sequences may also cause disease. Such expansions show extensive variability concerning the repeat positions in the genes, the sizes of the expanded sequence blocks and the sequence motifs concerned. Apparently similarly heterogeneous are the known pathophysiological mechanisms involved. As a common denominator, this disease group shares phenomena of reduced gene expression at the transcriptional and/or translational levels. The fragile XA and the rare fragile XE mental retardation syndromes are X-chromosomal (recessively) inherited. The (CGG)_n and (GCC)_n repeats are localized in the 5' untranslated regions of *FMR1* and *FMR2* genes, respectively. Repeat expansions entail increased methylation of neighboring (promoter) DNA sequences and lead to inactivation of *FMR* gene

Repeat Expansion Diseases, Dynamic Mutations Cause (Neurological) Model Disorders. Table 1 repeat expansion diseases: dynamic mutations cause (neurological) model disorders

Disease	Gene designation	Inheritance	Protein	Chromosomal locus	Repeat motif	Localization in respective gene	Physiolog. Repeat range	Expanded Repeat range
Dentatorubral-pallidoluysian atrophy (DRPLA)	<i>DRPLA</i>	autos. dominant	atrophin-1	13p13.31	CAG	coding	6–35	49–88
Progressive myoclonus epilepsy (EPM1)	<i>EPM1</i>	autos. recessive	CSTB	21q22.3	CCCCGCCCGCCG	5'UTR	2–3	50–80
Fragile XA-syndrome (FRAXA)	<i>FMR1</i>	X chromosome	FMR-1	Xq27.3	CGG	5'UTR	6–53	>230
Fragile XE-syndrome (FRAXE)	<i>FMR2</i>	X chromosome	FMR-2	Xq28	GCC	5'UTR	6–35	>200
Friedreich ataxia (FRDA)	<i>frataxin</i>	autos. recessive	<i>frataxin</i>	9q13-21.1	GAA	intronic	7–34	>100
Huntington disease (HD)	huntingtin	autos. dominant	huntingtin	4p16.3	CaG	coding	4–35	36–250
Huntington disease-like 2 (HDL2)	<i>JPH-3</i>	autos. dominant	junctophilin	16q23-24	CTG	coding?	6–27	>40–60
Spinobulbar musc. atrophy, Kennedy d. (SMAX1)	AR	X chromosome	androg. rec.	Xq13-21	CAG	coding	9–36	38–62
Myotonic dystrophy 1 (MD1)	DM1	autos. dominant	DMPK/SIX5	19q3	CTG	3'UTR	5–37	>50
Myotonic dystrophy 2 (MD2/ PROMM)	DM2	autos. dominant	ZNF9/?	3q13.3-24	CCTG	intronic	~26	75– >11000
Spinocerebellar ataxia 1 (SCA1)	SCA1	autos. dominant	ataxin-1	6p63	CAG	coding	6–44	39–82
Spinocerebellar ataxia 2 (SCA2)	SCA2	autos. dominant	ataxin-2	12q24.1	CAG	coding	15–31	36–63

Repeat Expansion Diseases, Dynamic Mutations Cause (Neurological) Model Disorders. Table 1 repeat expansion diseases: dynamic mutations cause (neurological) model disorders (Continued)

Disease	Gene designation	Inheritance	Protein	Chromosomal locus	Repeat motif	Localization in respective gene	Physiolog. Repeat range	Expanded Repeat range
Spinocerebellar ataxia 3 / Machado-Joseph d. (SCA3)	SCA3	autos. dominant	ataxin-3	14q32.1	CAG	coding	12–40	55–84
Spinocerebellar ataxia 6 (SCA6)	SCA6	autos. dominant	CACNA1A	19p13	CAG	coding	4–18	21–33
Spinocerebellar ataxia 7 (SCA7)	SCA7	autos. dominant	ataxin-7	13p12-13	CAG	coding	4–35	37–306
Spinocerebellar ataxia 8 (SCA8)	SCA8	autos. dominant	not known	13q21	CTG	3'UTR	16–37	110–250
Spinocerebellar ataxia 10 (SCA10)	SCA10	autos. dominant	ataxin-10	22q13	ATTCT	intronic	10–22	800–4600
Spinocerebellar ataxia 12 (SCA12)	SCA12	autos. dominant	PP2A-PR55B	5q31-33	CAG	promotor	7–28	66–78
Spinocerebellar ataxia 17 (SCA17)	SCA17	autos. dominant	TBP	6q27	CAG	coding	25–42	47–63
Oculopharyngeal muscular dystrophy (OPMD)	PABPN1	autos. dominant	PABPN1	14q11.2-13	GCG	coding	6	7–13
Synpolydactyly, type II (SPD)	HOXD13	autos. dominant	HOXD13	2q31-32	GCN	coding	15	22–24
[Pseudo]chondroplastic dysplasia (PSACH)	COMP	autos. dominant / het.	COMP	19p13.1	GAC	coding	5	6–7]

expression (1) resulting in critical perturbations in regular psychomotor development.

An ▶**intronic** expansion of a $(GAA)_n$ repeat in autosomal-recessively inherited ▶**Friedreich ataxia** prevents expression of frataxin, a protein maintaining iron homeostasis and therefore mitochondrial function. Myotonic dystrophy 1 (DM1), on the other hand, is a disorder characterized mainly by muscle wasting and myotonia. It is caused by $(CTG)_n$ expansions in the 3'-untranslated region of the *DMPK* gene. ▶**Dominant** inheritance suggests a dominant negative effect of the expansion on gene ▶**transcription**, eventually influencing the expression of neighboring genes. DM2/PROMM is the first repeat expansion disease found to be caused by tetranucleotide motifs (3). Certain forms of SCAs have expansions of a $(CTG)_n$ repeat in the 3' untranslated region (SCA8), an intronic $(ATTCT)_n$ block (SCA10) and a $(CAG)_n$ tract in the promoter region of the disease gene (SCA12), respectively. The $(ATTCT)_n$ expansion in the *SCA10* gene is the sole pathogenic pentanucleotide known to date.

Theoretically, additional simple tandem repeats may also cause expansion diseases like elongated ▶**minisatellite** alleles with their longer motifs/periodicities. The ▶**multifactorial** nature of ▶**common diseases** like ▶**diabetes mellitus** and certain ▶**malignancies** includes genetic contributions of e.g. given minisatellite loci and their specific alleles (2). These common diseases characteristically lack classical Mendelian inheritance patterns with high ▶**penetrance** making the slight variations (and hence effects) of repeat allele lengths conceptually optimal candidates with the required differential properties to be responsible for genetic contributions, probably exclusively in certain combinations, together with additional, thus far elusive, factors.

Cellular and Molecular Regulation

A number of large-scale repeat expansions had previously been shown to cause cytogenetically detectable 'fragile' sites demonstrable in metaphase ▶**chromosomes**. Yet this phenomenon is not inevitably pathogenic, except for critical expansions in the abovementioned *FMR* genes. Approximately 8% of human genes bear repeat ▶**polymorphisms** and are candidates for pathogenic expansions (Wren et al., 2000). Two general mechanisms are discussed to account for repeat instability. i) 'Slippage' of DNA polymerase during mitotic and meiotic ▶**DNA replication** (Sinden, 2001). But 'slippage' explains only low degree length variations observed in the physiological polymorphism ranges of these frequent loci due to mutations of 1(-2) repeat units. It may result from imprecise DNA replication during error-prone cell divisions, especially during gametogenesis. ii) Potentially, disease-associated triplet repeats form stable 'alternative' ▶**DNA conformations** like triplex and

quadruplex helices and hairpins and promote intermediate to large range expansions during DNA replication. Finally, ineffective ▶**DNA repair** mechanisms may account for some additional instability phenomena.

Depending on repeat class, different mechanisms are responsible for disease development.

- A potential common denominator of polyglutamine disorders includes distorted degradation and/or accumulation of mutant protein complexes. Yet formally, the relevance of the microscopically visible protein inclusions and of their cellular localization etc. remains to be proven. Additional disorder-specific molecular complexities behind these rather general rules are not yet explicable for the different disease entities.
- $(CGG)_n$, $(GCC)_n$ expansions in ▶**fragile X syndromes** and the intronic expansions in *FRDA* [$(GAA)_n$] and *SCA10* [$(ATCTT)_n$, Matsuura et al., 2000] down-regulate or abolish gene expression on the transcriptional level *via* nearly complete ▶**loss of function** (1). The pathophysiological effects of these expansions appear explicable solely by the reduced mRNA levels of the affected gene.
- Additional pathomechanisms behind untranslated $(CTG)_n$ and $(CCTG)_n$ expansions in the myotonic dystrophy forms DM1 and DM2/PROMM as well as *SCA8* are only gradually being understood, like the complex sequelae of $(CAG)_n$ expansions in the 5' UTR of the *SCA12* gene. In addition to the demonstrably reduced expression of the respective genes, the apparent dominant negative effect with accumulations of mutant RNAs in nuclear foci offers a novel rational explanation. Furthermore, expression of additional neighboring genes appears to be critically affected.
- $(GCG)_n$ expansions like in the coding region of the *PABPN2* gene are translated into polyalanine tracts which may cause mutated PABPN2 peptide oligomers to accumulate as filamentary inclusions in the nuclei of certain cell populations, finally leading to the rare oculopharyngeal ▶**muscular dystrophy**. This dominant negative effect may cause the proteins to be sequestered and degraded at an increased rate.

Clinical Relevance

Molecular biological investigations into repeat expansion diseases have already revealed unprecedented insights into the genetics and pathogenesis of these ▶**disease models** in man. The ramifications for other diseases cannot be underestimated, potentially also for common multifactorial disorders, where non-Mendelian, modulating genetic effects of variant repeats are to be expected. Yet it is also likely that more repeat expansion diseases will be discovered in the future for (rare) Mendelian traits, especially with regard to expansions of the many

additional repeat motifs demonstrably present in the
► [human genome](#).

References

1. Jin P, Warren ST (2000) Understanding the molecular basis of fragile X syndrome. *Hum Mol Genet* 9:901–908
2. Le Stunff C et al (2001) Paternal transmission of the very common class I INS VNTR alleles predisposes to childhood obesity. *Nature Genet* 29:96–99
3. Liquori CL et al (2001) Myotonic dystrophy type 2 caused by a CCTG expansion in intron 1 of ZNF9. *Science* 293:864–867
4. Richards RI (2001) Dynamic mutations: a decade of unstable expanded repeats in human genetic disease. *Mol Hum Genet* 10:2187–2194
5. Sherman MY, Goldberg AL (2001) Cellular defenses against unfolded proteins: a cell biologist thinks about neurodegenerative diseases. *Neuron* 29:15–32

Repetitive DNA

► [Repetitive Elements](#)

Repetitive Elements (Repetitive DNA, Repeats)

Definition

Repetitive elements are DNA fragments that are present in multiple homologous copies in the genome without a clearly assigned biological function.

Replication

Definition

Replication is the duplication of genomic DNA as part of the reproductive cycle of a cell.

► [DNA Replication Initiation](#)
► [Replication Fork](#)

Replication Bubble

Definition

Replication bubble designates the appearance at the electron microscope of a portion of just duplicated DNA near a replication origin, immediately following origin activation.

► [Replication Origins](#)

Replication Foci

Definition

Replication foci are subnuclear sites where DNA replication takes place. During S-phase, the newly synthesized DNA, and most of the proteins involved in DNA replication, accumulate in these sites.

► [DNA Ligases](#)

Replication Fork

ULRICH HÜBSCHER, IGOR SHEVELEV

Institute of Veterinary Biochemistry and Molecular Biology, University of Zürich, Zürich, Switzerland
hubscher@vetbio.unizh.ch

Definition

Each mammalian cell contains about 2 meters of genetic material, called DNA. The DNA consists of four bases; adenine (A), guanine (G), cytosine (C) and thymine (T), which are attached to a pentose sugar and connected by a phosphate backbone. There are about 5×10^9 bases in each mammalian cell. Prior to cell division, the process of ► [DNA replication](#) must faithfully duplicate the five billion bases with an absolute minimum of mistakes. Less than one mistake is allowed per replication round. DNA replication is an extremely complex cellular event involving many actors to coordinate this difficult task in a very short time (usually 30 min). These actors work in a big complex called the ► [replisome](#). Part of this replisome constitutes the enzymatic machinery around the replication fork. The enzymes that synthesize DNA according to the Watson-Crick base-pairing rule A opposite T and C opposite G and *vice versa* are called ► [DNA polymerases](#). DNA polymerases, of which a mammalian cell contains more than 18 species [reviewed in (1)], can use the energy

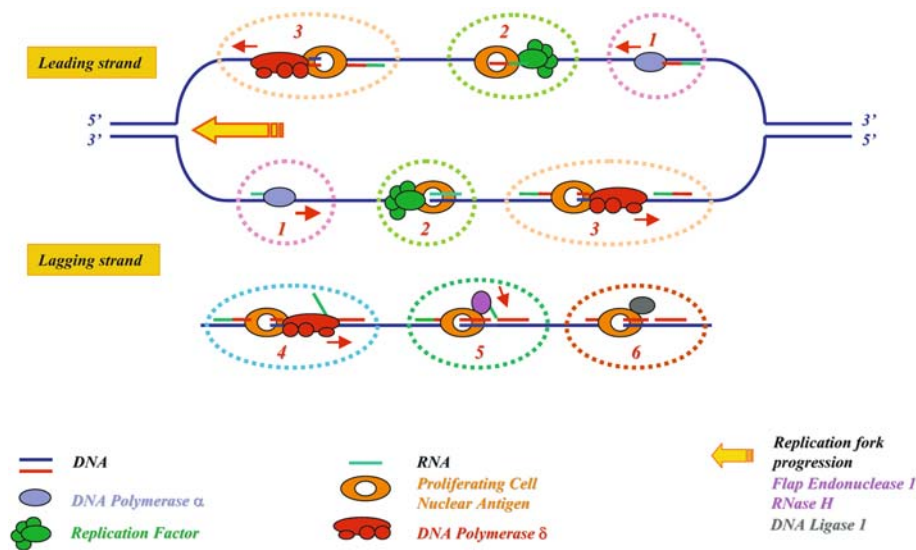
of the four deoxyribonucleoside 5'-triphosphates to incorporate deoxyribonucleoside 5'-monophosphates onto a growing DNA strand. DNA is an **▶antiparallel** double-strand molecule with polarities of 5'→3' and 3'→5', respectively. Since all DNA polymerases exclusively synthesize DNA in the 5'→3' direction the antiparallel DNA molecule presents a logistical problem for uni-directional replication. To circumvent this problem one strand, called the **▶leading strand**, is synthesized continuously and the other, called the **▶lagging strand**, discontinuously. Discontinuity means that the lagging strand is synthesized in short pieces of about 200 bases that are called **▶Okazaki fragments**. Discontinuity furthermore involves the necessity for frequent initiations of replication (10^7 per replication round) that are coupled to the subsequent processing of the Okazaki fragments in order to obtain a completely replicated DNA molecule. Here again the complex process includes many enzymes at the replication fork. The two original strands are called mother strands onto which the two strands are synthesized. DNA replication is therefore a semiconservative process in which the final replicated product contains an original and a newly synthesized DNA strand.

Characteristics

Maintenance of genetic stability is a key issue for any form of life. Consequently highly sophisticated mechanisms for maintaining of the integrity of the genome were well established before the three kingdoms of life (Prokaryotes, Archaea and Eukaryotes) separated. DNA replication in eukaryotic cells, including human, requires a large set of proteins. In all of these events, complexes of more than 30 polypeptides appear to function in highly organized and structurally well defined enzymatic machines. It is the replisome that gives the replication fork the characteristic structure for replication of DNA. DNA replication is the event leading to the duplication of DNA in advance of mitosis (or meiosis) and cell division. It occurs *in vivo* in an ordered and highly organized way, in which all enzymes and proteins involved have their exact roles in the replisome, which itself is located in so called nuclear replication factories. Models have been proposed as to how the enzymatic machinery might be spatially arranged at replication. They are based on the idea that DNA polymerases dimerize and that the lagging strand forms DNA loops that orient the DNA in such a way that the "directionality" for the DNA polymerases is the same. If one postulates that the replisome is fixed to structures in the nuclear replication factories, the DNA polymerases would thread the DNA through itself. The assembly and the events in a replisome might occur as follows [reviewed in (2)].

An initiator protein complex called the origin recognition complex (ORC) is bound to an origin of DNA

replication. The ORC has to be activated by other proteins, such as minichromosomal maintenance proteins (MCMs) and cell division cycle proteins (Cdc6 and Cdk7/Dbf4), by mechanisms such as phosphorylation or possibly other posttranslational protein modifications. This leads to the formation of an initiation complex that is able to alter DNA structures locally, presumably by activating the intrinsic helicase activity of an MCM subcomplex or by attracting other **▶DNA helicases** to the origins. The single stranded DNA produced must be protected and stabilized by the **▶single-stranded DNA binding protein**, called replication protein A (RP-A). RP-A helps to unwind the DNA by its intrinsic unwinding activity and possibly through its interactions with DNA helicases and DNA polymerase α /primase that acts as the initiating DNA polymerase (Fig. 1). In order to initiate DNA synthesis, **▶DNA primases** must synthesize short RNA stretches of 5–15 nucleotides. The short RNA stretches serve as primers for each Okazaki fragment and moreover as primers for leading strand replication. After very limited DNA synthesis, a **▶DNA polymerase switch** from DNA polymerase α /primase to the processive form, **▶DNA polymerase δ holoenzyme** occurs. This is most probably mediated by the DNA polymerase δ auxiliary protein replication factor C. DNA polymerase α /primase itself subsequently acts at the discontinuously synthesized lagging strand, where it initiates Okazaki fragments. While the DNA polymerase δ holoenzyme (DNA polymerase δ , proliferating cell nuclear antigen and replication factor C) is engaged in processive leading strand DNA synthesis, the situation at the lagging strand is more complex. At the lagging strand, a second DNA polymerase (DNA polymerase δ or ϵ) holoenzyme is formed for completion of each Okazaki fragment initiated by DNA polymerase α /primase. DNA syntheses on the leading and on the lagging strand are probably coordinated by dimerization of the two processive DNA polymerase holoenzymes. To complete DNA replication on the lagging strand, dimerization possibly occurs either through a direct physical interaction of two DNA polymerases or *via* a clamp factor. The initiator RNA between Okazaki fragments is removed by flap endonuclease 1 and/or by the Dna 2 endonuclease. After complete synthesis the Okazaki fragments are sealed by DNA ligase I. Topological constraints are released by **▶DNA topoisomerase I** and the replicated DNA can finally be separated by DNA topoisomerase II. Proofreading activity of DNA polymerases is an absolute requirement for accurate DNA synthesis, and it can be achieved by 3'→5' exonucleases that preferentially remove incorrectly incorporated bases e.g. guanine opposite a thymine. There are more than ten such exonucleases in a mammalian cell and some of them



Replication Fork. Figure 1 The replication fork proteins and their tasks. 1: primer synthesis by DNA polymerase α /primase; 2: Replication factor C displacement of DNA polymerase α and recruitment of proliferating cell nuclear antigen; 3: elongation by the newly recruited DNA polymerase δ holoenzyme; 4: strand displacement by DNA polymerase δ ; 5: cutting of the 5' displaced flap by flap endonuclease 1. 6: sealing by DNA ligase I. Reproduced with permission of Nature Publishing Group from ref. 7. For details see text.

are part of the DNA polymerase molecule e.g. DNA polymerases δ and ϵ [reviewed in (3)]. Cells containing a mutated 3'→5' exonuclease are prone to cancer (see Clinical Relevance).

Clinical Relevance

DNA replication is a conserved process required not only by mammals but also lower organisms such as bacteria

and viruses to replicate their DNA. Bacteria have their own replisome while viruses have evolved strategies in which they can either fully replicate their DNA (e.g. herpes viruses) or can partly do it (e.g. HI-virus, hepatitis virus) independently of a host. Although many replication proteins are extremely conserved from viruses to human, even subtle differences between the replication machineries have served as a basis for successful drug

Replication Fork. Table 1 Replication Fork Proteins

Protein	Function at the replication fork
MCM DNA helicase	Opens double-strand DNA in front of DNA polymerases
Replication protein A	Stabilizes single-strand DNA created by MCM DNA helicase
DNA polymerase α /primase	Initiator DNA polymerase, can synthesize RNA and DNA
DNA polymerase δ	Elongating DNA polymerase
DNA polymerase ϵ	Elongating DNA polymerase
3'→5' exonuclease	Proofreading for DNA polymerases
Proliferating cell nuclear antigen	Moving platform for DNA polymerases
Replication factor C	Loader of proliferating cell nuclear antigen
Flap endonuclease 1	Removes initiator RNA and DNA at the lagging strand of the replication fork
Dna2 endonuclease	Removes initiator RNA at the lagging strand of the replication fork
DNA topoisomerases I and II	Solve topological problems of the DNA in advance of and after DNA replication
DNA ligase I	Seals the remaining nick at the DNA to connect Okazaki fragments

development. It would be beyond such a short essay to cover in detail all the clinically relevant drugs; however we generate a short list below (4).

Antibacterial Drugs

Antibacterial drugs include inhibitors of DNA gyrase, a special form of bacterial DNA topoisomerase. So far these inhibitors have been used for treatment of more than a hundred million patients. Novel strategies include the development of DNA polymerase (DNA polymerase III) and DNA helicase (Dna B helicase) inhibitors.

Antiviral Drugs

Azidothymidine is the prototype of many nucleoside inhibitors that have been successfully applied as HIV reverse transcriptase (a special form of a viral DNA polymerase) inhibitors. Moreover non-nucleoside inhibitors such as Nevirapine were successfully used for recent combination drug therapy.

Anticancer Drugs

Anthracyclines are drugs that inhibit DNA topoisomerases. They have been, among other applications, used to preferentially inhibit fast replicating cancer cells. However since they also inhibit normal cells (but to a lesser extent) they are rather toxic and are far from being optimal.

Diseases due to Replication Fork Proteins

When the ►proofreading exonuclease of DNA polymerase δ is mutated in the active centre, mice containing such a mutation develop cancer much earlier than those containing the wild type proofreading activity. This suggests that DNA polymerase δ without the “Tippex” control is sloppier in DNA replication, which eventually leads to unfaithful DNA replication, to mutations of certain genes and to cancer (5). Moreover, ►haploinsufficiency of flap endonuclease 1 leads to rapid tumor progression in mice (6) suggesting that flap endonuclease 1 might act as a tumor suppressor protein.

Diagnostic

Detailed knowledge that has accumulated over four decades [the first DNA polymerase was discovered by A. Kornberg in 1956, reviewed in (1)] led to the invention in the mid 1980 of DNA polymerases that can be used to amplify DNA several million fold. This technology is called polymerase chain reaction and is known as PCR. The property of the DNA polymerase to amplify DNA is now universally used to test and to quantify viral, bacterial and parasitic infections, to test for genetic diseases and for paternity cases. Furthermore,

modern forensic medicine would be impossible without PCR.

In conclusion the detailed knowledge of the human replication fork is of paramount interest for the development of novel drugs against bacterial and viral infections. Moreover the therapy of various types of cancer will be improved because of knowledge of the replication fork. Finally, PCR diagnostics have already found their ways into practical medicine.

References

1. Hübscher U, Maga G, Spadari S (2002) Eukaryotic DNA polymerases. *Annu Rev Biochem* 71:133–163
2. Bell SP, Dutta A (2002) DNA replication in eukaryotes. *Annu Rev Biochem* 71:333–374
3. Shevelev I, Hübscher U (2002) The 3'→5' exonuclease. *Nat Rev Cell Biol* 3:364–375
4. Hübscher U, Spadari S (1994) DNA replication and chemotherapy. *Physiol Rev* 74:259–304
5. Goldsby RE, Lawrence NA, Hays LE et al (2001) Defective DNA polymerase-delta proofreading causes cancer susceptibility in mice. *Nat Med* 7:638–639
6. Kucherlapati M, Yang K, Kuraguchi M et al (2002) Haploinsufficiency of Flap endonuclease 1 (Fen 1) leads to rapid tumour progression. *Proc Natl Acad Sci USA* 99:9924–9929
7. Henneke G, Koundrioukoff S, Hübscher U (2003) Multiple roles for kinases in DNA replication. *EMBO Rep* 4:252–256

Replication Origins

ARTURO FALASCHI¹, GIUSEPPE BIAMONTI²,
SILVANO RIVA²

¹International Centre for Genetic Engineering and Biotechnology, Trieste, Italy

²Istituto di Genetica Molecolare, CNR, Pavia, Italy
riva@igm.cnr.it

Synonyms

Replicator; a DNA sequence on which regulation of DNA replication occurs, origin of bidirectional synthesis.

Definitions

DNA replication occupies a central position in the cell cycle and also in the development and life of multicellular organisms. During the last 10 years, our comprehension of this important process has considerably improved. Although the mechanisms that coordinate DNA replication with the other moments of the

cell cycle are not yet full understood, it is known that the regulation of the process operates by modulating the activation of DNA ►replication origins i.e. the sites where DNA replication starts.

The Replicon Model

In the early sixties, Jacob, Brenner and Cuzin proposed the ►replicon model for the replication of the bacterial chromosome. In this model, the replicon is a genetic element replicated from a single origin of DNA replication or ►replicator, which is recognized by a specific positive regulatory protein, the ►initiator. Since then this model has been validated in numerous prokaryotic and viral systems.

Unlike the single bacterial chromosome, the eukaryotic genome is organized as several chromosomes, each formed by numerous replication units that are named, by analogy with the prokaryotic models, replicons. Each replicon (10–300 kb in length) is characterized by an origin from which two oppositely moving forks issue. Origin activation is a tightly controlled event occurring in a well defined interval of the cell cycle called S-phase. Moreover, origin firing follows a precise and reproducible spatio-temporal program; origins embedded in transcribed portions of the genome are activated in early S-phase while origins associated with heterochromatic transcriptionally silent domains fire in mid and late S-phase.

The Replication Origins

Computation of the ratio between replicon and genome sizes, as well as calculation of the rate of advancement of the ►replication fork, led to the conclusion that about 30,000 origins should exist in mammalian cells. However, less than 50 mammalian origins have been identified so far, and even fewer have been characterized in detail. Fundamental questions, such as how origins are spaced in the genome, which are their functionally essential sequence features and how they are regulated remain largely unanswered.

Historically, the first insights into eukaryotic DNA replication derived from studies of viruses that replicate their DNA in mammalian nuclei (SV40 and Epstein Barr). The genome of these viruses is organized into circular mini-chromosomes and its replication requires a number of proteins borrowed from the replication machinery of the host in conjunction with a trans-acting initiator protein encoded by the virus itself. The DNA replication origins of these viruses are quite simple, consisting of a well-defined core and of regulatory regions composed of several auxiliary elements. A completely viral origin enables autonomous replication of plasmid DNA, when the cognate initiator protein and other necessary replication factors are provided.

An invaluable contribution to the characterization of eukaryotic DNA replication origins was made by the budding yeast *Saccharomyces cerevisiae* model. DNA replication origins in *S. cerevisiae* were identified as genomic regions able to drive autonomous replication of plasmid DNA molecules and were, therefore, called *ARS* for *autonomously replicating sequences*. A complete ARS element (100–200 bp) contains a highly conserved 11 bp AT-rich sequence (*ACS*, *ARS consensus sequence*), and two or three auxiliary motifs (*B domains*) that although divergent in sequence are functionally conserved. B-domains are involved both in the assembly of the pre-replication complex and in the unwinding of the double helix that is required for the assembly of the replication machinery (►replisome).

The relatively easy identification and characterization of replication origins in viruses, bacteria and yeast raised many expectations as to the possibility of isolating the mammalian origins of replication. However, this task has proved much more difficult than expected. The main reason for the elusiveness of origin sequences in animal cells is probably related to the lack of a successful functional assay. Following the approach used for the identification of yeast ARS elements, several groups have tried to obtain autonomous replication of a reporter plasmid driven by specific mammalian DNA sequences. Despite sporadic positive reports the overall outcome of these attempts has been negative.

The lack of simple genetic assays led to the development of methods for the identification and mapping of replication initiation sites in the chromosomal context. All these methods are based on some conserved features of replication forks. Initiation of DNA synthesis generates specific structures (►replication bubbles) and proceeds bidirectionally in a semidiscontinuous manner. Thus, an origin of DNA replication can be operationally defined as the site where: i) transition between continuous DNA synthesis of the ►leading strand and discontinuous DNA synthesis of the ►lagging strand occurs, ii) nascent DNA synthesis occurs and iii) replication bubbles are created. These three features have all been utilized in recent years for origin mapping in eukaryotic chromosomes.

A partial list of identified mammalian replication origins is shown in Table 1.

Characteristics

Pre-Replication and Post-Replication Complexes

In the last few years, our understanding of the molecular events regulating origin firing has significantly improved thanks to studies in *S. cerevisiae* that led to the identification of the numerous proteins involved. According to the current model a complex of six polypeptides – ►origin recognition complex, ORC– recognizes and binds to the yeast origins. In

Replication Origins. Table 1

Region studied	Organism	Sequence properties	Method
<i>TOP1</i> gene promoter	Human	CpG tracts and AT-rich element	Interaction with Orc2 in ChIP ¹ experiments.
<i>MCM4</i> gene promoter	Human	CpG tracts and AT-rich element	Interaction with Orc1 and Orc2 in ChIP experiments.
β -globin gene	Human	Gene promoter	Analysis of nascent DNA abundance
Lamin B2 gene	Human	CpG island and AT-rich elements	Analysis of nascent DNA abundance by competitive PCR.
c-myc gene promoter	Human	Gene promoter	Analysis of nascent DNA abundance
HPRT gene promoter	Human	High GC content.	Analysis of nascent DNA abundance by competitive PCR.
Ig heavy chain locus	Mouse	<i>IgH</i> regulatory gene region AT rich segment of 47 bp flanked by B2 SINE sequences. Multiple initiation sites	Identified by 2D gel Semiquantitative PCR
DHFR locus	Hamster	Initiation zone in a 55Kb intergenic region Three fixed origins localized at a 0.5 – 8 kb resolution	Interaction with MCM complexes in ChIP experiments Identified by 2D gel, by analysis of nascent DNA abundance by competitive PCR
APRT gene	Hamster	CpG island	Analysis of nascent DNA abundance
GADD gene	Hamster	CpG island	Analysis of nascent DNA abundance
TK gene	Hamster	CpG island	Analysis of nascent DNA abundance
ori II/9A	<i>Sciara coprophila</i>	1 kb region 80 bp ORC binding site	Replication start site mapped adjacent to the ORC binding site.

¹ChIP: chromatin immunoprecipitation

G1 phase, ORC proteins cooperate with the products of the Cdc6 – in *S. cerevisiae* – or Cdc18 – in *S. pombe* – genes to load a complex of six proteins – called MCMs (Mini Chromosome Maintenance). In *Schizosaccharomyces pombe* and in higher eukaryotes this process requires the activity of another protein, called Cdt1. The final result is the assembly on the origin of a pre-replicative complex. After this event ORC, CDC6 and Cdt1 proteins are no longer required. Following origin firing, the concordant removal of CDC6/CDC18 prevents the loading of MCM proteins onto the origin-bound ORC in G2, thereby preventing re-replication of DNA. The pre-replication complex formed in G1 is ready to initiate DNA replication upon the activation of S-phase promoting factors like cyclin-CDKs and CDC7/DBF4 kinases that are part of the pre-replication complex.

Interestingly the ORC, MCM, CDC6, CDC7, DBF4 and CDC45 proteins are quite conserved in evolution,

and mammalian homologues have been identified and cloned. However, very scanty evidence so far exists about the interactions of these proteins with any origin-specific sequences in mammalian DNA.

The DNA Origin Associated to the Human Lamin B2 Gene

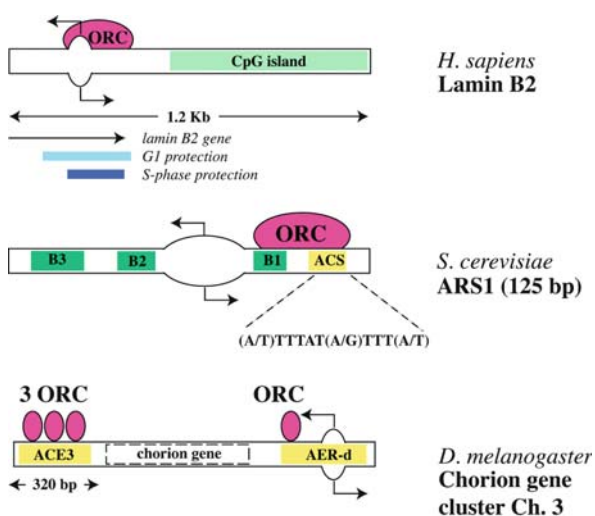
The human DNA replication origin for which most details concerning the *cis*-acting sequences involved in origin activity are currently available is the one located in a short intergenic region immediately downstream of the human lamin B2 gene.

The origin was localized using the competitive polymerase chain reaction (PCR) assay to quantify the abundance of adjacent genomic DNA fragments in the pool of newly synthesized DNA fragments. In this assay, fragments proximal to the origin are more abundant than distal ones. The origin has been mapped in an approximately 500 bp region corresponding to the

3'-end of the lamin B2 gene and the promoter of the TIMM13 gene. This origin is activated in the first two minutes of S-phase in HL-60 cells, and is active in all the cell types tested so far. A distinguishing feature of this origin, revealed by *in vivo* genomic footprinting experiments, is a prominent protected region overlapping the replication initiation site. The extension of this protection undergoes remarkable changes during the cell cycle. In quiescent cells, no protection is detectable. As the cells enter the cell-cycle a footprint appears and keeps growing throughout G1 to reach its maximal extension (over 100 bp) at the G1/S border. After origin firing, the protection shrinks to 70 bp and remains unchanged until G2/M. In mitosis, the protection totally disappears. A similar phenomenon takes place on ARS elements in yeast and reflects the assembly of the pre- and post-replication complexes. The proteins responsible for these interactions are currently the subject of intensive investigation. The precise initiation start sites of DNA replication have been recently mapped within the protected region at the single nucleotide level. Fig. 1 schematically illustrates the structure and the mode of activation of the human lamin B2 origin along with that of other dissected eukaryotic origins.

Relationship Between Origins of DNA Replication and Replicators

DNA sequences of four of the most characterized mammalian origins of DNA replication (β -globin,



Replication Origins. Figure 1 Schematic representation of three eukaryotic DNA replication origins. The positions of ORC binding sites relative to start sites of DNA replication (arrows passing through a bubble) are indicated.

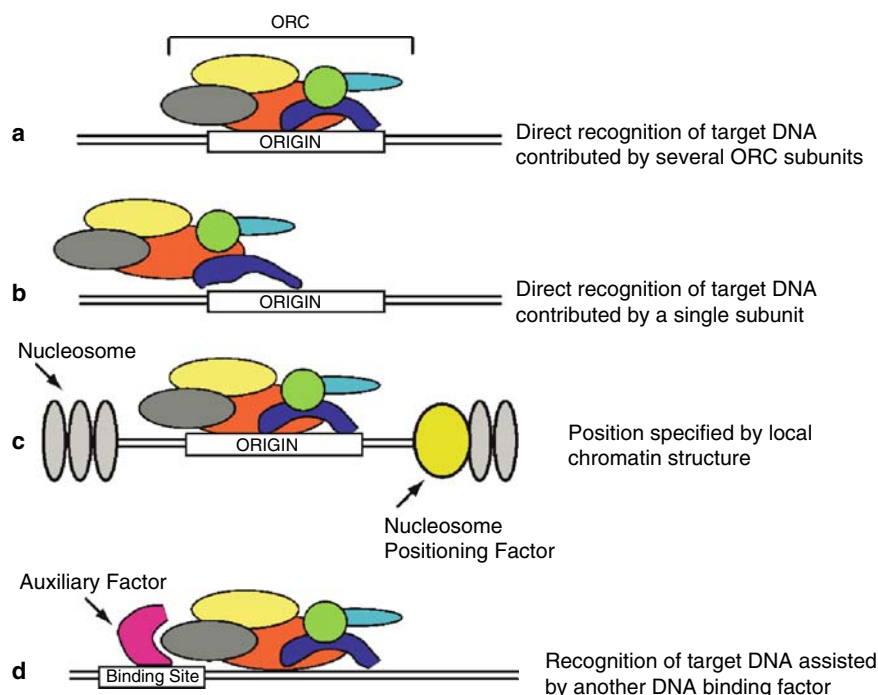
lamin B2, c-myc and DHFR) have been either mutated in their original locations or cloned in plasmid vectors and stably transferred to ectopic sites of the genome. Altogether the results of these experiments support the notions that: 1) mammalian cell origins might consist of relatively short regions, the activities of which are primarily specified by DNA sequence, by analogy with all replicators from simple organisms, 2) these origins probably have a modular organization of different genetic elements and 3) the genetic elements determining replication function correspond or lie in close proximity to the initiation region. Whether origin specification by these replicators might occur by primary DNA sequence recognition by initiator proteins (most probably members of the ORC complex) or by other epigenetic features favoring the recruitment of initiation proteins (attachment to matrix? chromatin modification? transcription factors?) begs further investigation.

Chromatin Structure, Epigenetics, and Origins

The absence of common consensus motifs in mammalian origins of replication, the low sequence-specificity of binding of the ORC complex to DNA and the observation that in early amphibian and *Drosophila* embryos DNA replication can start from any DNA sequence have led to the proposal that origin specification might be influenced by epigenetic mechanisms rather than solely by DNA sequence information. The notion that chromatin conformation is a strong determinant for origin activity is unquestionable. Even in yeast, initiation activity of ARS elements in the chromosomal context depends not only on the ARS sequence but also on chromatin structure and chromosomal position. For example, origins near telomeres tend to be activated late in S phase and the activation time of early-firing origins is considerably delayed upon translocation into a sub-telomeric position.

One of the possible mechanisms by which chromatin regulates initiation of DNA replication is by restricting access of DNA to direct or indirect ORC binding. Thus, a strict correlation has begun to be envisaged between ORC proteins, origins of DNA replication and regulators of chromatin modeling in metazoan cells. In this regard, it was proposed that CG-rich regions of the chromosome, approximately 1 kb long, free of methylation and often containing transcription promoters (the so called CpG islands), might specify the location of at least a subset of DNA replication origins. CpG islands might mark chromosomal sites that are more accessible to initiation proteins (ORC) as well as to transcription factors.

Some possible modes of ORC recruitment to eukaryotic origins are depicted in Fig. 2.



Replication Origins. Figure 2 Possible modes of ORC recruitment to origins. (a) ORC directly contacts target DNA with several ORC subunits. The paradigm for this mode is the interaction of ORC to ARS sequences in *S. cerevisiae*. ORC contacts both the ACS and the auxiliary B1 element. (b) The specificity of binding is conferred only by one ORC subunit. This is the case for *S. pombe* ORC in which Orc4 has an N-terminal extension containing nine repeats of an AT-hook motif that recognizes stretches of AT-rich DNA in the ARS element. (c) Specificity of ORC binding to DNA is very relaxed in term of primary DNA sequence and the interaction is mostly determined by chromatin conformation. (d) ORC directly binds DNA but the specificity of binding is provided by the association with another factor bound to neighboring DNA sequences. This may be the case of the ACE3 element of *Drosophila* chorion genes. Here ORC binding requires transcription factor E2F.

Origins of Replication and Cell-Cycle Control

Another important topic of investigation is related to the understanding of the molecular events linking cell-cycle control, replication regulatory proteins such as ORC, CDC6 and MCMs and origins of DNA replication. As in budding yeast so also in mammalian cells, the proteins involved in origin regulation are most probably the targets on which cell-cycle control mechanisms converge. This was indirectly suggested by the study of protein-DNA interactions at the lamin B2 origin, which showed a cell-cycle regulated pattern and by experiments in which *Xenopus* egg cytosol was used to stimulate hamster nuclei to enter S phase, which demonstrated that origin specification occurs at a distinct time in G1 phase. The existence of a specific network of interactions between cell-cycle regulation and origin activation can be also inferred from the elucidation of the mechanisms of S-phase checkpoint in yeast cells. In *S. cerevisiae*, the induction of DNA damage by genotoxic agents or the stalling of replication forks are sensed by the Rad 53

protein and DNA replication is blocked by specifically preventing late origin firing. Similarly, in *S. pombe* ORP2, the homologue of *S. cerevisiae* ORC2, is required for the response to checkpoint signals, preventing progression of the cell-cycle after DNA damage. Most probably, these biochemical and functional interactions requiring replication initiation proteins will also turn out to be conserved in higher eukaryotic cells.

Clinical Relevance

The definition of DNA replication origins and their timely and coordinate activation are key factors in the faithful replication of genomes and in the maintenance of their integrity. Regulatory circuits controlling cell proliferation and differentiation must generate signals that dictate origin activation/inactivation through the cell-cycle. It is to be expected that alterations of the programmed activation of origins of replication might have pathological consequences. Evidence of this is

starting to emerge. These issues, however, are greatly complicated by the realization that in higher eukaryotes initiation of DNA replication does not simply depend on sequence elements, but also on additional elements such as chromatin structure, transcriptional activity and possibly also the sub-nuclear location of DNA domains. The elucidation of the interplay between replication factors, transcription factors and proteins modulating chromatin structure and nuclear architecture remains one of the major challenges in molecular biology. A thorough understanding of the positive and negative elements regulating this process will offer essential insights into the modulation of DNA proliferation, assuring the harmonious development and differentiation of tissues and organs from the fertilized oocyte to the healthy adult organism; furthermore, a definition of the structural and mechanistic details of the molecules involved and of the reactions they perform may offer in the future indications for the production of drugs capable of interfering specifically with the uncontrolled cell proliferation typical of cancer.

- ▶ DNA Replication Initiation
- ▶ Recombinant Protein Expression in Bacteria
- ▶ Replication Fork
- ▶ YAC and PAC Maps

References

1. Abdurashidova G, Deganuto M, Klima R et al (2000) Start sites of bidirectional DNA synthesis at the human lamin B2 origin. *Science* 287:2023–2026
2. Bell SP, Dutta A (2002) DNA replication in eukaryotic cells. *Annu Rev Biochem* 71:333–374
3. Gilbert DM (2001) Making sense of eukaryotic DNA replication origins. *Science* 294:96–100
4. Kelly TJ, Brown GW (2000) Regulation of chromosome replication. *Annu Rev Biochem* 69:829–880
5. Mechali M (2001) DNA replication origins: from sequence specificity to epigenetics. *Nat Rev Genet* 2:640–645

Replicative Senescence

Definition

Normal human cells exhibit a restricted replicative lifespan when propagated in culture. This replicative barrier is called replicative senescence or mortality stage 1 (M1).

- ▶ Proteomics of Aging
- ▶ Telomerase

Replicator

Definition

The DNA sequence corresponding to the origin of DNA replication is called replicator.

- ▶ Replication Origins

Replicon

Definition

In prokaryotes, replicon defines a DNA molecule capable of autonomous replication; in eukaryotes, replicon denotes a DNA sequence (100 to 300 kb long) organized in tandem with other replicons on the chromosome, which contains in its middle a replication origin that is activated by specific proteins during the S phase of the cell cycle.

- ▶ Replication Origins

Replisome

Definition

Replisome is a multiprotein complex performing the advancement of semi-conservative DNA synthesis at a replication fork (e.g. *E.coli* DNA Polymerase III holoenzyme is composed of at least 17 polypeptides).

- ▶ Replication Fork
- ▶ Replication Origins

Repolarization

Definition

Repolarization is the phase of the action potential that reconstitutes the resting membrane potential.

- ▶ Ion Channels/Excitable Membranes
- ▶ Neurons

Reporter Genes

Definition

Reporter genes are used as genetic tools to visualize transgene expression. When introduced into a genomic locus via homologous recombination, reporter genes are under the transcriptional control of the mutant gene, which allows assessment of its spatio-temporal expression pattern (knock-in). Commonly used reporter genes are LacZ encoding β -galactosidase gene from *E. coli*, GFP Green Fluorescent Protein gene from marine jelly fish, and hPLAP human Placental Alkaline Phosphatase gene.

- [Mouse Genomics](#)
- [Protein Interaction Analysis: Variations of the Yeast Two-Hybrid System](#)
- [Transgenic and Knockout Animals](#)
- [Two-Hybrid System](#)

Resolution

Definition

In X-ray structure analysis, resolution describes the degree of order of the crystal that determines the maximum angle of diffraction from the crystal, and from Bragg's law, the level of detail resolvable by the diffraction data. The accuracy of the structure is typically much higher than the resolution of the diffraction data, depending on the applicability of chemical geometry information used to refine the crystal structure. In mass spectrometry, resolution (R) is defined as the differentiation of two neighbouring masses with $R=m/\Delta m$. In chromatography, resolution is the differentiation of two neighbouring co-eluting peaks.

- [3D Structure Determination by X-Ray](#)
- [Mass Spectrometry: Quantitation](#)

Repressor

Definition

Repressor is a protein that binds to a specific region of DNA to prevent transcription of an adjacent gene.

- [Two-Hybrid System](#)

Repressor Element–1 Transcription Factor

- [REST](#)

Respiratory Burst

Definition

Respiratory burst describes a sudden rise in oxygen consumption of phagocytes upon contact with micro-organisms or stimulation with, e.g. anaphylatoxins, immune complexes or proinflammatory cytokines. It is the physiological equivalent to superoxide anion production by NADPH oxidase and is essential to initiate the killing reaction during phagocytosis. Any inability to adequately respond with respiratory burst leads to inefficient clearance of pathogens (chronic granulomatous disease). Ubiquitous strong and persistent activation of the respiratory burst contributes to the pathology of septicemia.

- [Free Radicals](#)

Residue Interface Propensity

Definition

Residue interface propensity quantifies the preference for certain amino acids to be in the interface site of two interacting molecules.

- [Protein-Protein Interaction](#)
- [Two Hybrid System](#)

Respiratory Chain

Definition

The respiratory chain consists of five enzyme complexes embedded in the mitochondrial inner membrane. The respiratory chain produces ATP through the process of oxidative phosphorylation, and its function is dependent on both nuclear and mtDNA-encoded protein subunits.

- [Mitochondrial Myopathies](#)

Response Elements

Definition

Response elements are precise sequences of nucleotides recognized by transcription factors. In nuclear receptor action, they are recognized by the receptor. They are pairs of half sites that may be arranged palindromically, or as tandems, or as reverse tandem repeats or even singly.

► [Nuclear Hormone Receptors: Regulators of Gene Transcription and Cellular Signaling](#)

REST

Definition

Repressor element-1 silencing transcription factor (REST), also known as neuron restrictive silencer factor (NRSF), is a C₂H₂ zinc finger protein that is related to the Gli-Krüppel family of transcriptional repressors. In non-neuronal tissues, REST/NRSF is part of a transcriptional repressor complex that binds to the neuron restrictive silencer element (NRSE), effectively silencing neuronal genes in non-neuronal tissues. Huntingtin interacts with REST/NRSF to modulate the transcription of NRSE-controlled neuronal genes such as BDNF or proenkephalin.

► [Polyglutamine Disease, the Emerging Role of Transcription Interference](#)

► [Transcriptional Repression](#)

Restriction Enzymes

Definition

Restriction enzymes (also called restriction endonucleases) in nature are part of a bacterium's antiviral defense system. These enzymes recognise four to eight base pairs in a DNA strand (recognition site), and then cleave both strands of DNA at or close to that specific site (restriction site). In the molecular biological laboratory these enzymes are a key tool for analysing DNA; they are used to cleave DNA at defined sequences. The cleavage sites are used as landmarks and the sizes of fragments give the distance between them, allowing physical maps to be built up.

► [Recombinant Protein Expression in Bacteria](#)

► [YAC and PAC Maps](#)

Restriction Fragment Length Polymorphism

Definition

Restriction fragment length polymorphisms (RFLPs) are variations between individuals in DNA fragment sizes after cutting by specific restriction enzymes. RFLPs are usually caused by mutation at an enzyme's cutting site. If even one nucleotide within a cutting site is altered (by mutation) then the enzyme will no longer cut the target DNA at that point. The resulting restriction fragments will be of a different length. RFLPs are a quick way to differentiate two alleles by the size of DNA fragments.

► [Common Diseases, Genetics](#)

► [Spinal Muscular Atrophy](#)

Restriction Mapping

Definition

Restriction mapping displays a physical mapping technique in which the position and spacing of restriction sites is deduced, often as an aid to construction of clone contigs. This can be applied to genomic or cloned DNA. Key techniques used are multiple enzymes singly and in combination, partial digestion with a single enzyme, and Southern transfers.

► [YAC and PAC Maps](#)

Restriction Point

Definition

Restriction point is a critical point in the G1 phase of the cell cycle when a growth factor has committed a cell to cell proliferation.

► [Cell Cycle – Overview](#)

► [Growth Factors](#)

Retinal Pigment Epithelium

Definition

The retinal pigment epithelium (RPE) is a monolayer of cells loaded with melanin granules. The RPE is located

just outside the retina between Bruch's membrane (basal side) and the photoreceptors (apical side). The RPE are important for phagocytosis of photoreceptors, transport and storage of retinoids, regulation of extracellular ion concentration, and protection of the photoreceptive system from short-wavelength light.

►Retinitis Pigmentosa (RP)

Retinal Protein

Definition

Retinal protein refers to a protein in which retinal is covalently bound as a prosthetic group via a Schiff base.

►Photoreceptors

Retinitis Pigmentosa

EBERHART ZRENNER

University Eye Hospital, Eberhard Karl University
Tübingen, Germany
ezrenner@uni-tuebingen.de

Definition

The term "retinitis pigmentosa" (RP) comprises a large group of genetically heterogeneous and clinically similar hereditary retinal dystrophies and degenerations (prevalence of 1:3000–1:5000). The term was coined by the Dutch ophthalmologist F.C. Donders (1855), implying an infectious disease, which turned out to be wrong. The term "retinitis pigmentosa" is the most common one; synonyms are retinopathia pigmentosa, photoreceptor dystrophy, hereditary retinal degeneration or tapeto-retinal degeneration. Due to the excessive clinical variability of RP and its enormous genetic heterogeneity, it has become common in recent years to differentiate these retinal diseases according to the gene symbol and its chromosomal locus (examples see Table 1). They can be classified according to their inheritance: autosomal dominant, autosomal recessive, x-chromosomal recessive and mitochondrial. Additionally there are numerous cases where different mutations in the same gene cause clinically quite heterogeneous patterns, stationary or progressive or where clinically similar terms have been applied to patients with

different mutations in different genes (1, 2). In all cases, the functions of retinal photoreceptors (rods or cones) or of the ►retinal pigment epithelium ("RPE" or "tapetum") or of structures supporting these cells (choroid, glial cells, inter-photoreceptor matrix) are primarily affected. There are spontaneous animal models and generated mouse mutants that suffer from several forms of "retinitis pigmentosa" due to mutations also found in man.

Characteristics

The typical clinical characteristics of RP are:

1. Bilateral involvement with onset in the first three decades of life.
2. Loss of peripheral vision if ►rods are primarily affected or central vision if ►cones are involved first.
3. Progressive loss of photoreceptor function, as assessed by electroretinography, a process that may last several decades.
4. Alterations at the level of the pigment epithelium.

At the stage where rod function deteriorates, the corresponding visual symptoms comprise night blindness or delayed adaptation to dim light, impaired orientation caused by the loss of visual field, abnormal sensitivity to glare and, eventually, difficulties with color discrimination as well as a decrease in visual acuity. Many patients become legally blind in the late stages of the disease.

Morphological Findings

There are characteristic clinical signs of typical RP (3, 4), yellow, wax-colored optic disks, narrowed blood vessels and almost always changes in the ►macula, such as wrinkling of the inner limiting membrane, diffuse atrophy or cystoid macular edema. In most cases, changes are found in the mid-peripheral retinal pigment epithelium, beginning with depigmentation, followed by dark pigments in a typical "bone spicule" configuration, reflecting an intraretinal pigment migration. Vitreous changes are always found at least in later stages and myopia is very frequent. In middle-aged patients subcapsular posterior cataract is very common. There are atypical cases without pigment changes or with localized deposits only. Additionally, there are a number of syndromes that may involve extra-ocular structures and functions.

Mendelian Genetics

In about 25% of patients, autosomal dominant inheritance can be established; autosomal recessive inheritance is found in approximately 20% and x-chromosomal recessive inheritance in about 8% of the cases. In the majority of all affected patients, no heredity is provable;

Retinitis Pigmentosa. Table 1 Genes involved in retinitis pigmentosa

Phototransduction			
Symbols	Location	Diseases;Protein	Comments
RHO , RP4	3q22.1	dominant RP; dominant CSNB; recessive RP; protein: rhodopsin	accounts for 30 to 40% of autosomal dominant RP; more than 100 distinct mutations but RhoPro23His causes 10% of adRP
PDE6B , CSNB3	4p16.3	recessive RP; dominant CSNB; protein: rod cGMP phosphodiesterase beta subunit	same as retinal degeneration in the <i>rd</i> mouse, <i>r</i> mouse and <i>rcd1</i> Irish Setter dog; accounts for 3 to 4% of recessive RP; photoreceptor rescue with calcium-channel blocker in mouse but not dog
RPGR , RP3	Xp11.4	X-linked RP, recessive; X-linked RP, dominant; X-linked CSNB; X-linked cone dystrophy 1; X-linked atrophic, recessive; protein: retinitis pigmentosa GTPase regulator	mutations are found in 70% of RP3 cases, with dominant mutations in ORF15; exceptionally heterogeneous, retina-specific alternate splicing; protein is similar to RCC1, and binds PDE-delta and RPGRIP1 with species-specific colocalization (rod/cone outer segments in humans, connecting cilia in mice)
Cation Channel			
CNGA3 , ACHM2, CNCG3, RMCH2	2q11.2	recessive achromatopsia; protein: cone photoreceptor cGMP-gated cation channel alpha subunit	total color blindness and other cone-related abnormalities (rod monochromacy); CNGA3 accounts for 20–30% of achromatopsia cases
CACNA1F , CSNB2, CSNBX2	Xp11.23	X-linked CSNB, incomplete; protein: L-type voltage-gated calcium channel alpha-1 subunit	CACNA1F mutations are found in 60 to 90% of X-linked incomplete-CSNB patients; retina-specific expression with synaptic localization of protein
Disc structure			
RDS , RP7	6p21.2	dominant RP; dominant MD; digenic RP with ROM1; dominant adult vitelliform MD; protein: peripherin 2	dominant mutations; accounts for 5% of dominant RP; same as <i>rd</i> s mouse; photoreceptor rescue in mouse model
Rhodopsin transport			
TULP1 , RP14	6p21.31	recessive RP; protein: tubby-like protein 1	protein localizes to developing and adult rods and cones; possibly involved in transport of rhodopsin from inner to outer segment; a similar gene in <i>tub</i> mouse causes obesity, deafness and retinal degeneration
RP1	8q12.1	dominant RP; protein: RP1 protein	causes 5 to 10% of adRP; highly variable expression; two common mutations, Arg677X and 2280del4; protein is photoreceptor-specific,
Retinoid cycle			
RPE65 , LCA2, RP20	1p31.2	recessive Leber congenital amaurosis; (LCA) recessive RP; protein: retinal pigment epithelium-specific 65 kD protein	accounts for 2% of recessive RP and 7 to 16% of LCA; protein is necessary for production of 11- <i>cis</i> -vitamin A; 9- <i>cis</i> -retinal restores visual function in mouse model; successful gene therapy in dog; in same pathway as LRAT

Retinitis Pigmentosa. Table 1 Genes involved in retinitis pigmentosa (Continued)

Phototransduction			
Symbols	Location	Diseases;Protein	Comments
ABCA4 , ABCR, RP19, STGD1	1p22.1	recessive Stargardt disease, juvenile and late onset; recessive MD; recessive RP; recessive cone-rod dystrophy; protein: ATP-binding cassette transporter - retinal	may be involved in age-related macular degeneration; expressed in rod outer segment and foveal cones; ABCA4 mutation may increase severity of STGD3; flippase for all- <i>trans</i> retinal and N-retinylidene-PE
LRAT	4q32.1	recessive RP, severe early-onset; protein: lecithin retinol acyltransferase	gene is expressed in RPE; protein catalyzes first step in visual cycle transforming vitamin A into 11- <i>cis</i> -retinol; same pathway as RPE 65
RBP4	10q23.33	recessive RPE degeneration; protein: retinol-binding protein 4	RPE atrophy with night blindness and reduced visual acuity; carrier protein for serum retinol
RDH5 , RDH1	12q13.2	recessive fundus albipunctatus; recessive cone dystrophy, late onset; protein: 11- <i>cis</i> retinol dehydrogenase 5	stationary night blindness with subretinal spots and delayed dark adaptation; protein is an RPE microsomal enzyme involved in converting 11- <i>cis</i> retinol to 11- <i>cis</i> retinal; extremely delayed rod and cone resensitization in null mutation
Ciliary Structure			
MYO7A , DFNB2, USH1B	11q13.5	recessive Usher syndrome, type 1; recessive congenital deafness without RP; recessive atypical Usher syndrome (USH3-like); protein: myosin VIIA	MYO7A, an unconventional myosin, is a common component of cilia and microvilli; same as <i>sh1</i> shaker-1 mouse (but no RP) and <i>mariner</i> zebrafish; possible digenic deafness with USH3A
Phagocytosis			
MERTK	2q13	recessive RP; protein: c-mer protooncogene receptor tyrosine kinase	human ortholog of the mouse <i>Mertk</i> gene; causes defective phagocytosis of photoreceptor outer segments by the RPE and retinal degeneration in the RCS rat; successful gene therapy in rat; expressed in multiple tissues including RPE/sclera
Extracellular Matrix			
USH2A	1q41	recessive Usher syndrome, type 2a; recessive RP;	usherin is a basement membrane protein, with laminin, EGF and fibronectin type III domains, found in many tissues, in retina and inner ear; Cys759Phe mutation found in 4 to 5% of recessive RP without hearing loss
Transcription Factor			
CRX , CORD2	19q13.32	dominant cone-rod dystrophy; recessive, dominant and <i>de novo</i> Leber congenital amaurosis; dominant RP; protein: cone-rod otx-like photoreceptor homeobox transcription factor	meiotic drive suggested; CRX also activates pineal genes; interacts with NRL; <i>Crx</i> -deficient mice have diminished circadian entrainment; causes approximately 3% of LCA

Retinitis Pigmentosa. Table 1 Genes involved in retinitis pigmentosa (Continued)

Phototransduction			
Symbols	Location	Diseases;Protein	Comments
Splicing			
PRPF31, RP11, PRP3	19q13.42	dominant RP; protein: human homolog of yeast pre-mRNA splicing factor	high frequency in British Isles (21%); highly variable severity, protein is a highly-conserved, ubiquitously-expressed member of the U4/U6-U5 tri-snRNP particle complex
Upstream metabolisms			
CHM	Xq21.2	choroideremia; protein: geranylgeranyl transferase Rab escort protein 1	ubiquitously expressed protein (REP2 can substitute); attaches isoprenoids to Rab proteins

CSNB, congenital stationary night blindness; MD, macular dystrophy

these isolated cases are simplex patients, some of which may be autosomal recessive cases. Heterozygotes can be identified quite often by tapetal reflexes, scattered discrete pigment atrophy or in ERG recordings, especially in the flicker electroretinogram.

The Natural Course of the Disease

According to Massof and Finkelstein (1987), the loss of visual field in typical RP occurs in two stages (5). Stage 1 is characterized by very slow progression up to a critical age, which is 32 years in the autosomal dominant form and about 20 years in the x-linked type, with autosomal recessive forms lying somewhere in between. After this critical age, the rate of visual field loss follows an exponential function, so that every year about 20% of the remaining visual field is lost. Various clinical studies suggest two forms of autosomal dominant RP with different natural courses.

Type 1 is characterized by a relatively early diffuse loss of rod sensitivity with increasing concentric visual field impairment, but cone function is preserved for a long time.

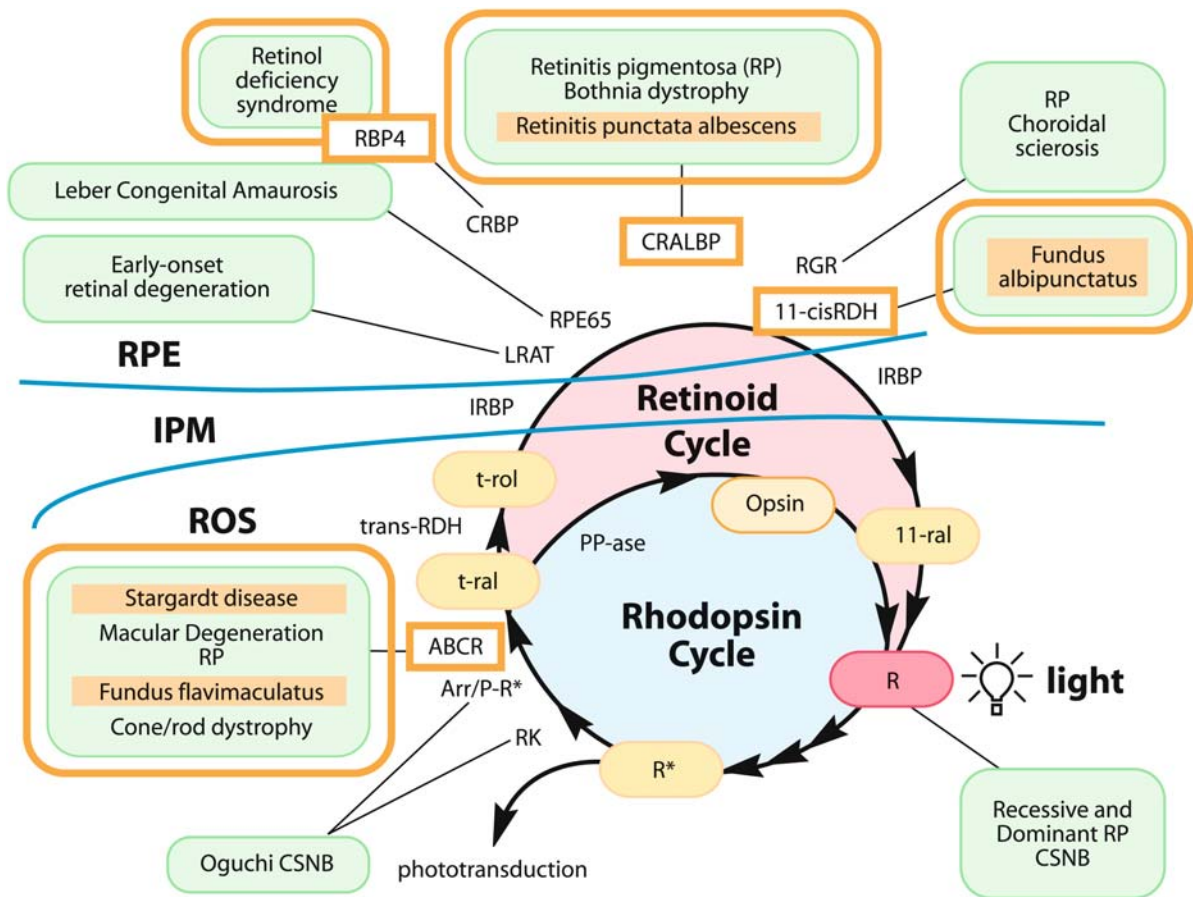
Type 2 is characterized by a more regional simultaneous loss of rod and cone function; the onset of night blindness occurs late and the loss of peripheral visual field progresses relatively slowly (6). A more refined differentiation is however possible only by molecular genetic differentiation. There is a large variation in the degree of function loss from very serious at birth to very mild at an advanced age.

Other Retinal Degenerations Related to RP

- ▶ **Choroideremia** is an x-linked disease, which manifests characteristic fundus findings of diffuse

choroidal and retinal atrophy. Its pattern of visual function loss is quite similar to that of typical x-linked retinitis pigmentosa.

- ▶ **Cone-rod dystrophies** differ clinically from RP in that photophobia, decrease in color vision and visual acuity loss occur before the onset of peripheral visual field loss and impaired dark adaptation. There are special forms of hereditary macular dystrophies that form an entity of their own (e.g. Stargardt's disease, Best's disease). They can be differentiated at an early stage by electroretinography (ERG) in conjunction with typical fundus appearances.
- ▶ **Leber's congenital amaurosis** (LCA) denotes a heterogeneous group of inherited retinal dystrophies which cause severe visual loss with non-recordable or a largely reduced ERG response either congenitally or in very early childhood. The fundus findings may be indistinguishable from those seen in typical RP.
- ▶ **Atrophia gyrata** is a rare autosomal recessive disorder with clinical signs similar to those seen in RP; it is however characterized by sharply defined garland shaped areas of chorioretinal atrophy. An elevation of plasma ornithine level caused by an ornithine aminotransferase deficiency is involved in this disease.
- Syndromic diseases with retinal dystrophy may be marked by typical or atypical RP. Frequent symptoms associated with RP are sensory neural hearing loss such as in Usher syndrome or various kinds of neurological involvement such as mental retardation, ataxia or other signs of degeneration of the central or peripheral nervous system, e.g. Laurence-Bardet-Biedl syndrome or Refsum's disease. Various metabolic disorders can occur together with retinal



Retinitis Pigmentosa. Figure 1 Hereditary retinal diseases can be caused by many different, usually monogenic mutations in either the phototransduction process and its support mechanisms ("Rhodopsin Cycle") or by involvement of the Retinal pigment epithelial cells that are important for the metabolic steps of visual purple regeneration ("Retinoid cycle").

degeneration such as paroxysmal disorders, mucopolysaccharidosis or abetalipoproteinemia.

- Phenocopies can mimic retinitis pigmentosa by their very similar fundus appearance in several non-genetic acquired disorders. Drug exposure, especially binding to melanin or neurotoxic drugs, as well as malnutrition and infectious diseases such as syphilis can mimic RP as well.

Genetic Differentiation

To the present at least 150 cloned and/or mapped genes causing retinal dystrophies and degenerations have been found (for surveys see 7). They are listed in specialized data banks such as RETNET: ► <http://www.sph.uth.tmc.edu/retnet/disease.htm>

At this Internet site the OMIM numbers, the genetic location, the known diseases and/or proteins, the mapping method and the reference given for each chromosome and for the mitochondrion are given. Table 1 shows the more common cloned genes of

hereditary retinal dystrophies and their contribution to such dystrophies.

Cellular and Molecular Mechanisms

In the following, only a subgroup of hereditary retinal dystrophies in which cellular and/or molecular mechanisms have been elucidated to some degree is discussed.

As also shown in the subheadings in Table 1, RP symptoms can be caused by mutations in practically every component of the phototransduction process from rhodopsin (RHO) to subunits of transducin, phosphodiesterase (PDE), the cyclic nucleotide gated cation channel (CNG3) or arrestin, to name just a few common examples. Thus the generation of an appropriate photoreceptor signal is hampered or a slow toxic increase of metabolic products is induced, leading finally to cell death, as can be studied in spontaneous animal models or generated mutants (Fig. 1).

Another set of causes for retinal dystrophies involves structural proteins responsible for the integrity of the membrane (e.g. ROM 1, cadherin etc.)

A third chain of events involves retinal pigment epithelium (RPE) cell function by mutations that disturb the photoreceptor pigment renewal performed in the retinoid cycle as shown in Fig. 1. The ►[retinoid cycle](#) is closely related to the ►[rhodopsin cycle](#) and involves the transport of retinoids from the rod outer segment (ROS) across the interphotoreceptor matrix (IPM) to the RPE and *vice versa*. Numerous clinically different retinal diseases, as shown in Fig. 1 stem from mutations in the various proteins involved in this cycle, especially enzymes such as retinol-kinase (RK) retinol dehydrogenase (RDH), lecithin-retinol-acetyltransferase (LRAT), isomerase controlling proteins (RPE 65), retinol binding proteins (RBP) and cellular retinol binding protein (CRALBP) serving as carrier for RDH5. Some mutations cause flecked retinal diseases by yellowish deposits of products of the altered visual cycle (disorders are marked by frames in Fig. 1). Other genes (TULP1) involve the ciliary structure, important e.g. for rhodopsin transport or the phagocytosis (MERTK) necessary for photoreceptor renewal. Transcription factors can also be affected (CRX) or splicing mechanisms (PRPF8) besides not fully upstream metabolisms such as the geranylgeranyl transferase in choroideremia.

Clinical Relevance

The fact that retinitis pigmentosa is the most common cause of blindness in early life and the fact that initial successful gene therapy approaches have been performed in dogs suffering from early onset RP (LCA, RPE65) make clear that the understanding of the various molecular mechanisms is a most important prerequisite for targeted therapeutic strategies in these diseases that are not yet curable.

Details and references can be found in the continuously updated RETNET data bank (see above).

References

1. Rivolta C, Sharon D, De Angelis MM et al (2002) Retinitis pigmentosa and allied diseases: numerous diseases, genes and inheritance patterns. *Hum Mol Genet* 11:1219–1227
2. Phelan JK, Bok D (2000) A brief review of retinitis pigmentosa and the identified retinitis pigmentosa genes. *Mol Vis* 6:116–124
3. Zrenner E, Apfelstedt-Sylla E, Rütther K (1995) Clinical aspects: retinitis pigmentosa. In: Djamgoz MBA, Archer SN, Vallerger S (eds) *Neurobiology and Clinical Aspects of the Outer Retina*. Chapman & Hall, London, pp 447–460

4. Zrenner E, Rütther K, Apfelstedt-Sylla E (1992), Retinitis Pigmentosa. *Ophthalmologie* 89:5–21
5. Massoff RW, Finkelstein (1987), A two stage hypothesis for the natural course of retinitis pigmentosa. In: Zrenner E, Krastel H, Goebel H-H (eds) *Research in Retinitis Pigmentosa*. Advances in Biosciences, vol 62, Pergamon Press, Oxford, New York, pp 29–58
6. Fishman GA, Alexander KR, Anderson RJ (1985) Autosomal dominant retinitis pigmentosa: a method of classification. *Arch Ophthalmol* 103:366–374
7. Rattner A, Sun H, Nathans J (1999) Molecular genetics of human retinal disease. *Annu Rev Genet* 33:89–131

Retinoblastoma

Definition

Retinoblastoma (RB) is a malignant tumor of the developing retina that occurs in children, usually before the age of five years. Retinoblastomas can occur both in hereditary and non-hereditary forms.

► [Tumor Suppressor Genes](#)

Retinoblastoma (Rb) Protein

Definition

Retinoblastoma (Rb) protein is the protein product of the ►[retinoblastoma gene](#) RB1.

► [Breast Cancer](#)

Retinoblastoma Gene

Definition

Retinoblastoma gene (RB1), located on chromosome 13q, is a classical tumor suppressor gene which is important for nervous system development, aging and replication of some viruses. RB1 is lost in familial retinoblastoma and many other malignant tumors. Physiologically, it represses the cell cycle by binding to an S-phase inducing transcription factor, E2F.

► [Breast Cancer](#)

► [Hyper- and Hypoparathyroidism](#)

Retinoic Acid

Definition

Retinoic acid designates the acid form of vitamin A. It is one of the retinoids; essential for growth, cell differentiation, and the immune system.

► [Nuclear Hormone Receptors: Regulators of Gene Transcription and Cellular Signaling](#)

Retinoid Cycle

Definition

Absorption of light by rhodopsin or cone pigments in photoreceptors triggers photoisomerization of their universal chromophore, 11-cis-retinal, to all-trans-retinal. This photoreaction is the initial step in phototransduction that ultimately leads to the sensation of vision. The process that restores rhodopsin and counterbalances the bleaching of rhodopsin, is called the retinoid cycle. In this process, visual pigments are regenerated by enzymatic isomerization of all-trans-retinal to 11-cis-retinal. Regeneration begins in rods and cones when all-trans-retinal is reduced to all-trans-retinol. The process continues in adjacent retinal pigment epithelial cells (RPE), where a complex set of reactions converts all-trans-retinol to 11-cis-retinal. A number of enzymes is involved in the transport, storage, and transformation of vitamin A and its derivatives; mutations in the genes encoding these proteins leads, in many cases, to retinal diseases, either cone-rod dystrophies or tapetoretinal degenerations of the retinitis pigmentosa group.

► [Retinitis Pigmentosa](#)

Retinopathy

Definition

Retinopathy is a disease of the retina in the eye. Retinal capillaries that provide blood to the retina are destroyed. With the onset of retinopathy, retinal capillaries weaken or bulge with microaneurysms that may hemorrhage, leaking blood or fluid into surrounding tissue. When new blood vessels grow on the retina

(and into the vitreous) they can cause blurred vision and even temporary blindness. Ultimately, scar tissue may form, detaching the retina from the back of the eye and often causing permanent loss of vision. Chronically elevated blood insulin and glucose levels can induce retinopathy.

► [Diabetes Mellitus, Genetics](#)

Retroposon

► [Retrotransposon](#)

Retropseudogene

► [Processed Pseudogene](#)

Retrotransposition

Definition

Retrotransposition describes a process that involves transcription, processing mRNA, translation, reverse transcription of the transcribed mRNA, and integration of the resulting cDNA into the host genome.

► [Repetitive DNA](#)

Retrotransposon

Definition

Retrotransposon (retroposon) defines a mobile (transposable) DNA element in eukaryotes that is reproduced via reverse transcription of an RNA intermediate. It can be split into two classes: ► [LTR](#) and non-LTR retrotransposons.

► [Repetitive DNA](#)

► [Retroviruses](#)

► [Transposons](#)

Retroviruses

JOHN M. COFFIN

Department of Molecular Biology and Microbiology,
Tufts University, Boston, MA, USA
john.coffin@tufts.edu

Definition

Retroviruses are a group of viruses that is widespread in nature, and contains members, identifiable as either infectious transmissible agents or as germline remnants of past infections, in all vertebrate species as well as many invertebrates. Related elements can also be found as ►retrotransposons in the genomes of plants, yeast and many other eukaryotic organisms. Effects of different retroviruses on the infected host vary greatly from asymptomatic infection to severe diseases such as immunodeficiency, malignancy, neurological disease and many others. Despite their wide distribution and diverse biology, retroviruses form a monophyletic group, in that all are more closely related to one another than any is to any other type of virus or transposable element.

Characteristics

Despite their wide distribution and varied biology, all retroviruses share a set of defining characteristics:

A ►virion (Fig. 1) consisting of:

- An inner core of uncertain symmetry made up of about 1500 copies of at least 3 proteins: ►MA (matrix); ►CA (capsid); and ►NC (nucleocapsid), products of the ►gag gene
- At least 3 viral enzymes: ►PR (protease) a product of the ►pro gene; ►RT (reverse transcriptase) and ►IN (integrase), products of the ►pol gene
- An ►envelope, derived from a cell membrane modified by the insertion of the trimeric combination of the ►SU (surface) and ►TM (transmembrane) proteins, products of the ►env gene

A genome (Fig. 2) with the following properties:

- Two identical molecules of plus (mRNA)-sense RNA molecules, 7–12 kb in length, which are capped and polyadenylated like cellular mRNA's
- Four genes, always in the order: *gag* (encoding internal structural proteins); *pro* (encoding the protease necessary for maturation of virions); *pol* (encoding reverse transcriptase and integrase) and *env* (encoding the virion SU and TM proteins

necessary for entry) as well as additional genes in some genera

- Associated molecules of a tRNA, specific for each virus group, which serve as primers for the initiation of ►reverse transcription.

A replication cycle (Fig. 3) that includes:

- Entry mediated by interaction with a host-cell receptor
- Reverse transcription of the RNA genomes into a single molecule of double stranded DNA characterized by ►long terminal repeats (LTRs) of about 0.5–1.5 kb
- ►Integration of viral into host cell DNA at more-or-less random sites to form the ►provirus
- Transcription of the provirus to yield new genomes as well as spliced and unspliced mRNAs that encode virion and (in some groups) regulatory proteins
- Assembly of the newly-made genomes and ►poly-protein precursors for internal (Gag), and enzymatic (Pro and Pol) proteins into capsids
- Budding of virions from the cell membranes and release into the extracellular space
- Proteolytic processing of the internal Gag, Pro and Pol proteins to yield mature, infectious virions

Based on details of genome organization and molecular relationships, retroviruses are classified into 7 genera (Table 1). All these genera have currently active members; some are also represented as ►endogenous proviruses, which are remnants of infection of the germ line of the host dating from recent times to 30 million or more years ago. In humans, there are about 80,000 such elements, as compared to about 30,000 genes. These proviruses constitute a valuable record of ancient host-virus interaction, as well as subsequent evolution.

The nature of retrovirus genome structure and replication has made them valuable as models for disease, particularly cancer, as well as vectors for gene delivery in both experimental and therapeutic settings. These virtues result from the facts that cis-acting signals for replication are found at the ends of the genome and that, in general, virus-encoded components necessary for replication are found in the virion. Therefore, genomes with non-viral sequences replacing some or all viral genes can be packaged into infectious virions when expressed in cells that express a full set of virion proteins.

Molecular Interactions

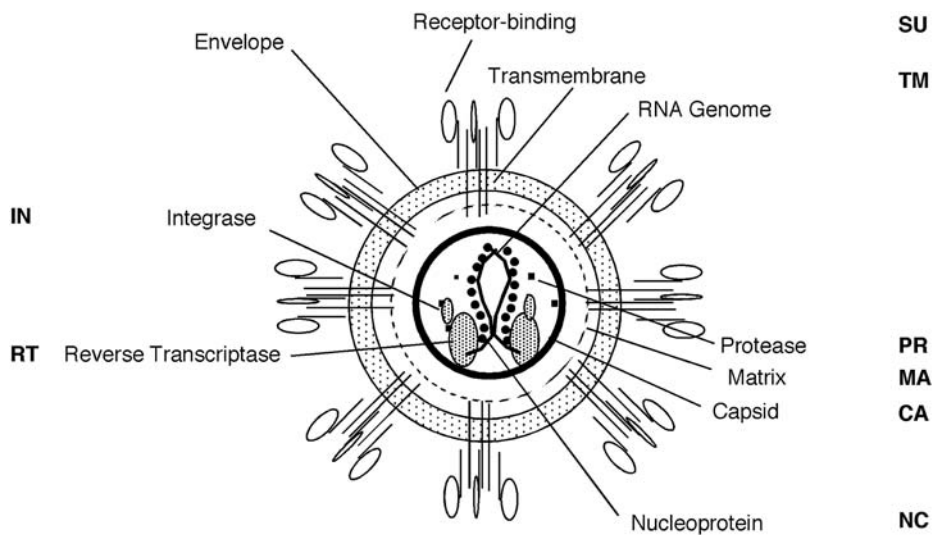
The replication cycle is characterized by a large number of interactions between and among viral and cellular proteins and nucleic acids, many of which are still

poorly characterized. Those central to the replication cycle are discussed here.

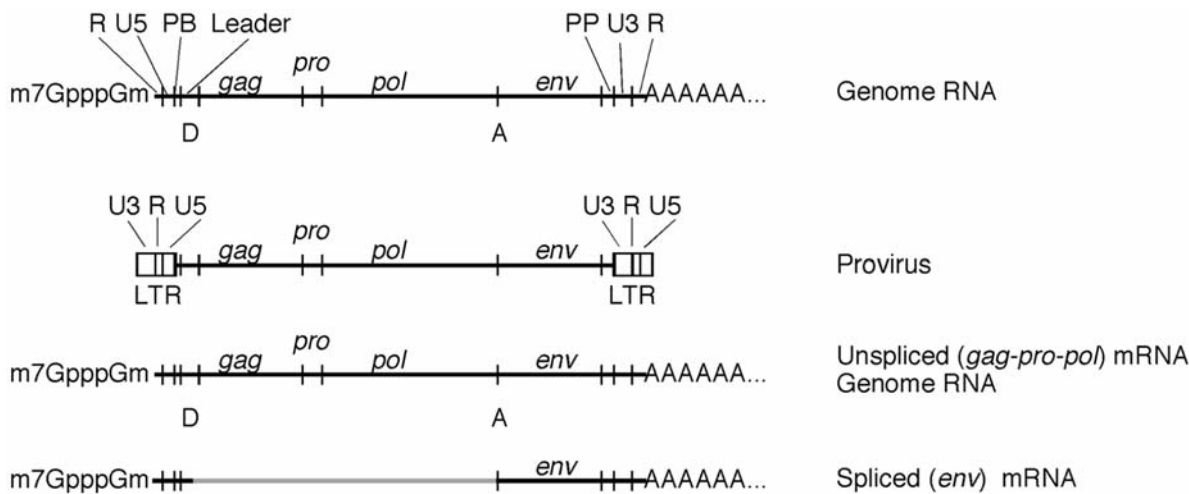
Entry

Entry of all retroviruses into cells involves fusion of virus and cell membranes after interaction of the Env protein with one or more specific receptors on the cell surface. This interaction is highly specific and a large number of different proteins is used as receptors by

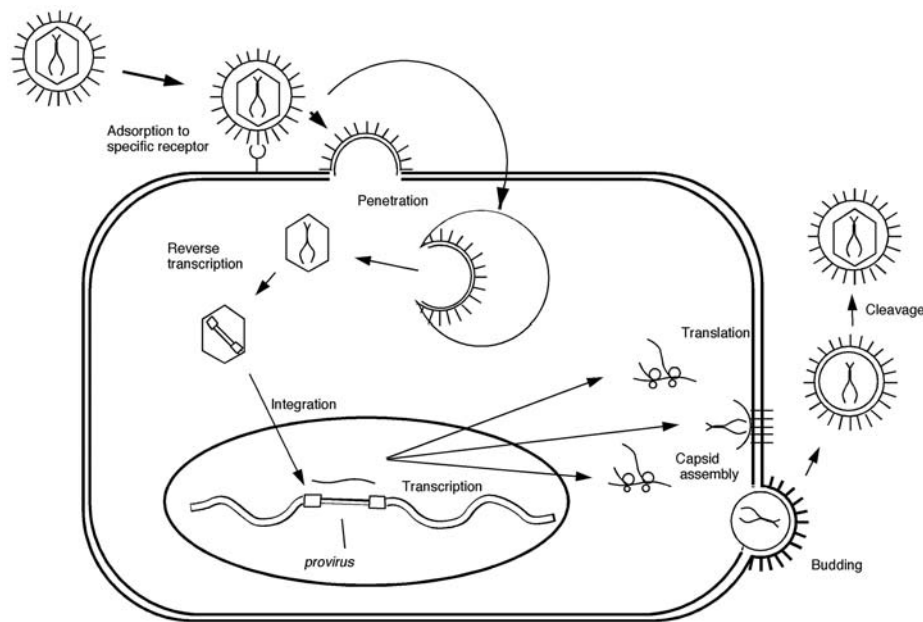
different viruses. The distribution of appropriate receptors is the most important determinant of retrovirus species and tissue tropism. For example, HIV-1 recognizes only the CD4 protein found on the T helper cells, macrophages and a few other cell types and then only the CD4 of humans and chimpanzees, limiting its infection to these specific cell types and species. Following binding of a specific site on the SU to the receptor, rearrangement of the SU and TM proteins



Retroviruses. Figure 1 A typical retrovirus virion.



Retroviruses. Figure 2 Minimal (simple) retrovirus genome, provirus and mRNAs. Symbols: R, a short terminally redundant sequence. U5, a unique sequence near the 5' end of the genome. U3, a unique sequence near the 3' end of the genome; PB, a site complementary to the 3' end of the RNA primer that provides the priming site for minus strand DNA synthesis. PP, the polypurine tract, which provides the initiation site for plus strand DNA synthesis. D, A donor and acceptor splice sites.



Retroviruses. Figure 3 The retrovirus replication cycle.

Retroviruses. Table 1 Retrovirus genera

Genus	Example Species	Simple/ Complex	Endogenous Members?
Alpha retrovirus	Avian Leucosis V. (ALV) Rous Sarcoma V. (RSV)	Simple	Yes
Beta retrovirus	Mouse Mammary Tumor V. (MMTV) Mason-Pfizer Monkey V. (MPMV)	Simple	yes
Gamma retrovirus	Murine Leukemia V. (MLV) Feline Leukemia V. (FeLV)	Simple	yes
Delta retrovirus	Human T-Cell Leukemia V. (HTLV) Bovine Leukemia V. (BLV)	Complex	No
Epsilon retrovirus	Walleye Dermal Sarcoma V. (WDSV)	Complex	?
Lentivirus	Human Immunodeficiency V. (HIV) Simian Immunodeficiency V. (SIV) Equine Infectious Anemia V. (EIAV)	Complex	No
Spumavirus	Human Foamy V. (HFV)	Complex	No

leads to exposure of a hydrophobic region of the TM protein that is buried in the native trimer. This region, known as the ►fusion domain, interacts with the target cell membrane and folding of two separate regions of the TM protein trimer into a coiled coil draws the membranes together until they fuse, causing the capsid to be introduced into the cytoplasm of the target cell. Although these steps are common to all retroviruses, different groups of viruses use different mechanisms to link receptor interaction to membrane fusion. With many viruses, including ►Gammaretroviruses like

MLV, fusion is activated by receptor binding itself, with no apparent intermediate steps. ►Alpharetroviruses (like ALV), by contrast, require endosomal uptake and the consequent reduction in pH for entry to occur. Unlike other viruses that use this pathway, Alpharetroviruses require receptor binding to activate the Env protein for low-pH mediated fusion, thus preventing infection of cells lacking receptors by chance endocytosis. ►Lentiviruses, particularly HIV and related viruses, require interaction with 2 cell surface molecules, the CD4 receptor and a ►coreceptor

molecule, usually one of the chemokine receptors CCR5 or CXCR4. Interaction with the receptor leads to rearrangement of the SU, exposing the binding site for the coreceptor, binding to which leads to rearrangement of the TM protein and fusion.

DNA Synthesis

Following entry into the cytoplasm, the viral RNA is copied into a double stranded DNA molecule by the action of the RT present in the virion. This process is covered in another chapter (Wainberg) and will not be discussed in detail here. It is important to note that reverse transcription takes place in a structure derived from the capsid and leads to a double stranded DNA containing long terminal repeats (LTRs) derived by copying the U3, R, and U5 regions at the ends of the RNA genome twice (Fig. 2).

Nuclear Transport

Following reverse transcription, the DNA-containing structure, often referred to as the **▶preintegration complex (PIC)**, must find its way to the nucleus. The process by which this transport happens is unclear and controversial and also seems to vary from virus to virus. MLV, for example, appears to have no active process for nuclear uptake of a structure as large as a PIC and requires breakdown of the nuclear membrane at mitosis to gain access to chromosomal DNA. This restriction imposes a requirement for division of the infected cell on the process of MLV replication. HIV, by contrast, can access an active cellular transport mechanism (which one is still unclear) for nuclear import of the PIC and therefore can infect many kinds of nondividing cells.

Integration

The covalent joining of viral to cell DNA is the central and defining event in the retrovirus life cycle. It is carried out in the context of the PIC, containing the viral DNA, IN, RT and most probably some capsid proteins as well as some cellular protein cofactors. Integration proceeds in 3 steps, the first two of which are carried out by IN: (1) removal of two bases from each 3' end of the viral DNA to expose a conserved CA-OH dinucleotide; (2) attack of the 3' OH at each end on internucleotide phosphates spaced 4–6 bases apart in the target DNA; (3) repair of the resulting gapped intermediate to join the opposite strands and create a characteristic 4–6 base pair duplication using cellular DNA repair systems. Although some viruses (including HIV) seem to exhibit a clear preference for integration into transcribed regions of host DNA and integration targets display a modest sequence preference, integration events are very widespread in the host genome and can, for many purposes, be considered to be randomly distributed.

Transcription

All retroviruses use host cell RNA polymerase and transcription factors for expression of the integrated provirus. For this reason, they must provide sites for transcription factor binding upstream of the initiation site of transcription at the U3-R border. Thus the U3 region of all retroviruses contains a complex combination of such sites and binds a number of transcription factors. As discussed in the next section, the pattern of transcription factors used by a virus can make an important contribution to regulation of retrovirus expression in certain cell types and certain differentiation states. Simple retroviruses, as a rule, do not actively participate in this process and all important functions are carried out by cellular machinery acting on viral DNA signals. Complex retroviruses encode **▶transactivating proteins** (e.g., **▶Tat** in HIV, **▶Tax** in HTLV) that are necessary for efficient transcription, yet work in very different ways. The role of these factors in regulation of the replication cycle is discussed below.

RNA Processing and Export

Retroviral transcripts are capped and polyadenylated by the same cellular systems as for cellular mRNAs. In addition, all viruses encode at least one gene (*env*) that must be expressed from a spliced mRNA. In the case of some complex viruses like HIV, many variously completely and partially spliced mRNA's are required for expression of all virus genes. Thus unlike most cell mRNAs, for which complete splicing is required to license RNA for transport, genome and *gag-pro-pol* mRNAs must be exported from the nucleus in an incompletely spliced form. Retroviruses use one of two mechanisms for this purpose. Some viruses encode a specific protein (**▶Rev** in HIV, **▶Rex** in HTLV) that binds to a highly structured region in the RNA and mediates its export in the absence of splicing. Other (simple) retroviruses contain a sequence (sometimes called the constitutive transport element, or CTE) that interacts directly with nuclear export machinery to bypass splicing of a fraction of genomes.

Translation

Translation of the *gag-pro-pol* mRNA takes place on free polysomes, while *env* mRNAs are translated on membrane-bound polysomes and cotranslationally secreted into the endoplasmic reticulum. A special feature of retroviral translation is the mechanism of synthesizing Gag, Pro, and Pol protein precursors from the same mRNA. Depending on the virus, there are terminators at the end of *gag* and/or *pro* that can be bypassed by either causing ribosomes to shift into the –1 reading frame (**▶frameshift suppression**), or by insertion of an amino acid opposite the termination codon (**▶read-through suppression**). Whatever the

mechanism, the result is the synthesis of Gag and Gag-Pro-Pol polyproteins at a ratio of 1:10 or so, as required for efficient assembly of virions with the correct ratio of structural proteins and enzymes.

Assembly

Synthesis of the viral gene products as polyproteins at relative amounts determined by efficiency of suppression of termination allows for efficient assembly of virions, since the sites required for the appropriate interactions need be present only once. Although a few retroviruses are assembled in the cytoplasm and subsequently associate with the plasma membrane, most retroviruses assemble at the plasma membrane, in a process involving simultaneous interaction between viral mRNA and proteins, viral proteins with one another and viral proteins and the cell membrane. Sites for Gag-Gag interaction are found on the NC and CA domains, for Gag-RNA interaction on NC, which contains both basic domains and zinc fingers essential for recognition of genome RNA and for Gag-membrane interaction on MA, which is usually myristoylated at its N-terminus. Assembly also requires RNA; genomes are preferentially packaged as a result of interaction of NC with a packaging (ψ) sequence near the 5' end of the genome. In the absence of genomes, cell RNAs are nonspecifically incorporated into virions. Pro and Pol proteins are included in proportions equal to their representation among the primary products of translation.

Release

As assembly proceeds, the envelope is formed by wrapping of the cell membrane, modified by insertion of the Env trimer, around the growing capsid. This process is apparently passive, driven only by association between the respective components. By contrast, the final step in budding – the release of virions from the cell surface by sealing of the neck attaching the nearly complete virion to the cell – requires the active participation of the cellular ►ESCRT machinery, normally involved in the formation of multi vesicular bodies by enabling budding of endosomal membranes. All retroviruses contain motifs that interact with various members of the ESCRT pathway, including tsg101, NEDD4 and AIP1 among others. All of these motifs are in the Gag protein, albeit in different locations in different viruses.

Processing

As budded, the immature virions are noninfectious, and require processing by the PR molecule initially embedded in an inactive form in the Gag-Pro-Pol polyprotein. Delay of processing and maturation until after budding ensures both that all gene products

remain in the virion and that infectious virions cannot find their way back into the already infected cell. PR is an aspartic protease whose active form is a dimer, with the active site formed by the dimer interface. Apparently, dimerization of the protein from within the precursor is a very slow step, accounting for the delay in processing, but once the first active dimer is formed, processing of the remaining polyprotein molecules occurs very rapidly, in a chain reaction, so that intermediates in virion maturation are not observed. The Env precursor also must be processed for infectivity; this step is accomplished by action of a cellular protease (►furin) found in the Golgi.

Regulatory Mechanisms

Simple Retroviruses and Passive Regulation

Regulation of the replication process for many retroviruses is accomplished in a passive manner determined by the specificity of the virus-host interaction. For example, the distribution of receptors for a given virus determines its target cell specificity. Another important type of passive regulation is the cell-type and temporal specificity determined by the pattern of transcription factor binding signals in the U3 portion of the LTR. For example, mammary tumor virus (►MMTV) is transmitted naturally from mother to newborn mice in milk. Soon after transmission, it establishes an infection in mammary gland tissue, where the provirus remains quiescent until lactation, at which time both the loss of negative control and the appearance of positive control elements, including glucocorticoid hormones, ensure expression at the correct place and time for efficient transmission, with minimal risk of collateral damage due to induction of transformation of other infected cell types (B and T cells, for example) by integration of active LTRs near oncogenes.

Complex Viruses and Active Regulation

In addition to these mechanisms, complex retroviruses like HIV have developed a mechanism for regulating expression of the provirus involving the transcriptional activator (Tat) as well as the RNA transport protein (Rev). Since both these proteins are translated from fully-spliced RNA's, their expression does not depend on the presence of the transport protein and a small amount of constitutive transcription from the newly integrated provirus in the absence of Tat can lead to expression of both Tat and Rev. Tat, through its binding to a region (TAR) near the 5' end of the nascent RNA and activation of phosphorylation of the C-terminal domain of RNA pol II, causes a large increase in transcription of the provirus, allowing expression of itself and other products of fully spliced mRNAs to be

rapidly ramped up in the absence of virion proteins, all of which are translated from partially-spliced RNAs. When concentrations of Rev increase to adequate amounts, nuclear export of Gag-Pro-Pol and Env mRNAs can occur and the synthesis of these “late” proteins commences. This control scheme permits a phased replication cycle in which the “early” proteins Tat and Rev set the stage for high level expression of the “late” virion proteins. HTLV exhibits a very similar pattern of regulation involving the analogous Tax and Rex proteins.

The benefit of such regulation to the virus is at least two fold. First, it prevents exposure of most virus antigens to the immune system until expression is maximal, permitting a rapid burst of virions to be produced before the infected cell is detected and killed. Second, it allows a means of balancing expression of virion proteins with other proteins depending on their mode of expression. HIV, for example, encodes several other proteins that help to overcome roadblocks thrown in its way by the cell and the Tat-Rev system allows them to be expressed in proper proportion to virions.

- ▶ Leucine Zipper Transcription Factors: bZIP Proteins
- ▶ Repetitive DNA
- ▶ Translational Frameshifting, Non-standard Reading of the Genetic Code
- ▶ tRNA
- ▶ Viral Oncogenesis

References

1. Coffin JM, Hughes SJ, Varmus HE (eds) (1997) Retroviruses. Cold Spring Laboratory Press, Cold Spring Harbor
2. Coffin JM, Hughes SH, Varmus HE (1997) Intermezzo: The interactions of retroviruses and their hosts. In: Coffin JM, Hughes SH, Varmus HE (eds) Retroviruses. Cold Spring Laboratory Press, Cold Spring Harbor, pp 335–342
3. Goff SP (2001) Retroviridae: The retroviruses and their replication, chapter 37. In: Knipe D, Howley P (eds) Fields Virology. Lippincott, Williams and Wilkins, Philadelphia

Rett Syndrome

HAYLEY ARCHER, ANGUS J. CLARKE
Institute of Medical Genetics, University Hospital of Wales, Heath Park, Cardiff, UK
archerhl@cardiff.ac.uk
clarkeaj@cardiff.ac.uk

Synonyms

Rett disorder

Definition

Rett syndrome (RTT) was first described by Andreas Rett in German in 1966 and in English in 1985. He described a distinctive nonprogressive, neurodevelopmental disorder in girls associated with a variable hyperammonaemia and characterised by deceleration of psychomotor development and the subsequent loss of acquired cognitive and motor skills. It is an X-linked dominant disorder, affecting 1 in 10,000 to 1 in 15,000 females and possibly 1 in 100,000 males, making it one of the most common causes of profound intellectual disability in girls. The clinical features and natural history of classical RTT have been well described. It is often associated with mutations in the *MECP2* gene located at Xq28. Variant or atypical RTT is a term used to describe cases with some, but not all of the features of classical RTT; this is much less commonly associated with *MECP2* mutations. RTT can occur without mutations in *MECP2* and not all those with mutations in *MECP2* have RTT.

Characteristics

Clinical Features

Rett syndrome is essentially a clinical diagnosis that is supported by mutational analysis of the *MECP2* gene in ~85% of patients with classic features of RTT. An international working group published modified guidelines in 2002 to aid the clinical diagnosis of RTT (Table 1). The phenotypes associated with mutations in the *MECP2* gene range from healthy, unaffected carrier females to severe neonatal ▶encephalopathy in males.

Classical RTT syndrome is described in those who have a normal prenatal and perinatal history and apparently normal development in the first 6 months of life. Subsequently there is a period of regression, with onset between the ages of 6 months and 3 years of age, with the loss of acquired motor skills and the development of hand stereotypies. There is postnatal deceleration of head growth. Profound cognitive impairment becomes apparent. In addition there are a number of supportive features including seizures, growth retardation, ventilatory irregularity and other features of autonomic dysregulation (Table 1). A variable hyperammonaemia and lactic acidosis have sometimes been reported.

Four stages in Classical RTT have been identified:

- stage 1 (first 6–18 months) hypotonia, disinterest in play.
- stage 2 (age 6 months up to 3 years) a period of agitation, withdrawal and rapid regression, often with some features resembling those found in autism.
- stage 3 (2–10 years) post-regression stabilisation with seizures, cognitive impairment, hand stereotypies, hyperventilation, air swallowing and teeth grinding

Rett Syndrome. Table 1 Revised diagnostic criteria for Rett Syndrome¹

Necessary criteria
1 Apparently normal prenatal and perinatal history
2 Psychomotor development largely normal for the first 6 months or may be delayed from birth
3 Normal head circumference at birth
4 Postnatal deceleration of head growth in the majority
5 Loss of achieved purposeful hand movements between ages 6 months-2 ½ years
6 Stereotypic hand movements such as hand wringing/squeezing, clapping/tapping, mouthing and washing/rubbing automatisms
7 Emerging social withdrawal, communication dysfunction, loss of learned words and cognitive impairment
8 Impaired (dyspraxic) or failing locomotion
Supportive criteria
1 Awake disturbances of breathing (hyperventilation, breath-holding, forced expulsion of air or saliva, air swallowing)
2 Bruxism (teeth grinding)
3 Impaired sleep pattern from early infancy
4 Abnormal muscle tone successively associated with muscle wasting and dystonia
5 Peripheral vasomotor disturbances
6 Scoliosis/kyphosis progressing through childhood
7 Growth retardation
8 Hypotrophic small and cold feet; small thin hands
Exclusion criteria
1 Organomegaly or other signs of storage disease
2 Retinopathy, optic atrophy or cataract
3 Evidence of perinatal or postnatal brain damage
4 Existence of identifiable metabolic or other progressive neurological disorder
5 Acquired neurological disorders resulting from severe infections or head trauma

- stage 4 (10 years+) scoliosis, rigidity, increased feeding difficulty, less mood disturbance, reduction in seizures and improved eye contact.

Variant RTT is used to describe those with some, but not all of the features of classical RTT. There are some more specific phenotypes within this category, congenital-onset RTT, the preserved speech variant and the delayed onset variant. The clinical diagnosis of variant RTT is based upon a person meeting 3 of 6 main criteria and 5 of 11 supportive criteria (Table 2). Congenital RTT is sometimes used to describe those whose development was never normal, but who have many features of the Rett disorder. However it is important to note that there are developmental problems in all RTT

girls, present from birth, but these are subtle and often missed. Delayed onset variant is used to describe those with regression occurring later than 3 years, and subsequently developing features consistent with variant Rett disorder. Preserved speech variant (PSV) describes those with Rett who are able to speak. They may only have a few words or even be able to converse. These patients are mildly affected and frequently have a previous diagnosis of atypical autism and speech delay. The Angelman syndrome-like RTT phenotype is seen in both males and females with RTT; the characteristic facial features include a long face, wide mouth and widely spaced teeth. Apart from this, and a few exceptional cases associated with various malformations,

Rett Syndrome. Table 2 Revised delineation of variant Rett phenotypes¹

Inclusion criteria	
1	Meet at least 3 of 6 main criteria
2	Meet at least 5 of 11 supportive criteria
Six main criteria	
1	Absence or reduction of hand skills
2	Reduction or loss of babble speech
3	Monotonous pattern to hand stereotypies
4	Reduction or loss of communication skills
5	Deceleration of head growth from first years of life
6	RS disease profile: a regression stage followed by a recovery of interaction contrasting with slow neuromotor regression
Eleven supportive criteria	
1	Breathing irregularities
2	Bloating/air swallowing
3	Bruxism (harsh sounding type)
4	Abnormal locomotion
5	Scoliosis/kyphosis
6	Lower limb amyotrophy
7	Cold, purplish feet, usually growth impaired
8	Sleep disturbances including night screaming outbursts
9	Laughing/screaming spells
10	Diminished response to pain
11	Intense eye contact/ eye pointing

the dysmorphology of RTT is subtle. Features described in females include a short, broad face, wide nasal bridge, hypotelorism, low hanging columella, short philtrum, everted upper lip and full lower lip. Most affected cases do not have a dysmorphic appearance.

The problems of RTT change with time. In teenage years and adult life, progressive scoliosis, feeding difficulties and hypertonia become increasingly prominent. Early features such as hyperventilation, seizures, poor eye contact and mood fluctuations improve. Most girls with RTT survive to adult life. However, the death rate is reported to be 1.2% per annum. Approximately 95% of affected women survive to 25 years. Early death occurs most commonly in those with severe RTT who are in a wasted state. Severe epilepsy or causes unrelated to RTT may also cause early demise. However, some deaths are

completely unexpected and may be related to the autonomic disturbance in RTT.

Males

The classical RTT phenotype arises in males most commonly in association with Klinefelter syndrome or with somatic mosaicism for a typical RTT-causing mutation in *MECP2*. When affected males arise in families with at least one affected female, the phenotype is often not typical – it is often more severe than in classical RTT and can range from classical RTT to neonatal encephalopathy. A small number of males have been reported with RTT-associated mutations in the *MECP2* gene. These often present with severe neonatal encephalopathy, hypotonia, seizures and apnoeic spells and usually die in early childhood. Dysmorphic features have been described in association with the encephalopathic presentation, as has bilateral perisylvian polymicrogyria.

Other mutations in the *MECP2* gene – apart from those associated with RTT in females – may be associated with X-linked recessive phenotypes in males, although it remains difficult to be certain whether some missense changes reported in association with non-syndromic mental retardation in males have any functional or pathogenetic significance. A phenotype of X-linked mental retardation with progressive spasticity is recognised. Developmental delay, hypotonia, sialorrhea, relatively large head circumference and complicated spastic paraparesis were apparent in affected males. Compared with cases of classical RTT, the features of growth retardation, the loss of hand skills and the acquired microcephaly are absent. Similarities with classical RTT include sialorrhea, spasticity, tooth grinding, ataxic gait and absence of speech. See below for discussion of A140V.

MECP2 Mutations

Mutations in *MECP2* are identified in more than 80% of classical RTT and up to 40% of less typical RTT. The low frequency of mutations in variant RTT suggests that this group may be heterogeneous. Given that at least 20% of classical cases do not have a mutation in the *MECP2* coding region and that there is a lower chance of identifying a *MECP2* mutation in families with affected sibs, it is likely that RTT is genetically heterogeneous. Some patients with features of Rett and onset of epilepsy or infantile spasms before 3 months of age have mutations in a second gene: *CDKL5* (cyclin dependent kinase like 5). In fact, a mutation in *MECP2* has not yet been reported in any patient with infantile spasms.

The *MECP2* gene has a very high mutation rate, perhaps 4×10^{-5} , as evidenced by the predominantly sporadic nature of RTT and its relatively high incidence

in the population. Sib recurrence risks are generally quoted as less than 1%. Mutations commonly arise at CpG hotspots and result from spontaneous deamination of methylated cytosine residues. There are eight recurrent mutations, which account for 70% of reported mutations and these arise at eight different CpG dinucleotide locations (Fig. 1).

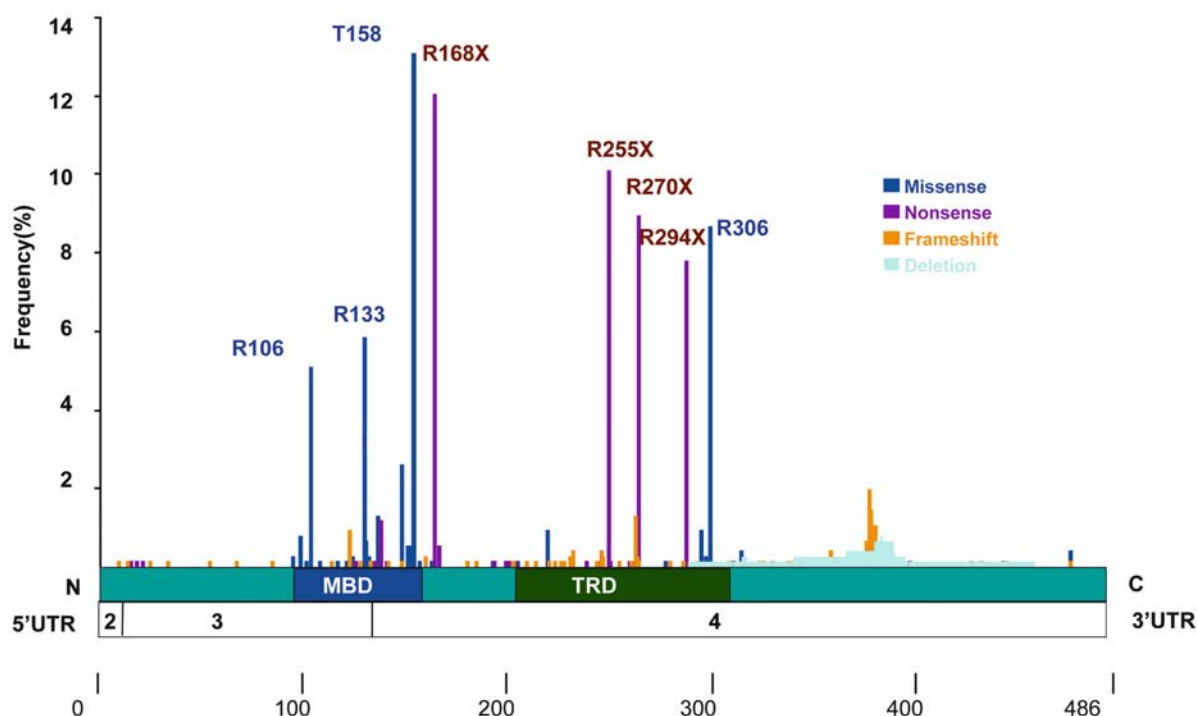
The female preponderance of RTT was previously thought to be the result of *in utero* lethality for affected male foetuses, but an additional factor is now known to operate – the origin of most mutations in spermatogenesis. This has been demonstrated by haplotype analyses indicating that over 95% mutations are paternally derived. This is consistent with other diseases where mutations tend to arise at CpG hotspots. Male germ cells are highly methylated and this, in combination with the increased number of mitotic divisions, may explain the largely paternal origin of *de novo* mutations.

Genotype-Phenotype Correlation

Many studies have been carried out to look at genotype-phenotype correlation in Rett syndrome. Two studies have shown that those with early truncating mutations are associated with a more severe phenotype than those with missense mutations. A recent study has shown that females with the missense mutation R133C have a relatively mild phenotype, with

better-preserved hand use, better ambulation and an increased likelihood of being able to use speech. The preserved speech variant (PSV) is associated with milder *MECP2* mutations i.e. missense and late truncating mutations. More recently large deletions of *MECP2* have been described, and it may be that deletions extending beyond the gene are associated with additional features.

X chromosome inactivation has an important role in modulating the phenotype of Rett syndrome. This is most clearly demonstrated in monozygotic twins, discordant for the Rett phenotype. In addition, clinically normal mothers with the same pathogenic mutation as their affected daughters have been described, in whom there has been highly skewed X-inactivation. There are some genotype-phenotype studies that have included X-inactivation and suggest that severe skewing is usually associated with a milder phenotype. Conversely, skewed X-inactivation has been associated with either a milder or more severe phenotype. Most girls with Rett Syndrome have random skewing and it is possible that those with severe skewing arose by chance. Further understanding of the influence of X inactivation on disease phenotype will require a knowledge of the parent of origin of the mutation and the phase of X inactivation. Since X-inactivation has recently been shown to be a leaky process, the development of functional analyses of the



Rett Syndrome. Figure 1 Site and nature of mutations in the *MeCP2* gene in Rett Syndrome.

MECP2 gene or MeCP2 protein may prove to be more revealing than X-inactivation studies.

The *MECP2* mutation, A140V was previously thought to be a benign variant, as it had no clinical effect in females. A number of phenotypes have now been reported in males with this mutation, mental retardation, childhood onset schizophrenia, mental retardation with abnormal gait and speech delay and PPM-X (a syndrome characterised by manic depressive psychosis, pyramidal signs, macroorchidism and mental retardation). This suggests either that the A140V mutation has a less severe effect on MeCP2 function than other mutations, or perhaps that it is a marker for a nearby genetic disorder.

Cellular and Molecular Regulation *MECP2*

Mutations in the *MECP2* gene were identified as the major cause of RTT in 1999 by Amir et al. This gene spans 76 kb, lying between the IRAK and rhodopsin loci at Xq28 and is subject to X-inactivation. It has four exons that are transcribed from telomere to centromere. It contains a 1461 nucleotide coding sequence within exons 2–4, translating a 486 amino acid protein. The MeCP2 translational initiation site was originally identified in exon 2. A second site has recently been identified in exon 1 which yields a new and predominant MeCP2 isoform. The MeCP2 protein contains a methyl binding domain, a transcription repression domain, two nuclear localisation signals and a proline rich C-terminal. The nuclear localisation signals are important for transporting MeCP2 to the nucleus. The C-terminus has a binding domain specific to nucleosomal DNA and may suggest that this domain reduces binding to methylated DNA and enhances protein stability.

Expression

MeCP2 protein is widely expressed throughout embryonic and adult tissues. Expression is low during early embryonic development. Levels of mRNA do not correlate well with MeCP2 protein levels in cells, suggesting that there may be tissue specific factors regulating MeCP2 expression. Recent studies show that MeCP2 follows the path of neuronal maturation.

MeCP2 Function

MeCP2 was first described in 1992 and has a role in gene regulation through methylation-dependent transcriptional repression and histone deacetylation. It binds symmetrically methylated CpG dinucleotides with its corepressor mSin3A. Histone deacetylases are recruited to the transcription repression domain,

contributing to the corepressor complex and play an important role in chromatin remodelling. Deacetylation of histones allows DNA to wrap more tightly around these proteins, preventing access of the transcription machinery to the promoter sequences. MeCP2 can act as a transcription repressor independently of histone deacetylases, again acting through its transcription repression domain. For example, MeCP2 binds to TF11B, a component of the basal transcription machinery.

MeCP2 also associates with pericentromeric heterochromatin. Consequently the loss of MeCP2 function may destabilise the process of cell division and possibly contribute to the disease phenotype. MeCP2 has a putative role as a matrix attachment protein, retaining silenced heterochromatin areas within the matrix. The flanking regions loop away from the matrix where they are more accessible to regulators of gene expression and chromatin structure. It is uncertain whether MeCP2 plays this matrix attachment role within neurons.

In the somatomotor cortex of RTT patients many genes involved in neuronal signal transduction show reduced expression and there is an increase in transcription in glial cells. Surprisingly, there has been no evidence of a generalised dysregulation of gene expression in animal models. The hypothesis that loss of MeCP2 function results in a generalised derepression of transcription in neurons is not supported by the available evidence.

Relationships are now being explored between *MECP2* and other genes with a role in brain growth and function including *BDNF*, *UBE3A*, *GABRB3*, *MAP2*, *NING1* and *DLX5*. That Ube3a expression is disturbed in patients with Rett is particularly interesting, since, in the early stages, those with RTT and Angelman can be difficult to clinically distinguish from one another.

Neuropathology

Brain weight is reduced but stable in RTT and histological studies show a reduction in neuronal size and increased cell packing density. There is a generalised reduction in both grey and white matter, but with regional differences. While forebrain areas are often markedly affected, the cerebellum is relatively spared. The substantia nigra is hypopigmented with reduced dendritic branching of the pyramidal neurones. It was believed that RTT was a neurodegenerative disorder, but more recently it has come to be thought of as a disorder of neuronal maturation and synaptogenesis.

Clinical Relevance

Rett syndrome is a distinctive and relatively common cause of profound intellectual disability, usually in

girls. Following the identification of the *MECP2* gene, the broad range of phenotypes associated with mutation in this gene is being elucidated. Study of the *MECP2* gene and the pathogenesis of RTT is likely to yield important insights into gene regulation and transcription repression through chromatin remodelling and other mechanisms.

► [Chromatin Acetylation](#)

► [Inherited Mental Retardation Syndromes](#)

References

1. Hagberg B, Hanefeld F, Percy A et al (2002) An update on clinically applicable diagnostic criteria in Rett Syndrome. *European Journal of Paediatric Neurology* 6:293–297
2. Amir RE, Veyer IB van de, Wan M et al (1999) Rett syndrome is caused by mutations in X-linked *MECP2*, encoding methyl-binding protein 2. *Nature Genetics* Oct 23:127–128
3. Kerr A, Engerstrom IW (eds) (2001) *Rett Disorder and the Developing Brain*. Oxford University Press, Oxford
4. Mnatzakanian GN, Lohi H, Munteanu I et al (2004) A previously unidentified *MECP2* open reading frame defines a new protein isoform relevant to Rett syndrome. *Nat Genet* 36:339–341
5. Laccone F, Junemann I, Whatley S et al (2004) Large deletions of the *MECP2* gene detected by gene dosage analysis in patients with Rett syndrome. *Hum Mutat* 23:234–244
6. Chen WG, Chang Q, Lin Y et al (2003) Derepression of BDNF transcription involves calcium-dependent phosphorylation of MeCP2. *Science* 302:885–889
7. Weaving LS, Christodoulou J, Williamson SL et al (2004) Mutations of *CDKL5* cause a severe neurodevelopmental disorder with infantile spasms and mental retardation. *Am J Hum Genet* 75:1079–93
8. Samaco RC, Hogart A, La Salle JM (2005) Epigenetic overlap in autism-spectrum neurodevelopmental disorders: *MECP2* deficiency causes reduced expression of *UBE3A* and *GABRB3*. *Hum Mol Genet* 14:483–492
9. Horike S, Cai S, Miyano M et al (2005) Loss of silent-chromatin looping and impaired imprinting of *DLX5* in Rett syndrome. *Nat Genet* 37:31–40
10. Vallender E, Pearson NM, Lahn BT (2005) The X chromosome, not just her brother's keeper. *Nat Genet* 37:343–345
11. Weaving LS, Ellaway CJ, Gecz J et al (2005) Rett syndrome: clinical review and genetic update. *J Med Genet* 42:1–7

Rev, Rex

Definition

Rev and Rex designate proteins of HIV (human immunodeficiency virus) and HTLV (Human T-cell

leukemia virus (HTLV), respectively, which interact with specific regions of the viral RNA to promote the export of incompletely spliced RNAs from the nucleus.

► [Retroviruses](#)

Rev-Erb

Definition

The orphan nuclear receptor Rev-erb-alpha is a major component of the circadian clock. While Rev-erb-alpha influences the period length and affects the phase-shifting properties of the clock, it is not required for circadian rhythm generation.

Reverse Genetics

Definition

Reverse genetics is an approach to discover the function of a gene. It begins with a cloned segment of DNA, or a protein sequence, and uses this knowledge to introduce programmed mutations (through directed mutagenesis) back into the genome in order to investigate gene and protein function. This is opposed to forward genetics, which begins with a mutant followed by a chromosomal mapping, and molecular identification of the gene by positional cloning.

► [C. Elegans as a Model Organism for Functional Genomics](#)

► [Drosophila as a Model Organism for Functional Genomics](#)

► [Morpholino Oligonucleotides, Functional Genomics by Gene 'Knock-down'](#)

► [RNA Interference in Mammalian Cells](#)

► [SLE Pathogenesis Genetic Dissection](#)

Reverse Transcriptase

► [Full Length cDNA Sequencing](#)

► [Repetitive DNA](#)

► [Telomerase](#)

► [Transposons](#)

Reverse Transcriptase, Human Immunodeficiency Virus Type-1

MARCO PETRELLA, MARK A. WAINBERG
McGill University, AIDS Centre, Lady Davis Institute
for Medical Research, Jewish General Hospital,
Montreal, QC, Canada
mark.wainberg@mcgill.ca

Definition

Reverse transcriptase (RT) is an ►enzyme that catalyzes the conversion of single-stranded RNA to double-stranded DNA. This process is known as reverse transcription or RNA-dependent DNA polymerization. RT is essential for the replication and evolution of all retroviruses that contain RNA genomes including the ►human immunodeficiency virus type-1 (HIV-1). In this essay, HIV-1 RT will be exemplified in order to discuss some important characteristics with regard to the structure and function of this enzyme and its role as a viral target for the development of antiretroviral drugs used for the treatment of patients with HIV infection.

RT is one of three enzymatic activities that is encoded by the HIV-1 ►polymerase (*pol*) gene open-reading frame. In mature viral particles, RT exists as a heterodimer that is made up of two protein subunits referred to as p51 and p66. The p51 subunit contains only the first 440 amino acid residues encoded by the RT gene. The larger p66 subunit includes the entire p51 amino acid sequence and, unlike the latter, contains all of the 560 amino acids normally found in the HIV-1 RT (1). The active site for the RNA- and DNA-dependent DNA ►polymerase catalytic activities of RT is partly contained within a tyrosine-methionine-aspartate-aspartate amino acid sequence between positions 183 and 186 in both the p51 and p66 subunits. This region of the RT enzyme is known as the YMDD motif and has been found to be highly conserved among most polymerases. However, in the p51 subunit, the active site is present in a catalytically incompetent conformation, and consequently only the p66 subunit is able to participate in the reverse transcription reaction. The carboxyl-terminal end of this subunit also contains the ribonuclease H (RNase H) domain, which is essential for retroviral DNA synthesis (1). Furthermore, phylogenetic studies have shown that the RT catalytic site is relatively well conserved among the different HIV-1 subtypes.

The three-dimensional structure of HIV-1 RT has been elucidated from X-ray crystallographic studies performed with a complex consisting of RT bound to

nevirapine, a non-competitive inhibitor of this enzyme. Interestingly, the molecular topography of RT has been described as resembling the structure of a right hand, and accordingly the major functional regions of RT are known as the fingers, palm, thumb and connection sub-domains. These sub-domains are present within the p51 and p66 subunits of RT (1).

Characteristics

The Reverse Transcription Process

Reverse transcription is a complex and highly regulated process that requires coordinated interaction between the enzyme's polymerase and RNase H activities with both RNA and DNA substrates in order to replicate the single-stranded RNA viral genome into double-stranded proviral DNA which is then integrated into host cell chromosomes. Briefly, reverse transcription is initiated with a cell-derived tRNA^{Lys3} primer, which provides a free hydroxyl group (-OH) that is required to covalently bind the incoming nucleoside triphosphate (dNTP). The primer-dNTP complex then binds to a primer-binding site (PBS) located in the genomic viral RNA strand and synthesis of the first product of the reverse transcription reaction, i.e. minus strand strong-stop DNA, then proceeds along the 5' direction. Contemporaneously, the RT-associated RNase H degrades the genomic RNA and, as a result, generates purine-rich RNA fragments that serve as primers for plus strand strong-stop DNA elongation using the minus DNA strand as a template for this second round of DNA synthesis. Both polymerization steps are also associated with strand transfer reactions that are needed in order to complete the production of a proviral pre-integrative complex which is later integrated into the host cell genome by integrase, another retroviral enzyme, which like RT, is also encoded by the HIV-1 *pol* gene (1).

Reverse Transcriptase and HIV-1 Genetic Variability

It is interesting to question whether the gains associated with the multifunctional nature of HIV-1 RT might come at the expense of some other important property such as the fidelity or error rate of this enzyme during DNA synthesis, although from an evolutionary perspective this may also be advantageous for HIV-1. A wide array of studies has confirmed the low fidelity of ►wild-type (WT) RT, which among other factors, is thought to be an important determinant for the high rate of spontaneous mutagenesis in HIV-1. Genotypic analyses of HIV-1 DNA have detected several types of mutations including base substitutions, additions and deletions. The frequency of spontaneous mutation for HIV-1 can vary considerably as a result of differences among viral strains studied *in vitro* and the specific methodology employed. In several studies, the forward

mutation rate for WT HIV-1 has been frequently estimated to be approximately 10^{-4} per nucleotide. The high level of genetic variation in HIV-1 has been largely attributed to the absence of 3'→5' exonuclease proof-reading activity in RT and the short viral generation time, although several other interdependent factors that include enzyme processivity, fitness, viral pool size and availability of target cells for new rounds of infection have also been shown to be important in this regard (2). Furthermore, as a result of the error-prone replication of HIV-1, resistance-conferring mutations to all known classes of **antiretroviral agents** (ARVs) arise spontaneously, and, in addition, are selected both *in vitro* and *in vivo* by pharmacological pressure. Therefore, it follows that the development of HIV-1 drug resistance can be considered to be both a consequence and limitation of antiretroviral therapy.

Clinical Relevance

Competitive and Non-Competitive Inhibitors of HIV-1 Reverse Transcriptase

In view of the central role that RT plays in the retroviral life cycle, this enzyme was the first major target for inhibition of HIV-1 replication to be identified and has been extensively exploited for the development of potent antiretroviral drugs for the treatment of HIV/AIDS. Two categories of antiretroviral agents (ARVs) have been developed to block RT; these are the nucleoside analog RT inhibitors (NRTIs) and the nonnucleoside RT inhibitors (NNRTIs) (3). These compounds are important components of highly active antiretroviral therapy (HAART) where their use in combination with other classes of ARVs such as the protease inhibitors (PI) has been associated with significant reductions in HIV/AIDS morbidity and mortality. Important aspects of the pharmacology and mechanism of action of both types of RT inhibitors will be discussed below.

NRTIs are pharmacologically inactive in their administered form and are phosphorylated to their active triphosphate form by host cellular kinases. These compounds are analogs of naturally occurring dNTPs and can effectively compete with these substrates for binding to RT and incorporation into proviral DNA (4). However, unlike dNTPs, NRTIs lack a 3'-hydroxyl group in their sugar moiety that is required for DNA polymerization and hence the mechanism of action underlying the antiviral activity of these compounds is based on premature termination of viral DNA strand synthesis (4). Zidovudine (ZDV, AZT), an NRTI, was the first antiretroviral drug to be licensed for the treatment of patients with HIV-1 infection and presently, including ZDV, there are seven NRTIs i.e., didanosine (ddI), dideoxycytidine (ddC), stavudine

(d4T), lamivudine (3TC), abacavir (ABC) and emtricitabine (FTC) that are commercially available for use in antiretroviral combination therapy regimens (3). Recently, tenofovir (TDF, PMPA), a novel antiviral drug from a related class of compounds known as nucleotide analog reverse transcriptase inhibitors (NtRTIs), has also become available for the treatment of HIV/AIDS. As is the case with NRTIs, TDF also functions as a chain terminator of DNA strand elongation. However, this new drug has a reduced requirement for metabolic activation because of the presence of a phosphonate group. This in turn is believed to impart certain advantages to TDF such as heightened antiviral potency and longer intracellular half-life compared to most existing NRTIs (4).

NNRTIs are non-competitive inhibitors of HIV-1 RT. The basis for their antiviral activity is not completely understood but is known to involve the binding of these compounds to a hydrophobic pocket (i.e. NNRTI binding pocket) that is located adjacent to the active site of RT in the p66 subunit. The NNRTI binding pocket also contains a few amino acid residues from the RT p51 subunit. It is thought that the inhibition of HIV-1 replication by NNRTIs may be mediated through allosteric effects that cause the RT active site to assume a configuration that is incompatible with DNA polymerization. NNRTIs are particularly active at template positions at which the RT enzyme naturally pauses. There are presently three NNRTIs, i.e. nevirapine (NVP), delavirdine (DLV) and efavirenz (EFV) that have been approved for use in antiretroviral combination therapy regimens (3).

Resistance-conferring **mutations** in RT can precede the initiation of antiretroviral therapy because of spontaneous mutagenesis and/or sexual and other forms of transmission. In addition, resistance-conferring **mutations** in RT are selected in patients following treatment with NRTI- or NNRTI-containing regimens. These mutations encode amino acid substitutions in RT at codons that are responsible for reduced antiviral susceptibility to these inhibitors and, in general, are also predictive of worse virological responses to antiretroviral therapy (5, 6). Drug resistance to NRTI and NNRTIs has been shown to be mediated by different mechanisms. For example, some resistance mutations in RT are associated with increased drug discrimination, a situation where the binding and/or incorporation of antiviral drugs by the mutated RT is significantly attenuated (5, 6). Other RT mutations known as **thymidine analog mutations** (TAMs) mediate resistance to ZDV and d4T as a result of increased levels of nucleotide primer uncoupling. With this mechanism, ZDV- or d4T-monophosphate is excised from the terminus of blocked primer templates by RT harboring TAMs, thus permitting DNA polymerization and HIV-1 replication to continue in the

Reverse Transcriptase, Human Immunodeficiency Virus Type-1. Table 1 Mutations in HIV-1 reverse transcriptase (RT) that are important for resistance to Nucleoside Analog RT Inhibitors (NRTIs) and Tenofovir (TDF)

Drug	Important NRTI/NtRTI Resistance Mutations
Zidovudine (ZDV)	M41L; E44A/D; D67N; K70R; V118I; L210W; T215Y/F; K219Q/E
Didanosine (ddI)	K65R; L74V
Dideoxycytidine (ddC)	K65R; T69D; L74V; M184V
Lamivudine (3TC)	K65R; M184I/V
Abacavir (ABC)	K65R; L74V; Y115F; M184V
Stavudine (d4T)	M41L; E44A/D; D67N; K70R; V118I; L210W; T215Y/F; K219Q/E
* Tenofovir (TDF)	K65R

* TDF is a nucleotide analog inhibitor of reverse transcriptase (NtRTI).

Source: Ref. 6.

Reverse Transcriptase, Human Immunodeficiency Virus Type-1. Table 2 Mutations in HIV-1 reverse transcriptase (RT) that are important for resistance to Nonnucleoside RT Inhibitors (NNRTIs)

Drug	Important NNRTI Resistance Mutations
Nevirapine (NVP)	L100I; K103N; V106A/M; V108I; Y181C/I; Y188C/L/H; G190A
Delavirdine (DLV)	K103N; V106M; Y181C; Y188L; P236L
Efavirenz (EFV)	L100I; K103N; V106M; V108I; Y181C/I; Y188L; G190S/A; P225H

Source: Ref. 6.

presence of these antiviral drugs (1, 5, 6). Tables 1 and 2 illustrate important mutations in RT that are known to confer resistance to NRTIs/NtRTIs and NNRTIs, respectively.

Resistance Mutations in Reverse Transcriptase and HIV-1 Fitness

It has been demonstrated that HIV-1 variants that harbor drug resistance mutations in RT display diminished fitness or replication competence compared to drug-sensitive viruses (7). Without antiretroviral pressure, WT virus is the fittest and most prevalent HIV-1 [▶quasispecies](#) in the replicating viral pool. The eventual outgrowth of drug-resistant variants arises due to their fitness advantage under conditions of selective drug pressure. Therefore, both the mutation rate and selection forces (i.e. drug pressure and/or immunological selection) are important factors for the evolution of HIV-1 quasispecies (2, 7). It has been shown that HIV-1 variants containing the M184V mutation in RT, which confers high-level phenotypic resistance to 3TC (i.e. 500–1000 fold increase in IC_{50}), demonstrate marked impairment of replication capacity relative to

WT viruses (8). In addition, M184V-mutated viruses also appear to have a lower propensity for further mutational accumulation and in addition exhibit several other alterations in RT function that may be associated with improved therapeutic outcomes (8). Interestingly, this situation with M184V may represent a unique case in which it may be advantageous to preserve a drug resistance mutation in therapeutic regimens; further research on this topic is required.

References

1. Gotte M, Wainberg MA (2000) Biochemical mechanisms involved in overcoming HIV resistance to nucleoside inhibitors of reverse transcriptase. *Drug Resist Updat* 3:30–38
2. Coffin JM (1995) HIV population dynamics *in vivo*: implication for genetic variation, pathogenesis, and therapy. *Science* 267:483–489
3. Yeni PG, Hammer SM, Carpenter CC et al (2002) Antiretroviral treatment for adult HIV infection in 2002: updated recommendations of the International AIDS Society-USA Panel. *JAMA* 288:222–235
4. Squires KE (2001) An introduction to nucleoside and nucleotide analogues. *Antivir Ther* 6:1–14

5. Miller V, Larder BA (2001) Mutational patterns in the HIV genome and cross-resistance following nucleoside and nucleotide analogue drug exposure. *Antivir Ther* 6:25–44
6. Johnson VA, Brun-Vezinet F, Clotet B et al (2005) Update of the Drug Resistance Mutations in HIV-1: 2005. *Top HIV Med* 13:51–57
7. Quiñones-Mateu ME, Weber J, Rangel HR et al (2001) HIV-1 Fitness and antiretroviral drug resistance. *AIDS Rev* 3:223–242
8. Petrella M, Wainberg MA (2002) Might the M184V substitution in HIV-1 RT confer clinical benefit? *AIDS Rev* 4:224–232

Reverse Transcription

Definition

Reverse transcription refers to the synthesis of a strand of DNA from RNA, which is used to make a complementary DNA copy of sample RNA. In retroviruses, reverse transcription refers to the process of copying the RNA genome into double-stranded DNA by the retrovirus enzyme transcriptase (a product of the ►[pol](#) gene).

- [DNA Chips](#)
- [Recombinant Protein Expression in Bacteria](#)
- [Retroviruses](#)
- [Reverse Transcriptase](#)
- [RT-PCR](#)
- [tRNA](#)

Reverse Transcription Polymerase Chain Reaction

Definition

mRNA from cell or tissue samples is reverse transcribed using oligo-dT or random primers, and the resulting complementary DNA is amplified by the polymerase chain reaction using gene-specific primers.

- [Marfan Syndrome](#)
- [Microarrays in Colorectal Cancer](#)

Reverse Transfection

Definition

Reverse transfection designates introduction of foreign nucleic acids by coating a solid surface with foreign

siRNA or DNA that has been treated with transfection reagent, and adding cells on top of the coated surface resulting in the confined uptake and expression of the genetic material by these cells.

- [High-Throughput Approaches to the Analysis of Gene Function in Mammalian Cells](#)

Reward System

Definition

One of the major targets of all drugs of abuse is a brain circuit called the reward system. The activation of this neuronal assembly reinforces behaviours. Responses of midbrain dopaminergic neurons represent learning signals. Rather than signaling pleasure, dopamine highlights significant stimuli and their motivational significance, suggesting that drug addiction might be a dopamine-dependent associative learning disorder.

- [Addiction, Molecular Biology](#)

Rex

- [Rev, Rex](#)

RFLP

- [Restriction Fragment Length Polymorphism](#)

RGD Motif

Definition

RGD motif refers to an arginine-glycine-aspartate tripeptide sequence, which is present in several ligands and directly bind integrin head domains.

- [Integrin Signalling](#)

RGEF

- Ras Guanyl Nucleotide Exchange Factor

RGR

Definition

RGR (retinal G protein-coupled receptor) refers to a G-protein coupled receptor in the retinal pigment epithelium, which is a protein with a potential role in photoisomerisation of all-trans-retinal to 11-cis-retinal.

- Photoreceptors

RGS

- Regulators of G Protein Signaling

Rhabdomyolysis

Definition

Rhabdomyolysis is the dissolution or disintegration of skeletal muscle fibers resulting in the release of muscle fiber contents into the circulation. Some of these are toxic to the kidney and frequently result in kidney damage.

- Mitochondrial Myopathies

Rhabdomyoma

Definition

Rhabdomyoma is a benign, extremely rare tumor derived from striated muscle. It consists of glycogen-filled myocytes called spider cells. The tumor occurs in the tongue, neck muscles, larynx, uvula, nasal cavity,

axilla, vulva, and the heart; commonly seen in tuberous sclerosis.

- Tuberous Sclerosis

Rhabdomyosarcoma

Definition

Rhabdomyosarcoma designates a highly malignant soft tissue tumour that develops from muscle in a number of regions of the body including the head and neck, testes, bladder and limbs.

- Hedgehog Signalling

Rheumatic Diseases

- Microarrays in Rheumatoid Diseases
- Rheumatoid Arthritis

Rheumatoid Arthritis

PETER C. TAYLOR, MARC FELDMANN
The Kennedy Institute of Rheumatology Division,
Faculty of Medicine, Imperial College London,
London, UK
peter.c.taylor@imperial.ac.uk,
m.feldmann@imperial.ac.uk

Definition

Rheumatoid arthritis (RA) is a chronic autoimmune inflammatory disease of unknown aetiology. However, current thinking favours the hypothesis that an interplay between genetic factors, sex hormones and possibly an infectious agent or another immune activating agent initiates an autoimmune pathogenic mechanism that culminates in a disease with inflammatory and destructive features.

To date there is no diagnostic test specific for rheumatoid arthritis. Revised classification criteria for the diagnosis of RA were formulated by the American College of Rheumatology (1) based on a hospital population of patients with established, active disease.

These criteria combine a constellation of clinical, serological and radiological features and have become widely accepted for epidemiological and clinical studies. By emphasising key features of the disease, the criteria help to differentiate RA from other forms of inflammatory arthritis with a diagnostic sensitivity and specificity of about 90% for active disease. However, these requirements have a much poorer sensitivity for a diagnosis of RA in the early stages of presentation at which time there may be an insufficient number of the features required to satisfy the classification criteria.

Characteristics

Once established, RA is characterised by deforming symmetrical polyarthritis of varying extent and severity, associated with synovitis of joint and tendon sheaths, articular cartilage loss, erosion of juxta-articular bone and, in a majority of patients, the presence of IgM ►**rheumatoid factor** in blood. In a proportion of patients systemic and extra-articular features may be observed during the course of the disease (and rarely prior to joint disease). These include anaemia, weight loss, vasculitis, serositis, nodules in subcutaneous, pulmonary and sclera tissues, mono-neuritis multiplex and interstitial inflammation in lungs as well as in exocrine salivary and lacrimal tissue. However, these systemic manifestations occur relatively late in the disease progression.

The clinical presentation of RA is heterogeneous with a wide spectrum of age of onset, degree of joint involvement and severity. Similarly, the disease course of RA is variable. This ranges from a brief, mild, self-limiting ►**oligoarticular** illness with minimal joint damage to a sustained ►**polyarticular**, synovial inflammation resulting in relentlessly progressive cartilage destruction, erosion of bone and ultimately changes in joint integrity with corresponding functional impairment. Up to 90% of patients with aggressive synovitis have radiological evidence of bone erosion within 2 years of diagnosis, despite treatment with traditional, so-called disease modifying agents (such as methotrexate, sulphasalazine and gold). The majority of patients with a more aggressive disease evolution become clinically disabled within 20 years. For those with severe disease or extra-articular features the mortality is equivalent to that of patients with three vessel coronary artery disease or stage IV Hodgkin's lymphoma.

Cellular and Molecular Regulation

Cellular Pathology

RA is characterised by chronic inflammation of synovial joints with synovial proliferation and infiltration by blood derived cells, in particular, memory T cells, macrophages and plasma cells, all of which show

signs of activation. Luxuriant vasculature is a prominent feature of RA synovitis, observed as a fine network of vessels over the rheumatoid synovium at arthroscopic inspection of RA joints. ►**Angiogenesis** is evident on microscopic examination of synovial biopsies from the earliest stages of disease development. Formation of new blood vessels permits a supply of nutrients and oxygen to the augmented inflammatory cell mass and so contributes to the perpetuation of synovitis. In the chronic phase of disease, capillaries and post capillary venules are particularly evident in the synovial sublining region. In histological sections mononuclear and polymorphonuclear leukocytes can sometimes be found in close apposition to vascular endothelium, probably in the process of margination and adhesion prior to migration into the inflamed tissue. The synovial tissue becomes markedly hyperplastic and locally invasive at the interface of cartilage and bone with progressive destruction of these tissues in the majority of cases. This invasive tissue is referred to as '►**pannus**', comprising mainly lining cells with the appearance of proliferating mesenchymal cells with very little sublining lymphocytic infiltration. The accompanying destruction of bone and cartilage is thought to be mediated principally by ►**cytokine**-induced degradative enzymes, notably the matrix metalloproteinases. In most cases the major manifestation of rheumatoid disease is in the synovial joint, but there is also evidence of systemic involvement, for example the up-regulation of acute phase proteins and, in severe cases, involvement of other organs.

Human Leucocyte Antigens

Genetic factors were originally implicated in the aetiopathogenesis of RA following the discovery that in population studies there is a slight increase in the frequency of RA in first-degree relatives of patients with this disease, especially if they are ►**seropositive** for rheumatoid factor. In hospital-based population studies of identical twins concordance rates of disease are around 30%, compared with 5% in non-identical twins. The figures are lower in community-based studies and although still supportive of the concept of a genetic contribution, argue against the proposition that RA is the result of a dominant single-gene disorder. These and other epidemiological studies have led to the conclusion that RA is a polygenic disease and that non-inherited factors are also of great importance (2).

Genes encoding particular class II human leucocyte antigens (►**HLA**) are among candidates for involvement in predisposition to RA. This discovery came about with the observation that 60–70% of Caucasian patients with RA are HLA DR4 positive by cellular or serological techniques compared with 20–25% of control populations. Furthermore, patients with more severe RA, especially those with extra-articular

complications such as vasculitis and Felty's syndrome, are even more likely to be HLA DR4 positive than patients with less aggressive disease confined to joints. The immunological significance of HLA class II molecules, expressed on the surface of antigen presenting cells, lies in their key role in presentation of processed linear peptide antigens of at least nine amino acids to T cells. Antigen is bound to the HLA antigen-binding cleft formed by the α and β chains of the HLA class II molecule. This trimolecular HLA-antigen complex binds in turn to the variable portion of the T-cell receptor.

Several laboratory techniques have been used in an attempt to define the amino acid sequences of HLA class II molecules common to all RA patients, including those expressing class II molecules other than HLA DR4. These include at the genetic level, for example, genotyping of DNA and nucleotide sequencing, use of the polymerase chain reaction and enzymatic digestion for restriction fragment length polymorphisms. At the level of expressed cell surface proteins, HLA epitopes have been sought by using monoclonal antibodies and alloreactive T-cell clones. Overall, these studies support the concept that susceptibility to RA is related to expression of a 'shared epitope' on the HLA molecules.

Nucleotide sequencing of HLA DR β , exons coding amino acid residues 70 to 74 has revealed that HLA DR4 subtypes Dw4, Dw14, and Dw15 share similarities with each other (with a conservative substitution of glutamine for lysine at position 71 in Dw4) and with HLA DR1. The sequence predicts susceptibility to RA and is associated with RA in 83% of Caucasian patients in the United Kingdom. In contrast, negative associations are observed in individuals who are DR4w10, in whom the charged basic amino acids glutamine and arginine in positions 70 and 71 are replaced by the acidic amino acids aspartic and glutamic acid. In Dw13 individuals, in whom a negative association is also observed, arginine is substituted for glutamic acid in position 74. Molecular modelling studies suggest that amino acid residues 70–74 are located in the α -helix forming the wall of the peptide-binding groove and thus likely to be involved in antigen binding and subsequent interaction with T-cell receptors. Acidic substitutions could profoundly alter protein structures and thereby alter affinity for peptide antigens. Serotyping with alloantisera and monoclonal antibodies as well as reactivity with homozygous T cells and T-cell clones support the prediction that protein structures on the HLA molecule are important in susceptibility to RA. However, molecular mechanisms accounting for susceptibility to RA remain to be elucidated. Possibilities include permissive binding of specific peptides such as those on autoantigens or on environmental antigens, initiation of disease by specific binding of

superantigens to HLA molecules or modulation of the T-cell repertoire by selection or tolerance.

However, the evidence that Dw4/Dw14 heterozygotes are more likely to develop RA than are Dw4 homozygotes and the evidence that individuals expressing DR4 (and not DR1) are more likely to have severe disease, challenges the concept that sequences 70–74 are the only HLA-D regions that influence disease expression. This is despite the sequence identity of DR1 and DR4w4 in these positions, and a conservative substitution in Dw14. It has been hypothesized that the severity of disease and extra-articular complications are related to homozygosity and the density of disease-associated MHC molecules that critically influence the selection of the T-cell repertoire and tolerance to antigens.

Other Candidate Genes

HLA genes are estimated to account for one-third of the genetic component of RA. To identify susceptibility genes outside the HLA region, studies have been undertaken using genome-wide microsatellite analyses in affected sibling pair families (3). Significant levels of linkage are reported at chromosomes 1, 4, 12, 16 and 17. In seropositive patients there is evidence of linkage with a locus on chromosome 12 at a level of significance comparable to the HLA locus. This marker lies within 10 cM of the tumour necrosis factor receptor 1 (*tnfrsf1*) locus. Differences in the findings of linkage analyses between study cohorts may reflect ethnic bias and genetic heterogeneity. However, common linkages at 16q24.1 are reported in American and European populations and at 17q22 in American and UK populations. Candidate genes mapping to the 17q22 locus include intracellular adhesion molecule-2 (ICAM-2) and platelet endothelial cell adhesion molecule-1 (PECAM-1).

Cytokines

Cytokines are small, short-lived proteins and important mediators of local intercellular communication. They play a key role in integrating responses to a variety of stimuli in immune and inflammatory processes. By binding their cognate receptors on target cells in their immediate vicinity, these molecules participate in many important biological activities including cell proliferation, activation, death and differentiation. In experimental systems, some cytokines are proinflammatory, such as ▶interleukin-1 (IL-1) and TNF α ; others, such as interleukin-10 (IL-10) and transforming growth factor β (TGF β) exert predominantly anti-inflammatory effects. However, it is now known that many cytokines, for example, interferon γ (IFN- γ), with chiefly proinflammatory activity can also in some instances have anti-inflammatory properties. Similarly, IL-10 and TGF β may also exhibit proinflammatory

properties under certain experimental conditions. Interleukin-6 (IL-6) does not have a classical pro-inflammatory action but has been implicated in the process of erosion of bone in inflammatory arthritis. ►Paracrine or ►autocrine pathways involving cytokines with either pro- or anti-inflammatory activity form complex networks determining whether chronic inflammation results.

Role of Cytokines in the Pathogenesis of RA

Cytokines derived from macrophages and fibroblasts are abundant in the rheumatoid synovium. These include IL-1, TNF α , granulocyte-macrophage colony stimulating factor (GM-CSF), IL-6 and numerous chemo-attractant cytokines known as chemokines (4). Many of these factors are of importance in regulating inflammatory cell migration and activation. By contrast, given the extent of synovial inflammation and lymphocytic infiltration, factors produced by T cells, for example, IFN γ , interleukin-2 (IL-2) and interleukin-4 (IL-4) are surprisingly sparsely expressed. However, there are a number of cytokines that cause costimulation of T helper cells including interleukin (IL)-7, IL-12, IL-15 and IL-18. There is a predominance of ►Th1 cell activity, as defined by IFN γ production and low ►Th2 cell activity as defined by IL-4 production.

Tissue expression of an extensive range of pro-inflammatory cytokines in human synovial samples, regardless of differences in donor disease duration, severity or even drug therapy, has been confirmed in studies from a number of laboratories. These observations imply that there is prolonged expression of cytokines in the rheumatoid joint, contrasting with the transient production induced by mitogenic stimulation. This hypothesis was confirmed when it was observed that proinflammatory cytokines are produced over several days in dissociated RA synovial membrane cell cultures in the absence of extrinsic stimulation, a finding suggesting that one or more soluble factors regulating prolonged cytokine synthesis were present within the RA synovial membrane cell cultures. These cultures comprise a heterogeneous population of cells producing numerous cytokine and other non-cytokine molecular messengers. A key observation to emerge employing these RA dissociated synovial cell culture systems was that addition of anti-TNF antibodies strikingly reduced the production of other pro-inflammatory cytokines, including IL-1, GM-CSF, IL-6 and IL-8. Furthermore, using the same RA synovial cell culture system, blockade of IL-1 by means of the IL-1 receptor antagonist results in reduced IL-6 and IL-8 production but not that of TNF α . These observations led to the concept that TNF α occupies a dominant position at the apex of a pro-inflammatory cytokine network. At the same time, in the early 1990's TNF α was known to be a pleiotropic cytokine with biological

properties including enhanced synovial proliferation and production of prostaglandins and metalloproteinases as well as regulation of other proinflammatory cytokines.

In support of the concept of a cytokine disequilibrium within the chronic inflammatory situation in rheumatoid synovium is the observation that multiple anti-inflammatory mediators are also upregulated, but at a level insufficient to suppress synovitis. Examples include the abundant expression of IL-10, IL-13 and TGF β both in latent and active form. Naturally occurring cytokine inhibitors, such as interleukin-1 receptor antagonists (IL-1ra) and soluble TNF receptors, the specific inhibitors of IL-1 and TNF α respectively, are also upregulated in the rheumatoid joint.

The predicted clinical success of anti-TNF therapy in RA was based on the demonstration of RA synovial tissue expression of TNF α and its receptors, in *in vitro* experiments employing dissociated synovial cell cultures and pre-clinical *in vivo* studies (5). A number of independent *in vivo* studies demonstrated that antibody therapies blocking bioactivity of TNF α , administered either during the induction phase of murine collagen-induced arthritis or, more importantly, after the onset of disease, were able to ameliorate clinical symptoms and prevent joint destruction. Furthermore, in a murine model, the over-expression of a human TNF α transgene modified at its 3' end to prevent degradation of its mRNA, was associated with the development of a destructive form of polyarthritis four to six weeks after birth. This could be prevented by administration of a human TNF-specific mAb.

Clinical Relevance

The treatment approach to rheumatoid arthritis has undergone a major evolutionary change in recent years. In part this is a consequence of a growing appreciation of the severity of this condition. For example, there is now evidence that treatment intervention with disease-modifying combination therapy early in the course of RA results in improved remission rates and sustained attenuation of the rate of radiographic progression and that this benefit is independent of subsequent anti-rheumatic therapy. However, the major impetus for development of new treatment paradigms has come from advances in molecular technology that have facilitated identification of cell subsets and cytokines contributing to the inflammatory and destructive components of the disease. These new insights have prompted development of specific targeted therapies directed at relevant disease molecules (6).

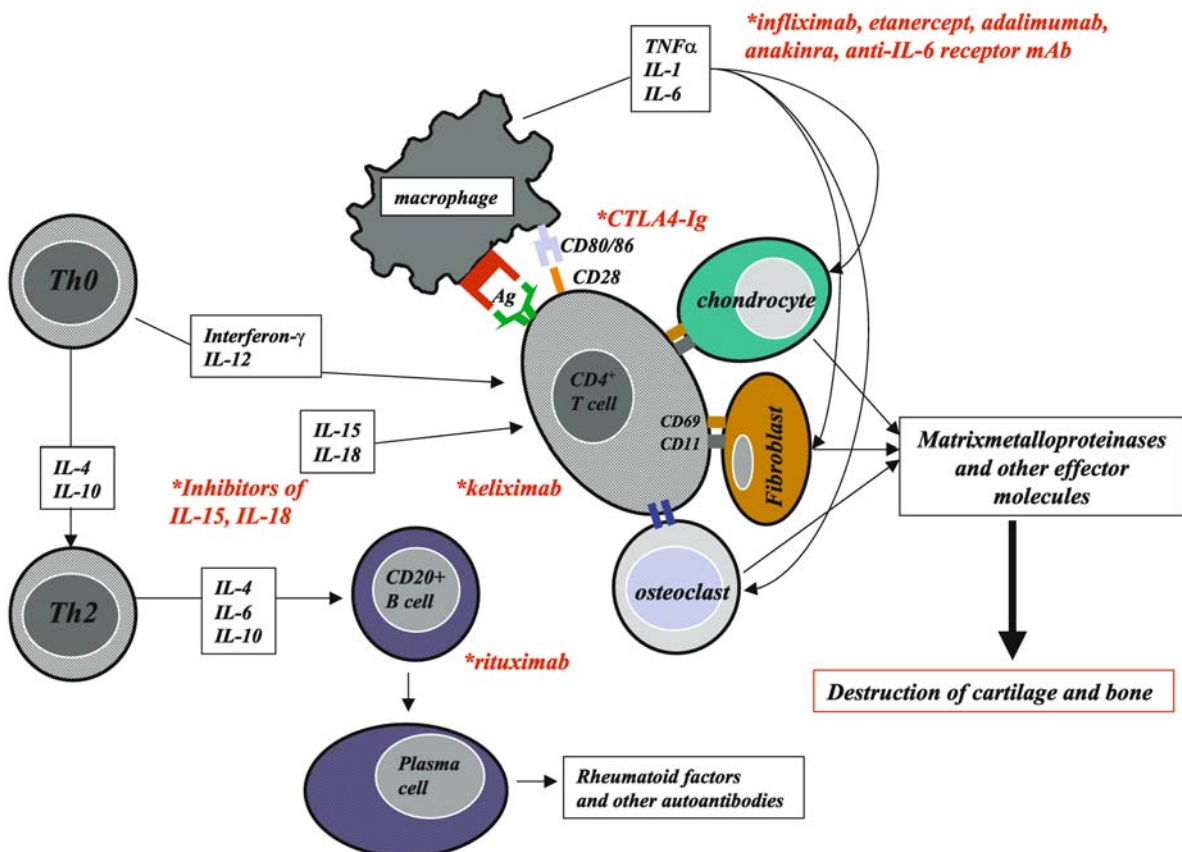
Very considerable progress in understanding the important role of cytokines in the immunopathogenesis of RA has led to two potential approaches to cytokine

modulation of rheumatoid synovitis; inhibition of dominant proinflammatory cytokines such as TNF α or IL-1 or augmenting the inadequate anti-inflammatory activity of certain cytokines or naturally occurring cytokine inhibitors as, for example, by administration of soluble TNF receptors or IL-1ra. It is now established that the long-term use of several biological agents targeting TNF α gives rise to sustained improvements in symptoms and signs of rheumatoid disease. It has recently been confirmed that anti-TNF therapy protects joints from structural damage as judged by a change in the rate of deterioration in radiographs of the hands and feet, assessed by a scoring system that assesses cartilage and bone loss separately; these two components are summed to give a total damage score. One year of combined therapy with methotrexate and the anti-TNF α antibody infliximab prevented the progressive joint damage characteristic of rheumatoid inflammation with zero median progression in joint damage over this time period. Furthermore, there was actually a reduction in the joint damage score in

40–55% of patients irrespective of whether a clinical response was achieved or not. In contrast, joint space narrowing and erosions progressed as anticipated in recipients of methotrexate alone.

Although anti-TNF α agents are well tolerated and have a good overall safety profile, pitfalls in the use of these drugs apparent with increasing clinical experience include infective complications and in particular, reactivation of tuberculosis. To date, no statistically significant increased rate of tumour occurrence over that expected has been noted although cases of lymphoma have been rarely reported in patients treated with TNF α blockade.

Clinical trials of interleukin-1 receptor antagonist show relatively modest anti-inflammatory efficacy but radiographic evidence indicative of retardation of joint damage. Several other pro-inflammatory cytokines represent potential therapeutic targets including interferon- β , interferon- γ , IL-6, IL-15, IL-17 and IL-18 and many biological interventions targeting these molecules are in clinical trials. For example, preliminary



Rheumatoid Arthritis. Figure 1 Cartoon illustrating the major cell types and cytokine pathways thought to be involved in the destruction of bone and cartilage, mediated by IL-1 and TNF α , within the rheumatoid joint. The various therapies discussed in the text target molecules marked by *.

data for blockade of the biological effect of IL-6 by means of intravenously administered antibodies to the IL-6 receptor demonstrate efficacy for reduction in symptoms and signs with acceptable safety.

In contrast to the dramatic success of TNF blockade as a new approach to RA therapy, early randomised, placebo-controlled clinical studies exploring the potential of biological therapies targeting T cells in the treatment of RA have had generally disappointing results. Some anti-T cell agents were non-efficacious whereas other preliminary trials demonstrating some clinical efficacy were terminated due to adverse events, particularly prolonged and profound T cell depletion. However, the primatized monoclonal anti-CD4 antibody keliximab results in dose-dependent clinical responses when administered once weekly over 4 consecutive weeks and the clinical response correlates with CD4⁺ T cell coating with keliximab rather than T cell depletion.

An alternative approach seeks not to deplete or inactivate T cells, but to modulate their function in such a way as to reduce their pathogenicity. For example, CTLA4-Ig, which blocks the interaction between the B7 receptor on antigen presenting cells and CD28 on T cells, has shown efficacy in the established phase of murine collagen-induced arthritis and clinical trials are underway in patients with rheumatoid arthritis. Preliminary reports indicate that in RA patients with an inadequate response to methotrexate, addition of CTLA4-Ig results in significant dose dependent clinical responses. Similarly, in patients with an inadequate response to the anti-TNF agent etanercept, addition of CTLA4-Ig therapy resulted in significant improvements in clinical and health related quality of life measures. It remains to be seen whether any of these biological agents targeting T cells have disease-modifying capability.

The potential of B lymphocyte depletion as an approach to therapy is under investigation using an anti-CD20 mAb, rituximab with encouraging preliminary data. The clinical benefit following antibody administration and time to disease relapse may be related to circulating antibodies, an interesting observation as immune complexes may be a regulatory factor in the production of TNF α .

In summary, recent years have seen the emergence of encouraging data for antibodies in clinical development that are directed against pro-inflammatory cytokines, inflammatory cells or accessory molecules. In particular, anti-TNF α therapies have set a new standard for symptom control in RA. Furthermore, by significantly retarding the rate of structural damage to joints, it is envisaged that long-term administration of biologics targeting TNF α in RA patients will preserve physical function.

► [Protein Interaction-Phage Display](#)

► [Rheumatism Related Genes, Identification](#)

References

1. Arnett FC et al (1988) The American Rheumatism Association. 1987 revised criteria for classification of rheumatoid arthritis. *Arthritis Rheum* 31:315–324
2. Maini RN (2003) Rheumatoid Arthritis. In: Oxford textbook of Medicine. 4th edition. Oxford University Press. Oxford. Volume 4:27–42
3. Abrahams S, Cope A, Taylor PC (2003) Assessment of disease activity in rheumatoid arthritis. In: Scott, Cope (eds) The year in Rheumatic disorders. Clinical publishing services, pp 103–126
4. Feldmann M et al (1996) Role of cytokines in rheumatoid arthritis. *Ann Rev Immunol* 14:397–440
5. Feldmann M, Maini RN (2001) Anti-TNF α therapy of rheumatoid arthritis: What have we learned? *Ann Rev Immunol* 19:163–196
6. Taylor PC (2003) Antibody therapy for rheumatoid arthritis. *Current Opin Pharmacol* 3:323–328

Rheumatoid Factors

Definition

Rheumatoid factors (RF) are autoantibodies that recognize the Fc portion of an IgG molecule as their antigens. Many conditions associated with chronic inflammation are associated with positive RF. In patients with rheumatoid arthritis, RF is an important predictor of a poor prognosis.

► [Rheumatoid Arthritis](#)

Rheumatism Related Genes, Identification

LARS C. HUBER, OLIVER DISTLER

Center of Experimental Rheumatology, University Hospital Zurich, Zurich, Switzerland
oliver.distler@usz.ch

Definition

► [Rheumatoid arthritis](#) (RA) is a chronic inflammatory disorder of mainly unknown etiology. It predominantly affects the joints and leads to their progressive destruction. Considering that the prevalence of RA in adults is estimated to be as high as 0.5–1% and that

work disability affects as much as half of the patients 10 years after onset of the disease, both the costs for public health systems and the social consequences for the individual patient are enormous.

The current model for RA pathogenesis proposes interaction between chronic inflammation, altered immune responses and synovial hyperplasia (previously called pannus). Synovial fibroblasts release large amounts of matrix-degrading enzymes and are thereby cellular key players in cartilage and bone destruction. The complex pathophysiology of joint destruction is still only partially understood, which limits the generation of specific therapies. Thus, the exponential development of new techniques in the field of molecular biology during the past decade has been a further milestone in our approach to understanding and treating RA. In particular, screening approaches such as gene arrays enable the fast and precise identification of genes that are differentially expressed in rheumatoid arthritis compared to the normal synovium and that might be of functional importance for the progression and perpetuation of the disease.

Characteristics

► Functional Genomics

With the accomplishment of the ►human genome project, researchers are challenged by identified sequences of approximately 30000 human genes, most of them coding for biological functions that remain so far undiscovered. The functional characterization of these genes in normal physiology as well as in the pathogenesis of diseases remains the main issue for biomedical research in the next few years.

The term “functional genomics” is defined as the use of molecular screening methods for the systematic analysis of the expression of a large number of genes. Since screening methods in general tend to generate a high number of false-positives, their results need to be confirmed by independent methods such as ►real-time polymerase chain reaction (real-time PCR) or Northern blot. The detection and quantification of differentially expressed genes in certain experimental conditions leads to the generation of a hypothesis as to their possible role in the pathogenesis of diseases. The postulated functions then have to be verified independently *in vitro* as well as *in vivo* on a functional level by using appropriate disease models. The typical procedure from the initial screening steps through the verification to the functional characterization is outlined in Fig. 1.

It is unquestionable that molecular screening procedures have the potential to identify key pathogenic pathways and, more generally, to lead to an exponential acceleration of biomedical research. Conversely, the highly sophisticated techniques and the specific

knowledge required from various scientific fields such as molecular biology, biophysics, medicine and bioinformatics carry the danger of exploitation and misinterpretation of data.

Screening Methods

Generally, a large number of molecular-biology-based screening methods have been developed including ►microarrays, subtractive ►hybridization, serial analysis of gene expression (SAGE) and RNA arbitrarily primed polymerase chain reaction (RAP-PCR). Among them, ►suppressive subtractive hybridization (SSH) and, in particular, microarrays are most often used to identify genes related to the pathogenesis of rheumatoid arthritis. Both methods are screening tools that are very useful for defining the differences in gene expression between any two related types of cells, or, more specifically, to distinguish between the expression profiles of cells derived from the rheumatoid synovium and the pattern of their healthy counterparts.

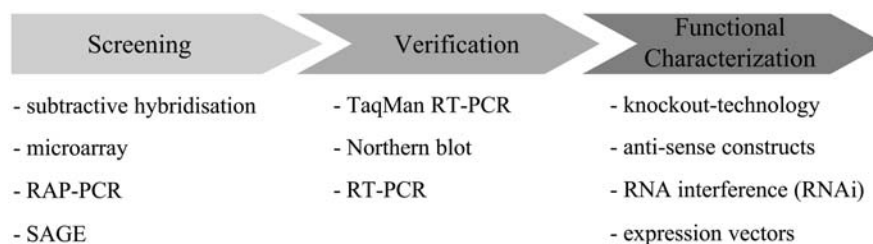
From mRNA to cDNA

The preparation steps are similarly performed in both screening methods. They consist of the extraction of messenger RNA (mRNA) from the cells or tissue of interest and the subsequent production of ►complementary DNA (cDNA).

In cell physiology, mRNA is transcribed by distinct enzymes from DNA and gets translated into functional proteins. The cytoplasmic levels of mRNA reflect the genes that are currently expressed. Normally, the proportion of expressed genes in a particular cell does not exceed 10–15%. After being transcribed, the pre-mRNA gets its final structure by the addition of a polyadenylate tail. This post-transcriptional modification converts the mRNA into its mature form, which is also referred to as Poly(A)⁺-RNA.

In a first step of the screening, total mRNA or Poly(A)⁺-RNA is extracted from the cells. The quality of the isolated RNA is critical for a successful experiment. The isolation of undamaged RNA from tissues may be especially challenging. The extracted RNA should be tested for integrity, e.g. by gel-electrophoresis, before continuing with the experiments. After isolation of mRNA molecules, cDNA copies are produced by the enzyme reverse transcriptase.

Human genes usually consist of short coding sequences (exons) and longer noncoding sequences (introns). The production of mRNA entails the removal of all introns in a process called splicing. Thus, cDNA generated from mRNA by reverse transcription contains only those regions of the genome that have been transcribed into mRNA. This is a major difference from genomic DNA, which mirrors a random sample of all DNA sequences in an organism and is indistinguishable



Rheumatism Related Genes, Identification. Figure 1 Classical procedure in molecular screening. Identification of differentially expressed genes (screening), followed by verification with independent methods and characterization of identified key molecules on a functional level (adapted and modified from Distler, *Molekularmedizinische Grundlagen von rheumatischen Erkrankungen* 2003).

between different cells. As the cells of different types of tissues produce distinct sets of mRNA, different cDNAs are obtained and can be used in further steps of the screening.

Suppressive Subtractive Hybridization (SSH)

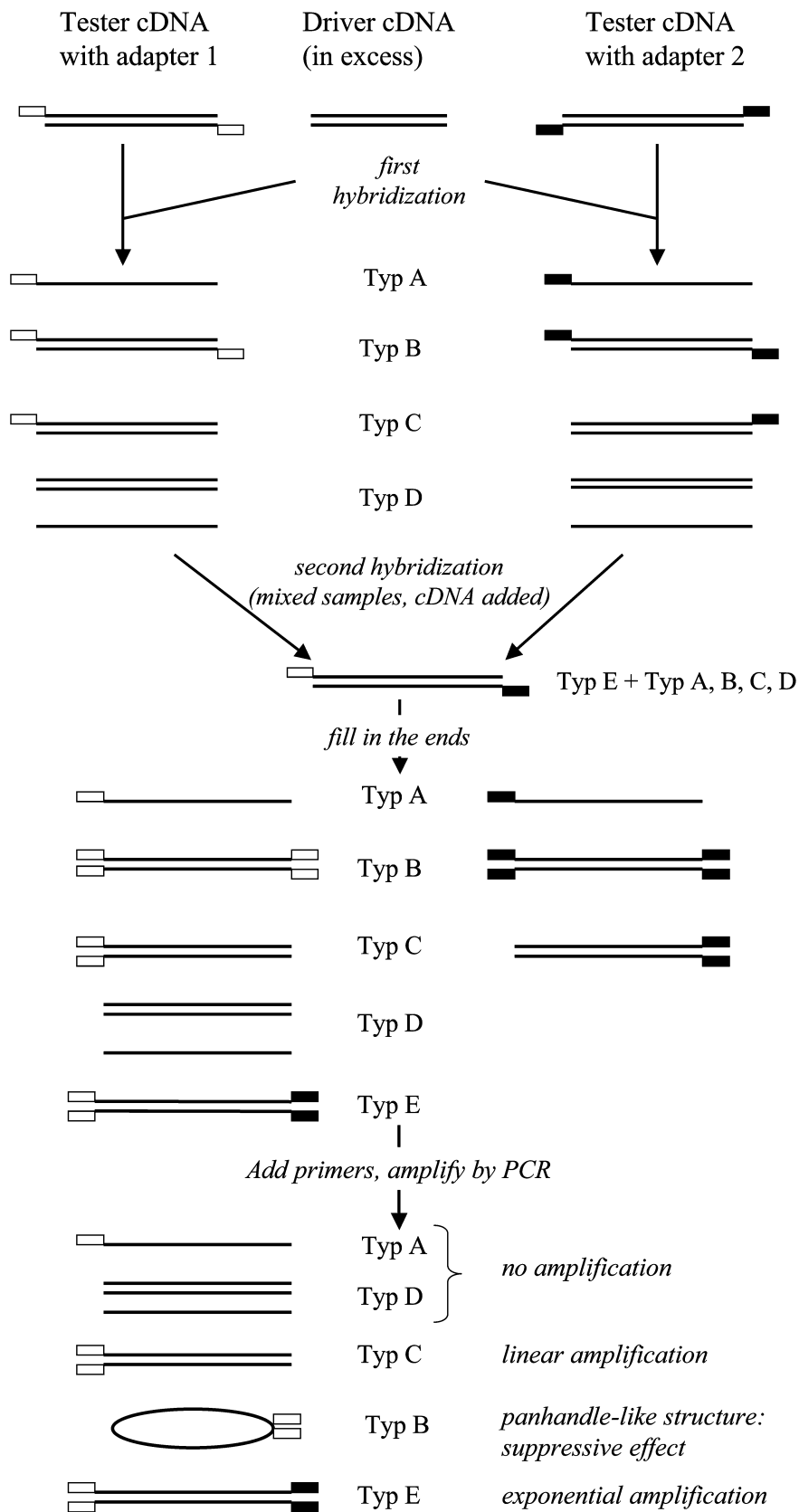
This method is technically based on the polymerase chain reaction (PCR). Generally, the cDNA population in which differentially expressed sequences will be detected, is termed tester cDNA (e.g. cDNA from rheumatoid synovium), whereas the cDNA population that acts as reference (e.g. cDNA from normal synovium) is named driver cDNA. The differentially expressed cDNAs are present in the “tester” but are absent (or present at lower levels) in the “driver”.

The underlying principle of subtractive hybridization consists of the hybridization of sequences that are expressed both in the test-probe (tester cDNA) as well as in the reference-probe (driver cDNA). The cDNA molecules in the test-probe are hybridized with a large excess of cDNA from the reference-probe and the double-stranded hybridized sequences are removed. Those cDNA sequences from the test-probe that fail to find their reference counterpart represent genes that are expressed exclusively in the cell type or tissue of interest.

The application of SSH was first described by Diatchenko and coworkers almost ten years ago. The molecular details of SSH are shown in Fig. 2. Briefly, the tester and driver double-stranded (ds) cDNAs are first digested with restriction enzymes. The tester cDNA fragments are then divided into two samples (1 and 2) and ligated with two different adapters (adapter 1 and adapter 2), resulting in two populations of tester, (1) and (2). The SSH technique uses two hybridization steps. First, an excess of driver is added to each tester population. The samples are then heat-denatured and allowed to anneal. In the second hybridization, the two samples from the first hybridization are mixed together. Only the remaining subtracted single-stranded (ss) tester cDNAs are able to reassociate to give (b), (c), and

new (e) hybrids. Addition of a second portion of denatured driver at this stage further enriches fraction (e) for differentially expressed genes. The newly formed (e) hybrids have an important feature that distinguishes them from hybrids (b) and (c) formed during first and second hybridizations. This feature is that they have different adapter sequences at their 5'-ends, one from sample 1 (adaptor 1) and the other from sample 2 (adaptor 2). The two sequences allow preferential amplification of the subtracted fraction (e) using PCR and a pair of primers, P1 and P2, which correspond to the outer parts of adapters 1 and 2 respectively. In all PCR cycles, exponential amplification can only occur with type (e) molecules, because intramolecular annealing of longer adapter sequences is both highly favored and more stable than intermolecular annealing of the much shorter PCR primers. This is the suppression PCR effect (suppressive subtractive hybridization). Furthermore, only type (e) molecules have different adapter sequence at their ends, which allows them to be exponentially amplified by PCR. The resulting cDNA library is highly enriched for differentially expressed genes and can be cloned and sequenced using standard protocols.

Subtractive hybridization offers the possibility of identifying yet unknown genes or genes with indefinite function. The creation of a complete expression profile by this method, however, is expensive in both money and time. Dependent on cell types or tissues used for the experiment, 300–500 cDNA clones have to be analyzed to obtain all the differentially expressed molecules. Many clones get picked multiple times showing a certain level of redundancy in the use of SSH. Since this hybridization technique uses probes from just a single patient and his healthy control, the differences in expression of the identified genes must be confirmed in additional samples to obtain definite conclusions. Finally, the efficiency of the SSH decreases rapidly if heterogeneous tissues with many different cell populations such as the highly heterogeneous synovium of rheumatoid arthritis are used.



Microarrays

The further development and optimization of gene-array technology is one of the most important technical achievements of the past few years. In principle, microarray technology allows scientists to detect the expression of numerous genes in the same sample simultaneously. By determining expression profiles with gene arrays under different experimental conditions, this technique can, for instance, be used to define biological features of genes with yet unknown functions. In addition, gene arrays are often used to search for key pathogenic molecules by comparing the expression profile of samples derived from patients (e.g. synovial fibroblasts derived from patients with rheumatoid arthritis) with the expression profile of samples derived from normal tissue (e.g. synovial fibroblasts from normal synovium).

Expression microarrays are devices containing up to tens of thousands of short DNA probes of known sequence, arrayed in a systematic fashion and bound to a flat surface, commonly a nylon membrane or a glass slide. Each spot on the array corresponds to a particular probe, and each probe corresponds to a short section of a gene. The hybridization of cDNAs present in the experimental sample to their corresponding sequences/DNA probes spotted on the array is the core reaction of microarray analysis. The complete experimental procedure, includes the making of the chip (microspotting, which is performed mechanically by a robot, a so-called arrayer), sample preparation, probe hybridization and finally analysis and interpretation of the data. Figure 3 shows the detailed procedures of a microarray experiment. Nowadays, a large number of different microarrays are commercially available and ready to use for probe hybridization.

In microarray experiments, the difficulty no longer lies with the technology, but rather with the handling and analysis of the large amount of data generated. In order to interpret these data, measurement of the raw data, ►normalization for different amounts of cDNA loaded on the array, strategies for the reduction of false-positive results and statistical analysis of the large amount of data have to be standardized in each experiment depending on the scientific question. For example, signals with low intensity (e.g. less than two-fold above background signals) representing genes with very low expression levels are often excluded from further analysis during raw data analysis, since results with these genes show a high variability and often cannot be reproduced. Normalization corrects for

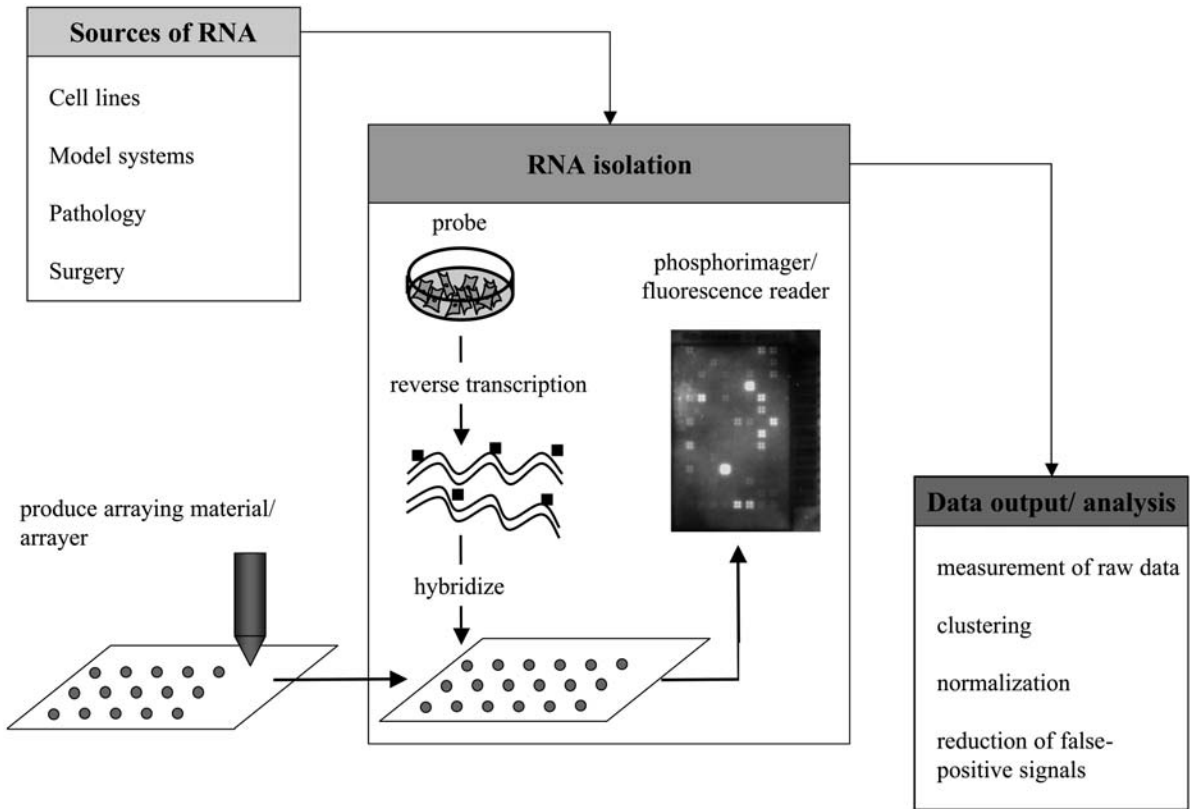
the different amounts of total cDNA in the samples that are compared. This can be achieved by measuring the signal intensity of so-called housekeeping genes, which are usually expressed at a stable level in all cells, independently of external factors. Differences in the expression levels of these housekeeping genes between two samples therefore represent differences in the amount of total cDNA extracted from the tissue/cells. To allow comparison between two samples, the expression levels of all non-housekeeping genes have to be corrected for the differences observed in the expression of housekeeping genes in the two samples. For the reduction of false-positive signals, genes with a less than two-fold difference between two samples are often excluded from further analysis (threshold-method). Although there are some limitations (e.g. genes with smaller differences in the expression level are automatically excluded from the analysis independent of their biological significance), this strategy is based on the fact that there is a strong inverse correlation between the percentage of false-positives and the level of differential expression. Finally, statistical methods need to be chosen to enable an efficient analysis and interpretation of the large amounts of expression data derived from the arrays. Among the various statistical methods available, cluster analysis is one of the most often used instruments. With this method, genes with a similar expression pattern under different conditions are grouped together to allow conclusions on similar biological functions or similar mechanisms of regulation between the clustered genes.

Microarray experiments can require significant costs, which often limits their academic use. Moreover, without bearing in mind the difficult interpretation and analysis of the microarray data, there is the danger of uncritical use of the technology. Guidelines have therefore been published for its utilization in rheumatology. These guidelines deal with the identification of false-positive results, the statistical analysis and the confirmation of differentially expressed genes by independent methods (7, 8).

Verification

Depending on the individual experiment, a major problem is the generation of false positive results. The proportion of these false positive results increases both when the number of differentially expressed molecules is low and when the quantitative differences of expression between the samples are small. Thus, the identified differences in the expression of key genes

Rheumatism Related Genes, Identification. Figure 2 Molecular details of suppressive subtractive hybridization The SSH technique uses two hybridizations between the driver and the tester cDNA. Nested PCR primers corresponding to the initially ligated adaptors allow preferential amplification of the subtracted fraction (adapted and modified from Diatchenko, Biochemistry 1996).



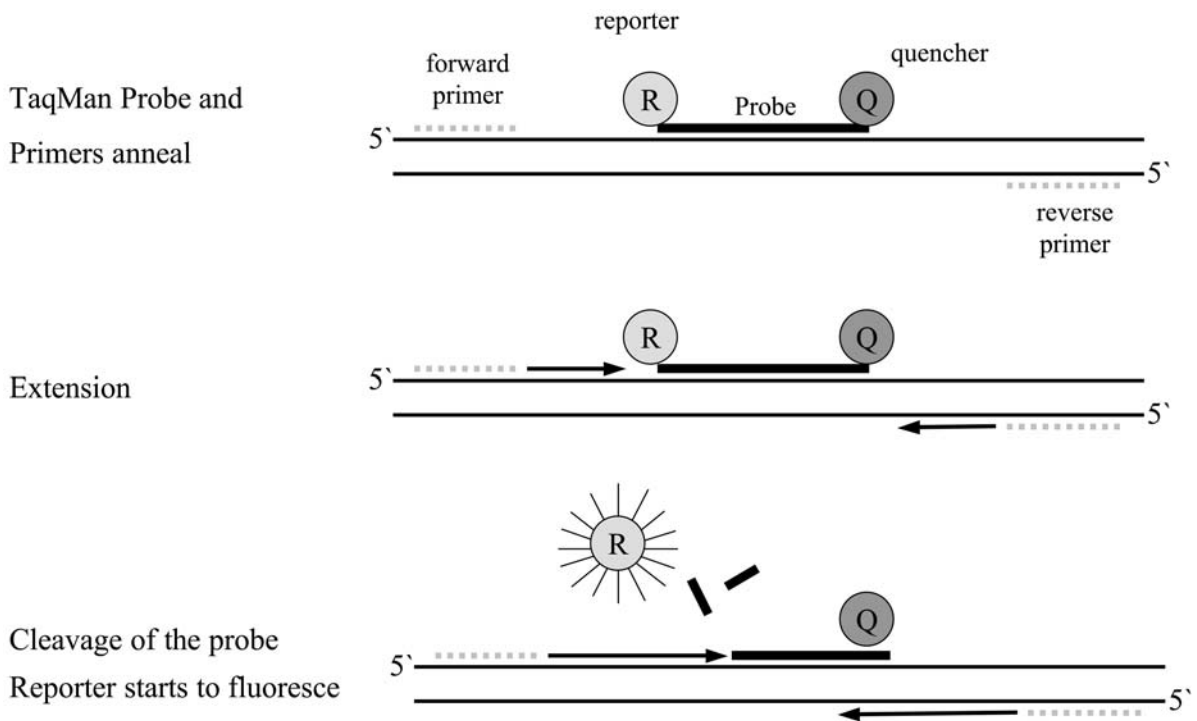
Rheumatism Related Genes, Identification. Figure 3 Details of the microarray technology Flow path scheme of the interacting components required for RNA-expression analysis using DNA – microarrays that are commercially available or generated by robotic arraying (adapted and modified from Bowtell, Nature Genetics Supplement 1999).

must be verified by an independent method. In addition to the conventional tools such as Northern blot or reverse transcriptase polymerase chain reaction (RT-PCR), the quantitative real-time PCR (e.g. TaqMan real-time PCR) is the new gold standard. Additionally to the forward and reverse primers that are used in normal PCR, in TaqMan real-time PCR another probe is introduced that binds specifically to the molecule of interest. It is labeled with a fluorogenic dye at its 5' end and with a quencher at its 3' end. The quencher suppresses the emission of the fluorogenic signal as long as the probe is intact. If the target sequence is present, the probe anneals downstream from one of the primer sites and is cleaved by the 5'-nuclease activity of Taq DNA polymerase as this primer is extended. The cleavage of the probe separates the reporter dye from the quencher dye and thus increases the reporter signal. Additional dye molecules are cleaved from their respective probes with each cycle, resulting in an increase in the fluorescence intensity proportional to the total of amplicons produced. The number of cDNA molecules produced within a single cycle of the

exponential phase of the PCR is directly correlated to the numbers of molecules in the starting solution, thereby allowing quantitative conclusions on the mRNA expression of the specific molecule. The principle of TaqMan-PCR is outlined in Fig. 4.

Functional Characterization

The term "functional genomics" includes the characterization of previously identified and verified genes at a functional level. Dependent on the experimental question, a gene can be over-expressed or inhibited by using for instance expression vectors or antisense constructs. It can then be investigated further for its biological functions in an appropriate disease model. The SCID-mouse coimplantation model that is highly disabled in its cellular immune response is such a model and is widely used in RA research. In this model, synovial fibroblasts that have been transfected with the expression vector or the anti-sense construct directed against the gene of interest are coimplanted with human cartilage under the renal capsule. After 60 days, the invasion of synovial fibroblasts into the cartilage can be



Rheumatism Related Genes, Identification. Figure 4 Principles of quantitative real-time PCR
 TaqMan real-time PCR is the most sensitive method available at the moment. It is based on the use of sequence-specific probes that are labeled with a fluorogenic dye. The amplification of the specific sequence cleaves the dye to emit a fluorogenic signal (adapted and modified from Distler, *Molekularmedizinische Grundlagen von rheumatischen Erkrankungen* 2003).

analyzed in a semi-quantitative fashion. In general, fibroblasts from RA patients show a strong invasion into the coimplanted cartilage, whereas synovial fibroblasts from healthy controls show no relevant invasiveness. The modulation of the invasion by transfection with expression vectors or with anti-sense constructs allows conclusions on the functional role of the gene of interest in cartilage degradation.

Clinical Relevance

The new molecular biology-based screening methods are a useful tool to identify target molecules for novel therapies in rheumatoid arthritis. While subtractive hybridization has been successfully used to identify yet unknown sequences and key molecules with yet unknown functions, microarrays are a powerful technique to obtain expression profiles of certain cell types or tissues under specific conditions within a comparatively short time. With both techniques, a significant number of false-positives are produced. The differential expression of key molecules identified by these methods therefore needs to be confirmed by independent methods such as real-time PCR. The term

“functional genomics” includes the functional characterization of the identified key molecules in appropriate disease models.

References

1. Distler O (2003) Genomics: Identifikation neuer und bekannter Gene. In: Ganten/ Ruckpaul (eds) *Molekularmedizinische Grundlagen von rheumatologischen Erkrankungen*, pp 1–17
2. Jantti J, Aho K, Kaarela K et al (1999) Work disability in an inception cohort of patients with seropositive rheumatoid arthritis: a 20 year study. *Rheumatology (Oxford)* 99:1138–1141
3. Alberts B, Bray D, Lewis J et al (1994) DNA cloning In: *Molecular biology of the cell*, 3rd edn. pp 308–312
4. Diatchenko L (1996) Suppression subtractive hybridization: a method for generating differentially regulated or tissue-specific cDNA probes and libraries. *Proc Natl Acad Sci USA* 93:6025–6030
5. Bowtell D (1999) Options available – from start to finish – for obtaining expression data by microarray. *Nat Genet* 21:25–32
6. Liu E, Karuturi K (2004) Microarrays and Clinical investigations. *N Engl J Med* 16:1595–97

7. Firestein GS, Pisetsky DS (2002) DNA microarrays: boundless technology or bound by technology? Guidelines for studies using microarray technology. *Arthritis Rheum* 46:859–861
8. Distler O, Gay S, Neumann E et al (2003) Minimum information about a microarray experiment: comment on the editorial by Firestein and Pisetsky. *Arthritis Rheum* 48:862

Rhizomelic Chondrodysplasia Punctata

Definition

Rhizomelic chondrodysplasia punctata (RCDP) is a rare peroxisome biogenesis disorder characterised by proximal shortening of the extremities, punctate calcifications in cartilage, mental retardation and dwarfism. Symptoms are usually present at birth or appear in the first few months of life. The majority of patients do not survive the first decade of life.

Type 1 RCDP (OMIM 215100) results from a defect in the PEX7 gene. In type 2 RCDP (RCDP2; OMIM 222765), patients show deficiency of the enzyme acyl-CoA: dihydroxyacetonephosphate acyl-transferase. RCDP3 (OMIM 60012) is caused by mutations in the alkyldihydroxyacetonephosphate synthase gene.

► [Peroxisomal Disorders](#)

proteins, including Rac1, Cdc42 and RhoA that cycle between an active GTP-bound and inactive GDP-bound states. They are involved in a variety of cellular processes such as cytokinesis, morphogenesis, cadherin-derived cell-cell contacts, polarity, transformation, cell-substratum contacts and migration. All are primarily controlled through modifications of the actin cytoskeleton.

► [Adherens Junctions](#)

► [Cardiac Signaling: Cellular, Molecular and Clinical Aspects](#)

► [Signal Transduction: Integrin-Mediated Pathways](#)

► [Rho, Rac, Cdc42](#)

Rho Proteins

Definition

Rho proteins are small monomeric GTPases implicated in actin organization and the interaction of the cytoskeleton with intracellular membranes. Rho proteins regulate a number of cellular functions (cell movement, cell adhesion, cytokinesis, cell growth). See also ► [Ras](#).

► [Epithelial Cells](#)

► [Tangier Disease](#)

Rho Family Small GTPases

Definition

The RHO genes code for a family of related proteins that can bind to and hydrolyze GTP. These proteins are similar to the Ras oncogenes and include RhoA, Rac1 and Cdc42. In the GTP-bound state, these proteins can bind to other effector proteins and modulate cell morphology.

► [Epithelial Cells](#)

Rho, Rac, Cdc42

ANNE RIDLEY, CLEMENS FRANZ

Ludwig Institute for Cancer Research, Royal Free and University College School of Medicine, Department of Biochemistry and Molecular Biology, University College London, London, UK
anne@ludwig.ucl.ac.uk

Definition

Rho ► [GTPases](#) form a subfamily of the Ras superfamily of small guanosine triphosphatases (► [GTPases](#)) and are molecular switches that control a wide variety of signal transduction pathways in all eukaryotic cells (1). Of the 20 members identified in mammals, RhoA, Rac1 and Cdc42 have been most widely studied and are best known for their role in regulating the actin

Rho GTPases

Definition

Rho stands for Ras homology. Rho GTPases are members of the Ras superfamily of GTP-binding

cytoskeleton, thereby affecting cell morphology and movement. However, they are also important regulators of cell-cell and cell-substratum adhesion, ►endocytosis, vesicle trafficking, ►phagocytosis and gene transcription. Furthermore, Rho GTPases play a role in a number of pathological processes, such as inflammation, cardiovascular diseases, bacterial/viral infection and epithelial tumour progression.

Characteristics

Cytoskeletal Rearrangement, Polarization and Migration

Rho proteins are key regulators of the actin cytoskeleton (2). Rho activation leads to the assembly of contractile ►actomyosin filaments known as ►stress fibres, Rac stimulates the formation of ►lamellipodia, sheet-like membrane extensions containing an extensive branching network of ►actin filaments, and Cdc42 induces the assembly of ►filopodia, finger-like protrusions in which actin filaments are arranged in parallel.

Cell migration requires membrane extension at the front of the cell, formation of new adhesions and cell body contraction (3). In all of these steps, cytoskeletal rearrangements caused by the coordinated activation of Rho GTPases are crucial. Actin polymerisation within Rac-induced lamellipodia provides the protrusive force for membrane extension in the direction of migration. In addition, Rac activation leads to the formation of small, integrin-containing ►focal complexes, which link lamellipodia to the ►extracellular matrix. Cell body contraction depends on Rho-mediated actomyosin contractility. To translate contraction into cell movement, the cell rear has to detach from the extracellular matrix through release of integrin-based attachments, and this is also induced by Rho.

Rho proteins are also emerging as key regulators of the microtubule network and in some cell types this is important for cell polarization and directed cell migration. Cdc42 can affect cell polarity by localizing the ►microtubule-organizing center (MTOC) and ►Golgi apparatus in front of the nucleus, oriented toward the leading edge, possibly facilitating microtubule growth into lamellae and providing vesicular transport into leading edge membrane extensions. Rac promotes extension of microtubules into the leading edge and both Cdc42 and Rac are involved in tethering microtubules to the plasma membrane (1).

Besides cell migration, Rho GTPases are involved in additional cellular functions that involve the actin cytoskeleton. For example, rearrangement of membrane-associated actin filaments leads to localized changes in the structure of the plasma membrane and in this way drives phagocytosis. Both Rac and Cdc42 are required for immunoglobulin-mediated (type I) phagocytosis, whereas complement-mediated (type II)

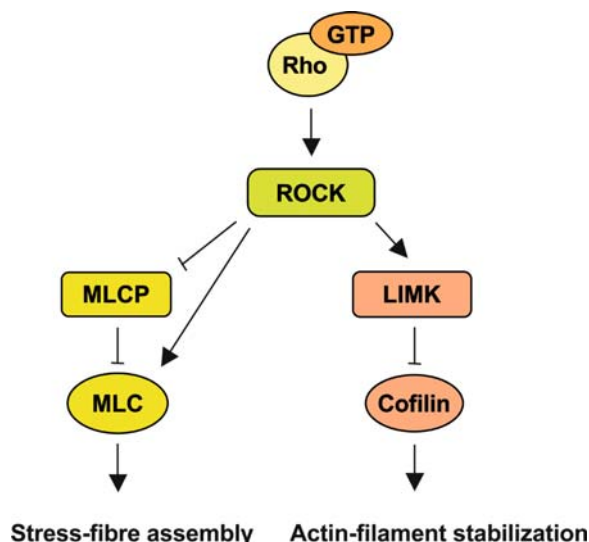
phagocytosis requires Rho. In addition, Rac and Cdc42 are required for the uptake of apoptotic cells.

Cell-Cell Adhesion

It is well established that Rho GTPases are required for the assembly of functional cell-cell contacts, as well as for their disassembly during motility. The activity of RhoA and Rac1 is required for establishment and maintenance of ►cadherin-dependent cell-cell adhesion (4, 5), although the underlying mechanisms are not well understood. Inhibition of Rac or Rho prevents the formation of cadherin-containing intercellular junctions between keratinocytes. Furthermore, inhibition of Rho leads to rapid removal of E-cadherin from mature junctions. Expression of activated Rac, on the other hand, increases the amount of E-cadherin accumulating at sites of cell-cell contact in epithelial cells and both Rac1 and Cdc42 are necessary to maintain E-cadherin-mediated cell-cell adhesion in these cells. Interestingly, Rac activity is also required during ►adherens junction disassembly, demonstrating that Rac1 is involved in positive as well as negative regulation of cadherin-mediated cell-cell adhesion. Rho proteins may regulate E-cadherin-mediated adhesion in ways that are more indirect; through regulating E-cadherin endocytosis, they may control the amount of E-cadherin able to engage in cell-cell adhesion at the plasma membrane.

Molecular Interactions

Rho proteins interact with a diverse range of downstream targets to induce cellular responses. The control of actomyosin-mediated contractility by Rho involves the activation of Rho-kinases (ROCKs) (6), which in turn increase the level of phosphorylated myosin light chain (MLC), the regulatory subunit of myosin II, both by inhibiting MLC phosphatase and by phosphorylating MLC (Fig. 1). Rho also regulates actin polymerization and microtubule organization through the Dia proteins. The ability of Rac to induce lamellipodia is believed to involve the activation of the ►Arp2/3 complex via ►Scar/Wave proteins (7, 8). The Arp2/3 complex then facilitates the formation of new actin filaments that branch off existing filaments. In addition, Rac is thought to stimulate actin polymerisation by promoting the uncapping of actin filaments. By activating ►phosphatidylinositol 4-phosphate 5-kinase (PIP 5-kinase), an enzyme which converts PtdIns(4)P (PIP) into PtdIns(4,5)P₂ (PIP₂), active Rac is able to increase the amount of PIP₂ at the plasma membrane. By binding to PIP₂, actin filament capping proteins, such as CapZ, are then recruiting away from the ends of actin filaments, allowing filament extension. Furthermore, Rac and Cdc42 may also be able to decrease the



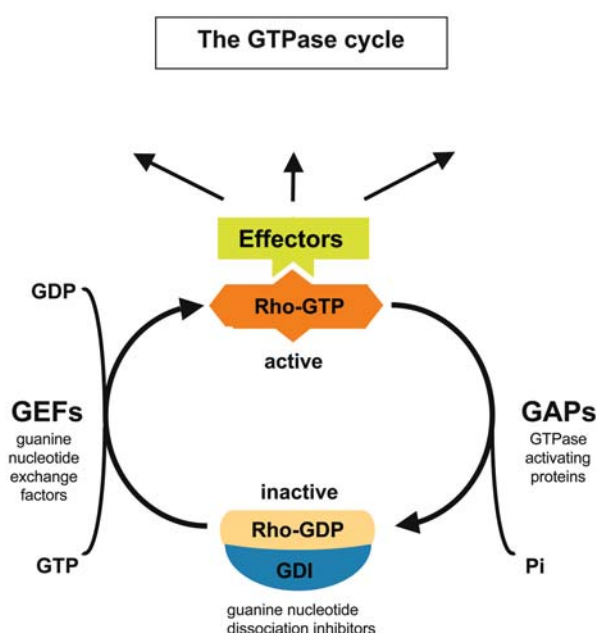
Rho, Rac, Cdc42. Figure 1 Rho regulates stress-fibre formation. GTP-bound Rho activates Rho-kinase (ROCK), which subsequently phosphorylates myosin light chain (MLC), leading to increased myosin II activity and stress-fibre formation and contractility. ROCK further increases myosin II activity by phosphorylating myosin light chain phosphatase (MLCK), thereby inhibiting MLC dephosphorylation. ROCK also phosphorylates LIM-kinase (LIMK), leading to actin-filament stabilization via inhibition of the actin-fibre depolymerizing protein cofilin.

rate of actin depolymerisation. *Via* their target p21-activated kinase (PAK), Rac and Cdc42 stimulate LIM-kinase, which in turn inactivates cofilin, a protein that can promote actin depolymerisation.

Rac and Cdc42 may modulate E-cadherin activity *via* IQGAP1, a scaffolding protein that negatively regulates E-cadherin-mediated adhesion. The RhoA targets ROCK and Dia act antagonistically to regulate cadherin-mediated cell-cell adhesion (4).

Regulatory Mechanisms

Generally, Rho proteins exist in an active, GTP-bound, or in an inactive, GDP-bound, state (Fig. 2). ▶**GTPase-activating proteins** (GAPs) stimulate the slow intrinsic GTPase activity of Rho proteins, thereby converting them from the GTP-bound form to the GDP-bound form and inactivating them. ▶**Guanine nucleotide exchange factors** (GEFs), on the other hand, promote the exchange of GDP for GTP and consequently activate Rho proteins (2). Upon activation Rho proteins usually translocate to the plasma membrane or other membrane compartments where they are able to interact with their different target proteins. The



Rho, Rac, Cdc42. Figure 2 Rho GTPases cycle between an inactive, GDP-bound and an active, GTP-bound state. Guanine nucleotide exchange factors (GEF) facilitate the exchange of GDP for GTP on Rho GTPases, thereby activating them, whereas GTPase activating proteins (GAPs) stimulate the intrinsic GTPase activity of Rho GTPases, reverting them into the inactive, GDP-bound state. In addition, Rho GTPases can associate with guanine nucleotide dissociation inhibitors (GDIs) to form an inactive complex. Effector proteins become activated upon binding specifically to the GTP-bound form of Rho GTPases.

membrane-association of Rho proteins is facilitated by prenylation at their C-terminus. In the absence of activating stimuli, Rho proteins are predominantly bound to ▶**guanine nucleotide dissociation inhibitors** (GDIs), which prevent them from binding membranes or interacting with their targets. The human genome contains over 60 GEFs and more than 70 GAPs for Rho proteins, indicating that Rho protein activities are subject to intricate regulatory pathways, to which additional mechanisms, such as the inhibition of RhoA activity by PKA-dependent phosphorylation or the regulation of Rho GTPase mRNA stability, add further complexity.

References

1. Etienne-Manneville S, Hall A (2002) Rho GTPases in cell biology. *Nature* 420:629–635
2. Hall A (1998) Rho GTPases and the actin cytoskeleton. *Science* 279:509–514

3. Ridley AJ, Schwartz MA, Burridge K et al (2003) Cell migration: integrating signals from front to back. *Science* 302:1704–1709
4. Braga VM (2002) Cell-cell adhesion and signalling. *Curr Opin Cell Biol* 14:546–556
5. Lozano E, Betson M, Braga VM (2003) Tumor progression: Small GTPases and loss of cell-cell adhesion. *Bioessays* 25:452–463
6. Riento K, Ridley AJ (2003) Rocks: multifunctional kinases in cell behaviour. *Nat Rev Mol Cell Biol* 4:446–456
7. Pollard TD, Blanchoin L, Mullins RD (2000) Molecular mechanisms controlling actin filament dynamics in nonmuscle cells. *Annu Rev Biophys Biomol Struct* 29:545–576
8. Cory GO, Ridley AJ (2002) Cell motility: braking WAVES. *Nature* 418:732–733

- [Spinal Muscular Atrophy](#)
- [Splicing](#)
- [Telomerase](#)

Ribose

Definition

Ribose is a five-carbon monosaccharide (pentose) that is commonly found in nucleic acid.

- [Glycosylation of Proteins](#)

Rhodopsins/Rhodopsin

Definition

Rhodopsins are proteins found in all three kingdoms of life. They form a family of three subgroups. Visual rhodopsins from eukaryotes are coupled to G-protein regulatory networks. Rhodopsin is the light sensitive pigment of the retina formed from retinal linked through a Schiff's base to opsin. It is located in the disks of the outer segments of rods and cones. Upon illumination with light, rhodopsin achieves an activated state due to isomerisation of 11cis retinal into the all-trans form (► [retinoid cycle](#)). Rhodopsin in the activated state triggers the phototransduction process, by activating transducin and phosphodiesterase via hydrolyzing cyclic nucleotides, thereby closing a cation channel in the cytoplasmic membrane of the outer segment of photoreceptors. Microbial rhodopsins, detected in eukaryotes, eubacteria, and archaea are either ion translocating proteins or involved in the sensory response of microbes. The newly discovered Channelopsins are light activated, proton- or other cation channels.

- [Photoreceptors](#)
- [Retinitis Pigmentosa](#)

Ribonucleoprotein

Definition

Ribonucleoprotein is composed of ribonucleic acid (RNA) and protein.

Ribosomal A Site

Definition

Ribosomal A site displays the site on the ribosome at which aminoacyl-tRNA binds and forms a codon/anticodon base pairing complex with the mRNA.

- [Translational Frameshifting, Non-standard Reading of the Genetic Code](#)

Ribosomal Binding Site

Definition

Ribosomal binding site denotes a sequence of DNA, where the ribosomes bind to initiate the translation of a protein on mRNA.

- [Recombinant Protein Expression in Bacteria](#)

Ribosomal Display

Definition

Ribosomal display refers to an *in vitro* affinity based selection method avoiding release of protein and RNA from the ribosome to achieve genotype/phenotype coupling.

- [Immunological Methods, the Use of Antibodies to Discover the Function of Novel Gene Products](#)

Ribosomal DNA

Definition

rDNA describes a part of the genome which encodes for ribosomal RNA (rRNA). It is noteworthy to mention that the abbreviation rDNA is often used for the term “recombinant DNA” as well.

- ▶ [Recombinant Protein Expression in Yeast](#)
- ▶ [RNA Polymerase I](#)

small subunit, each of which is an assembly of ribosomal RNA and ribosomal proteins. The rRNAs comprise the catalytic function of the ribosome.

- ▶ [Cap-Independent Translational Control](#)
- ▶ [RNA Polymerase I](#)
- ▶ [tRNA](#)
- ▶ [Translational Control in Eukaryotes](#)

Ribosomal P Site

Definition

Ribosomal P site displays the site on the ribosome to which the peptidyl-tRNA/mRNA complex binds after transfer of the growing peptide.

- ▶ [Translational Frameshifting, Non-standard Reading of the Genetic Code](#)

Ribosomal RNA

Definition

Eukaryotic ribosomes are constructed from four types of ribosomal RNA. Three of these species (18S, 5.8S, and 28S) are derived from the processing of a single large precursor RNA, 45S rRNA, which is transcribed by RNA Polymerase I from tandem arrays of rDNA repeats. The remaining rRNA, 5S, is produced by RNA Polymerase III from a different gene cluster.

- ▶ [Nuclear Compartments](#)
- ▶ [RNA Polymnerase I](#)

WOLFGANG WINTERMEYER, MARINA V. RODNINA
Institutes of Molecular Biology and Physical
Biochemistry, University of Witten/Herdecke,
Germany
winterme@uni-wh.de
rodnina@uni-wh.de

Definition

The amino acid sequence of proteins is encoded in the nucleotide sequence of the DNA. During gene expression, one strand of the DNA is transcribed into the complementary RNA copy (messenger RNA, ▶ [mRNA](#)) that is subsequently translated into protein by polymerizing amino acids in the sequence specified by the nucleotide sequence of the mRNA. The substrates of protein synthesis are aminoacylated ▶ [tRNAs](#) (Aa-tRNAs) that at their 3' ends carry specific amino acids in an energy-rich ester bond. Aa-tRNAs are produced by the action of aminoacyl-tRNA synthetases which each specifically recognize one amino acid and the tRNA that is specific for this amino acid as it has the anticodon triplet matching the respective codon triplet. Thus, the tRNA serves as an adaptor between the amino acid and the codon on the mRNA. Protein synthesis takes place on ribosomes that are large macromolecular assemblies consisting of several RNAs and numerous proteins. The ribosome presents the mRNA in such a way that Aa-tRNAs can bind to their respective codon and take part in peptide bond formation catalyzed by the ribosome. In their functions, ribosomes are assisted by a number of accessory translation factors.

Ribosome

Definition

The ribosome is the machinery that translates messenger RNAs into proteins, and it consists of a large and a

Characteristics

The Genetic Code

Three each of the four bases in RNA (guanine, G; adenine, A; cytosine, C; uracil, U) form one codon that specifies one particular amino acid. From the total of 64

(4³) code words, 61 stand for the 20 standard amino acids to be incorporated into proteins during translation (sense codons); the remaining three serve as termination signals that specify the end of the coding sequence (nonsense or stop codons). Because the number of codons is much larger than the number of amino acids to be coded for, most amino acids (except methionine and tryptophan) are encoded by two or more codons. One codon, AUG, stands for methionine; in a particular mRNA context, it specifies the start of the coding sequence, implying that the first amino acid of newly synthesized proteins is methionine. The codon triplets of the mRNA are decoded by the formation of three complementary base pairs with the anticodons of tRNAs. Internal codons are decoded by elongator tRNAs. Initiation codons are decoded by a particular methionine-specific tRNA, initiator tRNA, which differs from elongator tRNA^{Met} used for decoding internal methionine codons. The three termination codons, UAA, UAG, UGA, stand for “Stop”; these codons are recognized by proteins, the termination (or release) factors, rather than by tRNAs. The genetic code is the same in all organisms. Slight deviations, i.e. a few codons with different meanings, were found in mitochondria and ciliates.

The Translational Apparatus

Translation of mRNAs into protein requires a complex apparatus that consists of a large number of proteins and RNAs. Transfer RNA (tRNA) functions as adapter between the mRNA codon and the amino acid specified by the codon. tRNAs are aminoacylated by aminoacyl-tRNA synthetases. Decoding and peptide synthesis takes place on ribosomes, large ribonucleoprotein complexes consisting of three or four RNAs and 50 to 80 proteins, depending on organism and organelle. All phases of protein synthesis, i.e. initiation, elongation and termination, require the action of translation factors, which interact with the ribosome at defined stages of translation.

Transfer RNA

tRNAs are small RNAs (75–85 nucleotides) that have very similar secondary and tertiary structures; they contain a large number of modified nucleosides. The functional centers of the tRNA molecule are the anticodon and the 3'-terminal CCA sequence common to all tRNAs. The amino acid is attached to the 3'-terminal ribose by an energy-rich ester bond.

Decoding of mRNA codons on the ribosome entails the formation of base pairs between three bases of the codon and complementary bases of the anticodon. Most amino acids are specified by more than one codon, in some cases by up to six. In such cases, there are two or more tRNAs that are charged with the same

amino acid (isoacceptors) and have different sequences, including different anticodons. In the first and second codon position, base pairing with the anticodon strictly follows the Watson-Crick complementarity rules, i.e. G pairs with C and A with U. Base pairing at the third codon position is less restrictive, because “wobble” pairs (G:U, U:G, or I:U,C,A; I stands for inosine) are also “allowed”, i.e. energetically favorable. As a consequence, there are tRNAs that can pair with two or more codons differing in the third base, which reduces the total number of different tRNAs required to decode all 61 sense codons. Most organisms have 40–50 different tRNAs; a set of 22 tRNAs suffices for protein synthesis in mitochondria of mammalian cells.

The attachment of amino acids to the 3' end of tRNAs (aminoacylation) is catalyzed by aminoacyl-tRNA (Aa-tRNA) synthetases (amino acid:tRNA ligases). Aa-tRNA synthetases are specific for their respective amino acid and tRNA (or group of isoaccepting tRNAs). Generally, organisms contain 20 different Aa-tRNA synthetases, one for every amino acid incorporated. Aminoacylation takes place in two steps that are both catalyzed by the same Aa-tRNA synthetase:

1. Amino acid + ATP = Aa – AMP + PPi,
2. Aa – AMP + tRNA = Aa – tRNA + AMP.

The first step is the activation of the amino acid with ATP by formation of a mixed anhydride formed between the carboxyl group of the amino acid and the phosphoryl group of AMP. The second reaction is the transfer of the aminoacyl residue from Aa-AMP to tRNA to form an energy-rich ester bond with a hydroxyl group of the 3'-terminal ribose.

Aa-tRNA synthetases vary in size and subunit composition; there are monomeric enzymes with molecular masses around 50 kD, homo- and heterodimeric enzymes (100 kD) and homo- and heterotetramers with over 200 kD. Two classes of Aa-tRNA synthetases can be distinguished that differ in the structure of the catalytic center and transfer the amino acid to either the 2' or the 3' hydroxyl group of the terminal ribose of the tRNA. By spontaneous migration of the aminoacyl residue the 2' derivative isomerizes to the 3' Aa-tRNA that is the substrate for subsequent steps of translation.

The aminoacylation reaction is highly accurate, i.e. the frequency by which an incorrect amino acid is attached to a tRNA is low (about 10^{−4}). Most amino acids can be discriminated with sufficient accuracy in a single binding step on the basis of structure, size or chemical character. In the few cases where these criteria do not suffice for high-accuracy discrimination (isosteric or chemically similar amino acids, such as threonine and

isoleucine or isoleucine and valine, respectively), the required low error level is attained by hydrolytically discarding incorrectly formed products (editing). Editing synthetases have a second catalytic site where incorrectly formed Aa-AMP or Aa-tRNA is hydrolyzed. The recognition of the tRNA substrate is brought about by specific interactions between the Aa-tRNA synthetase and structural elements at various positions in the tRNA molecule (identity elements), frequently including residues in the acceptor arm and the anticodon. Isoaccepting tRNAs that are aminoacylated by the same synthetase possess the same pattern of identity elements.

Ribosomes

Ribosomes are large (diameter 20 nm; molecular weight several MD) ▶[ribonucleoprotein](#) particles (RNPs) that consist of two subunits of different size. They contain several molecules of ▶[ribosomal RNA](#) (rRNA) and many proteins that are mostly small (10–20 kD) and basic (Table 1). Ribosomes from bacteria and eukaryotes have similar architecture, although eukaryotic ribosomes are larger and have a higher molecular mass because they have larger rRNAs and more proteins. Mitochondrial ribosomes resemble bacterial ribosomes, but contain smaller rRNAs and many more proteins.

The structures of 30S and 50S ribosomal subunits are known at atomic resolution. In both subunits, the structure is determined by the tertiary structure of the

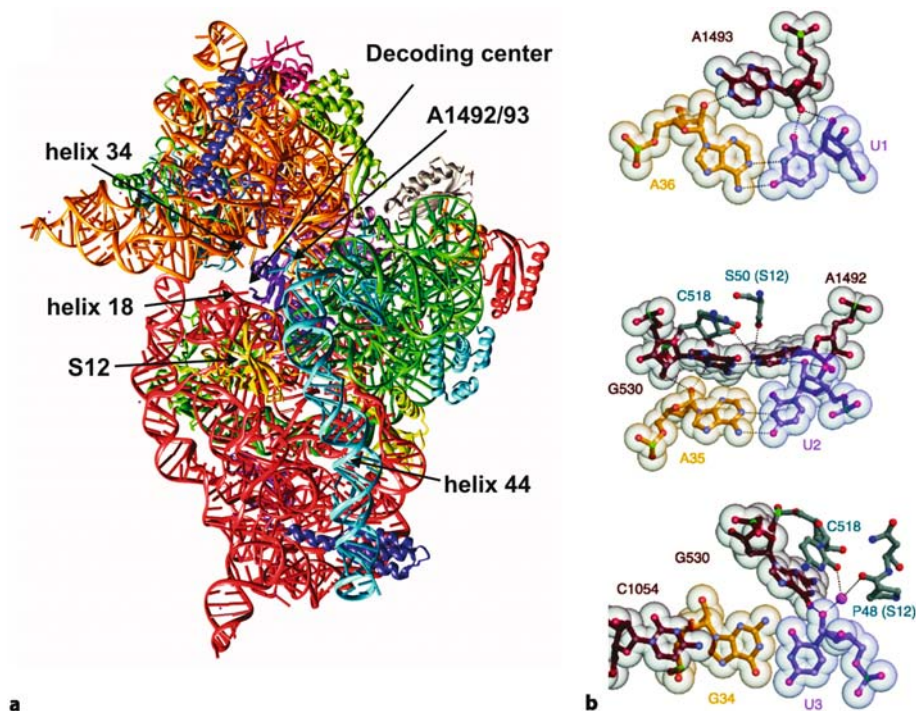
RNA to which the ribosomal proteins are attached. The three domains of 16S rRNA are represented in the body, the platform and the head domain of the 30S particle (Fig. 1a). The decoding center is built of parts of helices 44, 18 and 34 of 16S rRNA, and the only protein that reaches into the decoding center is S12. 16S rRNA forms contacts with the codon-anticodon complex that are essential for the accuracy of decoding. Monitoring the quality of codon-anticodon interaction is performed by helix 44 through A1493 and A1492 that contact base pairs at the 1st and 2nd positions respectively, in the minor groove of the codon-anticodon duplex (A-minor interactions); the 3rd position is monitored less stringently by a contact from G530 and other residues (Fig. 1b). The A-minor interactions formed by A1493/A1492 are sequence-independent, yet specific for Watson-Crick base pairs, providing the molecular basis for the induced-fit mechanism of decoding (see below).

Major landmarks of the 50S subunit (Fig. 2a) are the L1 region, the central protuberance that contains 5S rRNA and the stalk region where proteins L11, L10 and L7/12 (not seen in the structure) form an important site for elongation factor binding. The peptidyl transferase center, as localized by an analogue mimicking the tetrahedral intermediate of the peptidyl transferase reaction, is made up of RNA only, suggesting that the ribosome is a ▶[ribozyme](#). The exit tunnel for the growing peptide extends from the peptidyl transferase center through the body to the back of the subunit where the peptide emerges (Fig. 2b).

Ribosomes. Table 1 Composition of Ribosomes

Organism	Ribosome	Small Subunit	Large Subunit
Eukaryotes (Mammalia)			
Size	80S	40S	60S
Mass	4.2 MD	1.4 MD	2.8 MD
rRNA's		18S rRNA (1874 Nt)	28S rRNA (4718 Nt) 5.8S rRNA (160 Nt) 5S rRNA (120 Nt)
Proteins		33 Proteins	49 Proteins
Prokaryotes (E. coli)			
Size	70S	30S	50S
Mass	2.5 MD	0.9 MD	1.6 MD
rRNA's		16S rRNA (1542 Nt)	23S rRNA (2904 Nt) 5S rRNA (120Nt)
Proteins		21 Proteins	34 Proteins

S, Sedimentation constant; MD, Megadalton; Nt, Nucleotide



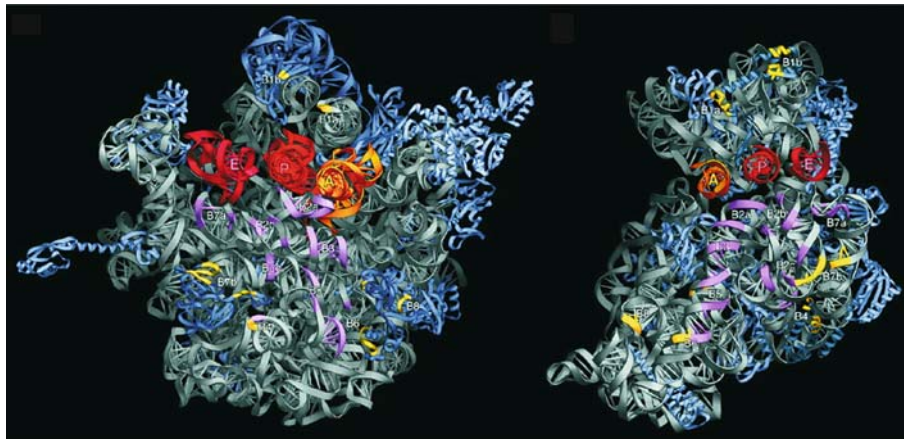
Ribosomes. Figure 1 Atomic structure of the 30S ribosomal subunit of ribosomes. (a) Overview of the 30S structure from *Thermus thermophilus* (ribbon representation). The three major domains of 16S rRNA are colored differently, and proteins are shown in various colors. (b) Recognition of codon-anticodon base pairs by ribosomal residues of the decoding center. Shown are the three anticodon bases (G34A35A36) bound to a codon triplet (U1U2U3) in the 30S crystal. Reprinted from Science 292 (2001) 897, with permission.

The structure of 70S ribosomes with three tRNA molecules bound in the binding sites for Aa-tRNA (A site), Pept-tRNA (P site) and the exit site for deacylated tRNA (E site) was determined by X-ray crystallography at 5.5 Å resolution (Fig. 3). The arrangement of the tRNAs in A and P sites with their anticodons and CCA ends respectively coming close together is clearly seen in the structure. There are a number of connections between the subunits, most of them made up of RNA. One prominent connection (bridge 2a) is formed of helix 69 of 23S rRNA and connects to the 30S decoding site (helix 44) and another (bridge 3) connects helix 44 (30S) and helix 71 (50S). Some bridges contact the A-site tRNA, e.g. helix 38 (bridge 1a, or “A-site finger”) from above and helix 69 from below, the latter also contacts the P-site tRNA. These interactions stabilize the tRNAs in their respective binding positions and presumably have to be released to allow tRNA movement during translocation. Structural and functional analyses suggest that rRNA is involved in the major functions of the ribosome, i.e. decoding, peptide bond formation and translocation, consistent with the high degree of conservation of the respective functional centers of rRNA in ribosomes from various organisms. Ribosomal proteins appear to

have important functional roles as well. Ribosomal proteins are required for rapid and accurate assembly of ribosomal particles *in vivo* and are probably essential for stabilizing certain structural elements of rRNA. For several ribosomal proteins a direct involvement in ribosome function has been demonstrated. ▶ **Mutations** in proteins S4, S5, S12 and others have strong effects on the fidelity of Aa-tRNA selection, suggesting an involvement of those proteins in decoding. In fact, all these proteins are located close to the decoding center. Proteins L7/12 that form the stalk of the 50S subunit of the bacterial ribosome and the corresponding P1/P2 proteins of eukaryotic ribosomes have an important role in the function of translation factors on the ribosome, including an involvement in the stimulation of the GTPase activity of those factors.

Translation Factors

All phases of ribosomal protein synthesis, except peptide bond formation itself, require that protein factors interact with the ribosome in order to proceed at physiologically relevant rates. There are factors for initiation, elongation and termination (Table 2). Several translation factors are GTPases that hydrolyze GTP at one point during their functional cycles. Atomic



Ribosomes. Figure 3 Crystal structure of 70S ribosomes from *Thermus thermophilus* with tRNAs bound to A, P, and E sites. Shown are the interface sides of 50S (left panel) and 30S (right panel) subunits. 23S and 16S rRNAs are shown in gray, 5S rRNA in blue at the top of the 50S subunit; proteins are in blue. tRNAs are in gold (A site), orange (P site), and red (E site). Connections between subunits are indicated in purple (RNA-RNA bridges) or yellow (RNA-protein bridges). Reprinted from Science 292 (2001) 883, with permission.

tRNA (N-formylmethionyl-tRNA^{fMet}, or fMet-tRNA^{fMet}) and the mRNA to the 30S ribosomal subunit. Usually the initiation site is found by base pairing between a short sequence in 16S rRNA and a partially or fully complementary sequence of the mRNA that is located about ten nucleotides upstream of the initiation codon (Shine-Dalgarno sequence). The initiation codon is recognized by the anticodon of the initiator tRNA bound to the 30S P site. 30S initiation complex formation in bacteria requires the function of three initiation factors, IF1, IF2 and IF3. IF3 promotes the dissociation of 70S ribosomes into subunits by binding to the 30S subunit and stimulates the binding of both initiator tRNA and mRNA. IF1 stimulates the function of the other two factors; it is bound to the 30S A site where it may help in directing initiator tRNA to the P site. IF2, a GTPase, is essential for the binding of initiator tRNA.

In the second phase of initiation, the 30S initiation complex is joined by the 50S ribosomal subunit. Subunit association triggers GTP hydrolysis by IF2 and subsequently all three initiation factors dissociate from the ribosome. The resulting 70S initiation complex contains initiator tRNA bound to the AUG start codon of the mRNA in the P site, and the following codon is presented in the A site. Binding of the respective Aa-tRNA to the 70S initiation complex initiates the first round of elongation.

The Elongation Cycle

The elongation of the peptide chain takes place in three main steps (Fig. 5). In the first step, bacterial EF-Tu (or eukaryotic eEF1 α) catalyzes the binding of Aa-tRNA to the A site of the ribosome. These factors are

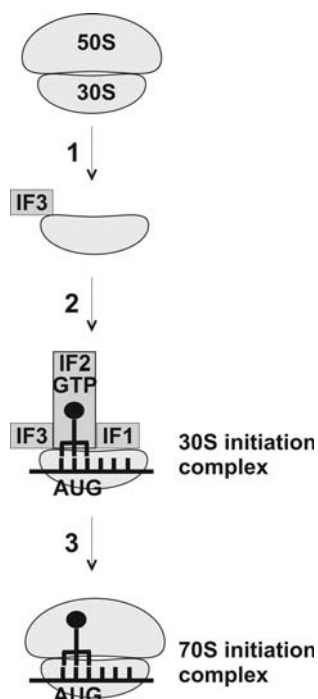
Ribosomes. Table 2 Protein Factors in Translation

	Prokaryotes	Eukaryotes
Initiation	IF1	~12 eIFs
	IF2	
	IF3	
Elongation	EF-Tu	EF1 α
	EF-Ts	EF1 β,γ,δ
	EF-G	EF2
Termination		EF3 (fungi)
	RF1	eRF1
	RF2	-
	RF3	eRF3
Ribosome Recycling	RRF	

► **GTPases** that in their GTP-bound conformation form a high-affinity ternary complex with Aa-tRNA, which, in turn, binds to the ribosome and, after GTP hydrolysis, releases Aa-tRNA to enter the peptidyl transferase center. Peptide bond formation between Aa-tRNA in the A site and Pept-tRNA in the P site is catalyzed by the ribosome. The final step is translocation, which is catalyzed by EF-G (eEF2 in eukaryotes), another GTPase.

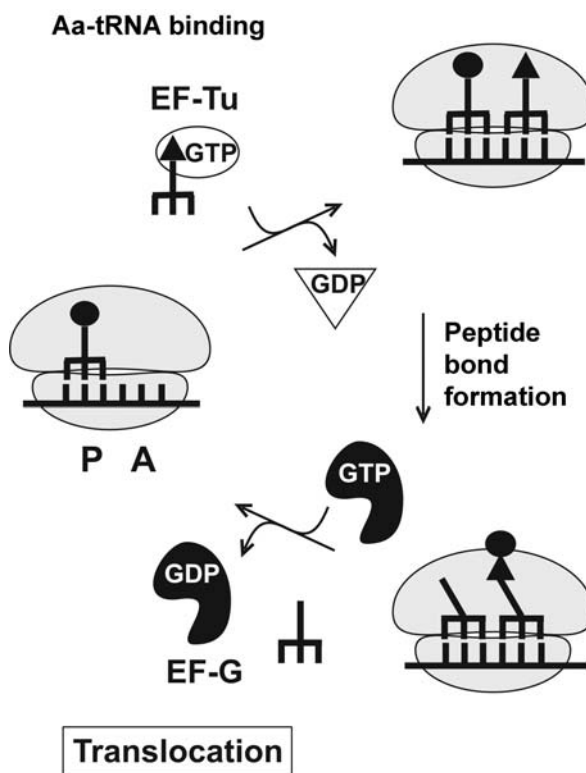
Aa-tRNA Binding

EF-Tu-dependent A-site binding of Aa-tRNA comprises a number of steps, as revealed by biochemical



Ribosomes. Figure 4 Initiation of translation in prokaryotes. Step 1: Subunit dissociation induced by IF3; step 2: Binding of fMet-tRNA, IF2-GTP, IF1, mRNA; step 3: 50S subunit joining, GTP hydrolysis, dissociation of Ifs.

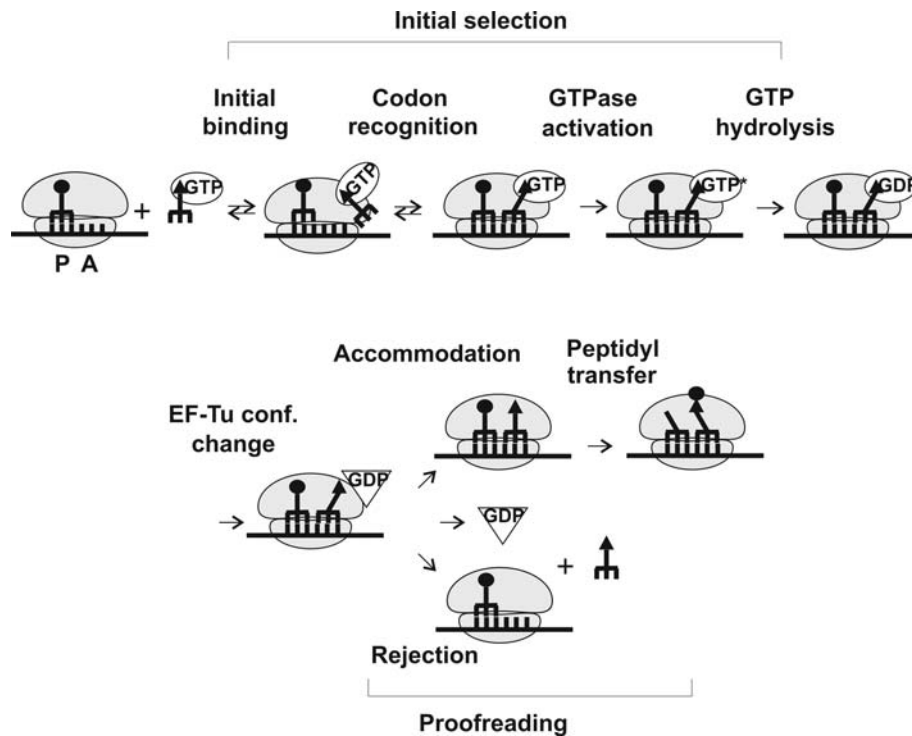
and rapid kinetic analysis (Fig. 6). The first step is the codon-independent formation of an unstable initial binding complex. During initial complex formation, EF-Tu-GTP-Aa-tRNA complexes interact with the ribosome in a stochastic fashion until one with the anticodon matching the A-site codon binds to form the codon-recognition complex. Codon recognition has two important consequences. One is stabilization of the complex by formation of base pairs between codon and anticodon as well as by ribosome interactions. The other is stimulation of the GTPase activity of EF-Tu by more than five orders of magnitude (GTPase activation). GTP hydrolysis leads to an extensive structural change in EF-Tu from the GTP- to the GDP-bound form, by which it loses the affinity for Aa-tRNA. The aminoacyl end of Aa-tRNA set free from EF-Tu-GDP moves into the peptidyl transferase center (accommodation) and takes part in peptide bond formation. EF-Tu-GDP dissociates from the ribosome and is reactivated by EF-Ts that acts as a guanine nucleotide exchange factor (GEF) and catalyzes the exchange of GDP for GTP. The functional cycle of eukaryotic eEF1 α is similar, with eEF1 β,γ acting as exchange factor. Because of the high concentrations of EF-Tu or eEF1 α present in the cell, practically all Aa-tRNA is



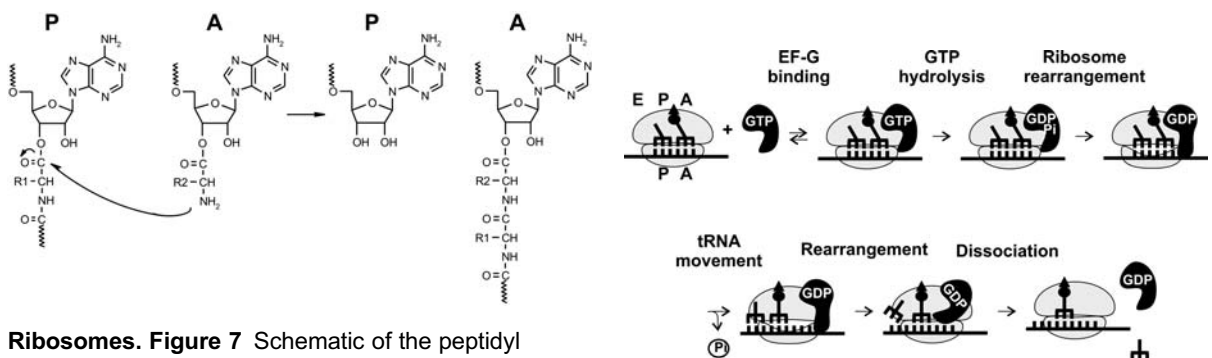
Ribosomes. Figure 5 Schematic of the elongation cycle.

bound in the ternary complex and immediately available for entering the ribosome.

The ribosome discriminates between correct and incorrect Aa-tRNAs or their complexes with EF-Tu-GTP with high accuracy. The average frequency of amino acid misincorporation *in vivo* is about 10^{-3} . This low level of missense errors is difficult to reconcile with the difference in the free energy of binding between codon-anticodon pairs that are correct, i.e. fully complementary, called cognates, and those that are nearly correct, having one mismatch, called near-cognates. These differences can be quite small (≤ 10 kJ/mol), predicting error frequencies of up to one out of 100 amino acids incorporated. Moreover, even this discrimination potential is unlikely to be used in full, because the high rates of the step(s) following codon recognition preclude the codon reading step reaching equilibrium. Thus, in order to reconcile an error frequency of at least 10^{-2} predicted for a single-step selection mechanism and the much lower error frequencies actually observed *in vivo*, the ribosome must either augment the precision of a single recognition step or use the same discriminatory interactions more than once in consecutive rejection steps that are separated by an irreversible, energy-dissipating



Ribosomes. Figure 6 Mechanism of Aa-tRNA binding to the ribosome.



Ribosomes. Figure 7 Schematic of the peptidyl transferase reaction between P site-bound Pept-tRNA and A site-bound Aa-tRNA.

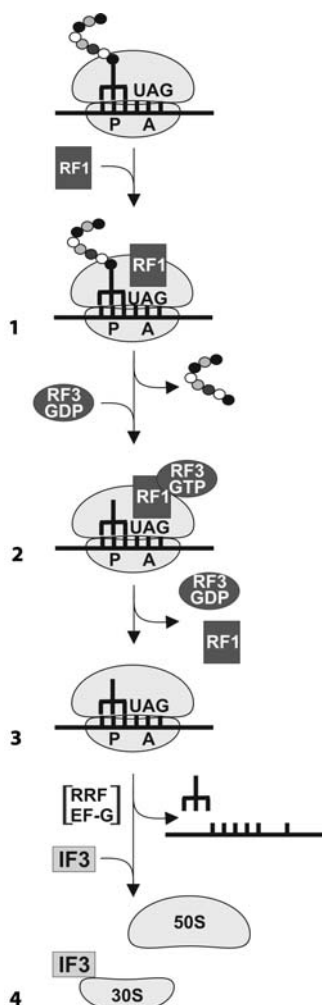
Ribosomes. Figure 8 Mechanism of translocation.

step, i.e. GTP hydrolysis. Both mechanisms operate on the ribosome in that there are two steps of selection (Fig. 6), i.e. initial selection of ternary complexes prior to GTP hydrolysis and proofreading selection of Aa-tRNA after GTP hydrolysis. In essence, Aa-tRNA selection is controlled kinetically rather than by the thermodynamics of codon-anticodon interaction. The mechanism involves induced fit. Correct codon recognition induces a conformational change in the decoding site that is driven by the formation of interactions between ribosomal residues and base pairs of the codon-anticodon duplex (Fig. 1B). The conformational change in the decoding site, in turn,

accelerates the rates of the chemical steps of A-site binding, i.e. GTP hydrolysis by EF-Tu and peptide bond formation.

Peptide Bond Formation

Catalysis of peptide bond formation is the enzymatic activity of the ribosome. The rate of the uncatalyzed reaction is extremely low and catalysis by the ribosome accelerates the reaction by seven orders of magnitude. The reaction proceeds *via* nucleophilic attack by the α -amino group of Aa-tRNA (A site) on the carbonyl carbon of the ester bond of Pept-tRNA (P site) and leads to a tetrahedral intermediate which subsequently



Ribosomes. Figure 9 Termination of prokaryotic translation. Step 1: RF1 binding and stop codon recognition; step 2: Peptidyl-tRNA hydrolysis, RF3 binding and GDP-GTP exchange; step 3: GTP hydrolysis, dissociation of RFs; step 4: tRNA-mRNA dissociation promoted by RRF-EF-G.

breaks down to form deacylated tRNA (P site) and a new, one amino acid longer, Pept-tRNA (A site) (Fig. 7). The peptidyl transferase (PT) center is located on the 50S subunit of the ribosome. 50S subunits largely depleted of protein retained some PT activity, suggesting that the activity might reside in 23S rRNA, although attempts to obtain protein-free 23S rRNA that was active in the PT reaction were unsuccessful. Atomic structures of the large ribosomal subunits show that the PT center is composed of RNA exclusively (Fig. 2), implying that the reaction is in fact catalyzed by RNA. Thus, the ribosome is a ribozyme, actually the largest known and the only natural one with polymerase activity.

Ribosomes. Table 3 Antibiotics acting on prokaryotic translation

Antibiotic	Action
Tetracycline	Inhibits Aa-tRNA binding
Neomycin	Induces misreading
Streptomycin	Induces misreading
Chloramphenicol	Inhibits peptidyl transferase
Fusidic acid	Inhibits translocation
Thiostrepton	Inhibits translocation
Erythromycin	Inhibits peptide elongation
Puromycin	Peptide release

The catalytic mechanism of peptide bond formation is not known in full detail. Positioning of the 3' CCA ends of both Pept-tRNA and Aa-tRNA by forming base pairs with 23S rRNA in the PT center has an important role, as shown by mutational and kinetic analysis. Several base pairs between CCA sequences of substrate analogues and bases in the P and A loops of 23S rRNA in the PT center were revealed by the crystal structures. Ribosomal residues may also contribute to chemical catalysis, e.g. general acid-base catalysis, although the evidence is not clear.

Translocation

The elongation cycle is completed by translocation, that is the synchronous movement of the two tRNAs and the mRNA on the ribosome by exactly one codon triplet. Following the movement, deacylated tRNA spontaneously dissociates from the E site, so that the ribosome in the posttranslocation state has Pept-tRNA in the P site and an empty A site presenting the codon to be decoded in the next round of elongation.

Translocation is catalyzed by elongation factor G in bacteria (eEF2 in eukaryotes) that hydrolyzes one molecule of GTP during the reaction. Rapid kinetic analysis has revealed the following reaction sequence (Fig. 8). Binding of EF-G-GTP to the pretranslocation complex triggers rapid GTP hydrolysis (unbound EF-G has no GTPase activity). Ribosomal residues involved in GTPase activation include proteins L7/L12 and possibly also elements of 23S rRNA, such as the sarcin loop. GTP hydrolysis induces a conformational change in the factor that is coupled to a rearrangement of the ribosome that leads to rapid tRNA movement and the release of Pi from ribosome-bound EF-G-GDP-Pi. The sequence is completed by further conformational changes and, finally, dissociation of EF-G-GDP and deacylated tRNA. When GTP is replaced with

non-hydrolysable GTP analogues, translocation is much slower. This suggests a mechanochemical function of EF-G, coupling the energy of GTP hydrolysis to conformational work on the ribosome.

Termination

The elongation cycle is repeated until a termination codon appears in the decoding site. This is the signal for termination, during which the completed protein is hydrolytically released from Pept-tRNA (Fig. 9). Termination is catalyzed by termination or release factors (RF). In bacteria, RF1 recognizes the termination codons UAG and UAA, while RF2 is specific for UGA and UAA (in eukaryotes, one factor, eRF1, recognizes all three termination codons). The specificity of the release factors in distinguishing termination codons from sense codons is extremely high, hence the very low frequency of premature termination on sense codons. Binding of RF1/2 to a termination codon in the A site leads to hydrolysis of Pept-tRNA. Apparently, the factor binds to the termination codon in the decoding site and, at the same time, contacts the peptidyl transferase center and induces the hydrolytic activity, possibly by taking part in the activation and/or positioning of the hydrolytic water molecule. After release of the peptide, another release factor, RF3, a GTPase, promotes the dissociation of RF1/2 from the ribosome in a process that requires GTP hydrolysis. The posttermination complex releases the deacylated tRNA from the P site and dissociates into subunits, assisted by initiation factor IF3 that binds to the small subunit and stabilizes it in the unbound form. In bacteria, ribosome recycling requires another factor, ribosome-recycling factor (RRF), that together with EF-G promotes the dissociation of the posttermination complex and the formation of free ribosomal subunits.

Clinical Relevance

The translation apparatus of bacteria is the target for many natural antibiotics that affect certain steps of ribosomal protein synthesis; examples are listed in Table 3. Most of these antibiotics inhibit ribosome function by binding to ribosomal RNA. For instance, certain aminoglycosides (neomycin, paromomycin, gentamicin) bind to 16S rRNA in the decoding site and impair the accuracy of decoding, thereby increasing amino acid misincorporation. Streptomycin, from another group of aminoglycosides, binds to another site in 16S rRNA, but has the same effect. Tetracycline also binds in this region and interferes with productive binding of Aa-tRNA. There are many inhibitors of the peptidyl transferase reaction (e.g. chloramphenicol) that bind to the peptidyl transferase center of 23S rRNA. The binding site of erythromycin is located in the peptide exit tunnel of the 50S subunit and binding of the antibiotic precludes peptide elongation.

Thiostrepton binds to a single site in 23S rRNA and inhibits translocation by blocking conformational transitions of the ribosome that are important for the reaction.

The therapeutic use of many antibiotics is restricted at present, because resistant pathogens have evolved. Much effort is therefore spent to develop new antibiotics that act on translation or other targets.

References

1. Stillman B (2002) The Ribosome. Cold Spring Harbor Symposia on Quantitative Biology LXVI, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY,
2. Ramakrishnan V (2002) Ribosome structure and the mechanism of translation. *Cell* 108:557–572
3. Rodnina MV, Wintermeyer W (2001) Fidelity of aminoacyl-tRNA selection on the ribosome: kinetic and structural mechanisms. *Annu Rev Biochem* 70:415–435
4. Moore PB, Steitz TA (2002) The involvement of RNA in ribosome function. *Nature* 418:229–235
5. Yusupov MM, Yusupova GZ, Baucom A et al (2001) Crystal structure of the ribosome at 5.5 Å resolution. *Science* 292:883–896

Ribozyme

Definition

The word “ribozyme” is derived from the words ribonucleic acid (RNA) and enzyme, and it denotes an RNA molecule with catalytic properties. Naturally occurring catalytic RNAs catalyze phosphoryl transfer reactions, which require the activation of a 2′-OH group (self-splicing introns; hammerhead ribozyme; hepatitis delta ribozyme; hairpin ribozyme) or a water molecule (RNase P) for nucleophilic attack of a neighboring phosphodiester bond. The ribosome constitutes the first example of a natural ribozyme with (amino acid) polymerase activity. Synthetic ribozymes catalyzing a variety of chemical reactions have been obtained by *in-vitro* selection from random-sequence RNA.

- Catalytic RNA
- Ribosomes

Rickets

Definition

Rickets is a childhood disorder that disturbs normal bone formation (ossification). Rickets is characterized

by deficient mineralization of osteoid tissue with associated skeletal deformities. Rickets is primarily caused by a lack of vitamin D and calcium. A familial form of rickets is characterized by low blood phosphate level, defective intestinal absorption of phosphate, and unresponsiveness to vitamin D.

► [Hyper- and Hypoparathyroidism](#)

Rigor

Definition

Specific force-generating actomyosin complex which forms in the absence of ATP.

► [Muscle Contraction](#)

RING Finger Domain

Definition

RING finger domain is defined structurally by two interleaved metalcoordinating sites. The consensus sequence for the RING finger is: CX₂CX(9–39)CX(1–3)HX(2–3)C/HX₂CX(4–48)CX₂C. The cysteines and histidines represent two metal-binding sites. The first, second, fifth and sixth of these bind one zinc ion and the third, fourth, seventh and eighth bind the second zinc ion. This zinc-finger containing structure is characteristic for ubiquitin-ligases. Relatives of this structural motif are the U-box and the PHD domain.

► [Adherens Junctions](#)

► [Sumoylation](#)

► [Ubiquitination](#)

RIP

Definition

RIP is the abbreviation for regulated intra-membranous proteolysis. During this process, ► [NEXT](#) gets cleaved in the transmembrane domain to release the intracellular domain of Notch.

► [Notch Pathway](#)

RISC

Definition

RISC stands for RNA-induced silencing complex. It is a multiprotein complex responsible for recognizing double-stranded RNA, target mRNA, and for cleavage of the target resulting in its degradation.

► [MicroRNA](#)

► [RNA Interference in Mammalian Cells](#)

► [RNA-Induced Silencing Complex](#)

RITS

RNA-induced initiation of transcriptional gene silencing

RLGS

Definition

RLGS stands for Restriction Landmark Genomic Scanning. RLGS is a two dimensional gel electrophoresis that allows the scanning of genomes for changes in DNA methylation patterns.

► [CpG Islands](#)

RMCE

Definition

‘Recombinase-mediated cassette exchange’ (RMCE) is a technique that allows for site-specific integration of exogenous DNA into the genome of ES cells. This method is most commonly used to mediate the exchange of a selectable marker pre-localized in the genome with a transgene, via Cre-mediated double reciprocal recombination between inverted pairs of *loxP* sites (present in the vector containing the transgene and targeted to the locus).

► [Cre/Lox P Strategies](#)

► [Mouse Genomics](#)

RNA

Definition

RNA stands for ribonucleic acid. It mostly occurs as a single stranded molecule, and is an intermediate between DNA (genes) and proteins, or functional as such. It is instable compared to DNA due to its biochemical specificities.

- ▶Fragile X Syndrome
- ▶Full Length cDNA Sequencing
- ▶mRNA
- ▶Nucleotide Biosynthesis
- ▶RNA Polymerase
- ▶rRNA
- ▶tRNA

RNA Capping

- ▶RNA Polymerase II Transcription

RNA Dominance

Definition

RNA dominance refers to a model explaining how a 3'-UTR expansion mutation results in a dominant neuromuscular disorder; transcripts containing long tri- or tetranucleotide repeats exert toxic effects on RNA metabolism (e.g. splicing and transport) such as observed in myotonic dystrophy (DM1) and DM2.

- ▶Myotonic Dystrophy

RNA Export

ROZANNE M. SANDRI-GOLDIN

Department of Microbiology and Molecular Genetics,
University of California, Irvine, CA, USA
rmsandri@uci.edu

Synonyms

RNA transport: nucleo-cytoplasmic trafficking

Definition

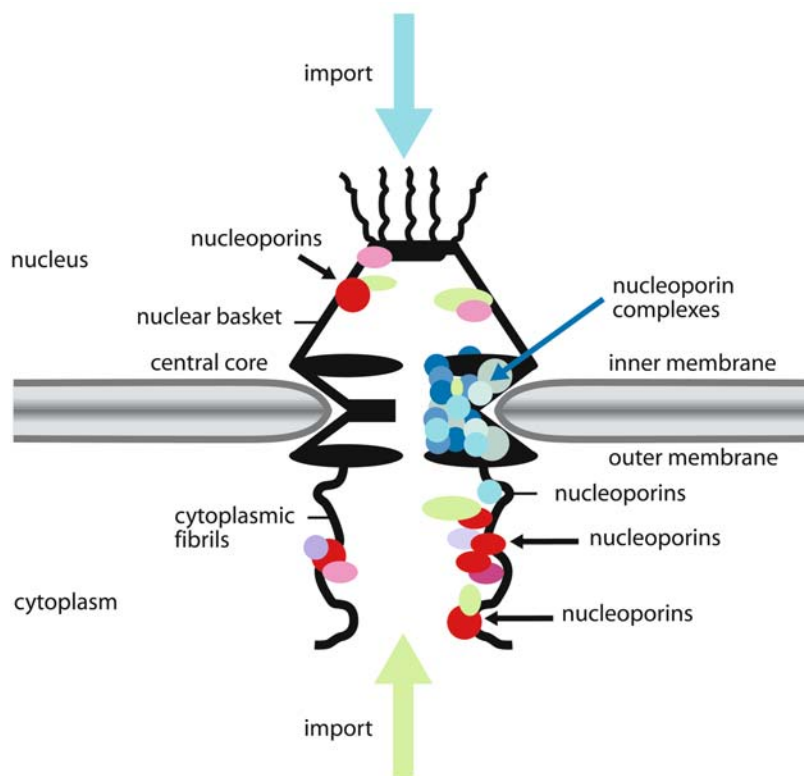
In eukaryotic cells, RNA synthesis and processing occur in the nucleus, whereas translation of RNA into protein occurs in the cytoplasm. The nucleus and cytoplasm are separated by the ▶nuclear envelope, which is comprised of two membrane lipid bilayers. ▶Nuclear pore complexes (NPC) are large protein assemblies that form channels across the nuclear envelope (1). Bidirectional transport of macromolecules occurs through the NPC and is mediated by soluble transport receptors (1, 2). In metazoans, the NPC is composed of about 30 or more different proteins that are termed ▶nucleoporins, which often contain clusters of hydrophobic phenylalanine-glycine (FG) dipeptide repeats (1). The FG repeats interact with nuclear transport receptors, which in turn are bound either directly or indirectly through adapter proteins to the RNA that is being transported and which is referred to as ▶cargo. The receptor-cargo complex binds to different nucleoporins as it is translocated across the channel of the NPC into the cytoplasm (1, 2, 3) (Fig. 1).

Characteristics

There are two major classes of RNAs, noncoding and coding. The noncoding RNAs include ▶ribosomal RNA (rRNA) and tRNA, which are involved in translation and small nuclear RNAs (snRNAs), which are involved in splicing. Coding RNA is also called messenger RNA or mRNA. Noncoding RNAs and mRNAs use different export pathways that are mediated by different ▶export receptors.

Export of Noncoding RNA

Noncoding RNAs have been found to depend on members of the karyopherin or importin/exportin family of transport receptors for their export (2). A characteristic of this group of export factors is their dependence on the small GTPase Ran. A gradient of ▶Ran-GTP is required to promote the directionality of export (2). In the nucleus, ▶karyopherins bind the cargo in the presence of the GTP-bound form of Ran, Ran-GTP. The Ran specific guanine nucleotide exchange factor, RanGEF is localized in the nucleus thus keeping levels of Ran-GTP high. The Ran-specific GTPase-activating protein, RanGAP is localized in the cytoplasm so that Ran exists there predominantly as inactive Ran-GDP. Conversion of Ran-GTP to Ran-GDP in the cytoplasm by RanGAP results in the release of the bound cargo and the karyopherin is recycled to the nucleus to bind and export more cargo RNA (2). Pre-rRNA is synthesized by RNA polymerase I after which pre-rRNA undergoes processing to generate the mature 18S, 5.8S and 28S rRNAs. The fourth rRNA is transcribed by RNA polymerase III. In a highly ordered process, the 28S, 5.8S and 5S rRNAs are assembled in



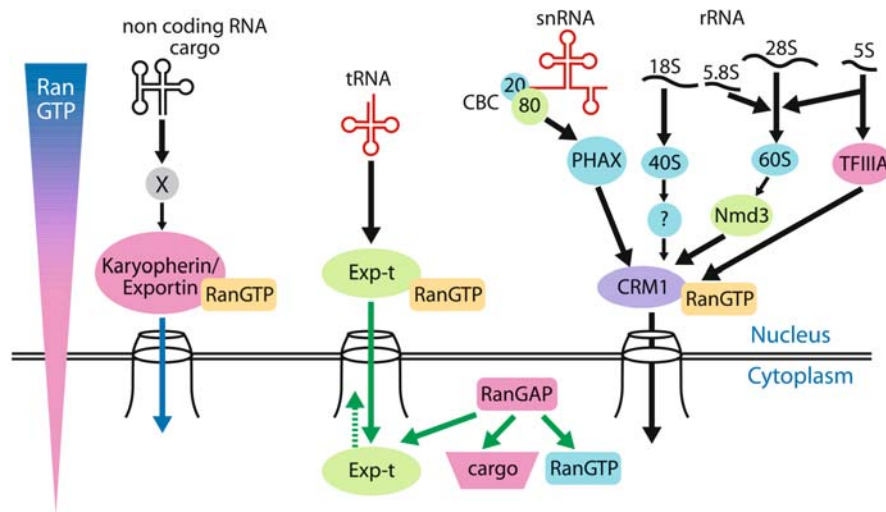
RNA Export. Figure 1 A cross section of the nuclear pore complex (NPC) is depicted, which shows the inner and outer membranes of the nuclear envelope. The nuclear basket and cytoplasmic fibrils are lined with nucleoporins. In higher eukaryotes, there are more than 30 different nucleoporins, many of which are present in multiple copies. Nucleoporin complexes are shown in the central core of the NPC.

a complex along with 50 ribosomal proteins. This complex is the 60S pre-ribosomal subunit. The 18S rRNA is assembled with 33 ribosomal proteins to yield the 40S pre-ribosomal subunit. The 60S and 40S subunits are exported separately to the cytoplasm where they form the translating ribosome. Export of both subunits is dependent on the karyopherin termed CRM1 or exportin 1 (2). CRM1 does not bind to the RNA cargo directly but instead interacts with export factors that bind to the RNA cargo and to CRM1 simultaneously. The interaction with CRM1 occurs through **nuclear export signals** (NES) on the export factors, which are comprised of short hydrophobic stretches of amino acids that are rich in leucine residues. TFIIIA, which contains an NES, has been shown to bind 5S rRNA to bridge it to CRM1. In yeast, a factor termed Nmd3 has been found to bring the 60S pre-ribosomal subunit to CRM1 but the human homologue of Nmd3 has not been identified. Further, while the 40S subunit also depends on CRM1 for its export, the factor responsible for bridging it to CRM1 has not yet been identified. CRM1 bound to Ran-GTP

has been shown to bind to a number of nucleoporins to translocate the pre-ribosomal subunits through the NPC to the cytoplasm. Upon Ran-GTP hydrolysis to Ran-GDP, the cargo is released and CRM1 shuttles back to the nucleus (2) (Fig. 2).

Pre-tRNA is transcribed by RNA polymerase III, after which a number of post-transcriptional modifications occur to generate mature tRNA. Export of tRNA is mediated by another karyopherin family member termed exportin-t (Exp-t) (2). Additional export factors are not required for tRNA export because Exp-t binds directly to highly structured, mature tRNAs but only in the presence of Ran-GTP. Like other karyopherins, Exp-t can bind to specific nucleoporins and can shuttle between the nucleus and cytoplasm (2) (Fig. 2).

Small nuclear ribonucleoprotein particles (snRNPs) are RNA-protein complexes that play critical roles in pre-mRNA splicing. The small nuclear RNA (snRNA) components of snRNPs are U-rich and thus are termed, U1, U2, U4, U5 and U6 snRNAs. All but U6 are transcribed by RNA polymerase II; U6 is transcribed by RNA polymerase III. Although snRNAs function in



RNA Export. Figure 2 Karyopherin/exportin family members mediate export of non-coding RNAs. The general scheme for export of non-coding RNAs is depicted on the left. The RNA molecule is bound by RNA adapter protein (s), labeled as factor X, which brings the RNA cargo to a member of the karyopherin/exportin family. The karyopherin/exportin must bind to Ran-GTP to bind the cargo. Upon entry to the cytoplasm, Ran GAP converts Ran-GTP to Ran-GDP and the cargo is released. The karyopherin/exportin is re-imported to the nucleus. For tRNA, the export receptor is exportin-t or Exp-t, which binds directly to the tRNA and does not require adapter proteins. In contrast, snRNAs are bound at the 5' cap by cap binding complex (CBC), a heterodimer that is comprised of 80 kD and 20 kD subunits. CBC in turn binds to phosphorylated adapter of RNA export (PHAX), which binds to CRM1, a karyopherin family member. Ribosomal RNAs (rRNA) transcribed by RNA polymerase I undergo processing to form the 18S, 5.8S and 28S rRNAs. A set of ribosomal proteins assembles on the 18S rRNA forming the 40S pre-ribosomal complex, which interacts with CRM1-Ran-GTP through an adapter that has not been identified. The 5.8S, 28S and 5S rRNAs form the 60S pre-ribosomal complex upon binding by 50 ribosomal proteins. In yeast, the adapter is Nmd3, but the human counterpart is not known. Some of the 5S rRNA binds to TFIIA, which bridges it to CRM1 for export.

pre-mRNA splicing in the nucleus, they must be exported to the cytoplasm to acquire the protein components of snRNPs. The assembled snRNPs are subsequently imported into the nucleus where they function in spliceosome assembly. Export of snRNAs to the cytoplasm also requires the karyopherin CRM1 (2). A factor termed PHAX, which stands for phosphorylated adapter for RNA export, was found to bind snRNAs and the Ran-GTP bound form of CRM1. It was further shown that a heterodimer that binds to the 7-methylguanosine cap that is added co-transcriptionally to all RNA polymerase II transcripts, binds to PHAX-CRM1-Ran-GTP. The heterodimer, termed cap-binding complex (CBC) is composed of a 20 kD protein (CBP20) and an 80 kD protein (CBP80). It has been demonstrated that CBC bound to the cap of snRNAs is required for their export, along with PHAX and CRM1 (2) (Fig. 2).

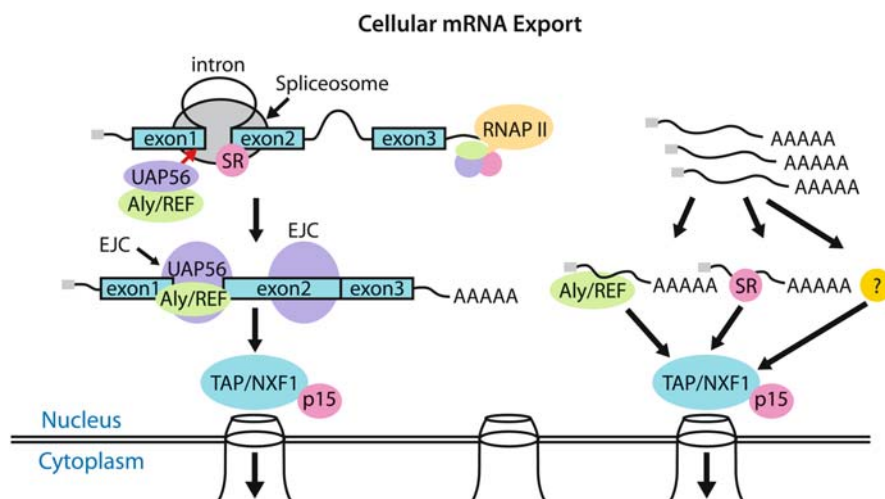
Export of Coding RNA

Export of cellular mRNA is independent of Ran and karyopherins. Instead, mRNA export is mediated by a conserved heterodimeric receptor that is structurally unrelated to the karyopherin family and notably, which

does not depend on Ran-GTP to bind cargo (3, 4). A preponderance of evidence has shown that the mRNA export receptor is a protein initially termed TAP and later termed NXF1 for nuclear export factor 1 (3, 4). The homologue of [TAP/NXF1](#) in yeast is called Mex67p. TAP/NXF1 interacts with p15, which is also called NXT1. The yeast orthologue is termed Mtr2p. Mutational analysis of Mex67p in yeast and knock-down of expression of TAP/NXF1 by RNA interference in metazoans have demonstrated that these are the receptors that are required for mRNA export. TAP/NXF1 interacts with nucleoporins in the NPC through a C-terminal domain and binding of TAP/NXF1 to its heterodimeric partner p15 appears to increase the interaction of TAP/NXF1 with the NPC. Further, TAP/NXF1-p15 does not bind directly to the mRNA cargo but instead requires RNA adapter proteins to bridge the interaction. Moreover, recent studies have revealed a functional coupling between mRNA export and upstream events in RNA biogenesis (4). Eukaryotic mRNAs are transcribed by RNA polymerase II. The pre-mRNAs are processed after synthesis in the nucleus by capping at the 5' end, cleavage and polyadenylation to form the 3' end and splicing to

remove intervening sequences. These events appear to occur co-transcriptionally and proteins required for each processing event interact with the highly conserved C-terminal domain (CTD) of RNA polymerase II. The CTD serves as a platform for RNA processing factors so that these factors can travel with the elongating polymerase until they reach the appropriate sequences on the nascent transcript where their action is required. A theme that has been emerging is that mRNA export is coupled to other aspects of mRNA biogenesis, probably to insure that only properly processed, mature mRNA exits the nucleus (4). This is because export of an incompletely processed pre-mRNA would have profound effects on the protein that was translated. In fact, unspliced mRNAs are retained in the nucleus. Nuclear retention is caused by interactions of splicing factors with splice site consensus sequences in partially processed transcripts and serves to ensure that translation occurs only on mature mRNAs. In metazoans, nuclear export of mature mRNAs has also been linked to pre-mRNA splicing (4, 5). Early reports showed that certain mRNAs transcribed from cDNAs failed to exit the nucleus and, therefore, did not express protein, whereas the same mRNAs expressed from intron-containing constructs could enter the cytoplasm and be efficiently translated. Further, spliced RNAs were more efficiently exported from *Xenopus* oocyte nuclei than identical RNAs transcribed from cDNAs (5). The basis of this

connection was uncovered with the discovery of a protein complex, which is deposited on pre-mRNAs undergoing splicing, at a specific position about 20 nucleotides upstream of exon-**▶exon junctions**. This exon junction complex or EJC consists of at least six proteins, which have been shown to function in pre-mRNA splicing, nuclear mRNA export, RNA localization and mRNA surveillance (5). One of these proteins, initially termed Aly and then subsequently called REF for RNA export factor, interacts directly with TAP/NXF1. **▶Aly/REF** is recruited to pre-mRNA sites near exon junctions by a helicase termed **▶UAP56**, which functions in spliceosome assembly and also appears to have role in mRNA export beyond the recruitment of Aly/REF (4, 5, 6). A role for Aly/REF in the export of mRNAs is supported by observations that Aly/REF remains tightly bound to the spliced mRNA, antibodies to Aly/REF that prevent its interaction with RNA reduced export of mRNA in *Xenopus* oocytes and excess Aly/REF increased the efficiency of mRNA export in mammalian cultured cells. However, knock-down of Aly/REF expression in *Drosophila* cells by RNA interference showed that Aly/REF was not absolutely required for the export of mRNAs, suggesting that there are additional **▶export adapters** with redundant functions. In contrast, knock-down of TAP/NXF1 showed that it was absolutely required for mRNA export (4, 5, 6) (Fig. 3).



RNA Export. Figure 3 Cellular mRNA export is mediated by the heterodimeric export receptor TAP/NXF1-p15. TAP/NXF1 is structurally unrelated to karyopherins and does not require Ran-GTP. Export of mRNA is linked to earlier steps in RNA biogenesis. Most metazoan mRNAs contain introns. Splicing occurs co-transcriptionally and this is depicted on a pre-mRNA that is being transcribed by RNA polymerase II (RNAP II). Processing factors interact with RNAP II C-terminal domain and travel with the elongating complex to sites where they are needed. An assembled spliceosome is shown and UAP56, a splicing protein recruits the export adapter Aly/REF to a position just upstream of the exon 1-exon 2 junction. Aly/REF is part of the exon junction complex (EJC). The mature mRNA, which is capped at the 5' end, polyadenylated at the 3' end and decorated with EJCs at each exon junction, is exported via TAP/NXF1-p15, which interacts with the NPC. Some cellular mRNAs that are intronless have been shown to be exported by Aly/REF. Others have been found to use SR proteins as adapters. Other export adapters probably also exist.

It should be noted that although splicing can enhance RNA export from intron-containing genes, it is not an unconditional requirement because export factors can interact with RNA independently of splicing. For example, naturally intronless transcripts, such as those encoding histones, contain specific sequences that recruit export factors independently of splicing. Interestingly, two members of a highly conserved family of serine-arginine-rich splicing factors, termed **SR proteins**, have been shown to interact with TAP/NXF1 and to be involved in the export of intronless mRNAs. SRp20 and 9G8, two SR proteins that play essential roles in splicing also appear to have roles in mRNA export (5) (Fig. 3). There may be other adapter proteins that bring mature mRNAs to TAP/NXF1. In contrast to metazoans, in yeast, where fewer than 5% of genes encode introns, recruitment of export proteins has been shown to occur co-transcriptionally. Mex67p, the homologue of TAP/NXF1 is the export receptor and Yra1, the homologue of Aly/REF is the export adapter that is required for yeast mRNA export (4).

Clinical Relevance

Because export of RNA to the cytoplasm is essential for the cell to make protein, there are no known human diseases that result from defects in RNA export. In fact, a general strategy for the replication of animal viruses is to disrupt the export of cellular RNA and to hijack components of export pathways to favor viral gene expression and replication. Thus, part of the pathogenesis of viral infection results from the shut off of host gene expression. Viruses with a DNA genome, such as herpesviruses, or which go through a DNA intermediate, such as lentiviruses and retroviruses, replicate in the nucleus and these viruses have evolved strategies that give an export advantage to viral RNAs over cellular RNAs. Human immunodeficiency virus (HIV), a lentivirus, and retroviruses encode essential viral proteins within introns of viral pre-mRNAs, which are transcribed by the cellular RNA polymerase II. Thus, unspliced and partially spliced transcripts must be exported to the cytoplasm. To circumvent retention in the nucleus, HIV encodes a protein termed Rev, which binds to a highly structured region within an intron in HIV pre-mRNA, termed the Rev response element or RRE. Rev contains an NES and is able to bind to CRM1 (2, 6). Thus, Rev facilitates the export of unspliced HIV transcripts through the CRM1 pathway rather than the TAP/NXF1 pathway used by cellular mRNA. In contrast, unspliced pre-mRNA in retroviruses is exported *via* the TAP/NXF1 pathway. A part of the intronic RNA forms a stem loop structure, termed the constitutive transport element or CTE and TAP/NXF1 binds directly to the CTE without the need

of an adapter protein (2, 5, 6). In fact, expression of excess CTE-containing RNA can interfere with export of cellular mRNA by saturating TAP/NXF1 (5). Herpes simplex virus type 1 (HSV-1) presents a different situation. Most of its mRNAs are intronless and thus do not undergo splicing and therefore do not acquire EJC. HSV-1 encodes a protein termed ICP27, which interacts with splicing complexes resulting in the inhibition of host splicing (6). Thus, host transcripts are retained in the nucleus. ICP27 then recruits Aly/REF to sites of viral transcription providing intronless viral mRNAs access to the TAP/NXF1 pathway (6). Although influenza virus has a RNA genome rather than a DNA genome, it replicates in the nucleus of infected cells. Influenza also interferes with cellular mRNA export through a viral protein called NS1. Inhibition of transport by NS1 appears to be due to its ability to impair 3' end formation of cellular mRNAs through interaction with two cellular factors that are required for polyadenylation (2). Finally, RNA viruses that replicate in the cytoplasm have also been shown to disrupt cellular nucleocytoplasmic trafficking. Poliovirus appears to target two nucleoporins in the NPC causing their degradation during infection. As a result receptor: cargo complexes do not dock at the NPC of poliovirus-infected cells (7). In this case, cellular proteins that are normally nuclear but which enhance poliovirus replication are trapped in the cytoplasm. In sum, RNA export to the cytoplasm is essential for gene expression. Complex and separate pathways that involve different soluble receptors and redundant export adapters have evolved to ensure that all types of cellular RNAs are efficiently exported to the cytoplasm.

References

1. Bednenko J, Cingolani G, Gerace L (2003) Nucleocytoplasmic transport: navigating the channel. *Traffic* 4:127–135
2. Cullen BR (2003) Nuclear RNA export. *J Cell Sci* 116:587–597
3. Izaurralde E (2002) A novel family of nuclear transport receptors mediates the export of messenger RNA to the cytoplasm. *Eur J Cell Biol* 81:577–584
4. DiMauro C, Ullman KS (2004) Nucleocytoplasmic transport: integrating mRNA production and turnover with export through the nuclear pore. *Mol Cell Biol* 24:3069–3076
5. LeHir H, Nott A, Moore MJ (2003) How introns influence and enhance eukaryotic gene expression. *Trends Biochem Sci* 28:215–220
6. Sandri-Goldin RM (2004) Viral regulation of mRNA export. *J Virol* 78:4389–4396
7. Gustin KE (2003) Inhibition of nucleocytoplasmic trafficking by RNA viruses: targeting the nuclear pore complex. *Virus Res* 95:35–44

RNA Expression Profiling

Definition

RNA expression profiling comprises the determination of the pattern of genes expressed i.e. transcribed, under specific circumstances or in a specific cell by quantitative analysis of their RNA expression level.

► [Protein Microarrays as a Tool for Protein Profiling and Functional Proteomics](#)

RNA Interference

Synonym

RNAi

Definition

RNA interference (RNAi) (also called “double-stranded RNA mediated interference”) refers to the mechanism by which double-stranded RNA (dsRNA) specifically suppresses the expression of a gene bearing its complementary sequence. The dsRNA is processed into small interfering RNAs (siRNAs), assembled in endoribonuclease-containing complexes known as RNA-induced silencing complexes (RISCs) that consequently cleave, thereby destroy the transcript of the silenced gene, namely the cognate RNA. The term “RNA interference” was first used by researchers studying the worm *Caenorhabditis Elegans*.

► [C. Elegans as a Model Organism for Functional Genomics](#)

► [Catalytic RNAs](#)

► [Drosophila as a Model Organism for Functional Genomics](#)

► [Drosophila Model of Cardiac Disease](#)

► [High-Throughput Approaches to the Analysis of Gene Function in Mammalian Cells](#)

► [Large-Scale Homologous Recombination Approaches in Mice](#)

► [Micro RNA's](#)

► [Morpholino Oligonucleotides, Functional Genomics by Gene ‘Knock-down’](#)

► [RNA Interference in Mammalian Cells](#)

► [Transposons](#)

RNA Interference in Mammalian Cells

REINHOLD SCHÄFER, ANJA SCHRAMME

Laboratories of Molecular Tumour Pathology and of Functional Genomics, Charité, University Medicine Berlin, Berlin, Germany
reinhold.schaefer@charite.de

Synonyms

Double-stranded RNA-mediated gene silencing; Small interfering RNA-mediated gene silencing; Sequence-specific post-transcriptional gene silencing by double-stranded RNA

Definition

► [RNA interference](#) (RNAi) describes a mechanism of gene silencing in which double-stranded RNA (dsRNA) with identical sequence to a target mRNA mediates the destruction of the target. Target mRNA degradation is catalyzed by the multi-protein complex ► [RISC](#) (RNA-induced silencing complex), which recognizes dsRNA and cognate mRNA. This results in a reduction of the encoded protein and in the suppression of gene function (for review see (1, 2)).

Characteristics

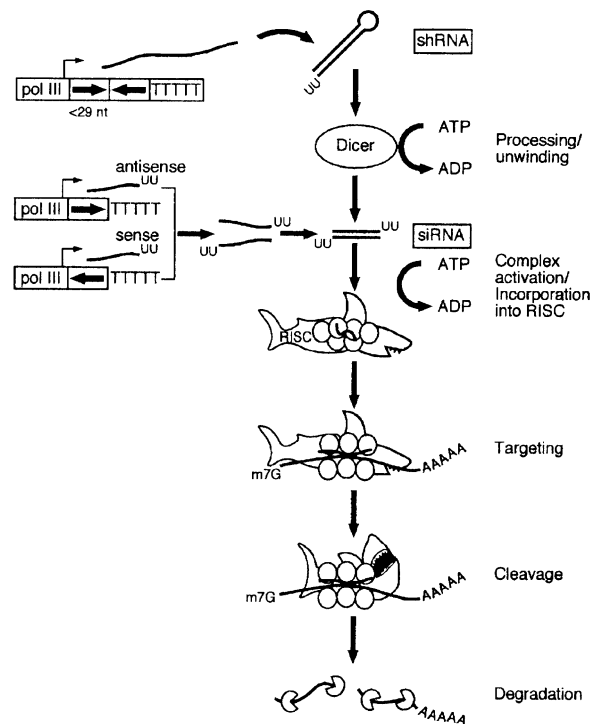
RNAi was first detected in the nematode ► [Caenorhabditis elegans](#) and in plants and serves as a defence mechanism for the protection of cells against parasitic nucleic acids such as viral genomes. Despite its original recognition as a biological oddity, RNAi was very rapidly applied as a powerful genetic tool and genome-wide RNAi screens of gene functions were performed in *C. elegans*. For example, the factors regulating fat metabolism were analyzed systematically by RNAi (3). RNAi has gained particular attention among geneticists, molecular biologists and scientists in the field of experimental therapy for two reasons. Firstly, after sequencing of the human, mouse and rat genomes, simple experimental approaches are required that permit functional analysis of the vast number of annotated genes and of yet ill-defined potential genes. Secondly, genome-wide screenings for genes specifically expressed in normal development and physiology and in disease based on ► [microarray technology](#) and on alternative approaches such as ► [serial analysis of gene expression](#) (SAGE) or differential expression cloning have yielded large numbers of disease-associated candidate genes as well as ► [gene clusters](#) that need to be tested for functional involvement.

RNA interference became particularly prominent in the field of molecular medicine following reports from Tuschl's group in Germany and from Morgan's group in the USA who showed that mammalian genes can be effectively silenced by virtue of incorporation of short 19–21 nucleotide dsRNAs into somatic cells. Earlier reports had already suggested that sequence-specific gene silencing occurred following injection of 500 nucleotide long dsRNA into mouse embryos and embryonal cell lines (for review see (1)). Unlike in invertebrates and plants however, dsRNA of >30 nucleotides in length provoked antiviral/interferon

responses in somatic mammalian cells and induced a generalized translational repression and cell death. This non-specific response occurred even in cells devoid of effector proteins of the interferon pathway such as ▶PKR (dsRNA-dependent protein kinase) and IRF1 (▶interferon-regulatory factor 1). dsRNA <30 nucleotides in length, transiently administered to somatic mammalian cells using commonly available transfection procedures, did not induce non-specific responses. Rather, small interfering RNAs (▶siRNAs) mediated efficient silencing of the cognate mRNA. The siRNAs were synthesized chemically as 19 nucleotide fragments with two 3'-overhangs on each strand. An alternative to the costly chemical synthesis is *in vitro* transcription from DNA oligonucleotides using T7 polymerase followed by annealing of the two RNA strands. Chemically or enzymatically synthesized siRNAs are transiently transfected into mammalian cells using lipofection or electroporation or a combination of both methods. Directly transfected siRNAs exhibit efficient silencing at concentrations ranging from 0.05 to 20 nM. A typical transient siRNA-mediated silencing effect may last for 2–7 days, depending on the abundance of the cognate mRNA and protein, their stability and the efficacy of *de novo* transcription. The silencing effect is also diluted by cell division in the manipulated cells. The transient gene knockdown suffices for analyzing phenotypic effects, provided that rapid and robust cellular readouts for gene function are available. The currently accepted mechanism of RNA interference in mammalian cells is depicted in Fig. 1.

An example for siRNA-mediated knockdown of three closely related mRNAs encoding Raf-kinase isoforms is shown in Fig. 2. Many growth factor-receptor-mediated signals converge on Raf effector proteins. Raf-kinases may be activated by mutation in cancer and are being exploited as targets for therapeutic intervention. The isoform-specific knockdown demonstrates the high specificity of RNAi-mediated gene silencing and emphasizes its power for functionally dissecting complex signal transduction pathways as well as its potential for experimental therapy.

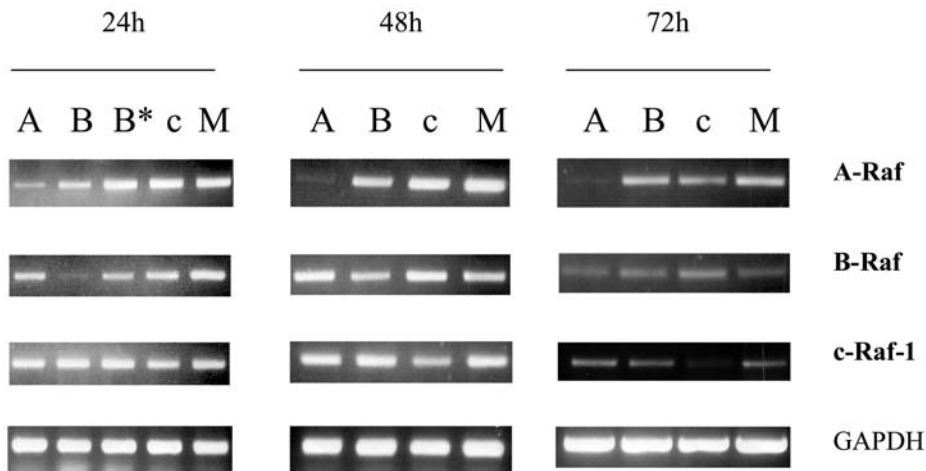
As an alternative to the incorporation of exogenous dsRNAs by common transfection procedures, siRNAs can be synthesized directly in cells through the expression of short hairpin RNAs (shRNAs). Such endogenously expressed shRNAs were initially detected in *C. elegans*. shRNAs fold back on themselves, creating a region of dsRNA and a loop structure. The mature siRNAs are produced by enzymatic removal of the loop. To overcome the disadvantages of transient siRNA expression, various retroviral expression systems have been developed. Most expression systems employ polymerase III promoters to drive the interfering RNA expression. It is noteworthy, however, that



RNA Interference in Mammalian Cells. Figure 1 A model for RNA interference in mammalian cells (ref.1). This model is deduced from dsRNA-induced silencing mechanisms which were commonly detected in plants, fungi and metazoans. The silencing mechanism begins with the processing of dsRNA silencing triggers into siRNAs by Dicer, a nuclease belonging to the family of RNase III proteins. The triggers are expressed from RNA polymerase III promoters. siRNAs bind to a multiprotein nuclease complex called RISC (RNA-induced silencing complex). RISC targets the cognate mRNA resulting in cleavage and degradation. Small interfering RNAs synthesised chemically or enzymatically by *in vitro* transcription using T7 polymerase from DNA templates can be incorporated directly into RISC, obviating Dicer activity. In contrast to the silencing mechanism in *C. elegans*, RNAi is not a self-propagating effect in mammalian cells. In nematodes, siRNAs can prime RNA-dependent RNA polymerases (RdRPs) along the targeted mRNA sequence to produce dsRNA homologous to upstream sequences of the initially targeted sequence. This results in an amplification of the RNAi effect throughout the nematode.

transient gene silencing may reveal the immediate effects of functional gene loss, while stable suppression of gene activity may allow adaptation processes in variant cell lines to occur.

Algorithms for designing siRNAs are available on-line from various public and commercial web sites. The initially established rules for siRNA selection had



RNA Interference in Mammalian Cells. Figure 2 Transient siRNA-mediated silencing of Raf-kinase isoforms in transformed mouse NIH/3T3 cells. Optimal silencing effects were detectable at 48h (**A-Raf**), 24 (**B-Raf**) and 48-72h (**c-Raf-1**) post-transfection. mRNA levels related to non-targeted isoforms remained unaffected at the indicated time intervals. **GAPDH**, RT-PCR control. siRNA duplexes were transfected twice in 24 h intervals using Oligofectamine. RNA was prepared from transfected cells 24-72 hr post-transfection and analysed by RT-PCR.

RNA oligonucleotides used to generate dsRNA:

A, A-Raf, sense: 5'-AACAGCGGGAGCGGAAGUCCUdTdT-3'

A-Raf antisense: 5'-AAGGACUUCGGCUCGCCUGUdTdT-3';

B, B-Raf sense: 5'-AACGTGTCCCAGCCAATTGGTCTGTCTC-3'

B-Raf antisense 5'-AAACCAATTGGCTGGGACACGCCTGTCTC-3'

B* (ineffective duplex) B-Raf sense: 5'- CCAGCUUGAGAAAACACUUGUdTdT-3'

B-Raf antisense: 5'-ACCAAGUGUUUCUCAAGCUGGdTdT-3',

c, c-Raf-1 sense: 5'-GCUGUCUACUCGGAUCGGGUCAdTdT-3'

c-Raf-1 antisense: 5'-UGACCCGAUCCGAGUAGACAGCdTdT-3',

Primers used for RT-PCR:

A-Raf 5': 5'-ACCATGCACAATTTTGTACGG-3' **3'**: 5'-ACT TTTTCTT TTCATCTGCC-3';

B-Raf 5': 5'-ACAACACACAACCTTTGTACGG-3' **3'**: 5'-GTTTATATGAACATTGGGAGC-3'

c-Raf-1 5': 5'-G TTTCCAAATTCCTACTGTTGG-3' **3'**: 5'-TTTTGTTTTTTCCTGGGTCC-3'

suggested not to target regions of the cognate mRNA that might bind to regulatory proteins, i. e. sequences close to the start site and to 5' and 3' untranslated regions (UTR). Twenty-three nucleotides conforming to the consensus sequence 5'-AA[N19]UU-3' (where N is any nucleotide) are selected from the mRNA. The selected sequence is subjected to [BLAST](#) analysis in order to exclude targeting of closely related mRNA sequences. The optimal GC-content of siRNA is between 30–70%. The siRNA is then constructed by designing sense and anti-sense N19 sequences, each with two 3' 2-deoxythymidine residues at the end. For choosing shRNA-targeting sequences, similar rules were initially applied (1). A targeting sequence of 5'-(N18-28)C-3' (where N is any nucleotide) is selected from the target mRNA. More recently, several authors have systematically investigated the sequence requirements for efficient RNAi by comparing multiple siRNAs targeting the same exogenous or endogenous genes. Additional rules were established including the positioning of A/U at the 5' end of the antisense strand, G/C at the 5' end of the sense strand, allowing at least

five A/U residues in the 5' terminal one-third of the antisense strand and avoiding any GC stretch of more than nine nucleotides in length (4). Others reported that duplexes targeting the middle of the coding sequence exhibited a poorer silencing effect, while targeting the 3' UTR was as efficient as targeting the coding sequence. Pooling of four or five duplexes per target gene was particularly efficient and efficient duplexes showed nucleotide preferences at positions 11 and 19 of the siRNA duplex (5). Yet another publication lists several characteristics for good siRNA functionality including low G/C content, lack of inverted repeats and sense strand base preferences affecting positions 3, 10, 13 and 19 (6).

RNA Interference – the Issue of Specificity

While efforts to optimize the efficacy of RNAi toward the target mRNA have quickly produced a substantial list of successfully silenced genes, “collateral effects” of siRNAs were initially neglected. The properties of siRNA-mediated gene silencing were analyzed on a genomic scale using the [enhanced green fluorescent](#)

protein (EGFP) of *Aequoria victoria* as a targeted transgene, because it has no known physiological role in the recipient cells (7). Neither the presence of EGFP nor its silencing was likely to perturb the transcriptome. To prove or disprove this assumption, cells retrovirally transduced with the EGFP-construct were transfected with siRNA targeting EGFP, with siRNAs representing scrambled EGFP sequences and mock transfected. 36K cDNA microarrays were interrogated with target RNAs prepared from transfected cells and the expression patterns were compared. No obvious changes in mRNA levels were observed indicating that “off-target” gene expression perturbation was not associated with siRNA-mediated gene silencing. In addition, Chi et al. did not find any evidence for “transitive” RNAi effects as has been shown in *C. elegans* (see Fig. 1). In contrast to this, other authors have reported off-target gene regulation by RNAi (8). siRNA directed against an exogenous luciferase gene regulated the expression of several endogenous genes despite the lack of a homologous target in the human genome. In experiments aimed at silencing mitogen-activated protein kinase 1 (p38 α), a signaling pathway effector kinase, the authors demonstrated that nine genes were silenced with similar kinetics, although none of them served a known functional role in the corresponding signaling pathway. In further kinetic experiments with siRNA molecules exhibiting various nucleotide exchanges deviating from the perfect targeting sequences, Jackson et al. found that “collateral” or off-target silencing may occur when as few as 11 contiguous nucleotides

matched the target mRNA. In conclusion, gene knockdown can occur as a result of degradation of mRNA targets with partial identity to the siRNA sequence in conjunction with the robust silencing of the desired target. Yet another important caveat has to be taken into account when the phenotypic effects of siRNA-mediated gene targeting are assessed. Although initial reports indicated that short interfering RNA duplexes do not elicit non-specific interferon-dependent cellular responses, a recent publication reports that efficient siRNA-mediated silencing of GAPDH, a housekeeping gene, induced activation of the Jak-STAT pathway and a global up-regulation of interferon-stimulated genes (9). This effect was not sequence-specific, since the silencing of five additional target genes had the same effect. In addition, the authors provided evidence that the off-target effect is mediated by the dsRNA-dependent protein kinase PKR. In conclusion, peer reviewers will ask future authors of publications based on RNA interference in mammalian cell cultures and mice, what kind of controls have been considered to exclude off-target silencing effects.

Genome-Wide Functional Screens in Mammalian Cells

Two very recent publications reveal the full power of RNAi application in mammalian cells. Two groups have generated retrovirus-based libraries comprising at least three shRNAs targeting about one third of human genes for suppression. One library was used to search for genes that modulate the function of the tumor

RNA Interference in Mammalian Cells. Table 1 Applications of RNAi in disease models

Model	Targeted gene	Phenotype
Hepatitis	<i>Fas</i> receptor	Prevention of infection and death
Acute liver failure	Caspase 8	Prevention of Fas-mediated cell death
HIV infection	HIV-encoded <i>Tat</i> and <i>Rev</i>	Inhibition of HIV-1 replication in cultured cells
Influenza infection	Influenza virus encoded nucleocapsid and reverse transcriptase	Prevention of virion RNA accumulation in cell lines and chicken eggs
Chronic myeloid leukemia	Bcr/abl	Decrease of bcr/abl expression and kinase activity
Drug resistance	<i>MDR1</i>	Decrease of drug resistance in several carcinoma cell lines
Oncogene-mediated tumorigenesis	<i>KRAS</i>	Loss of anchorage independence and tumorigenicity
Familial cylindromatosis	<i>CYLD</i> (among 50 de-ubiquitination genes)	Detection of gene in a cancer-related pathway Enhancement of NF κ B activation and prevention of cell death
Amyotrophic lateral sclerosis	Cu, Zn superoxide dismutase (<i>SOD1</i>) mutant	Protection of cultured neuroblastoma cells against cyclosporine A-induced cell death

suppressor p53 in a cell system indicative for bypassing p53-dependent growth inhibition. Screening of approx. 8,000 human genes identified 6 genes, including 53 itself, whose knockdown confers resistance to growth arrest (10). The second library comprising shRNAs targeting approx. 10,000 human genes was screened to identify components of the proteasome, which is responsible for regulated protein degradation in the cell. As a read-out for suppressed proteasome function mediated by shRNA, the authors used a reporter assay based on mouse ornithine decarboxylase coupled to a fluorescent protein. When the shRNA library was co-transfected with the reporter, transfectants exhibiting reporter protein accumulation were recovered, indicating that proteasomal function was impaired. Fifteen genes encoding known proteasome subunits were identified in this way (11). Thus, systematic [reverse genetics](#), a powerful approach for dissecting gene function in *C. elegans*, is no longer a desirable option for mammalian cells, but has become a reality. RNAi may indeed revolutionize reverse genetic approaches provided that the collateral or off-target effects can be controlled properly.

Clinical Relevance

Very quickly after the reports on successful RNAi in cultured mammalian cells, sequence-specific silencing of transgenes was demonstrated in mice. Transgene expression was suppressed following hydrodynamic transfection of synthetic siRNAs into the liver (12) or by injection into the tail vein (13). Several reports describe the use of sequence-selective siRNAs in various *in vitro* and *in vivo* models for cancer, viral infection and genetic disorders. Some selected examples of siRNA application in model systems are shown in Table 1 (for review see (14)).

References

1. Paddison PJ, Hannon GJ (2002) RNA interference: the new somatic cell genetics? *Cancer Cell* 2:17–23
2. Tijsterman M, Ketting RF, Plasterk RH (2002) The genetics of RNA silencing. *Annu Rev Genet* 36:489–519
3. Kamath RS, Fraser AG, Dong Y et al (2003) Systematic functional analysis of the *Caenorhabditis elegans* genome using RNAi. *Nature* 421:231–237
4. Ui-Tei K, Naito Y, Takahashi F et al (2004) Guidelines for the selection of highly effective siRNA sequences for mammalian and chick RNA interference. *Nucleic Acids Res* 32:936–948
5. Hsieh AC, Bo R, Manola J et al (2004) A library of siRNA duplexes targeting the phosphoinositide 3-kinase pathway: determinants of gene silencing for use in cell-based screens. *Nucleic Acids Res* 32:893–901
6. Reynolds A, Leake D, Boese Q et al (2004) Rational siRNA design for RNA interference. *Nature Biotechnology* 22:326–330

7. Chi JT, Chang HY, Wang NN et al (2003) Genome wide view of gene silencing by small interfering RNAs. *Proc Natl Acad Sci USA* 100:6343–6346
8. Jackson AL, Bartz SR, Schelter J et al (2003) Expression profiling reveals off-target gene regulation by RNAi. *Nat Biotechnol* 21:635–637
9. Sledz CA, Holko M, de Veer MJ et al (2003) Activation of the interferon system by short-interfering RNAs. *Nat Cell Biol* 5:834–839
10. Berns K, Hijmans EM, Mullenders J et al (2004) A large-scale RNAi screen in human cells identifies new components of the p53 pathway. *Nature* 428:431–437
11. Paddison PJ, Silva JM, Conklin DS et al (2004) A resource for large-scale RNA-interference-based screens in mammals. *Nature* 428:427–431
12. McCaffrey AP, Meuse L, Pham TT et al (2002) RNA interference in adult mice. *Nature* 418:38–39
13. Lewis DL, Hagstrom JE, Loomis AG et al (2002) Efficient delivery of siRNA for inhibition of gene expression in postnatal mice. *Nat Genet* 32:107–108
14. Wall NR, Shi Y (2003) Small RNA: can RNA interference be exploited for therapy? *Lancet* 362:1401–1403

RNA Maturation

Definition

The primary transcript of a gene (called heterogeneous nuclear RNA; hnRNA) is processed in the nucleus to produce the mature mRNA, e.g. the 3' end is modified by adding a stretch of adenosine nucleotides, and the non-coding introns are excised in a process called splicing.

- Full Length cDNA Sequencing
- Heteronuclear RNP Proteins

RNA Polymerase

Definition

The enzyme RNA Polymerase I is responsible for the synthesis of RNA from a DNA template ([► transcription](#)). Prokaryotic gene expression involves one type of RNA polymerase. In eukaryotes, three different types of RNA polymerase exist that each transcribe specific sets of genes.

1. RNA Polymerase I is located in the nucleolus and synthesizes only a single type of product, the

ribosomal RNA (rRNA), RNA Polymerase I accounts for the majority of RNA synthesis within the cell, transcribing all genes coding for rRNA.

2. RNA Polymerase II (POL II) is located in the nucleoplasm and synthesizes the messenger RNA (mRNA) precursor, heterogeneous nuclear RNA (hnRNA). All protein coding genes of the cell, whether they are constitutively expressed, developmentally regulated or restricted to certain types of tissues, are transcribed by RNA Polymerase II.
3. RNA Polymerase III (POL III) synthesizes transfer RNA (rRNA) and other small RNAs which are involved in the genetic translation of mRNA into the amino acid sequences of protein molecular (amino acyl tRNA). Some viral genes, such as some adenovirus and ▶Epstein-Barr virus genes, are also transcribed by RNA Polymerase III. RNA Polymerase III transcripts are generally very short (less than 300 nucleotides), and never encode any proteins.

RNA Polymerase I, II and III can be distinguished by their α -amanitin sensitivity, which is not sensitive, sensitive and species-specific, respectively.

- ▶Catalytic RNA
- ▶Core Promoters
- ▶Polyadenylation
- ▶RNA Polymerase I Transcription
- ▶RNA Polymerase II Transcription
- ▶RNA Polymerase III
- ▶RNA Stability
- ▶Splicing
- ▶Transcription Elongation
- ▶Transcription Factors and Regulation of Gene Expression
- ▶Two-Hybrid System

RNA Polymerase I Transcription

TACIANA KASCIUKOVIC,
JOOST C.B.M. ZOMERDIJK
Division of Gene Regulation and Expression,
Wellcome Trust Biocentre, University of Dundee, UK
j.zomerdijs@dundee.ac.uk

Synonyms

DNA-dependent RNA polymerase I (Pol I or RNAP I); in *S. cerevisiae* also referred to as RNA polymerase A (Pol A). For the purpose of this overview, we will primarily discuss mammalian RNA polymerase I transcription.

Definition

DNA-dependent ▶RNA polymerases catalyse the sequential incorporation of ribonucleotides complementary to a DNA template in a 5' to 3' direction to produce a primary RNA molecule or ▶transcript, a process referred to as ▶transcription. Eukaryotes have three nuclear RNA polymerases that are multi-subunit complexes with core subunits evolutionarily conserved with those of the DNA-dependent RNA polymerases from *Archaea* and prokaryotes. The RNA polymerase I (Pol I) enzyme complex in most eukaryotes is solely responsible for transcription of the ▶ribosomal DNA genes, which are found in multiple copies in eukaryotic cells. Transcription of rDNA by Pol I occurs in a distinct subcompartment of the nucleus, the ▶nucleolus, and yields a primary ribosomal RNA transcript, which is processed into the mature 18S, 5.8S and 28S rRNAs (1). Following RNA modification, the rRNAs are assembled together with ribosomal proteins into pre-ribosomal subunits. These then exit the nucleolus and are exported from the nucleus to sites in the cytoplasm or endoplasmic reticulum where, with the help of tRNAs, the ▶ribosomes translate mRNAs into proteins. The rRNAs produced by Pol I are not simply a structural component of the ribosome but in fact catalyse peptide bond formation during protein synthesis on the ribosome (2).

Characteristics

RNA Polymerase I Transcription of rRNA Genes and Ribosome Biogenesis

On average a eukaryotic cell contains between 2 and 10 million ribosomes, the protein synthesis machinery. While these RNA-protein complexes are relatively stable, cell growth and subsequent division requires a doubling of the number of ribosomes. Transcription of the rDNA genes by RNA Pol I drives this production of ribosomes. In an actively growing eukaryotic cell, a large proportion of nascent RNA synthesis (more than 50%) is accounted for by the activity of Pol I, which constitutes a considerable energy expenditure to the cell. Therefore, several mechanisms have evolved to control transcription tightly and to balance the production of ribosomes with demand during cell growth and division (3, 4, 5, 6). Moreover, control of Pol I transcription is coordinated with the activities of the other two nuclear RNA polymerases, whose products contribute to ribosome biogenesis directly (i.e. the 5S ribosomal RNA synthesized by Pol III and the messenger RNAs synthesized by Pol II encoding the many ribosomal proteins) or to protein synthesis in general (i.e. the transfer RNAs synthesized by Pol III and the mRNAs synthesized by Pol II).

Structural and Spatial Organisation of Genes Encoding Ribosomal RNAs

There are at least 100 copies of the rDNA gene in a eukaryotic genome (in the human genome there are an estimated 400 copies). Presumably these levels are necessary to meet the demands of ribosome production in actively growing and dividing cells, though interestingly only about 50% of the genes are actively transcribed at any one time. There is an inverse correlation between densely packed nucleosomes and the transcriptional activity of the rDNA genes, as inferred from electron microscopy and psoralen photo-cross-linking experiments. Concerted changes in DNA methylation and histone modification mark the switch between active and non-active gene status (7).

The multiple copies of the rDNA genes are organised as tandem head-to-tail repeat units and in humans these rDNA clusters are located on the short arms of the five **▶acrocentric chromosomes** 13, 14, 15, 21 and 22. These chromosomal rDNA loci are referred to as the **▶nucleolar organising regions** (NORs) around which the nucleolus is formed. Each unit (in humans ~43 kb) contains an intergenic spacer (IGS)(~30 kb) and sequences transcribed by Pol I that encode the pre-rRNA (~13 kb). The repetitive nature of the rDNA is discernable in electron micrographs of nuclear **▶chromatin** spreads (Fig. 1). Non-transcribed DNA strands are interspersed with regions of densely packed polymerases engaged in transcription of the rDNA giving a 'Christmas tree' pattern as the primary transcripts emanating from Pol I complexes (the branches) grow with the distance the polymerases have travelled along the rDNA (the trunk) from the start site of transcription. In yeast, each repeat also contains a copy of the 5S rRNA gene in the IGS, transcribed by Pol III in the opposite direction to rDNA transcription by Pol I. The transcribed regions of the repeats encoding the major rRNAs are highly conserved and the alignment of encoded rRNAs in a pre-rRNA gene is always in the same order, 18S, 5.8S and 28S, and dictates the order of pre-rRNA processing and assembly of pre-ribosomal subunits. The non-transcribed spacer regions vary considerably between species. These contain the regulatory elements such as the **▶promoter** DNA region, which directs initiation of transcription, **▶enhancer** sequences that stimulate transcription initiation at the promoter and **▶terminator** elements 3' of the rRNA gene cluster that direct termination of transcription. The spacers also contain various other sequence elements including an origin of DNA replication, a DNA replication fork barrier and many repetitive sequence elements of unknown function.

Active transcription by Pol I of the rRNA genes is essential for the formation of nucleoli. The nucleolus

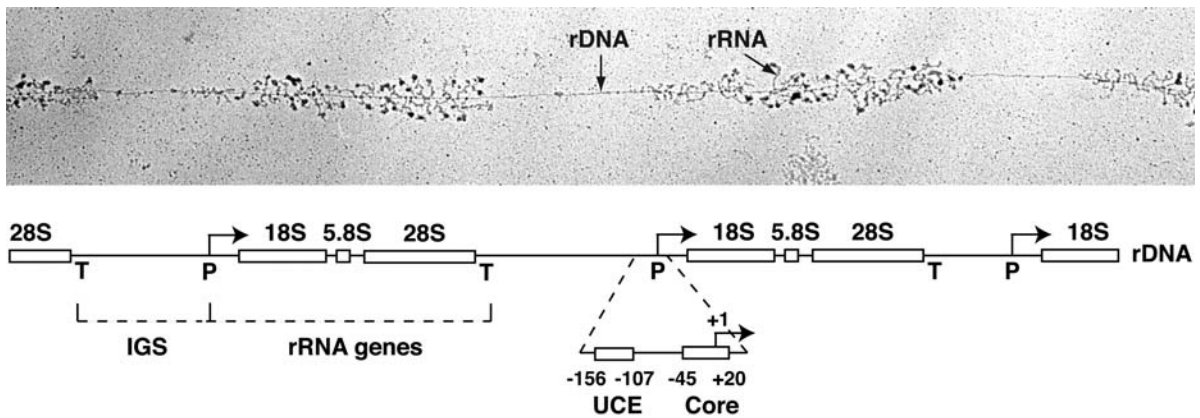
displays distinct structural features that may be linked to different functions. The fibrillar centres (FC) contain Pol I and the transcription factors important for transcription of the rDNA repeats. Transcription of the rDNA is thought to occur at the boundary of the FC and the surrounding dense fibrillar component (DFC). The DFC is where the pre-rRNAs are thought to be processed (endonucleolytic cleavages and exonucleolytic trimming) and modified (pseudo-uridylation and 2'-O-methylation), guided by a large number of small nucleolar ribonucleoprotein particles (snoRNPs) and, in part, this may occur co-transcriptionally. Surrounding these domains is the granular component where the assembled pre-ribosomal particles are found (8).

Pol I transcription is subject to cell-cycle control. Pol I transcription ceases when cells enter mitosis, concomitant with nucleolar disassembly. Some of the components of the Pol I transcription machinery stay attached to the rDNA in the condensed chromosomes throughout mitosis, whereas the Pol I enzyme itself appears to be transiently released. Upon completion of cell division and chromosome decondensation, transcription resumes and the nucleoli reform around transcriptionally active NORs. The emerging small multiple nucleoli grow and then fuse to form larger nucleoli, the size and number of which is in part dependent on the total level of Pol I activity in the cell.

RNA Polymerase I Transcription Machinery: Pol I and Transcription Factors

The core Pol I enzyme contains 14 subunits in *S. cerevisiae*, which have been cloned, and, with one exception, homologues for all of these have been identified in mammals. Seven of the subunits are unique to Pol I, two are shared between Pol I and Pol III and five are common to all three nuclear polymerases. It has been speculated that the shared subunits allow the organism to exert coordinate control over transcription by the three nuclear RNA polymerases, whereas the unique subunits bestow the core polymerase with the specificity required for control of the expression of specific subsets of genes. The two largest subunits of Pol I form the catalytic core of the enzyme and, while distinct for this polymerase, these subunits share considerable homology with the largest subunits of all other eukaryotic RNA polymerases.

Pol I itself, as with all eukaryotic nuclear RNA polymerases, is unable to recognise promoter sequences. **▶Transcription factors** with promoter sequence-specific DNA binding activity recruit Pol I to the start site of transcription, which is defined by the promoter DNA elements. A eukaryotic rRNA gene promoter consists of two elements (Fig. 1), the core element (in humans from -45 to +20, relative to the transcription start site at +1), which is essential and



RNA Polymerase I Transcription. Figure 1 Ribosomal DNA repeat organization. The electron micrograph (kindly provided by Dr Ann Beyer, University of Virginia, USA) shows two complete rDNA repeats in a chromatin spread from yeast cells. Transcription in these rDNA repeats proceeds from left to right, and the rRNA emanating from the RNA polymerase I molecules (themselves not clearly visible) is indicated. The dense particles associated with the rRNA are rRNA processing complexes. Underneath is a schematic representation of the two rDNA repeats. The rRNA genes encode the 18S, 5.8S and 28S rRNAs, which are produced by Pol I as a single transcript. The transcription start site (arrow) lies within the promoter (P). Transcription stops at the terminator (T). The non-coding and generally non-transcribed regulatory region of the repeat is referred to as the intergenic spacer (IGS). The human rDNA promoter region, with the upstream control element (UCE from -156 to -107) and the core (from -45 to +20, relative to the transcription start site at +1), is shown enlarged beneath one of the rDNA repeats.

directs specific initiation of transcription and the upstream control element (UCE, from -156 to -107). The exact spacing between these elements is important for the UCE to fulfil its role in stimulating transcription from the core. While the primary sequences are not conserved from human to yeast, the overall organisation of the two promoter elements does appear to be conserved.

The transcription factors required to achieve efficient promoter-specific initiation of Pol I transcription include ►SL1 (selectivity factor 1) (or murine TIF-IB) and ►UBF (upstream binding factor) in mammals and core factor (CF), TBP (TATA-binding protein), Hmo1 and upstream activation factor (UAF) in yeast (9, 10).

UBF is involved in the activation of Pol I transcription. Mammalian UBF, a 97 kD protein, contains several ►HMG (high mobility group) boxes, which are structural motifs of non-histone chromosomal HMG proteins, involved in binding the minor groove of the DNA and in DNA bending. UBF contains an amino terminal dimerisation domain important for activation of transcription. A dimer of UBF can organise approximately 140 bp of DNA into a single 360° loop and it has been suggested that UBF functions as an architectural protein at the promoter, binding the UCE and core elements and leading to a stereo-specific alignment of these elements. Due to the relaxed sequence specificity of binding by UBF, UBF distribution in the rDNA repeats is not restricted to these

regulatory elements however, suggesting it may have a more global role in influencing DNA topology throughout rDNA repeats. UBF can displace a linker histone H1 from a DNA-nucleosome assembly, suggesting a possible role in chromatin remodelling and antirepression. UBF also activates Pol I transcription by increasing productive transcription initiation events. Direct interactions of UBF with SL1 and Pol I subunits may be important for this. In one model, this activation was suggested to occur as a result of increased ►preinitiation complex formation, but evidence is emerging that UBF may have a role in stimulating events occurring subsequent to the formation of preinitiation complexes. In yeast, a single HMG-box containing protein, Hmo1, stimulates Pol I transcription, perhaps functioning similarly to UBF in vertebrates.

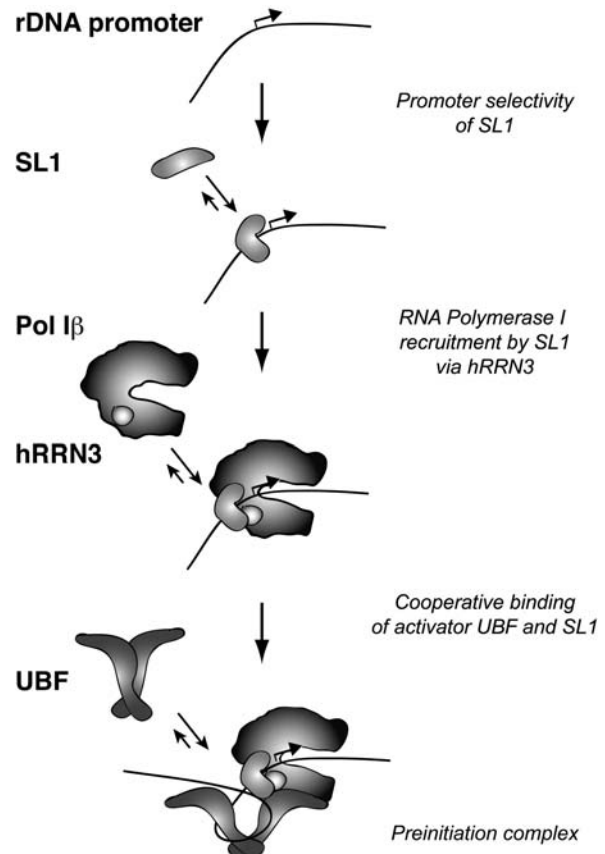
SL1 is capable of the independent recognition and binding of the mammalian rRNA gene core promoter. SL1 is a multimeric peptide complex of ~300 kD comprised of TBP (►TATA box binding protein) and at least three TAFs (►TBP-associated factors) specific to the Pol I transcription machinery, including TAF_{110/95}, TAF_{63/68} and TAF₄₈. TBP has a role in the initiation of transcription by all three nuclear polymerases in complex with subsets of TAFs that define its polymerase specificity. The TAFs of the SL1 complex confer upon SL1 the ability to position itself precisely at the core element of the rRNA gene promoter. Unlike the polymerase and UBF components of the Pol I

transcription machinery, SL1 is not interchangeable between species and indeed was named for its selective role in directing species-specific transcription. The promoter sequences of the rRNA genes are divergent between species and the TAFs in SL1, which have only limited sequence similarity between species, are responsible for sequence-specific promoter recognition. In *S. cerevisiae*, CF (core factor) appears to function like SL1, binding the core promoter element. Although, unlike SL1, CF is not stably associated with TBP, it does contain three subunits, Rrn6p, 7p and 11p, two of which have limited homology to the TAFs of SL1. TBP in yeast is found associated with upstream activating factor (UAF), a multiprotein complex consisting of Rrn5p, 9p, 10p, Uaf30p and histones H3 and H4. UAF stably interacts with the upstream DNA element in the yeast rDNA promoter. In mammals, UBF and SL1 remain stably associated with the promoter following transcription initiation by Pol I, potentially providing a re-initiation scaffold, whereas in yeast, UAF remains promoter-bound and CF dissociates from the promoter. At present no UAF-like complex has been identified in mammalian cells. The RNA polymerase II basal factor TFIID has also been shown to play a role in transcription by Pol I, although its precise function in this process remains unknown.

Different forms of the Pol I complex have been isolated in which the core enzyme (~600 kD) is associated with different subsets of polypeptides to form complexes of >1 MD. The majority of these Pol I complexes (Pol I α) are incapable of directing the initiation of transcription specifically from the rRNA gene promoter. Initiation competent Pol I complexes comprise only about 10% of the Pol I complexes of the cell (Pol I β) and are distinguished by the presence of the polypeptide hRRN3 (or murine TIF-IA), which is functionally conserved from human to yeast. SL1, or CF in yeast, recruits the initiation competent Pol I to the promoter *via* the interaction of its subunits with RRN3 in the Pol I complex. Following transcription initiation, the inactivated hRRN3/TIF-IA dissociates from Pol I, with conversion of the polymerase from the initiating to the elongating form. Other polypeptides associated with the Pol I core enzyme include kinases, components of the DNA repair and DNA replication machineries and proteins involved in chromatin remodelling. Their precise roles in Pol I have yet to be defined.

The Mammalian Pol I Transcription Cycle

Pre-initiation complex formation is an important and highly regulated step in Pol I transcription, involving the recruitment by SL1 of initiation competent Pol I containing the hRRN3/TIF-IA protein to the promoter (Fig. 2). SL1 also interacts with UBF, resulting in cooperativity in promoter DNA binding by these



RNA Polymerase I Transcription. Figure 2 RNA polymerase I and transcription factors SL1 and UBF in the formation of a transcription preinitiation complex at the human rDNA promoter. Selectivity factor SL1, a TBP-TAF complex, binds the rDNA core promoter element. SL1 interacts with hRRN3 associated with Pol I and consequently recruits and positions the initiation-competent form of Pol I (Pol I β). Transcription activator upstream binding factor (UBF) binds the rDNA promoter cooperatively with SL1, bending the promoter region substantially. Transcription initiation follows preinitiation complex formation.

factors. After promoter opening and conformational changes in the pre-initiation complex, Pol I incorporates the first few nucleotides and escapes the promoter. This rate-limiting step may be stimulated by the transcriptional activator UBF. The transition from the initiating to elongating form of the polymerase involves conformational as well as compositional changes of the Pol I enzyme complex. Elongation might require nucleosome displacement or rearrangement and the chromatin remodelling activities associated with the polymerase. Elongation may also be facilitated by TIF-IC, a factor associated with murine Pol I, as well as TFIIS, a known Pol II elongation factor. DNA topoisomerase activity may be required for relief of

the torsional strain accumulated during transcription. Transcription terminates at a T-rich sequence followed by an array of terminator sequences, to which TTF-1 (transcription termination factor-1) protein is bound. Dissociation of the transcripts from the terminated Pol I complex is mediated by PTRF (polymerase I and transcript-release factor). TTF-1 has also been implicated in the recruitment of chromatin remodelling activities to a terminator site just upstream of the UCE of the rDNA promoter, which thereby modulate chromatin structure and accessibility of the promoter.

Regulation of rDNA Transcription

Cell growth and division is dependent on the capacity of the cell to synthesize proteins, and this is related to the number of ribosomes. Under conditions of active growth and division, for example during early embryonic development, a cell can augment the efficiency of rDNA transcription, which drives ribosome biogenesis. There are two ways for cells to modulate the level of rDNA transcription, to change the frequency of loading of Pol I molecules onto already active genes and/or to engage more of the rDNA genes in transcription. Signal transduction pathways activated by growth factors, such as those including phosphatidylinositol 3-kinase (PI3K) and mitogen-activated protein kinase (MAPK), or influenced by nutrient availability of the cell, such as the mammalian target of rapamycin (mTOR) pathway, have been demonstrated to regulate Pol I transcription. For example, EGF or serum-induced activation of the MAPK pathway and nutrient regulated activation of the mTOR pathway lead to phosphorylation and modulation of the activities of UBF and Pol I-associated TIF-IA (RRN3). Phosphorylation of yRrn3 influences Pol I recruitment to the rDNA promoter. Growth factor stimulation of cells has also been demonstrated to increase the occupancy of the rDNA promoter by SL1. These same signalling pathways are known to regulate the activity of chromatin remodelling activities, which may be relevant for the activation of previously silent rRNA genes. The level of ribosome biogenesis appropriate to the external and internal cellular conditions can therefore be achieved through the control of rRNA synthesis *via* a multitude of signalling pathways and mechanisms.

Shutdown of Pol I transcription during mitosis and its reactivation upon entry into interphase is also regulated by the phosphorylation and dephosphorylation of components of the transcription machinery. SL1 and UBF, for instance, are phosphorylated by Cdk/cyclin complexes. Additionally, Pol I transcription can be directly modulated by casein kinase II (CK2) and cell-cycle regulators including the tumour suppressor proteins p53 and retinoblastoma (RB).

Clinical Relevance

Normal cell growth and proliferation require tight regulation of protein synthesis at the global level of the protein synthesis capacity of the cell as well as at the level of expression of individual protein-coding genes. The net result of up-regulated protein synthesis is the ability of cells to support runaway growth and proliferation. In most, if not all tumour cells, Pol I transcription and ribosome production appear elevated and there are also examples of tumour cells with increased levels of the proteins involved in Pol I transcription, such as UBF. The size and number of nucleoli, detectable following silver staining of tissue samples, reflects Pol I transcription levels and has been used for many years by tumour pathologists as a prognostic marker for the proliferative capacity of cells. Many proto-oncogenes and tumour suppressor proteins can regulate rRNA synthesis and so it is possible that deregulated Pol I transcription can be a predisposition for cancer susceptibility (11). Pol I synthesis of the rRNAs permits cell growth and division through ribosome biogenesis, but it is also likely that Pol I transcription itself regulates cell growth under certain circumstances. It is therefore crucial to understand the mechanisms and signalling pathways that control Pol I transcription of the rRNA genes. This may lead to the development of procedures to regulate Pol I transcription in different disease states.

The link between up-regulation of Pol I transcription and cancer and the lack of mutations associated with human diseases in components of the Pol I machinery both reflect the requirement for stringent control of Pol I transcription to achieve normal cellular growth.

References

1. Paule MR (1998) Transcription of Ribosomal RNA Genes by Eukaryotic RNA Polymerase I. Springer, Berlin Heidelberg New York, p 321
2. Moore PB, Steitz TA (2002) The involvement of RNA in ribosome function. *Nature* 418:229–235
3. Moss T (2004) At the crossroads of growth control; making ribosomal RNA. *Curr Opin Genet Dev* 14:210–217
4. Grummt I (2003) Life on a planet of its own: regulation of RNA polymerase I transcription in the nucleolus. *Genes Dev* 17:1691–1702
5. Warner JR (1999) The economics of ribosome biosynthesis in yeast. *Trends Biochem Sci* 24:437–440
6. Russell J, Zomerdijs JCBM (2005) RNA-polymerase-I-directed rDNA transcription, life and works. *Trends Biochem Sci* 30:87–96
7. Grummt I, Pikaard CS (2003) Epigenetic silencing of RNA polymerase I transcription. *Nat Rev Mol Cell Biol* 4:641–649
8. Olson MO, Hingorani K, Szebeni A (2002) Conventional and nonconventional roles of the nucleolus. *Int Rev Cytol* 219:199–266

9. Zomerdijk JCBM, Tjian R (1998) Initiation of transcription on human rRNA genes. In: Paule MR (ed) Transcription of eukaryotic ribosomal RNA genes by RNA Polymerase I. Springer, Berlin Heidelberg New York
10. Nomura M (2001) Ribosomal RNA genes, RNA polymerases, nucleolar structures, and synthesis of rRNA in the yeast *Saccharomyces cerevisiae*. Cold Spring Harb Symp Quant Biol 66:555–565
11. Ruggero D, Pandolfi PP (2003) Does the ribosome translate cancer? Nat Rev Cancer 3:179–192

RNA Polymerase II Transcription

OLIVIER BENSAUDE, BENOÎT PALANCADE

Régulation de l'Expression Génétique, Ecole Normale Supérieure (ENS), Paris, France
bensaude@wotan.ens.fr

Definition

DNA-dependent RNA polymerases (RNAP) catalyze the synthesis of an RNA copy (transcript) of a DNA template. The α -phosphate of an incoming nucleotide triphosphate is ligated to the ribose 3'-OH of the last nucleotide incorporated in the elongating chain and a pyrophosphate molecule is released. Eukaryotic cells use three distinct but related enzymes, each specialized in transcription of a distinct subset of genes, RNAP I, RNAP II and RNAP III. RNAP II transcribes into pre-messenger RNAs (pre-mRNA) all genes that code proteins as well as many genes that specify small non-coding RNAs (snRNA). It is usual to distinguish three major steps in the synthesis of transcripts, initiation, elongation and termination. Following their synthesis, the transcripts undergo several modifications (processing or maturation). The 5' end of messenger RNAs (**►mRNAs**) is protected by addition of 7-methyl guanosine on the protruding triphosphate (capping). Intervening sequences or introns are excised and exons religated (splicing). The primary transcript is cleaved at a specific site (cleavage) followed by the addition of a polyadenylic tail extending over hundreds nucleotides (**►polyadenylation**). The coupling between transcription and processing of the primary transcript is an important characteristic of class II gene transcription. Processing and transcription occur within discrete nuclear structures, the transcription factories.

Characteristics

Structure of RNA Polymerase II

The yeast and mammalian RNA polymerases comprise 12 protein subunits ranging from 10 to 210 kD in

molecular weight thus forming complexes of 500 kD. Most subunits are specific for each polymerase but some are common to RNAP I, RNAP II and RNAP III. Five subunits are orthologues of the prokaryotic RNA polymerase subunits. Rpb1, the largest RNAP II subunit, possesses a unique carboxyl-terminal domain (CTD) made of multiple repeats of a seven amino-acid motif (Tyr₁-Ser₂-Pro₃-Thr₄-Ser₅-Pro₆-Ser₇) that has been conserved throughout evolution.

The yeast RNAP II has been crystallized and the resolution of its structure at 2.8 Å has been a *tour de force*. The transcribed double-stranded DNA tunnels through the enzyme and, in the vicinity of catalytic site, the complementary base pairs are disrupted forming a transcription “bubble” (Fig. 1). RNA is synthesized complementary to one DNA strand in the bubble. The CTD is localized next to the RNA exit site and plays a critical role in controlling the initiation and the elongation steps as well as in coordinating the maturation events.

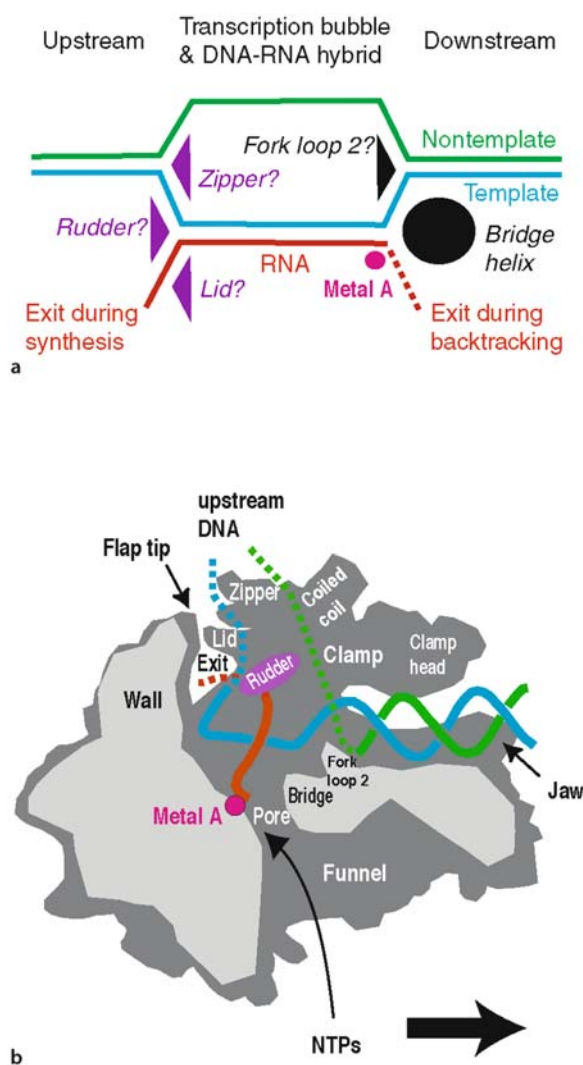
Sequential Steps in the Transcription Cycle

Preinitiation and Initiation Steps

RNAP II is recruited on promoters through interactions with regulatory transcription factors, general transcription factors and coactivators. The regulatory transcription factors recognize characteristic DNA sequences and determine the pattern of gene expression as a function of cell type and extracellular signals. The general transcription factors (TFIIA, B, D, E, F, H) assist the assembly and the precise positioning of RNAP II within a “pre-initiation complex” (PIC) of transcription. The coactivators such as the mediator and/or chromatin modifying complexes relay the regulatory factor signaling to the general transcription factors. The core RNAP II may bind coactivators such as the mediator complex and form a holoenzyme prior to interacting with the promoter region. The double-stranded DNA opens within the PIC to generate a transcription bubble (Fig. 1). As transcription initiates, the polymerase dissociates from the preinitiation complex to proceed (promoter clearance). Formation of the transcription bubble and promoter clearance require the ATP-dependent DNA helicase and kinase activities of the general transcription initiation factor **►TFIIH**.

Protein Phosphorylation Governs the Entry into the Elongation Step

The CTD is not phosphorylated in pre-initiation complexes and its phosphorylation contributes to the disruption of the interactions with the mediator complex and with the general transcription factor TFIID. The CTD is phosphorylated first on the serines at position 5 in the repeated motif by CDK7, the kinase subunit of TFIH. Elongation of transcription beyond



RNA Polymerase II Transcription. Figure 1 The yeast RNAP II elongation complex. (a) Schematic presentation of the arrangement of nucleic acids during RNA chain elongation. The DNA template and non-template strands are in blue and green, respectively, and the RNA is in red. The active site metal ion A is indicated by a pink sphere. Protein elements that are proposed to be involved in the maintenance of the arrangement of nucleic acids are indicated. (b) Cutaway view of the RNAP elongation complex. Cut surfaces are lightly shaded. During transcription, DNA enters the enzyme from the right (the polymerase moves to the right). Structural features that appear to be important for function are labelled. Colouring of nucleic acids is as in (a). Exiting RNA and DNA strands are not revealed in the electron density map, but their anticipated locations are indicated by dashed lines. In this view, the clamp swings over the active center from back to front. Only one of the jaws (the lower jaw) is visible in the cutaway view. Taken from CRAMER P. (2002) Multisubunit RNA polymerases. *Curr. Opin. in Struct. Biol.* 12:89–97.

circa a hundred nucleotides is blocked by the negative elongation factor (NELF) and by the DRB sensitivity-inducing factor (DSIF). Phosphorylation of the CTD on the serines at position 2 and of the largest subunit of the DSIF by the CDK9 subunit of P-TEFb, the positive transcription elongation factor, is generally critical in relieving this block and allowing a productive transcription.

Factors that Favor Processive Elongation Through Chromatin Templates

Along the DNA matrix, RNAP II encounters numerous obstacles that slow down its progression and may provoke pausing or premature termination. Elongation factors such as TFIIS assist the paused RNAP II molecules in resuming transcription. Other factors enhance the catalytic rate of elongation by affecting the K_m and/or V_{max} of RNAP II and include TFIIF, the elongins, and the ELL family of proteins. Furthermore, the transcription process generates supercoils that slow down the enzyme progression and are removed by topoisomerases. Chromatin remains a major obstacle, which is overcome by chromatin remodeling factors that displace the nucleosomal components or by enzymes that covalently modify histones such as acetylases, methylases, ubiquitin ligases and kinases.

Recycling RNA Polymerase

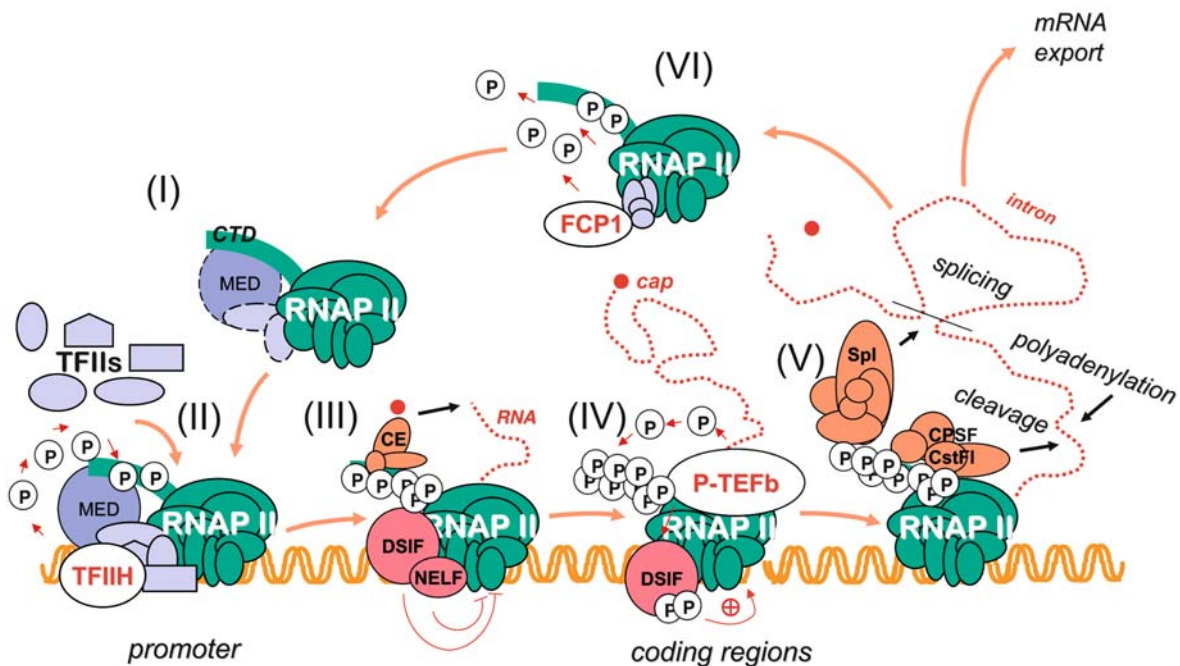
Transcription terminates when cis-elements at the 3' end of the pre-mRNA sequence trigger cleavage/polyadenylation of the primary transcript by the cleavage/polyadenylation factors (CPSF/CstF) and signal the polymerase to fall off its DNA template. To reinitiate a transcription cycle, RNAP II must be dephosphorylated by FCP1, a highly specific TFIIF-dependent CTD phosphatase.

Integration of Pre-mRNA Processing with Transcription

Transcription and pre-mRNA processing are not separated in time and space. Capping occurs first on nascent transcripts roughly 20 nucleotides in length, while NELF and DSIF prevent elongation of the transcript. Splicing occurs next on the nascent transcript. Finally, termination of transcription is determined by cleavage and polyadenylation of the transcript.

The CTD Acts as a Platform for RNA Processing

The phosphorylated CTD binds to the capping enzyme, the spliceosome components and the cleavage/polyadenylation factors and stimulates their catalytic activities. Sequential phosphorylation of the serines at position 5 and of those at position 2 in the CTD



RNA Polymerase II Transcription. Figure 2 CTD phosphorylation and the transcription cycle. (I) Recycling. The “free” RNAP II core enzyme is not phosphorylated on the CTD. It may assemble with coactivators such as the mediator complex (Med) thus forming a holoenzyme. (II) Preinitiation. The unphosphorylated RNAP II core or holoenzyme assemble onto the promoter sequences with general transcription factors such as TFIID, TFIIA, TFIIB, TFIIE, TFIIH and TFIIH thus forming a preinitiation complex of transcription. The CTD is phosphorylated on serines at position 5 (5P) by the CDK7 subunit of TFIIH. (III) Initiation. Transcription begins. The phosphorylated CTD recruits the capping enzymes and the nascent transcript is capped at its 5' end. (IV) Elongation. Phosphorylation of the CTD on serines at position 2 (2P) by the CDK9 subunit of the positive transcription elongation factor (P-TEFb) is required to remove to block due to the DSIF/NELF factors and elongate transcription. (V) RNA Processing and transcription termination. The phosphorylated CTD recruits the splicing machinery (Spl) to remove introns and, finally recruits the cleavage and polyadenylation factors (CPSF and CstF) that cleave the transcript and add a polyadenylic tail at its 3' end. This step signals transcription to terminate and RNAP II falls off its DNA template. The resulting mRNA is then exported to the cytoplasm. (VI) To be recycled for another transcription round, RNAP II is dephosphorylated by the FCP1 CTD phosphatase.

repeated motifs contributes to the timed recruitment of the RNA processing machinery. The capping enzymes are recruited onto the CTD that have been phosphorylated by CDK7 and remain localized at the 5' ends of the genes. The triphosphatase component of the capping enzymes delays reinitiation by a novel RNAP II molecule and contributes to the recruitment of the P-TEFb that neutralizes the block imposed by NELF and DSIF. This transcriptional checkpoint may allow only capped mRNA to undergo processive elongation. Unexpectedly, RNA processing factors such as CPSF already interact with the preinitiation complex on the promoter. They are loaded later on RNAP II through the phosphorylated CTD.

Transcription, Splicing and mRNA Export Cross-talk

On the one hand, transcription and splicing factors share common components: U1 snRNA stimulates the

TFIIH initiation factor and the Tat-SF1 elongation factor binds to U2 snRNA. In some cases, the presence of splice sites enhances transcription levels. On the other hand, promoter and transcription elongation rate may determine the balance between the alternatively spliced variants. Indeed, a slow elongation rate favors the use of weak splicing donor and acceptor sites. Transcription and splicing are also connected to mature mRNA export into the cytoplasm through the nuclear pore. Indeed, proteins involved in RNA packaging and export from the nucleus are loaded on the nascent transcript through physical interactions with the RNAP II transcription elongation complex.

Integration of mRNA Quality Control with Transcription

Cotranscriptional mRNA quality control is a fascinating concept that has recently gained support in

eukaryotes a long time after being established in bacteria. The quality control is exerted at various steps from the incorporation of nucleotides into the nascent transcript to a test of proper translatability into protein.

RNAP II Proofreading

Misincorporation of a nucleotide arrests RNAP II, as base-pairing mismatch of this nucleotide prevents the progression of the transcription bubble. The incorrect nucleotide is then removed by the intrinsic 3'-exonuclease activity of the stalled RNAP II. This activity is stimulated by TFIIIS.

Non-sense Codon Mediated Quality Control

Premature stop codons or absence of stop codons may have detrimental consequences due to the synthesis of short truncated or carboxy-terminal extended abnormal proteins, respectively. Premature stop codons or the absence of stop codons are consequences of mutations or natural processes such as the DNA sequence shuffling responsible for immunoglobulin and T-cell receptor gene diversity. A non-sense mediated quality control eliminates the deficient mRNAs. In many cases, this phenomenon occurs in the nucleus through exosome-mediated degradation, alternative [▶splicing](#) or aggregation near the site of transcription. Non-sense stop codons are detected as premature if they precede the junction between two exons (exon splice junction). This detection involves the progression of a ribosome translating the mRNA into protein to remove the so-called exon-junction complex that remains left over after the splicing process.

Clinical Relevance

Diversion of Transcription Factors by Viral Proteins

The Tat protein of the human immunodeficiency virus directly recruits P-TEFb to phosphorylate RNAP II and activate viral transcription. The HDAG protein of hepatitis δ virus mimics the NELF largest subunit but, unlike its cellular counterpart, it activates RNAP II-mediated viral transcription.

Inhibition of RNAP II Transcription by Natural and Pharmacological Compounds

α -Amanitin from the fungus *Amanita phalloides* is a highly specific and potent inhibitor of RNAP II. It binds irreversibly to Rpb1, the largest subunit and blocks elongation of the transcript. Numerous anti-tumor drugs are DNA intercalators that also inhibit transcription.

Human Genetic Diseases Associated with Defects in RNAP II Transcription

Cockayne syndrome and trichothiodystrophy are cancer-free multisystem disorders associated with

deficiencies in the helicase activities of the TFIIH transcription initiation factor. The Von Hippel Lindau syndrome is associated with mutations in the Von Hippel-Lindau (VHL) tumor suppressor gene. Binding of the mutated VHL gene product to elongins B and C is reduced. Consequently, the VHL defect causes misregulation in the expression of a number of genes involved in oncogenesis. The human ELL gene can undergo translocation with the MLL gene, generating a MLL-ELL fusion protein in patients affected by acute myeloid leukemia. Congenital cataracts facial dysmorphism neuropathy (CCFDN) syndrome (OMIM 604168) is associated with a partial deficiency in the CTD phosphatase FCP1.

References

1. Meinhart A, Kamenski T, Hoepfner S et al (2005) A structural perspective of CTD function. *Genes Dev* 19:1401–1415
2. Armache KJ, Kettenberger H, Cramer P (2005) The dynamic machinery of mRNA elongation. *Curr Opin Struct Biol* 15:197–203
3. Guthrie C, Steitz J (2005) Nucleus and gene expression Coordinated nuclear events regulate mRNA synthesis, processing, export and turnover. *Curr Opin Cell Biol* 17:239–241 and following reviews in the same issue
4. Shilatifard A (2004) Transcriptional elongation control by RNA polymerase II: a new frontier. *Biochim Biophys Acta* 1677:79–86
5. Conaway JW, Conaway RC (1999) Transcription elongation and human disease. *Annu Rev Biochem* 68:301–319

RNA Polymerase III

SARAH J. GOODFELLOW, ROBERT J. WHITE

Institute of Biomedical and Life Sciences, University of Glasgow, Glasgow, UK
rwhite@udcf.gla.ac.uk

Synonyms

Synonyms used for RNA polymerase III and some of its associated [▶transcription factors](#) are listed in Table 1.

Definition

RNA polymerase III is one of three highly related [▶RNA polymerases](#) found within the nuclei of eukaryotic cells. RNA polymerases are large, complex enzymes that synthesise RNA using specific DNA templates in a process known as gene [▶transcription](#)

RNA Polymerase III. Table 1 Synonyms used for the basal RNA Polymerase III transcriptional apparatus

Name used in essay	Synonyms
RNA Polymerase (pol) III	RNA Polymerase C, RNAP III
BDP1	B", TFIIIB90 (yeast), TFIIIB150 (human), TAF3B1, TFNR
BRF1	TFIIIB70 (yeast), hTFIIIB90 (human), hBRF (human), TAF3C, GTF3B, TAF3B2
SNAPc	PBP, PTF

(1). Each of the three eukaryotic nuclear RNA polymerases is dedicated to the transcription of a different set of genes. RNA polymerase III is responsible for approximately 10% of all nuclear transcription and synthesises a variety of small, untranslated RNAs, many of which have roles in cellular metabolism (1, 2). ▶**RNA polymerase I** transcribes genes encoding the large ▶**ribosomal** (rRNAs), which account for around 70–80% of cellular RNA synthesis (1), and ▶**RNA polymerase II** produces messenger RNAs that code for cellular proteins (1). The templates transcribed by RNA polymerases I, II and III, often referred to as class I, II and III genes respectively, are distinguished by distinct ▶**promoter** sequences. The polymerases themselves cannot recognise these promoters and so rely on accessory proteins known as transcription factors. Each of the RNA polymerases employs a distinct set of transcription factors, which bind to specific DNA elements within promoter regions and then recruit the appropriate RNA polymerase. Once the polymerase has been recruited, transcription can ensue. Features of the RNA polymerase (pol) III transcriptional apparatus and the genes transcribed by this enzyme are discussed below.

Characteristics

RNA Polymerase III

Pol III is a huge 600–700 kD complex composed of 17 subunits in yeast and humans, making it the largest of the three eukaryotic nuclear ▶**RNA polymerases** (2, 3, 4). Genetic studies in yeast have indicated that at least 16 of these 17 subunits are essential for pol III function and yeast viability (2, 3, 4). Five of the 17 subunits are shared by all three RNA polymerases, a further two are common to both pols I and III and ten subunits are unique to the pol III complex (2, 3, 4). All three polymerases perform the same fundamental function; they catalyse the covalent attachment of ribonucleotides to form an RNA chain that is complementary to the gene template being transcribed. Therefore, it is not

RNA Polymerase III. Table 2 Summary of pol III gene products and their functions

Pol III Product	Known Functions
tRNA	Involved in protein synthesis as a translational adaptor
5S rRNA	Involved in protein synthesis as a component of ribosomes
U6 snRNA	Involved in mRNA splicing
H1 RNA	Involved in tRNA processing (RNase P component)
MRP RNA	Involved in rRNA splicing
7SL RNA	Involved in intracellular protein transport (component of signal recognition particle)
7SK RNA	Involved in controlling transcriptional elongation by pol II
SINE transcripts	Unknown
VA RNA	Involved in adenovirus translational control
EBER RNA	Thought to be involved in Epstein-Barr virus translational control

surprising that these enzymes share some common subunits. However, the polymerases transcribe distinct genes with different characteristics and so they also have several unique properties. The subunits unique to pol III are presumed to contribute to pol III-specific properties, such as its distinct nuclear localisation, its interaction with specific transcription initiation factors and its ability to perform elongation and termination of transcription independently of accessory factors.

Genes Transcribed by RNA Polymerase III

Pol III transcribes genes that encode a variety of small (usually less than 300 bp), stable RNAs that do not get translated. As summarised in Table 2, many of these RNAs, including 5S ▶**rRNA** and transfer (t)RNAs, have essential functions in cellular metabolism, being involved in protein synthesis and transport and RNA processing (1, 2).

Promoters of Genes Transcribed by RNA Polymerase III

Gene promoters contain specific sequence elements, which direct the recruitment of the appropriate transcription factors and RNA polymerase through multiple protein-DNA and protein-protein interactions. One of the most notable features of the majority of class III gene promoters is that the crucial sequence elements are found downstream of the transcription start site,

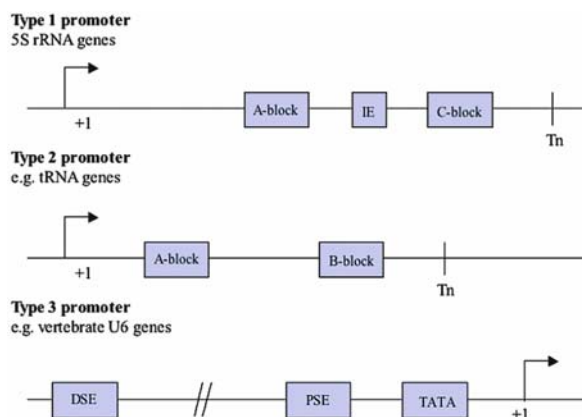
within the transcribed region (as shown in Fig. 1), rather than upstream as with class I and II genes. Promoters with such a structure are referred to as type 1 or 2. Type 1 promoters are unique to 5S rRNA genes and require three internal promoter elements for efficient transcription, an A-block, an intermediate element (IE) and a C-block (1, 2, 3, 4). The spacing between these elements is relatively restricted.

The majority of pol III transcribed genes, including those for **tRNAs** and adenovirus VA RNAs, have type 2 promoters. These consist of two essential, highly conserved sequence elements, an A-block (homologous to the type 1 promoter A-block) and a B-block. The spacing between the A- and B-blocks is not as restricted as the spacing between type 1 promoter sequence elements and in fact the position of the B-block is highly variable (1, 2, 3, 4).

A minority of pol III-transcribed genes in vertebrates have a type 3 promoter. Genes with type 3 promoters lack any requirement for internal promoter regions, with important sequence elements being found upstream of the transcription start site. For example, efficient transcription of the vertebrate U6 genes requires three extragenic promoter elements, a TATA box, a proximal sequence element (PSE) and a distal sequence element (DSE) (1, 2, 3, 4). The three types of promoter used by pol III are illustrated in Fig. 1.

Transcription by RNA Polymerase III

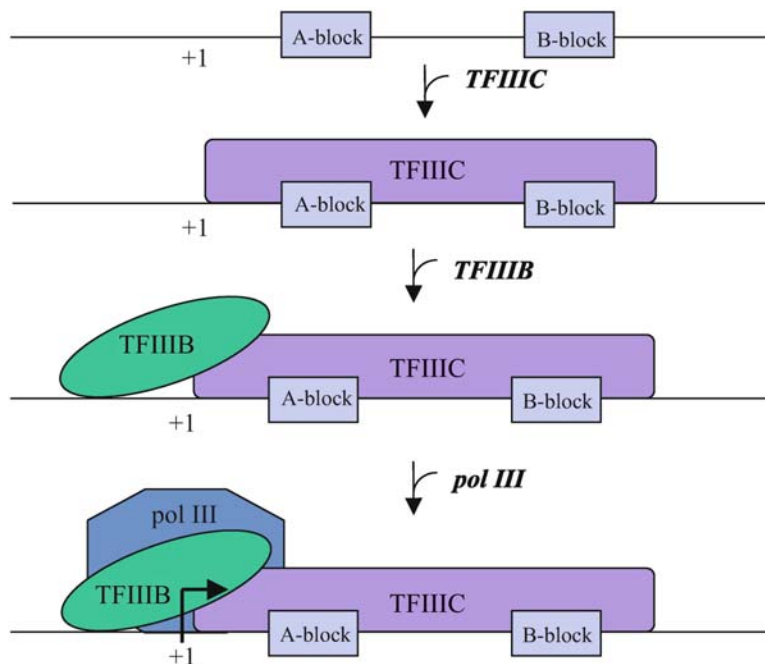
The transcription of a class III gene can be divided into three main stages, initiation, elongation and



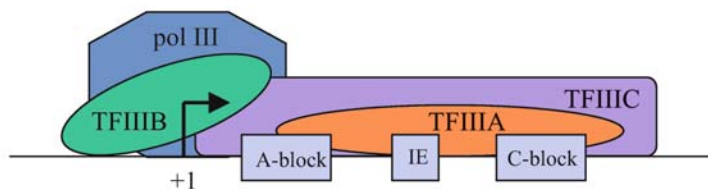
RNA Polymerase III. Figure 1 Promoter structures utilised by pol III. The transcription start site on each promoter is indicated by '+1' and 'Tn' indicates the termination site. The important sequence elements of each promoter type are represented by coloured boxes. These are the A-, B-, and C-blocks, the intermediate element (IE), the TATA box (TATA), the proximal sequence element (PSE) and the distal sequence element (DSE).

termination. Initiation of transcription begins with the assembly of a transcription initiation complex at the promoter. Regardless of the promoter type, pol III itself has little affinity for promoter sequence elements and so relies on the specific transcription factors which form this transcription initiation complex for its accurate recruitment. Figure 2a shows a simplified illustration of the formation of a transcription initiation complex on the most commonly used pol III promoter type (type 2). This process requires two pol III-specific, basal transcription factors, the DNA-binding factor TFIIC and the polymerase recruitment factor TFIIB. TFIIC is one of the most complex transcription factors ever studied, having six subunits in yeast with an aggregate mass of 520 kD. Human TFIIC is even larger and more complex than the yeast factor and, in fact, there is little homology between the TFIIC subunits from these two species (2, 3, 4). Nonetheless, regardless of the species, TFIIC binds the A- and B-blocks of type 2 promoter DNA directly and then serves to recruit TFIIB. TFIIB is essential for the recruitment of pol III to all class III genes and consists of at least three proteins, TATA binding protein (TBP) and two TBP-associated factors, known as BRF1 and BDP1. Both BRF1 and BDP1 have been shown to make direct contacts with TFIIC, thus allowing TFIIB recruitment to the promoter (3). Each of the TFIIB components is essential for subsequent polymerase recruitment; however, only BRF1 has been shown to make direct contacts with the polymerase itself.

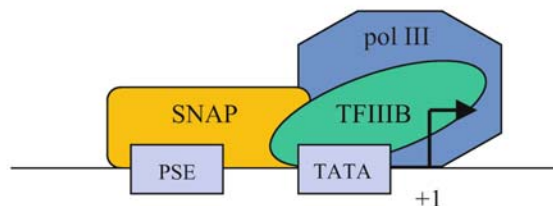
Figures 2b and c show transcription initiation complexes on type 1 and 3 promoters (4). As with type 2 promoters, TFIIB is directly involved in the recruitment of pol III. However, these promoter types rely on additional, gene-specific factors for efficient TFIIB recruitment. For example, the gene-specific transcription factor TFIIA is required for the productive recruitment of TFIIC, and hence TFIIB, to 5S rRNA (type 1) promoters (1, 2, 3, 4). Binding of TFIIB to type 3 promoters in vertebrates does not rely on TFIIC, but on an alternative factor known as SNAPc (1, 2, 3, 4). SNAPc binds to the PSE of type 3 promoters and facilitates the recruitment of TFIIB to the TATA box. The DSE of type 3 promoters binds an additional transcription factor known as Oct1, which further enhances SNAPc/TFIIB binding to promoter DNA. A form of TFIIB unique to vertebrates is required for pol III transcription from type 3 promoters (4). Thus, although the route to TFIIB binding differs for each promoter type, the outcome is the same; TFIIB brings pol III into the vicinity of the transcription start site, thereby facilitating transcription initiation. Table 3 summarises the basal transcription factors utilised by pol III and outlines their functions.



a *Assembly of a *pol III* transcription initiation complex on a type 2 promoter.*



b *Pol III transcription initiation complex on a type 1 promoter.*



c *Pol III transcription initiation complex on a type 3 promoter in the vicinity of the transcription start site.*

RNA Polymerase III. Figure 2 Transcription initiation complexes utilised by class III genes. The transcription start site on each promoter is indicated by '+1'. The important sequence elements of each promoter type are represented by boxes. These are the A-, B-, and C-blocks, the intermediate element (IE), the TATA box (TATA), the proximal sequence element (PSE) and the distal sequence element (DSE). *Pol III* and its associated transcription factors (TFIIB, TFIIC, TFIIIA and SNAPc) are shown as coloured shapes. On type 2 promoters (a), TFIIC recognises and binds the A- and B-blocks directly. Once bound to the DNA, TFIIC then recruits TFIIB via protein-protein interactions. TFIIB can then recruit *pol III* and transcription is initiated. Recruitment of TFIIC, and hence TFIIB and *pol III*, to type 1 promoters requires TFIIIA (b). TFIIB recruitment to type 3 promoters requires SNAPc (c).

RNA Polymerase III. Table 3 Summary of basal pol III transcription factors and their functions

Basal Pol III Transcription Initiation Factor	Known Functions
TFIIIB	Essential for the initiation of transcription of all class III genes. Recruits pol III directly.
TFIIIC	Binds directly to type 2 promoters, and to TFIIIA-bound type 1 promoters. Recruits TFIIIB.
TFIIIA	Binds specifically to type 1 promoters and recruits TFIIIC.
SNAPc	Binds PSE of type 3 promoters in vertebrates. Involved in the recruitment of TFIIIB to type 3 promoters.
Oct-1	Binds DSE of type 3 promoters in vertebrates. Necessary for the productive recruitment of SNAPc to the type 3 promoter PSE.

Following pol III recruitment, the two strands of DNA around the transcription start site are separated to form a transcription bubble (2, 3). This melting of the DNA helix allows the polymerase to access the template strand, and is required before transcription can proceed. DNA melting is performed by the polymerase, although the BRF1 and BDP1 components of TFIIIB also play an active role. Thus, TFIIIB serves not only to recruit pol III, but also participates in the formation of an open promoter complex (2, 3, 4).

Once the DNA strands have been separated, RNA synthesis can be initiated and the polymerase progresses into the gene and dissociates from promoter-bound TFIIIB (2). The transcription bubble moves downstream with the elongating polymerase. Although TFIIIC assembles within the transcribed regions of the majority of class III genes, this large factor is not dissociated from promoters during elongation (2). It is unclear at present how the polymerase passes DNA-bound TFIIIC during transcription. However, unlike pols I and II, pol III does not require any accessory factors for efficient chain elongation (2, 3, 4). Furthermore, termination by pol III also occurs independently of other factors; four or more T residues within the template strand of a class III gene are sufficient to signal the accurate and efficient termination of transcription (2, 3, 4).

Following the synthesis of the first transcript, pol III is known to be recycled on the same DNA template for several further rounds of transcription. As a

consequence, the slow initial step of polymerase recruitment is avoided, making the production of subsequent RNAs by pol III more efficient (1, 2).

Regulation of Transcription by RNA Polymerase III

As detailed in Table 2, many pol III products are known to be involved in protein synthesis, and hence are essential determinants of the biosynthetic capacity of cells. As a consequence, pol III transcription is tightly regulated, allowing cells to control the rate of protein synthesis in response to changing metabolic needs. Numerous studies in eukaryotic systems have demonstrated that pol III transcription is regulated under a variety of cellular and environmental conditions (2). For example, pol III transcription is low when nutrients or mitogens are limiting, but up-regulated when the availability of these growth stimuli increases. Pol III transcription is also regulated in response to a range of cellular stresses and during many fundamental cellular processes, such as during metazoan development and [cell differentiation](#), and during the [cell cycle](#) (2). This stringent control of pol III transcription is achieved by several regulatory proteins, which can interact with or modify components of the basal pol III transcriptional machinery and either activate or repress transcription. Some of these regulators are also responsible for coordinating the activities of all three nuclear RNA polymerases. Examples of pol III regulators and their physiological relevance are discussed below. However, several other molecules, not discussed here, have also been shown to regulate pol III transcription in various organisms. For a more complete review of these regulators see (2).

Negative Regulators of Pol III Transcription

Proteins known to act as negative regulators of pol III transcription include Dr1, Maf1 the [retinoblastoma protein](#) ([pRB](#)), and [p53](#) (1, 2, 3, 4, 5). Dr1 is a small nuclear protein, highly conserved in all eukaryotes. It represses pol III transcription by disrupting the interaction between TBP and BRF1, thereby inactivating TFIIIB, and preventing pol III recruitment to class III genes. Dr1 is also known to inhibit pol II transcription and can therefore coregulate the activities of both pols II and III (1, 2).

Maf1 was recently identified in yeast and shown to be a negative effector of pol III transcription in response to a variety of cellular stresses such as nutrient limitation and DNA damage (4). Although homologues of Maf1 have been identified in higher eukaryotes, a role for Maf1 in the regulation of pol III transcription has thus far only been demonstrated in yeast.

Unlike Dr1 and Maf1, RB and p53 are restricted to metazoan systems. RB is an important regulator of the mammalian cell-cycle, ensuring cells only progress through the cycle and proliferate in response to

appropriate external stimuli. Loss of RB function results in aberrant cellular proliferation and tumour formation. RB is therefore known as a ►**tumour suppressor** protein. In resting cells, RB binds and inactivates TFIIB, thereby restraining the unnecessary production of tRNAs, 5S rRNA and other pol III products (5). However, in response to growth stimuli, RB becomes inactivated by hyperphosphorylation. This leads to the release of TFIIB and the activation of pol III transcription, thus permitting an increase in protein synthesis and growth. This is a prerequisite for cell division, as cells must duplicate their contents prior to dividing in order to maintain a constant size. RB also regulates the transcription by pol II of certain genes whose products are involved in DNA synthesis, and transcription by pol I, which synthesises the large rRNAs. Therefore, this common regulatory mechanism allows components of the DNA and protein synthetic machinery to be produced in a highly coordinated manner during the cell-cycle.

p53 is also an important cell-cycle regulatory protein, mediating cell-cycle arrest or cell death in response to a variety of stresses including radiation, hypoxia and oncogenic stimuli. As with RB, p53 is a tumour suppressor protein whose function is thought to be compromised in all human tumours. Although unrelated to RB, p53 also represses pol III transcription by binding and inactivating TFIIB (5).

Positive Regulators of Pol III Transcription

Mitogenic stimulation of eukaryotic cells results in the activation of several intracellular signalling cascades, which promote growth and proliferation. Some of these cascades have been shown to lead to the activation of pol III transcription. In many instances, stimulation of pol III transcription results from an activation of TFIIB. As already discussed, derepression of TFIIB by RB in response to growth stimuli is an important mechanism for pol III activation in mammalian cells. However, several other mechanisms for the up-regulation of pol III transcription also exist. For example, ►**CK2**, a highly conserved eukaryotic protein kinase activated in response to growth stimulation, has recently been shown to phosphorylate and activate TFIIB, thus increasing pol III transcription in yeast and mammalian cells (4). Furthermore, the mitogen-activated protein kinase ►**ERK**, which is activated *via* the ►**ras** signalling pathway in mammalian cells, can also induce pol III transcription through phosphorylation and activation of TFIIB. In addition to these kinases, the ►**proto-oncogene** product ►**c-Myc** is a crucial regulator of cell growth in metazoans. c-Myc specifically activates pol III transcription again by targeting TFIIB, leading to an increase in tRNA and 5S rRNA production (5). Thus, it is clear that the pol III transcriptional machinery is a common target for

numerous growth-promoting pathways within eukaryotes.

Stimulation of pol III transcription in mammalian cells also occurs in response to viral infection (2, 5). For example, adenovirus, simian virus 40, polyomavirus, herpes simplex virus, hepatitis B virus (HBV), human T-cell leukaemia virus type 1, human papilloma virus (HPV), and pseudorabies virus have all been associated with pol III activation (2, 5). The resulting increase in the cellular biosynthetic capacity may be a general requirement for efficient viral propagation. In some cases, expression of certain viral genes also requires the host pol III transcriptional machinery directly. Multiple mechanisms, which target both TFIIB and TFIIC (2, 5), are thought to contribute to this increase in pol III transcription.

Clinical Relevance

The regulation of pol III transcription is clearly a crucial feature in normal cellular growth and proliferation. In mammals, pol III transcription can be restrained by two key tumour suppressors, p53 and RB and activated by known proto-oncogene products including CK2, ras and c-Myc (5). Inactivation of p53 and RB and abnormal activation of proto-oncogenes are common features of most human cancers, causing uncontrolled cell growth and division. High levels of pol III transcription are likely to contribute to the rapid growth that is characteristic of tumour cells. Indeed, it is well established that pol III transcripts are elevated in many tumour types, including ovarian and breast carcinomas (2, 5). Furthermore, viruses associated with human cancers are known to stimulate pol III transcription. For example, HPV, the causative agent in cervical carcinoma, and HBV, which can cause hepatocellular carcinoma, both activate TFIIB and hence pol III transcription (2, 5). Deregulation of TFIIB activity through viral infection, loss of tumour suppressors and/or activation of oncogenes is likely to play a role in the stimulation of pol III transcription in tumour cells. Abnormally elevated expression of TFIIC may also contribute to increased pol III transcription in tumours, as TFIIC subunits are consistently over-expressed in certain tumour types and virally transformed cell lines (5). Understanding the mechanistic basis and functional significance of the aberrant activation of pol III transcription in cancers may lead to the development of novel prognostic tools and therapies.

References

1. White RJ (2001) Gene Transcription. Mechanisms and Control. Blackwell Science Ltd., Oxford, UK
2. White RJ (2002) RNA polymerase III transcription 3rd Edition. Landes Bioscience, Austin. <http://www.eurekah.com>

3. Geiduschek EP, Kassavetis GA (2001) The RNA polymerase III transcription Apparatus. *Journal of Molecular Biology* 310:1–26
4. Schramm L, Hernandez N (2002) Recruitment of RNA polymerase III to its target promoters. *Genes and Development* 16:2593–2620
5. Felton-Edkins ZA, Kenneth NS, Brown TRP et al (2003) Direct regulation of RNA polymerases III transcription by RB, p53 and c-Myc. *Cell Cycle* 2:181–184

RNA Profiling

Definition

RNA profiling refers to the determination of the change in RNA expression (on, off, strong, weak, etc) of all or a majority of genes due to an experimental manipulation (genetic, surgical, environmental, etc.). Typically done with high-density DNA microarrays.

- ▶ [Drosophila Model of Cardiac Disease](#)
- ▶ [Proteomics in Human-Pathogen Interactions](#)
- ▶ [RNA Expression Profiling](#)

RNA Transport, Nucleo-Cytoplasmic Trafficking

- ▶ [RNA Export](#)

RNA Viruses

Definition

RNA viruses are viruses that use ribonucleic acid (RNA; either single or double stranded) as their genome. ▶ [Retroviruses](#) are also RNA viruses.

- ▶ [Cap-Independent Translational Control](#)

RNA-Induced Silencing Complex

Definition

RNA-Induced Silencing Complex (RISC) is a multi-component nuclease that cleaves messenger RNAs homologous to the silencing trigger. RNAi is mediated by ▶ [RISC](#).

- ▶ [Catalytic RNA](#)

RNAP I

- ▶ [RNA Polymerase I Transcription](#)

RNAse

Definition

A class of enzymes that rapidly degrade RNA molecules

- ▶ [Microarrays in Pancreatic Cancer](#)

RNP Complexes

Definition

There are numerous small nuclear (sn) and small nucleolar (sno) RNAs that complex with specific proteins to form ribonucleoprotein particles (RNPs) which are involved in RNA processing.

- ▶ [Nuclear Compartments](#)

RNS

Definition

The term reactive nitrogen species (RNS) is frequently used to describe the nitroxyl radical ($\cdot\text{NO}$), and radical or non radical species derived from molecules such as NO_x and peroxy nitrite.

- ▶ [Free Radicals](#)

Rods

Definition

Rods are elongated photoreceptors of high light sensitivity, which contain the photopigment rhodopsin in a stack of disks that forms their outer segment. They transform light via the phototransduction process into

an electrical signal that modulates the release of transmitter (glutamate) in the inner segment. In the human retina, rods – in contrast to ►cones – contain only a single photopigment with a spectral peak at 500 nm.

►Retinitis Pigmentosa

ROI

►Region of Interest

Rolling Circle Mechanism

Definition

The rolling circle mechanism of DNA replication starts from circular DNA and is used in the replication of the DNA of some bacteriophages, and in the amplification of DNA coding for ribosomal RNA in amphibia. The circle is copied many times (as if it were rolling), giving rise to a long DNA strand that consists of many linked copies of the circular DNA.

►DNA Amplification

ROR Response Element

Definition

ROR response element refers to DNA recognition sequences for members of the ►REV-ERB and ROR orphan nuclear receptor families. The consensus sequence is WAWNTRGGTCA (where W = A or T and R = A or G). ROR stands for “retinoic acid receptor-related orphan receptor”

►Circadian Clocks

ROS

►Reactive Oxygen Species

Rosetta Stone Proteins

Definition

Some interacting proteins such as acetate CoA transferase in *E. coli* are fused into a single chain in other organisms. The homologous protein in human consists of two subunits, succinyl-CoA transferase a and b. The fused protein is called a Rosetta stone protein after the historical Rosetta stone of ancient Egypt, which contained the same text in three languages. The presence of Rosetta stone protein implies that the two independent proteins may be functionally linked with each other.

►Protein-Protein Interaction

►Two Hybrid System

Rostrocaudal

Definition

Rostrocaudal designates the longitudinal axis of the body from head to tail.

►Muscle Development

RPE

►Retinal Pigment Epithelium

RRF

Definition

Ragged red fibers (RRF) are muscle fiber segments with large accumulations of mitochondria. The mitochondrial accumulations are unevenly distributed and can be stained red with particular tissue stains (Gomori trichrome stain), hence the name, ragged red fibers. RRF are considered a hallmark of mitochondrial myopathy.

►Mitochondrial Myopathies

rRNA

Definition

- Ribosomal RNA
- RNA Polymerase III

rRNA Processing Events

Definition

The 45S ribosomal RNA (rRNA) primary transcript undergoes extensive modification prior to its cleavage to generate rRNAs. These modifications include approximately 100 nucleotide methylations, and 100 isomerisations of uridine to pseudouridine. Small nucleolar RNAs (snoRNAs) are important in rRNA processing, with each snoRNA acting as “guide molecule” directing a specific modification.

- [Nuclear Compartments](#)

RSK

Definition

Receptor serine/threonine kinases (RSK) comprise of a small family of kinases localized at the plasma membrane. They mediate inhibitory as well as stimulatory signals for growth and differentiation by binding to members of the transforming growth factor- β (TGF- β) superfamily. They are generally subdivided to Type I and Type II RSKs.

- [Receptor Serine/Threonine kinases](#)

RSTS

- [Rubinstein-Taybi Syndrome](#)

RT

- [Reverse Transcription](#)

RTase

- [Reverse Transcriptase](#)

RTK/RTK1

Definition

Receptor tyrosine kinases (RTK) comprise of a large family of cell surface receptors for polypeptide growth factors. RTKs are membrane spanning proteins with an extracellular, ligand-binding domain and an intracellular kinase domain. There are many subfamilies of RTKs; each subfamily has characteristic ligand-binding and kinase domains and is activated by distinct ligands, for example PDGFs, FGFs or ephrins. Upon ligand binding, the extracellular ligand binding domain phosphorylates certain tyrosine residues in other cytosolic proteins or adjacent RTK molecules (substrates) autocatalytically.

- [Peutz-Jeghers Syndrome](#)
- [Receptor Serine/Threonine Kinases](#)
- [Tyrosine Kinases](#)

RTKs

Receptor Tyrosine Kinases

- [Tyrosine Kinases](#)

RT-PCR

- [Reverse Transcription-Polymerase Chain Reaction](#)

Rubinstein-Taybi Syndrome

Definition

Rubinstein-Taybi syndrome is a well recognizable disorder (1:100000 newborns) that is caused by mutation in the gene encoding the transcriptional coactivator CREB-binding protein (CREBBP); the functional loss of one CREBBP allele is due to mutation in 90% of the cases, in about 10% it is caused by deletion at chromosome 16p13.3. The disorder is characterized by short stature, microcephaly, characteristic face, broad thumbs and great toes

with radial or tibial deviation and moderate mental retardation.

- ▶ [Chromatin Acetylation](#)
- ▶ [Microdeletion Syndromes](#)

RxRE

- ▶ [REE, RxRE](#)
- ▶ [Rev, Rex](#)

S Phase

Definition

S phase describes the cell cycle phase during which DNA replication occurs and centrosomes duplicate.

► [Cell Cycle – Overview](#)

Saltatory Nerve Conduction

Definition

Saltatory nerve conduction refers to transmission of impulses along myelinated axons in vertebrates, with action potentials restricted to the ► [Nodes of Ranvier](#).

► [Glial Cells and Myelination](#)

S-Acylation

► [Thioester-Linkage](#)

SAD

► [Single-Wavelength Anomalous Diffraction](#)

SAGA

Definition

SAGA is a yeast histone acetyltransferases complex Spt/Ada/Gcn5.

► [Polyglutamine Disease, the Emerging Role of Transcription Interference](#)

SAGE

► [Serial Analysis of Gene Expression](#)

Salvage Pathway

Definition

Salvage pathway refers to a short sequence of enzyme-catalyzed reactions in the purine and pyrimidine metabolism which uses preformed purine or pyrimidine bases, or nucleosides to form nucleotides.

► [Nucleotide Biosynthesis](#)

SAMP Repeats

Definition

SAMP repeats are found three times in the tumor suppressor APC, and serve as binding sites for axin/conductin. SAMP repeats are about 30 amino acids long and contain at their core the invariant Ser-Ala-Met-Pro sequence.

► [Colorectal Cancer](#)

Sampling

Definition

Sampling designates the process of generating configurations for a given system.

► [Molecular Dynamics Simulation in Drug Design](#)

Sanger Sequencing Method

Definition

In contrast to the Maxam/Gilbert-Method of DNA-sequencing that is based on base-specific chemical modification, cleavage of the DNA at the modified sites, electrophoresis of the fragments, and autoradiography of the sequencing, the method of Sanger (also known as the dideoxycchain termination method) is based on hybridisation by a short primer, synthesis by DNA-polymerase In the presence of desoxynucleosidtriphosphates, stop of synthesis by addition of 2'3'-didesoxynucleosidtriphosphates, electrophoreses of the cleavage products, and autoradiography. The method of Sanger developed by Frederick Sanger and co-workers is the most common DNA sequencing method.

► [SNP Detection and Mass Spectrometry](#)

Sanglifehrin

Definition

Sanglifehrin is a 22-membered macrocycle isolated from the actinomyocyte *Streptomyces* sp. A92–308110. It shows strong affinity for members of the ► [cyclophilin](#) family with peptidyl-prolyl isomerase activity (PPIase). It has immunosuppressive properties but does not inhibit calcineurin activity.

► [Peptidyl Prolyl Cis/Trans Isomerases](#)

SAR1

Definition

SAR1 designates a small GTPase of the ► [Ras](#) super-family involved in recruiting COP II coat proteins to endoplasmic reticulum membranes for selective export of newly synthesized, properly folded proteins in COP II-coated vesicles.

► [Vesicular Traffic](#)

SARA

Definition

► [Smad](#) anchor for receptor activation (SARA) is a protein that binds to phospholipids in the membrane of the early endosome via its FYVE domain, and also binds to Smad2, Smad3 and TβRI, thus mediating the phosphorylation of Smads by Type I ► [RSKs](#). SARA is also thought to act as an anchor of Smads in the cytoplasm, inhibiting their intrinsic ability to translocate into the nucleus.

► [Receptor Serine/Threonine Kinases](#)

SAR-by NMR

Definition

Structure activity relationship (SAR) by NMR is a drug discovery and design strategy developed by Fesik and co-workers, which involves NMR-based screening of compound libraries followed by chemically combining weakly interacting compounds.

► [3D Structure by NMR](#)

Sarcolemma

Definition

Sarcolemma is the plasma membrane of the skeletal muscle fiber.

► [Cardiac Signaling: Cellular, Molecular and Clinical Aspects](#)

► [Limb Girdle Muscular Dystrophies](#)

► [Muscle Contraction](#)

► [Muscle Development](#)

Sarcomas

Definition

Sarcomas are cancers arising from connective tissue or muscle cells.

► [Tumor Suppressor Genes](#)

Sarcomere

Definition

Sarcomere is the repetitive contractile unit of the myofibril of skeletal muscle delimited by two Z-discs, and constituted by the thick and thin filaments.

- [Cardiac Signaling: Cellular, Molecular and Clinical Aspects](#)
- [Heart](#)
- [Limb Girdle Muscular Dystrophies](#)
- [Muscle Contraction](#)
- [Muscle Development](#)

Sarcoplasmic Reticulum

Definition

Endoplasmic reticulum found in striated muscle fibers is designated as sarcoplasmic reticulum. It consists of longitudinal tubules and terminal cisternae.

- [Cardiac Signaling: Cellular, Molecular and Clinical Aspects](#)
- [Heart](#)
- [Muscle](#)

Satellite

Definition

The term satellite designates an array of tandemly repeated DNA units spanning up to 10^6 bp, which are mainly present in centromeric and paracentromeric heterochromatin. Satellite DNA usually forms a prominent “satellite band”, separable from the rest of genomic DNA by density gradient centrifugation.

- [Repetitive DNA](#)

Satellite Cell

Definition

Satellite cell is a quiescent myogenic stem cell associated with muscle fibres in adult muscle. These cells have a limited proliferative capacity and contribute to fibre growth and muscle repair following injury.

- [Muscle Development](#)

Saturation Point

Definition

During the preparation of suitable crystals of membrane proteins for electron microscopy investigations a number of steps have to be performed. One step is the extraction of the desired protein from its native environment in the membrane. In this context, the saturation point defines the point in the solubilization process, where the biological membrane is saturated with detergent molecules intercalating between lipids and proteins.

- [Two-dimensional Crystallization of Membrane Proteins](#)

SBMA

- [Spinobulbar Muscular Atrophy](#)

SCA

Definition

Spinocerebellar ataxia (SCA) is a heterogeneous group of diseases characterized by progressive degeneration in the cerebellum and spinal cord. The basic defect in most types of spinocerebellar ataxia is expansion of a CAG triplet repeat. In this way, it is similar to fragile-X syndrome, Huntington disease and myotonic dystrophy (DM1), all of which exhibit a triplet repeat expansion of a gene.

- [Histone Acetylation in Polyglutamine Disease](#)
- [Huntington's Disease](#)
- [Polyglutamine Disease, the Emerging Role of Transcription Interference](#)
- [Repeat Expansion Diseases](#)

Scaffolding Protein

Definition

Scaffolding proteins harbour several modular domains (e.g. PDZ domains) that mediate the interaction with other molecules like proteins or DNA. Scaffolding proteins are involved in the assembly of multi-molecular complexes.

- [Microvilli](#)

Scanning Tunneling Microscope

Definition

The scanning tunneling microscope (STM) uses a metal tip to which a voltage is applied while scanning conducting surfaces. A so called tunneling current between tip and surface already runs a few angstroms before the tip makes contact. Its amperage depends with extreme sensitivity on the tip-surface distance, and can therefore be used as a distance control that allows the exploration of the surface roughness. This instrument made it possible to image single atoms for the first time.

► [Atomic Force Microscopy](#)

Scar

Definition

Scar is a mark left by the healing of injured tissue. The scar is characterized by a dense fibrotic tissue covered by the epidermis as well as by a lack of all appendages.

► [Wound Healing](#)

Scar/Wave Proteins

Definition

Scar/Wave proteins are members of the Wiscott-Aldrich Syndrome protein (WASP) family, which stimulate actin polymerization by coupling Rho-GTPases and other signalling proteins to the Arp2/3 complex.

► [Rho](#), [Rac](#), [Cdc42](#)

Scavenger Receptor(s)

Definition

Scavenger receptor(s) comprise of cellular receptors that recognize modified LDL due to oxidation or experimental acetylation.

► [High-HDL Syndrome](#)

SCC

► [Squamous Cell Carcinoma](#)

SCE

► [Sister Chromatid Exchange](#)

SCF Ubiquitin Ligase

Definition

The SCF (Skp -Cullin -F-box) ubiquitin-ligases are a group of highly diverse multi-subunit ubiquitin ligases. SCF triggers the degradation of various proteins including CDK inhibitors and cyclin E. The core subunits of the complex are always Skp1 and a Cullin, while the variable adaptor – the F-box protein – provides the variability to recognize a specific target molecule. The target molecule is then fixated in close proximity to the E2-enzyme which catalyses the transfer of ubiquitin onto the substrate.

► [Cell Cycle – Overview](#)

► [Ubiquitination](#)

Schiff Base

Definition

A Schiff Base is a compound containing a C = N-double bond, which is obtained by the condensation of an aldehyde or keton with a primary amine. In rhodopsins the Schiff base is formed by the reaction of the ε-amino group of Lys residue located in the middle of the seventh helix (H-VII; H-G) with the aldehyde group of retinal.

► [Photoreceptors](#)

Schizophrenia, Genetics

NADINE NORTON, MICHAEL J OWEN

Department of Psychological Medicine, Wales School of Medicine, Cardiff University, Cardiff, UK

nortonn@cardiff.ac.uk

owenmj@cardiff.ac.uk

Synonyms

Dementia praecox

Definition

Schizophrenia is a severe and disabling psychiatric disorder often with symptoms of hallucinations, delusions and thought disorder. Symptoms usually begin in late adolescence or early adulthood and are heterogeneous. Some psychiatric disorders have a specific identifiable pathology. However, it has not been possible to identify changes in the brain or other tissues that are “diagnostic” of schizophrenia. Thus diagnosis is purely clinical and is specified in the fourth revision of the Diagnostic and Statistical Manual of Mental disorders (DSM IV).

Characteristics

The prevalence of the disorder in the general population is 0.5–1% with compelling evidence from family, twin and adoption studies for genetic factors in its aetiology.

1. Family studies compare the prevalence of the disorder in relatives of affected individuals with the general population. The lifetime risk of 1% in the general population contrasts with the average risk to siblings (9%) and offspring of affected individuals (13%). The offspring of dual matings (between two affected parents) have a risk for the disorder of 46%.
2. Twin studies compare the concordance of the disorder in ►dizygotic (sharing on average 50% genes) and ►monozygotic (sharing 100% genes) twins. A dizygotic twin with an affected co-twin has a risk for the disorder of 17% and a monozygotic twin with an affected co-twin has a risk of 48%.
3. Adoption studies using three designs (►adoptive studies, ►cross-fostering studies, and ►adoptive family studies) have demonstrated a greater risk of schizophrenia in biological relatives of cases (where genes are shared but not family environment and experiences) than in control groups, further corroborating evidence of a genetic contribution.

Mode of Inheritance

Despite there being a clear genetic contribution to schizophrenia, ►discordance between monozygotic twins indicates that what is inherited is not the certainty of disease accompanying a particular ►genotype but rather a predisposition to developing the disorder, and certainly that other factors (such as environment) are involved. Furthermore, studies on different classes of relatives show that the recurrence risk decreases too rapidly with decreasing relatedness to an affected individual for schizophrenia to be a single-gene disorder or a collection of single-gene disorders. The

most likely mode of transmission is ►oligogenic or ►polygenic, also observed in other common disorders such as diabetes and asthma. This means that schizophrenia susceptibility is most likely the combined effect of the inheritance of a combination, or combinations, of different susceptibility and protective ►alleles in a number of (possibly many) genes. Until such genes have been identified, it is not possible to determine the number of susceptibility loci, the disease risk conferred by each locus or the degree of interaction among loci.

Cellular and Molecular Regulation

There are four main approaches to identifying disease genes in human populations.

Chromosomal Abnormalities

Identification of chromosomal abnormalities that co-segregate with schizophrenia may provide valuable clues to the location of susceptibility genes. The first example of this phenomenon came from a large Scottish family with 47% prevalence of mental illnesses including schizophrenia. A balanced reciprocal ►translocation between chromosomes 1 and 11 co-segregated with illness. Subsequent mapping of the translocation breakpoint has identified three disrupted genes of unknown function on chromosome 1. Unfortunately, such abnormalities have not been identified in other families and the mechanism by which the translocation confers disease susceptibility is so far unknown.

A second example of cytogenetic abnormality in schizophrenia is the association with chromosome 22q11 deletion syndrome (also known as DiGeorge syndrome and velocardiofacial syndrome [VCFS]). 22q11 deletion syndrome occurs in 1/4000 live births and results in a complex ►phenotype with multiple congenital abnormalities, characteristic dysmorphism, cleft palate, cardiac defects, learning disabilities and high rates of psychiatric disorder. Rate of schizophrenia in adults carrying the deletion is as high as 24%. Deletion of 22q11 has been detected in 1–2% of individuals with schizophrenia and only accounts for a small proportion of the risk of developing schizophrenia in the general population. However, mutations and ►polymorphisms in genes within the deleted region may make a more general widespread contribution to schizophrenia susceptibility and there is some evidence for genetic linkage and association to this region.

Genetic Linkage

Linkage studies aim to identify chromosomal regions that are co-transmitted with the disease in families with two or more affected individuals. Replicated positive linkages to several chromosomal regions are now

accumulating with evidence for linkage on chromosomes 1q21-q22, 5q21-q31, 6p24-p22, 6q, 8p22-p21, 10p15-p11, 13q14.1-q32 and 22q11-q22. With the advent of the human genome sequence, it is probably only a matter of time before susceptibility genes are identified by association studies focusing on genes within these regions.

Association Studies

Association studies aim to detect variations in genes (alleles) that are more (or less) common in cases than in the control population. Such studies are widely used on genes that are hypothesized to play a role in schizophrenia on the basis of their function. Such genes are termed “candidate genes”. To date, genes involved in dopaminergic and serotonergic ►[neurotransmission](#) have received a great deal of attention based on neurochemical evidence that antipsychotic drugs bind to dopamine and serotonin receptors. Two positive associations with schizophrenia are polymorphisms causing an amino acid variation (Ser9Gly) in ►[exon 1](#) of the dopamine D₃ receptor ►[gene](#) and a thymine/cytosine polymorphism at nucleotide 102 of the 5HT_{2A} gene but there are conflicting data for both findings. The catechol-O-methyltransferase (COMT) gene located on chromosome 22q11 is both a strong positional and functional candidate for schizophrenia. The gene encoding *COMT* is located within the VCFS deleted region and the protein metabolizes both dopamine and serotonin. An amino acid variation (Val/Met) in *COMT* has been associated with variation in enzyme activity. A recent study has identified a significant association with several variants (including the Val/Met polymorphism) that make up a ►[haplotype](#) for disease susceptibility. The Icelandic deCODE Genetics group showed association of the neuregulin (*NRG1*) gene (located on chromosome 8p22-p21) to schizophrenia identifying a significant core haplotype for disease susceptibility (relative risk 2.1). This finding has since been replicated in a Scottish population. A second positive association has come from the dysbindin (*DTNBP1*) gene on chromosome 6p. Initial association was with a rare haplotype in Irish families and replication has since been achieved in a German population.

Recent association findings in the *COMT*, *NRG1* and *DTNBP1* genes are potentially very important but should be viewed with caution. For each gene, it is the combination of several variants that show association to schizophrenia, and it has not been possible to identify a particular genetic variant that is responsible for increased susceptibility.

Microarray Studies

Microarrays are a new tool in molecular genetics for gene expression analysis. A recent study identified the

expression of several ►[myelination](#) related genes to be down-regulated in schizophrenic patients. Association studies of the genes identified by such analyses will undoubtedly follow.

Clinical Relevance

Schizophrenia is a severe and disabling psychiatric disorder. Once symptoms occur they persist for the entire lifetime of the patient. Not all patients respond well to current medications and unpleasant side effects often result. Genetic studies aim to identify susceptibility genes for schizophrenia so that we can better understand disease mechanisms, design more effective drugs with fewer side effects and provide earlier diagnosis and improved treatment.

References

1. Gottesman II (1991) Schizophrenia Genesis. The Origins of Madness. Freeman, New York
2. McGuffin P, Owen MJ, Gottesman II (2002) Psychiatric Genetics and Genomics. Oxford University Press, Oxford, New York
3. Murphy KC, Owen MJ (2001) Velo-cardio-facial syndrome (VCFS): a model for understanding the genetics and pathogenesis of schizophrenia. Br J Psychiatry 179:397–402
4. Harrison PJ, Owen MJ (2003) Genes for schizophrenia? Recent findings and their pathological implications. Lancet 361:417–419
5. Stefansson H, Sigurdsson E, Steinthorsdottir V et al (2002) Neuregulin 1 and susceptibility to schizophrenia. Am J Hum Genet 71:877–892

Schwachman-Diamond Syndrome

Definition

Schwachman-Diamond Syndrome denotes an autosomal recessive disorder that is characterized by pancreatic exocrine insufficiency, hematological dysfunction, and skeletal abnormalities due to a defect in the SBDS gene on chromosome 7q11.

►[Methylation of Proteins](#)

Schwann Cells

Definition

Schwann cells (named after Theodor Schwann) are glial cells of the peripheral nervous system. Schwann

cells separate and insulate peripheral nerve cells, they enwrap the axons and form the myelin sheath. During development and regeneration of the nervous system, they play an important role in guiding the outgrowing axons. They are derived from neural crest cells.

- ▶ [Glial Cells and Myelination](#)
- ▶ [Neural Crest Cells and their Derivatives](#)
- ▶ [Neurofibromatosis Type 1 \(NF1\), Genetics](#)

Sclerotome

Definition

Sclerotome designates a mesenchymal cell mass located in the medial region of a somite, from which the axial skeleton derives.

- ▶ [Bone Disease and Skeletal Disorders, Genetics](#)
- ▶ [Somitogenesis](#)

Scoring Function

Definition

Scoring function refers to the empirical relationship to quantify the quality of ligand binding modes and to rank different ligands (binding to the same receptor) according to their affinity. It usually consists of the sum of several empirical terms describing properties that are important for binding, such as surface and chemical complementarity.

- ▶ [Molecular Docking](#)
- ▶ [Molecular Dynamics Simulation in Drug Design](#)

Scrambled Disulfide Species

Definition

Scrambled disulfide species is a fully oxidized protein with non-native disulfide bonds.

- ▶ [Protein Disulfide Bonds](#)

Scrapie

Definition

The prion disease scrapie is naturally occurring in sheep.

- ▶ [Prion Diseases](#)

SCTAT

Definition

SCTAT refers to ovarian sex cord tumors with annular tubules that sometimes produce oestrogen and trigger precocious puberty.

- ▶ [Peutz-Jeghers Syndrome](#)

SDAT

- ▶ [Senile Dementia/Alzheimer's Type](#)

SDH

Definition

Succinate dehydrogenase (SDH) is an enzyme of the respiratory chain whose protein subunits are all encoded by the nuclear DNA and then transported into the mitochondria.

- ▶ [Mitochondrial Myopathies](#)

SDS

Definition

SDS stands for sodium dodecyl sulfate (lauryl sulfate). It is a detergent that is used in polyacrylamide gel electrophoresis.

- ▶ [Recombinant Protein Expression in Bacteria](#)

SDS-PAGE

Definition

Proteins are separated by sodium dodecyl sulfate (SDS) polyacrylamide gel electrophoresis (PAGE) according to their molecular mass. The proteins are surrounded by positively charged SDS giving all the proteins the same charge. They are then separated, mainly according to their molecular weight (MW). The gel normally consists of a separating gel and a stacking gel, where the proteins are stacked with the aim that they start at the same position to run into the separation gel.

► [Two-Dimensional Gel Electrophoresis](#)

Seborrheic Keratosis

Definition

Seborrheic keratosis describes a condition of the skin that causes oily reddish and flaking skin.

► [Parkinson's Disease: Insights from Genetic Causes](#)

Second Site Suppression

Definition

Sometimes two proteins coded by distinct genes are required for an activity. If two proteins interact with one another, then any mutations in one of the proteins which alter amino-acids involved in the interaction will destroy the activity. However, a second mutation in the other protein can sometimes allow the interaction to take place and restore activity. Second site suppression defines a mutation in a second gene that restores the wild-type phenotype of an already mutant individual.

► [Protein Interaction Analysis: Suppressor Hunting](#)

Secondary Hyperparathyroidism

Definition

Secondary hyperparathyroidism is a hyperparathyroidism that arises as a result of disordered metabolism. In secondary hyperparathyroidism, the parathyroid glands themselves are normal, but a problem such as kidney failure makes the body resistant to the action of parathyroid hormone.

► [Hyper- and Hypoparathyroidism](#)

Secondary Hypertension

Definition

Secondary hypertension is high blood pressure that occurs because of other disorders such as kidney disorders (arteriolar nephrosclerosis), or can be caused by a response to medication (drug-induced hypertension).

► [Mendelian Forms of Human Hypertension and Mechanisms of Disease](#)

Secondary Structure

Definition

Secondary structure describes the folded, coiled, or twisted shape of a polypeptide that results from hydrogen bonding between parts of a molecule. There are two types of secondary structure: alpha helix and beta pleated sheet.

► [Protein Databases](#)

Second-Generation MDR Modulators

Definition

Second-generation MDR (multi drug resistance) modulators are molecules that hinder the drug transport out of a multi drug resistant cell by competitive binding, and thereby inhibit one or more ► [ABC](#) transporters. Intracellular drug concentration increases and thus chemosensitivity of the cell is particularly restored. They are used as a reversal strategy for cells and tumors possessing the multi drug resistance phenotype ► [MDR](#).

► [Multi-Drug Resistance](#)

Secreted Frizzled Related Proteins

Definition

Secreted Frizzled Related Proteins (FzB) are secreted proteins which bind Wnt and have similarity to the CRD of Frizzled.

► [Wnt/Beta-Catenin Signaling Pathway](#)

Secretion

Definition

Secretion is the process by which material is synthesised and then transported to the outside of a cell.

- Endocytosis
- Vesicular Traffic

Secretory Pathway

Definition

The secretory pathway comprises of a series of membrane-bounded compartments, including the endoplasmic reticulum (ER), vesicular-tubular clusters (VTCs), and the Golgi complex, through which a large number of newly synthesized proteins pass en route to the extracellular space or to other final, intracellular destinations.

- Vesicular Traffic

Secretory Vesicle

Definition

Secretory vesicle is a membrane bounded compartment in which secretory products accumulate.

- Cell Polarity
- Endocytosis
- Vesicular Traffic

Sedimentation Coefficient

Definition

Sedimentation coefficient denotes the ratio of the speed of molecular migration (► [sedimentation velocity](#)) in the centrifugal field, to the magnitude of the centrifugal field. This is a molecular constant dependent on the size and shape of the molecule, measured in units of Svedbergs ($1\text{ S} = 10^{-13}\text{ sec}$).

- Analytical Ultracentrifugation

Sedimentation Equilibrium

Definition

Sedimentation equilibrium describes the steady-state between sedimentation, diffusion, and chemical

reactions attained after sedimentation in the limit of long times.

- Analytical Ultracentrifugation

Sedimentation Velocity

Definition

Usually refers to the experiments on the hydrodynamic process of macromolecular sedimentation, by observing the evolution of the concentration distributions with time.

- Analytical Ultracentrifugation

Segment Duplication

- Gene Duplications

Segmentation

Definition

In the context of image processing, segmentation refers to an image processing method that is used to isolate a feature of interest from its context. It can be based on absolute differences in data signal and/or data gradient information.

- Electron Tomography

Segregation Analysis

Definition

Segregation analysis refers to a statistical method of inferring the genetic model underlying a particular phenotype (i.e. number of genes involved, allele frequencies, mode of inheritance, mutation rate etc.).

- Cleft Lip Palate
- Diabetes Mellitus, Genetics
- Genetic Epidemiology

SELDI

► Mass Spectrometry: SELDI

SELDI-ProteinChip Technology

► Mass Spectrometry: SELDI

SELDI-Tof

Definition

The surface-enhanced laser desorption/ionization time of flight (SELDI-Tof) is combining chromatography on a surface with an analyzer (Tof) for mass spectrometry. This technology is suitable for protein profiling; complex biological samples can be analysed and compared in a high throughput manner to detect differentially expressed proteins in the samples.

► Mass Spectrometry: SELDI

► Proteomics in Cancer

Selection, Selective Advantage

Definition

Selective advantage selection refers to a process that favours the survival of a given genotype in a population. Selection can be subdivided into natural selection, which operates in a natural ecosystem, and artificial selection, such as may occur in a laboratory or livestock breeding programme.

► Gene Duplications

Selection/Selectable Marker

Definition

Selection marker defines a fragment of DNA that can be built into a cloning vector or other construct, which

allows only those host cells carrying it to survive in a given medium. For example, selection markers are essential components of ES-cell based genome manipulation. Integration of foreign DNA into the ES-cell genome by homologous recombination normally occurs at very low frequencies. Therefore, those cells in which correct integration has occurred (positive selection), or in which DNA has integrated in an undesired non-homologous manner (negative selection), must be identified and selected for. Positive selectable markers, widely used for ES-cell based genome engineering and for bacterial hosts, are neomycin phosphotransferase neo, puromycin resistance gene puro and hygromycin resistance gene hygro. Negative selectable markers frequently used are herpes simplex, virus-1 thymidine kinase gene HSV-tk, and diphtheria toxin A dtA. For yeast, selection markers are mostly amino acid biosynthetic genes, which are selectable in auxotrophic hosts on defined media.

► Mouse Genomics

► Recombinant Protein Expression in Yeast

► Transgenic and Knockout Animals

► YAC and PAC Maps

Selectivity

Definition

The selectivity of chemical analysis describes the degree to which chemical detection can distinguish between different metabolites, which are present within the same sample.

► Metabolomics

Selenocysteine

Definition

Selenocysteine is formed from serine by enzymatic transformation of a particular ►tRNA charged with serine to form Sec-tRNA^{Sec} (Sec, selenocysteine). An important example of codon redefinition is the incorporation of selenocysteine at UGA codons, which normally function as termination codons. With the help of a specialized translation factor (SelB in bacteria), Sec-tRNA^{Sec} binds to the ribosome and recognizes UGA when a particular structural element of the mRNA, the selenocysteine insertion sequence (SECIS), which recruits SelB is present, either within the coding

sequence downstream of the UGA (bacteria) or further away in the 3′-untranslated region (3′-UTR) of the mRNA (eukaryotes).

► [Ribosomes](#)

► [Selenoproteins](#)

Selenoenzymes

Selenoproteins

Selenoproteins

DIETRICH BEHNE, ANTONIOS KYRIAKOPOULOS
Department of Molecular Trace Element Research in
the Life Sciences, Hahn Meitner Institute Berlin,
Berlin, Germany
behne@hmi.de

Synonyms

Selenoenzymes

Definition

Selenoproteins are defined as selenocysteine-containing proteins into which ► [selenocysteine](#) is incorporated by means of a specific genetic code. Most of the selenoproteins are redox-active enzymes, with the selenocysteine responsible for their catalytic functions.

Characteristics

Biological Significance of Selenocysteine

Selenocysteine is an amino acid analogous to cysteine with selenium in place of sulfur. The trace element selenium is also found in proteins in the form of selenomethionine that is incorporated non-specifically into methionine-containing proteins by accidental replacement of methionine.

Most of the chemical and physical properties of the analogous compounds of selenium and sulfur are similar. A difference that is significant for the importance of selenium in biological systems is, however, the fact that the selenohydryl groups of selenocysteine-containing proteins are anionic at physiological pH, while the sulfhydryl groups in cysteine-containing proteins are mainly protonated under these conditions. The experimental replacement of selenocysteine in the active center of a selenoenzyme with

cysteine therefore results in a drastic decrease in its catalytic activity.

Formation of the Selenoproteins

With the concentration of sulfur in living matter exceeding that of selenium by several orders of magnitude and with the similarity in the properties of the two elements, highly selective mechanisms are necessary to prevent substitution of selenocysteine by cysteine during protein formation. One of the unique features in the incorporation of selenocysteine is the use of the UGA codon that normally acts as a termination codon. Another is the biosynthesis of selenocysteine, which takes place on its ► [tRNA](#).

Selenocysteine insertion by means of the UGA codon requires a specific mRNA stem loop structure as well as four unique gene products, SelA, SelB, SelC and SelD. The UGA codon is recognized by the anticodon of a specific tRNA^{(Ser)Sec} (SelC) which is first loaded with serine. Selenocysteine synthase (SelA) then catalyzes the replacement of the side-chain oxygen of serine by selenium and thus the conversion of seryl- tRNA^{(Ser)Sec} into selenocysteyl-tRNA^{(Ser)Sec}. The selenium donor molecule for this reaction is selenophosphate, which is produced from selenide by means of selenophosphate synthetase (SelD). Selenocysteyl-tRNA^{(Ser)Sec} and a unique elongation factor SelB, homologous to the elongation factor EF-Tu for all other amino acid-tRNAs, form a complex which binds to the stem loop structure of the selenoprotein mRNA and mediates the incorporation of selenocysteine into the protein at the ribosome.

Eukaryotes differ from bacteria with regard to the position of the mRNA stem loop structure, named SECIS (selenocysteine insertion sequence) element, which is not situated immediately downstream of the UGA codon as in bacteria but in the 3′ untranslated region.

Hierarchy in Selenoprotein Expression

Another unique characteristic is the hierarchy in the selenium distribution among the selenoproteins in periods of insufficient selenium intake. With the formation of the selenoproteins depending on an adequate supply of the element, their concentrations are diminished in dietary selenium deficiency, but they do not decrease to the same degree in the different body compartments. During selenium depletion, regulation mechanisms strive to maintain the selenium levels in certain tissues, such as the central nervous system, the endocrine and the reproductive organs, and within these tissues the levels of certain selenoproteins. The liver, the skeletal muscle, the red blood cells and the blood plasma are at the bottom of this ranking order and show a relatively rapid loss, but here too, the concentrations of certain selenoproteins are more

affected than those of others. The hierarchy between the different tissues is achieved by the preferential supply of the element to the most important target organs. The intracellular hierarchy in the expression of selenoproteins may be regulated mainly by differences in the stability of their mRNAs in selenium deficiency. Due to these regulation mechanisms, it was virtually impossible to produce a severe decrease in the concentrations of the preferentially supplied selenoproteins in high priority organs such as the brain and the pituitary, even during extreme experimental depletion of laboratory animals.

The Individual Selenoproteins

Selenoprotein research is far from complete. Nearly 50 years after the discovery of the essentiality of selenium, we still do not know all the selenoproteins present in the human body, far less their biochemical functions and their biological roles. Earlier work on the identification of the individual selenoproteins was made difficult by the fact that, unlike many other protein families, their common characteristic is not found in their functions or their sites of action but in the selenocysteine residue and the genetic code responsible for its insertion. Accordingly, the first studies were carried out by means of protein purification and analysis of selenium and selenocysteine. The first selenoenzyme to be identified was glutathione peroxidase, in 1973. This was followed by the detection of a second enzymatic function in the form of the type I iodothyronine deiodinase in 1990. Several other selenoproteins have been identified in this way, including the thioredoxin reductases, another important enzyme family. By labeling rats with radioactive ^{75}Se and determining the tracer distribution in the proteins separated by gel electrophoretic methods, it was shown that selenium is present in more than thirty specific mammalian proteins or protein subunits. This finding was the basis for the detection of further selenocysteine-containing selenoproteins. The decoding of the mammalian genomes then offered the possibility of identifying the selenoprotein genes by computational screening methods and in this way 25 different human selenoprotein genes were identified. According to our current stage of knowledge, the selenoproteins can be divided into two groups, those with known biochemical functions and those with functions still unknown.

Selenoproteins with Known Functions

All the selenoproteins in this group are enzymes, with the selenocysteine residue responsible for their catalytic functions.

Glutathione Peroxidases

The glutathione peroxidases are either monomers with one selenocysteine residue or homotetramers of

selenocysteine containing subunits with molecular masses around 22 kD. They catalyze the reduction of hydrogen peroxide and various organic peroxides and thus protect cells from oxidative damage. Glutathione normally serves as the electron donor but there are cases where other thiols are oxidized in order to fulfill a specific biological role.

Cytosolic Glutathione Peroxidase (GPx1, cGPx)

is a tetramer, cytosolic and ubiquitous, with highest levels in liver and erythrocytes.

Gastrointestinal Glutathione Peroxidase (GPx2, GI-GPx)

is a tetramer, cytosolic and found only in the gastrointestinal tract and the liver.

Plasma Glutathione Peroxidase (GPx3, pGPx)

is a tetramer, extracellular glycoprotein, expressed in various tissues with the kidney as the main production site, which can also use glutaredoxin and thioredoxin as electron donors.

Human Glutathione Peroxidase 6 (GPx6)

is a tetramer, identified by computational screening of the human genome. GPx6 mRNA is only found in embryos and in the olfactory epithelium.

Phospholipid Hydroperoxide Glutathione Peroxidase (GPx4, PHGPx)

is a 20 kD monomer, present in tissues in both cytosolic and membrane-associated forms. The mitochondrial form is due to a precursor with a specific N-terminal leader sequence. The enzyme can reduce phospholipid and cholesterol hydroperoxides and may use other thiols as electron donors. It has a structural function as an inactive matrix component of the mitochondrial capsule of the sperm flagellum.

Sperm Nuclei Glutathione Peroxidase (snGPx)

is a 34 kD monomer, the only selenoprotein in sperm nuclei, which differs from GPx4 in its N-terminal sequence by having a nuclear localization signal encoded for by an alternative exon in the first intron of the gene. snGPx acts as a protamine thiol peroxidase responsible for disulfide cross-linking and stabilization of the condensed nuclei.

Iodothyronine Deiodinases

The iodothyronine deiodinases catalyze the monodeiodination of the iodothyronines at the 5'-position of the phenolic ring and/or at the 5-position of the tyrosyl ring. In this way they are responsible for the activation and inactivation of the thyroid hormones that regulate various metabolic processes and are indispensable for the normal development of the fetal brain.

Type 1 Deiodinase (D1)

is a membrane-associated homodimer of 27 kD subunits located mainly in the thyroid, liver, kidney and pituitary, which catalyzes 5'- and 5-monodeiodination. Its main functions are production of T_3 from T_4 and inactivation of T_4 to reverse T_3 and of T_3 to T_2 .

Type 2 Deiodinase (D2)

is a membrane-associated 31 kD monomer expressed mainly in the brain, brown adipose tissue, pituitary and placenta. It catalyzes 5'-monodeiodination, its main function being tissue-specific intracellular production of T_3 from plasma T_4 .

Type 3 Deiodinase (D3)

is a membrane-associated 32 kD monomer, predominantly found in the central nervous system, placenta and skin, which catalyzes 5-deiodination, regulates the supply of T_4 and T_3 from mother to fetus and protects the developing brain from excessive amounts of T_3 by inactivation of T_4 to reverse T_3 and of T_3 to T_2 .

Thioredoxin Reductases

Thioredoxin reductases, present in various tissues, were named for their ability to catalyze the reduction of oxidized thioredoxin. Reduced thioredoxin (Trx) is an electron donor for various redox-dependent systems (e.g. ribonucleotide reductase essential for DNA synthesis) and for the redox regulation of transcription factors and is involved in other cellular processes such as secretion, regulation of cell growth and protection against ►oxidative stress and ►apoptosis. Mammalian thioredoxin reductases are able to use other substrates including hydroperoxides and various enzymes and proteins. Three human thioredoxin reductase isoenzymes encoded for by different genes have been identified:

Thioredoxin Reductase 1 (TrxR1, TR1)

is a cytosolic homodimer of 56 kD subunits, which constitutes, together with Trx1, the cytosolic thioredoxin system.

Thioredoxin Reductase 2 (TrxR2, TR3)

is a homodimer of 56 kD subunits with an N-terminal mitochondrial leader sequence. Together with Trx2, it forms a mitochondrial thioredoxin system.

Thioredoxin-Glutathione Reductase (TGR, TR2)

is a microsomal homodimer of 65 kD subunits, mainly present in the testis. It differs from TrxR1 and TrxR2 by an additional N-terminal glutaredoxin domain and catalyzes the reduction of both oxidized thioredoxin and oxidized glutathione.

Further Thioredoxin Reductases

Multiple mRNA forms of TrxR1 and TrxR2 suggest the existence of further thioredoxin reductase species, which may differ with regard to their distribution among tissues and subcellular compartments and may have specific biological roles.

Other Selenoenzymes**Selenophosphate Synthetase 2 (SPS2)**

is a cytosolic 50 kD enzyme, present in various tissues, which catalyzes the production of selenophosphate and, as a selenoprotein involved in selenoprotein synthesis, is of special interest with regard to the regulation of selenium metabolism.

Methionine-R-sulfoxide Reductase (SelR, SelX, MsrB1)

is a cytosolic (and nuclear in some tissues) zinc-containing 12.6 kD enzyme that catalyzes the reduction of methionine-R-sulfoxide in a diastereomeric mixture of this sulfoxide and methionine-S-sulfoxide, present after oxidation of methionine by reactive oxygen species.

Selenoproteins with Functions Still Unknown**Selenoprotein P (SelP)**

is an extracellular heparin-binding 43 kD glycoprotein with 10 selenocysteine residues per molecule, produced in most tissues but with the liver as the main production site. It is secreted into plasma or interstitial fluids. Several studies suggest that it may act as a selenium transport protein but it may also have antioxidant properties.

Selenoprotein W (SelW)

is a cytosolic 10 kD protein found in various tissues but enriched in skeletal muscle, heart, brain, testis and spleen. Some studies suggest a redox function.

15 kD Selenoprotein (Sel15, Sep15)

is a selenocysteine-containing 15 kD subunit of a cytosolic 240 kD protein, found in various tissues but especially enriched in the prostatic epithelium. Some studies suggest anticarcinogenic properties.

18 kD Selenoprotein (Sel18)

is present in the mitochondrial membranes and is one of the selenoproteins most preferentially supplied with selenium.

Selenoproteins Identified in Silico

Computational screening of the human genome led to the detection of several selenoproteins with functions not yet known. They include (approximate molecular masses calculated from their cDNA sequences) SelH

(14 kD), SelI (48 kD), SelK (10.5 kD), SelM (16 kD), SelN (60 kD), SelO (73.5 kD), SelS (20.8 kD), SelT (20 kD) and SelV (38 kD). With the exception of SelV, which is expressed only in the testis, they are present in various tissues. SelK is a plasma protein, SelM a perinuclear protein and SelS a plasma membrane protein related to glucose metabolism. SelN is an endoplasmic reticulum glycoprotein that is expressed mainly in fetal tissues and developing cells and may be involved in early muscle formation.

⁷⁵Se-labeling studies in rats showed the existence of further forms of selenoproteins, which may be due to alternative splicing or posttranslational modification, but there is also the possibility that there may be further selenoprotein genes not yet detected by the computational screening methods.

Clinical Relevance

A vast number of publications deal with possible relationships between changes in selenium and selenoprotein status and diseases, but in many cases evidence for a cause-effect relationship has not yet been found. Unlike other proteins, selenoprotein expression depends not only on genetic factors but also on the dietary selenium supply and most of the information on the effects of diminished selenoprotein expression has so far been obtained in controlled selenium deficiency studies. Animal tissues affected in selenium depletion include cardiac muscle, erythrocytes, eye, kidney, liver, pancreas, skeletal muscle, skin, spermatozoa and testis. In humans selenium deficiency was found to be a pathogenic factor in cardiomyopathy, muscular disorders and thyroid hormone related diseases. Protective effects of selenium administration have been observed in carcinogenesis, several viral infections such as hepatitis and HIV and epileptic seizures.

Due to the hierarchy in selenium distribution, selenium depletion does not affect the most preferentially supplied tissues and selenoproteins to such a degree that lesions appear. Genetic changes, however, may affect any selenoprotein in any tissue.

Single nucleotide polymorphisms have been found in some selenoproteins. A variant allele for GPx1 was less responsive to the stimulation of its enzymatic activity by selenium supplementation and was associated with an increased risk of lung and breast cancer. Similarly two alleles of Sel15 were identified which responded differently to selenium supplementation and may be involved in cancer development or risk. Genetic alterations of GPx2 and SelP have been investigated in relation to colon cancer. Mutations of the SelN gene are involved in rigid spine muscular dystrophy, a neuromuscular disorder. It is very likely that numerous further variants will be found. With SelP, which is involved in selenium transport, genetic changes may lead to an insufficient supply to the high priority organs such as the central

nervous system. Alterations in GPx4 and snGPx may lead to infertile spermatozoa. As the selenoproteins have a wide spectrum of enzymatic and regulatory functions, polymorphisms may affect various metabolic processes and may thus play a role in a multitude of diseases.

References

1. Behne D, Kyriakopoulos A (2001) Mammalian selenium-containing proteins. *Annu Rev Nutr* 21:453–473
2. Hatfield DL, Gladyshev VN (2002) How selenium has altered our understanding of the genetic code. *Mol Cell Biol* 22:3365–3376
3. Kryukov GV, Castellano S, Novoselov SV et al (2003) Characterization of mammalian selenoproteomes. *Science* 300:1439–1443
4. Rayman MP (2000) The importance of selenium to human health. *Lancet* 356:233–241
5. Taylor PR, Parnes HL, Lippman SM (2004) Science peels the onion of selenium effects on prostate carcinogenesis (2004) *J Natl Cancer Inst* 96:645–647

Self-Assembled Monolayer

Definition

Self-assembled monolayer is a monolayer of hydrophobic molecules with a functional head group, for example (-SH), which shows a strong and specific interaction with a surface of noble metals like gold, silver and platinum. Interaction leads to an array of ordered and oriented molecules.

► [Surface Plasmon Resonance](#)

Self-Renewal

Definition

Self-renewal is a key characteristic of stem cells. When stem cells divide they produce, by definition, at least one progeny that is identical to themselves.

► [Neural Stem Cells](#)

► [Stem Cells: an Overview](#)

Semidominant Allele

Definition

Semidominant allele is one of a series of terms applied to the phenotypic effect of a particular allele in

reference to another allele (usually the standard wild-type allele) with respect to a given trait. An allele “A” is said to be semidominant with respect to the allele “a” if the A/A homozygote has a mutant phenotype, the A/a heterozygote represents a less severe phenotype, while the a/a homozygote is a wild-type.

► [Large-Scale ENU Mutagenesis in Mice](#)

Senescence

Definition

Senescence describes a stage in which the cell, after many cell generations, withdraws from the cell cycle and ceases dividing. It is a safeguard against uncontrolled cell proliferation of abnormal cells, protecting us from cancer.

► [Cellular Senescence](#)
 ► [Tumor Suppressor Genes](#)

Senile Dementia/Alzheimers Type

► [Alzheimer's Disease](#)

Senile Plaques

Definition

Senile Plaques (SP) – synonym amyloid plaque – are one of the histopathologic hallmarks of Alzheimer disease (AD). SPs consist of a dense core of compact

► [β-amyloid](#) (Aβ) fibrils surrounded by porous Aβ and degenerating neurites.

► [Alzheimer's Disease](#)

Sensitivity

Definition

Sensitivity is the likelihood of a positive test result in patients with disease. It measures how well the test detects the disease. It is the complement of the false-negative rate (ie, the false-negative rate plus the sensitivity = 100%).

► [Mass Spectrometry: SELDI](#)

► [Metabolomics](#)

► [Protein Microarrays as a Tool for Protein Profiling and Functional Proteomics](#)

Sensory Ganglia

Definition

Sensory ganglia comprise of a conglomerate of neurons and glial cells that transmit sensory information (temperature, nociception, touch, pressure, proprioception, etc) from the periphery to the central nervous system.

► [Neural Crest Cells and their Derivatives](#)

Septum Pellucidum

Definition

Septum Pellucidum is a two-layered thin wall in the brain consisting of glial-like elements. This midline brain structure is located inferior to the corpus callosum, and separates the right and the left lateral ventricle (► [Anterior Horn Cells](#)). The septum pellucidum probably has no special functional importance.

► [Brain](#)

► [Hypothalamic and Pituitary Diseases Genetics](#)

Sequence Alignment

Definition

Sequence alignment refers to the arrangement of two or more amino acid or base sequences from an organism or organisms, to align sequence areas sharing common properties.

► [Protein Databases](#)

Sequence Annotation in Evolution

STEFFEN HENNIG

RZPD - Deutsches Ressourcenzentrum für Genomforschung, Berlin, Germany
 hennig@rzpd.de

Definition

The history of life as far as we know at present spans the enormous time period of around 3.5 billion years. The ‘tree of life’ concept originally starting with Darwin's pioneering ideas was for more than a century mainly based on findings from the fossil record and morphological comparison of living species. With the availability of the DNA sequences of several complete genomes and the sequences of genes and their products (proteins), we now enter an era where the ‘tree of life’, which shows the main phyla as they arose from their respective ancestors, can be inspected directly by genome comparison between species.

Practical applications of sequence annotation in the context of evolution are, e.g. the development of the major histocompatibility complex (MHC) between primates (1) or over long time scales (2) or studies on the origin of vertebrates (3).

Characteristics

The main difference between ‘classical’ sequence annotation and annotation with an evolutionary focus is the proper definition of the genes and proteins related between species, i.e. the identification of ►[orthologs](#) and ►[paralogs](#) is essential. Evolutionary sequence annotation tries to find the root through the tree of species in order to gain insights either into the general ►[phylogeny](#) (e.g. how and when vertebrates arose from the invertebrate lineage) or focuses on specific gene families conserved between species. Finally it has to be noted that with the abundance of genome and protein data generated by the large-scale genome projects, we have, for the first time, the chance to investigate the mechanisms of evolution from first principles – by studying the dynamics of small changes in the genome and how they are transferred into functional genes and proteins.

A Survey of the Available Data

Table 1 lists the milestones in the public sequencing efforts carried out by the large sequencing centres and by various smaller groups. In 1995 the genome sequence of the bacterium *Haemophilus influenzae* comprising about 1.8 Mb was the first example of a completely sequenced genome. Since then the world-wide sequencing effort has scaled up a lot, the most spectacular results being the almost completed genomes of *Homo sapiens*, *Mus musculus* and several invertebrates (*D. melanogaster*, *C. elegans*, *S. cerevisiae*) and plants (*A. thaliana*). Besides the genomic sequencing projects, several groups have focused on the generation of full-length mRNA sequences, which allow determination of the corresponding protein sequences. For instance in SwissProt (by beginning of 2006) around 200000 protein sequences from over 9500 species are stored and functionally characterized.

Sequence Annotation in Evolution. Table 1 History of genome sequencing starting with the first completely sequenced genome of *H.influenzae*. Data are rounded and mainly reflect the correct order of magnitude. The number of genes and proteins is taken as equal for simplicity. SwissProt is one of the largest public collections of protein data for all species

organism	size [Mb]	year	no genes/ proteins
<i>Haemophilus influenzae</i>	1.8	1995	1700
<i>Escherichia coli</i>	4.6	1997	4400
<i>Saccharomyces cerevisiae</i>	12	1997	6200
<i>Caenorhabditis elegans</i>	102	1998	20000
<i>Arabidopsis thaliana</i>	114	2000	5300
<i>Drosophila melanogaster</i>	180	2000	13500
<i>Homo sapiens</i>	2860	2001	33000
<i>Mus musculus</i>	3000	2002	33000

Protein data: SwissProt (200000 proteins from 9500 species)

Proteins are the main subject of evolutionary studies based on sequence analysis, since even over distances of hundreds of million years, related proteins in different species can be identified by sequence comparison, while the corresponding mRNA sequences will have diverged too much for proper analysis. On the other hand, for species close in evolution (e.g. primates or rodents) analysis on the DNA level provides direct insights into the molecular mechanisms of evolution.

Methods of Analysis I: Evolution at the DNA Level

It is a widely accepted hypothesis that all living species have one common ancestor, who emerged around 3.8 billion years ago, although it is still unclear if LUCA (the last universal common ancestor) belonged to archaea or bacteria or represented another yet unknown phylum. The main argument for the existence of LUCA is the surprising number of proteins conserved from ancient bacteria up to higher vertebrates and the universal genetic code. Many details of the tree of life (Table 2) can be deciphered by comparative analysis of genome and protein sequences, which also gives important insights into the dynamics of evolution on the molecular level.

There are several processes which cause changes in the genome and therefore might be responsible for evolutionary changes: a) single base substitutions or

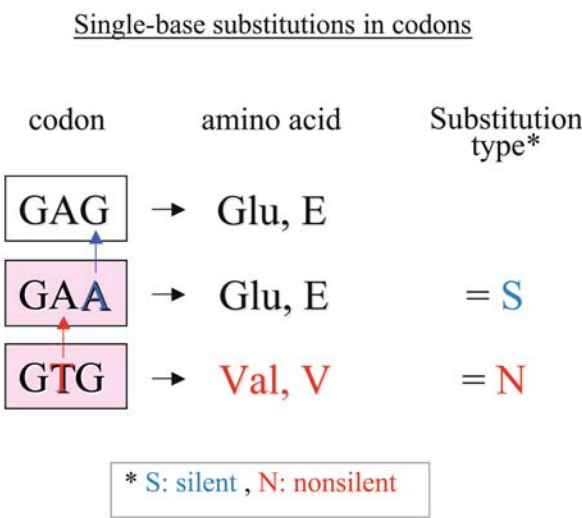
Sequence Annotation in Evolution. Table 2 A collection of very useful on-line resources and tools in the field of evolutionary biology and sequence analysis

subject	source
tree of life	▶http://tolweb.org/tree/phylogeny.html very comprehensive overview on all details of the tree of species
Hovergen	▶http://pbil.univ-lyon1.fr/databases/hovergen.html database of homologous vertebrate genes, includes tools for phylogenetic analysis
PAML package	▶http://abacus.gene.ucl.ac.uk/software/paml.html software package for phylogenetic analysis.
COG database	▶http://www.ncbi.nlm.nih.gov/COG/ Clusters of groups of genes orthologous between species, maintained at NCBI
phylogeny tools	▶http://evolution.genetics.washington.edu/phylip/software.html overview on public tools for phylogenetic analysis, maintained by J.Felsenstein, the author of PHYLIP
CLUSTAL	▶http://www-igbmc.u-strasbg.fr/BioInfo/ClustalX/Top.html very popular tool for multiple sequence alignment and construction of phylogenetic trees

insertions/deletions, b) chromosomal rearrangements or duplications and c) horizontal gene transfer. The latter is known to happen frequently in microorganisms, where transfer of genes from one species to another has been reported in many publications. There are strong indications that especially in the early stages of life, horizontal gene transfer was an essential mechanism for the creation of new species, although the underlying principles are not yet understood. a) and b) are fundamental processes probably acting in all higher eukaryotes and are regarded by many scientists as the main driving forces of evolution.

The genetic code is not unique. There are 61 triplets of nucleotides (codons) coding for altogether not more than 20 amino acids. As a consequence there are base substitutions, which do not affect the resulting amino acid (▶**silent substitution**), mostly changes at the third codon position, while changes at the first or second position in almost all cases change the resulting amino acid (▶**non-silent substitution**). Figure 1 gives an example of a codon where a single base is substituted, which in one case doesn't change the resulting amino acid (glutamic acid) while in the other case it does (valine). In fact, the non-silent mutation shown in Fig. 1 is a classical example, since the same kind of single-base mutation affecting the β -globin chain of hemoglobin causes sickle-cell anemia.

The number of silent (S) or non-silent (N) substitutions observed between a pair of genes in two different species, or in one species if we consider duplicated genes, is an indicator of their relative age and also



Sequence Annotation in Evolution. Figure 1 Schematic presentation of single base substitution types in codons resulting in an unchanged or changed amino acid.

provides information about their evolutionary status. If we define $\omega = dN/dS$ as a cumulative ratio of non-silent/silent substitution rates per codon site, then $\omega > 1$ means that the corresponding protein is under strong evolutionary pressure (i.e. high rate of functional relevant amino acid changes), $\omega < 1$ indicates strong evolutionary constraints for amino acid changes (▶**purifying selection**) and $\omega \approx 1$ means a fully relaxed

status where both kinds of substitutions occur at the same frequency. Conery and Lynch (4) have studied the temporal behavior of ω for 9 different species with almost completely sequenced genomes. They identified duplicated genes within each species by sequence comparison of all available open reading frames, which resulted in a very rich set of duplicate pairs. The numbers of silent and non-silent substitutions per duplicate were counted. From this exhaustive study, they deduced that duplicated genes undergo different phases of selective pressure. For young duplicates, where the absolute number of silent substitutions per site is small ($S < 0.05$), they found in all species ω in the range of relaxed or strong evolutionary pressure ($\omega \geq 1$), while for older duplicates, non-silent substitutions are generally reduced ($\omega < 1$), indicating strong selective constraints.

On the assumption of the inherently stochastic nature of molecular evolution, the rates of silent and non-silent base substitutions also allow estimates of the relative age of the respective gene pairs. One of the most general mathematical models up to now has been developed by Yang and Nielsen (5). Their algorithm, which is implemented in the publicly available PAML package (Table 2), takes into account that for gene pairs in species separated by long evolutionary times, multiple steps of substitution at the same site might have occurred and therefore a proper model has to weight all possible paths of substitution steps leading from codon A to codon B. The different rates between non-silent substitutions changing the amino acids either from purine to purine (►transition) or from purine to pyrimidine (►transversion) are also explicitly included in Yang and Nielsen's algorithm. It takes as input two properly aligned nucleotide sequences and gives as output the estimated values for rates dS, dN, ω and the relative time distance t .

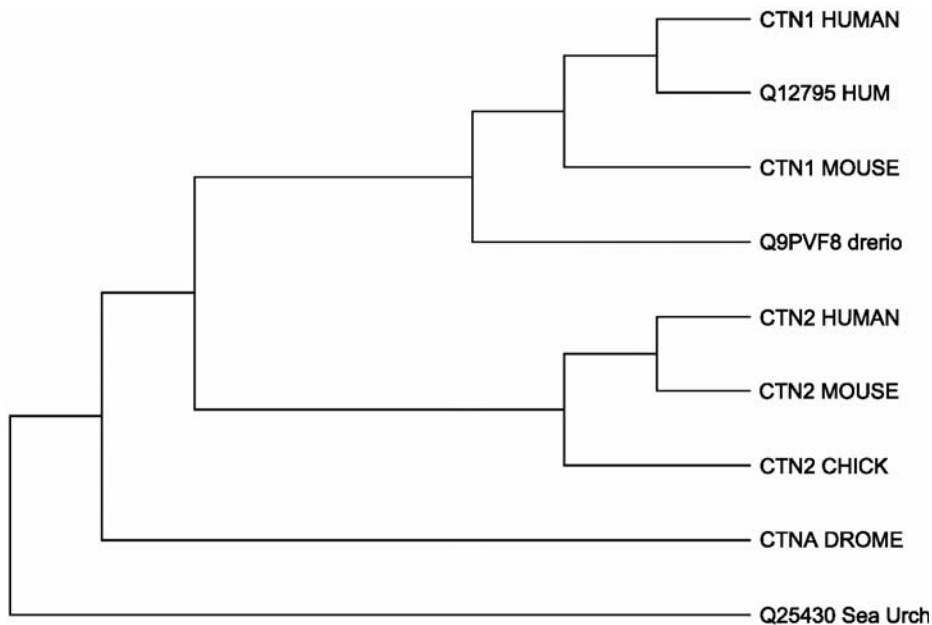
Methods of Analysis II: Evolution at the Protein Level

Typically, for species separated by hundreds of millions of years of evolution the divergence of DNA sequences is high. Analysis of mutations based on the methods described above is very difficult and has to deal with saturation effects. Protein sequences show a much higher degree of conservation during evolution, so that evolutionary studies of very distant species are normally performed by protein comparison. At this point it is helpful to introduce the concepts of orthology and paralogy. If two proteins from different species arose from one common ancestor protein they are described as orthologs, if a protein was the result of gene duplication within a single species it is described as a paralog. In all sequence based studies of evolution between species, it is very important to focus on the orthologous proteins, since we know e.g. from the results of Conery and Lynch (4) that the mutation rates

of duplicated genes within a species (paralogs) might be different from those of orthologs. An easy way to find the orthologous proteins for different species was proposed by Tatusov et al. (6). Given the complete protein sets for two species, the orthologs are found by an all-against-all sequence comparison, i.e. all proteins of species A are compared against all of species B and *vice versa*, followed by a ranking procedure, which finds those pairs of proteins in A and B with highest scoring hits with each other. The resulting best-hit pairs are the most likely orthologs between species A and B. Since extension of the algorithm to multiple species comparison is straightforward it has been widely used for the construction of groups of orthologous proteins for various species (3), e.g. in the COG database (Table 2). It should be noted that this algorithm is valid only under the condition that the compared protein sets are completely known, which is nearly fulfilled for many bacterial genomes and might also be a good assumption for the completely sequenced higher eukaryotes like yeast, *C.elegans*, *Drosophila*, *A.thaliana*, mouse and *H.sapiens*. For species with only few known proteins the method of Tatusov is not applicable.

When studying families of proteins with more than two members, the first step in building the corresponding phylogenetic tree, which is the representation of their relationship, is usually a multiple sequence alignment (MSA). There are various methods and tools available to construct an MSA, the most famous one may be the CLUSTAL program, which is freely distributed by the authors (Table 2). A comprehensive introduction to the field can be found in Durbin et al. (7), where all important concepts and algorithms for phylogenetic calculations are also treated in detail. The MSA is the input for the algorithm that builds the phylogenetic tree. CLUSTAL, for example, has an integrated neighbor-joining algorithm, which performs the tree construction directly from the MSA.

There are numerous methods to calculate phylogenetic trees, e.g. ►neighbor-joining (NJ) and ►parsimony. For a comprehensive overview the book of Durbin et al. (7) and the web page of J. Felsenstein (Table 2) are highly recommended. Figure 2 gives an example of a tree constructed by NJ for the α -catenin family of proteins, which are found in invertebrates and vertebrates. It was drawn with the additional constraint that the sea urchin protein (Q25430) should be the outgroup, since sea urchin is known (from the fossil record) to be the oldest species in the set. The tree nicely separates the two subgroups $\alpha 1$ and $\alpha 2$ of catenins in mammals (human, mouse), while in invertebrates (sea urchin, *Drosophila*) only one group is found. $\alpha 1$ catenin in humans is located on chromosome 5 (q31.2) and is important for cell adhesion. Mutations in $\alpha 1$ catenin are involved in the process of cancer invasion and metastasis. The human



Sequence Annotation in Evolution. Figure 2 Phylogenetic tree of the α -catenin class of proteins. CTN1/2 represent the α 1/ α 2 subgroups of the protein family in human and mouse. The tree was constructed by the Neighbour-Joining algorithm. The sea urchin protein was manually selected as the most likely outgroup.

α 2 *catenin* is found at a different locus (chrom.2p12) and probably has similar function.

A standard method for testing the quality of the tree is **▶bootstrapping**: In the final MSA single columns of the alignment are randomly shuffled and the tree-building algorithm is started again. This is repeated many times (1000 \times) and the most likely solution is just the tree that was found in most of the cases. Bootstrapping therefore gives a statistical measure for the resulting tree (score), which is very useful for automated quality tests.

The length of the edges, i.e. the lines connecting the nodes in the tree with the end-nodes representing the proteins, can be regarded as evolutionary distance on an abstract time scale. In the **▶molecular clock** hypothesis one assumes that amino acid substitution on an average happened at a constant rate over all branches of the tree. If the molecular clock is valid, knowledge about the true time distance t between two nodes in the tree is sufficient to convert all distances to the respective numbers of years. Molecular clock approximations have been used extensively in the last few years for estimates about the age of specific members of the human MHC family of genes (2) or for dating of the invertebrate-vertebrate split (3).

A Large Scale Study on Ancient Genome Duplications

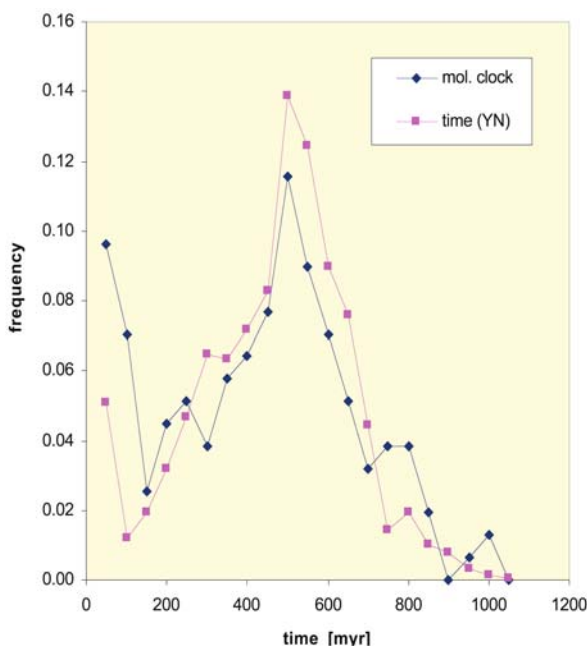
The 2R hypothesis originally proposed by Ohno in 1970 postulates two rounds of whole genome duplications at

the origin of vertebrates to explain the increased complexity of vertebrates with respect to their invertebrate relatives. Ohno's work was mainly ignored for more than 20 years by the scientific community, but since the availability of several complete invertebrate and mammalian genomes has gained more and more attention. Several publications based on selected small gene families addressed the 2R hypothesis, but the final results gave no consistent picture. Recently, in an elaborate study I was involved in, several thousand protein families (3) from 4 invertebrate species (yeast, *C. elegans*, *Drosophila*, *Amphioxus*) were compared against the complete human protein set. The analysis steps can be summarized as follows:

- Clusters of proteins orthologous between all invertebrate species were constructed by the method of Tatusov and Koonin (see above). Only those proteins were taken which existed in not more than one copy in each invertebrate species.
- The resulting ≈ 3000 orthologs were compared against all human and mouse proteins, which were attached to the orthologous sets and classified into orthologs or paralogs. In total around 10,000 human proteins could be clearly attached to the invertebrate sets. Extra copies of human proteins (paralogs) were counted per group and the final statistics showed a tremendous increase of extra copies in mammals. Compared to the invertebrate orthologs (1 copy per definition), there are 3–4 copies in human.

- c) For all proteins in the respective groups, phylogenetic trees were constructed and the time distance between the human duplicates measured using the molecular clock model.
- d) The human gene sequences corresponding to the duplicated proteins were analyzed by the method of Yang (see above) for occurrence of silent and non-silent substitutions, which gave an additional measure of the time point of duplication.

In Fig. 3 the results of two different methods for time measurement are shown. As mentioned above the data contributing to the plotted bars were human duplicated proteins (genes), which were proven to be orthologs of invertebrate proteins existing only in one copy. The correspondence of both methods is remarkable. Both gave a time distribution around 500–550 million years ago (mya), which coincides quite well with estimates about the split between vertebrates and invertebrates. Moreover, by analyzing the positions of the human duplicates along the chromosomes, it could be shown that duplicated genes are found with high significance



Sequence Annotation in Evolution.

Figure 3 Distribution of duplication times for human genes, which have a known orthology to 3 invertebrate species. The methods used were molecular clock (blue boxes) and Yang and Nielsen's (YN) method (red boxes) as described in the text. Both methods give a relatively sharp maximum around 500–550 million years (myr), where a genome duplication might have happened. The second peak around 50 myr can be attributed to recent duplications.

in specific chromosomal segments, which can be interpreted as traces of an ancient genome wide duplication event.

Thus, by a combined approach using all the techniques and models described above, it was possible to identify an ancient duplication event centered around 550 mya, which might explain the increased genome complexity in higher vertebrates. In parts the hypothesis of genome duplication seems to be justified, with the restriction that only evidence for one round of genome duplication could be found.

Clinical Relevance

While the example above demonstrates mainly the potential of evolutionary sequence analysis for answering fundamental questions about the tree of life, there are various applications of the same methods with direct medical relevance. Evolutionary analysis of specific gene families like the MHC class contributed significantly to a deeper understanding of the basic mechanisms of the immune system. Comparative analysis of e.g. the MHC class II region in human, rhesus macaque and rodents showed that several genes in that region are primate specific (1) and seem to be under purifying selection as indicated by analysis of substitution rates at silent and non-silent codon sites. Furthermore, a deep phylogenetic analysis of MHC class I genes in vertebrates and their orthologs in invertebrates supports the hypothesis that many of the primate MHC genes are very young and subject to frequent changes in a birth-and-death like process, where single genes are duplicated, modified and frequently deleted again (2).

References

1. Sudbrak R, Reinhardt R, Hennig S et al (2003) Comparative and evolutionary analysis of the rhesus macaque extended MHC class II region. *Immunogenetics* 54:699–704
2. Piontkivska H, Nei M (2003) Birth-and-Death Evolution in Primate MHC Class I Genes: Divergence Time Estimates. *Mol Biol Evol* 20:601–609
3. Panopoulou G, Hennig S, Groth D et al (2003) New evidence for genome-wide duplications at the origin of vertebrates using an amphioxus gene set and completed animal genomes. *Genome Res* 13:1056–1066
4. Lynch M, Conery JS (2000) The evolutionary fate and consequences of duplicate genes. *Science* 290:1151–1155
5. Yang Z, Nielsen R (2000) Estimating synonymous and nonsynonymous substitution rates under realistic evolutionary models. *Mol Biol Evol* 17:32–43
6. Tatusov RL, Koonin EV, Lipman DJ (1997) A genomic perspective on protein families. *Science* 278:631–637
7. Durbin R, Eddy S, Krogh A et al (1999) *Biological Sequence Analysis: Probabilistic Models of Proteins and Nucleic Acids*. Cambridge University Press, Cambridge

Sequence Comparison

MARTIN VINGRON

Max Planck Institute of Molecular Genetics, Berlin,
Germany
vingron@molgen.mpg.de

Definition

DNA and proteins are usually denoted as strings of letters representing the nucleotides or amino acids, respectively. Due to the common evolutionary origin of sets of genes these strings may display a marked similarity. Sequence comparison methods are employed for the detection of this similarity and for the identification of similar sequences in large [sequence databases](#).

Description

Sequence Alignment and Dot-Plots

Similarity among two or more sequences is displayed in the form of an alignment, which is an arrangement of the sequences on top of each other such that matching or conserved residues are aligned with each other, forming a column. This usually requires the introduction of gap characters to make up for insertions or deletions in the sequences. Figure 1 shows the alignment of the amino acid sequences of human alpha and beta chains of hemoglobin. An alignment of two sequences is called a pairwise alignment; an alignment of several sequences is called a multiple alignment. Generally it is assumed that a certain level of similarity among sequences is an indicator of homology, i.e. of a common evolutionary origin.

For two sequences, the exact definition of sequence similarity, its quantification and the algorithm for computing an alignment are tightly linked (1). The general goal is to find an alignment that introduces just enough gaps to pinpoint the conserved residues among the sequences. To this end, weights are assigned to the matching of letters and to the gaps. A reasonable scheme for nucleotide matching would e.g. assign +1 for pairing up two identical residues, -0.7 for pairing up two unequal residues. While good matches lead to a positive score, any gap that is needed gets penalized, say with the subtraction of 2 score points per gap

character. Frequently, a series of contiguous gap characters is assigned a score that is composed of a “gap-open” penalty plus a “gap-extension” penalty that is multiplied by the length of the contiguous gapped stretch. A dynamic programming algorithm is used to compute an alignment optimizing this score and at the same time the score itself. The most widely used variant of sequence alignment algorithm is the Smith-Waterman algorithm (2), which identifies those regions from the two sequences that match up best.

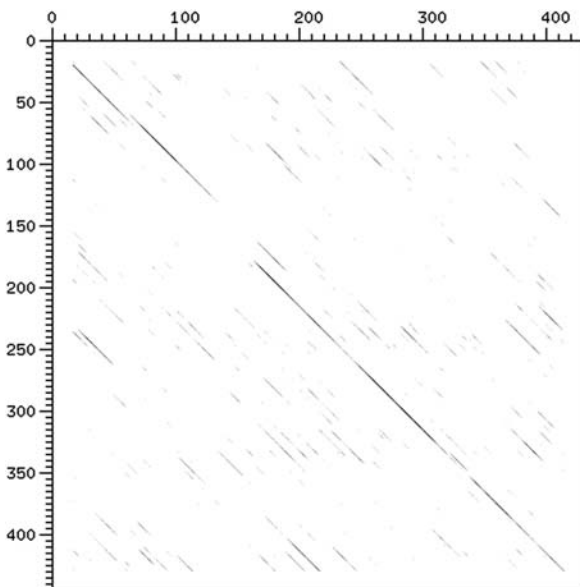
Graphically the similarity between two sequences becomes apparent when depicting stretches of matching residues in a rectangle spanned by the two sequences. Figure 2 shows an example of such a dot-plot. There, the coding DNA sequences of the alpha chain of human hemoglobin are compared to those of the beta chain of human hemoglobin. Sequence windows of length 31 were inspected. Within these windows a score is calculated where +5 was assigned for pairing up equal residues and -4 for unequal residues. The grey values of the dots scale with the similarity of the two windows. One can clearly discern a sequence of diagonal stretches along the length of the two sequences. Note that linking these diagonal stretches requires the introduction of jumps. These jumps correspond to positions where in an optimal alignment a gap would be introduced.

The alignment of proteins is usually based on more sophisticated scoring schemes for pairs of amino acids. Each possible amino acid pair is assigned a value that is derived from an analysis of observed exchange frequencies among amino acids in homologous proteins. Among the widely used examples of such amino acid exchange matrices are the PAM120, PAM250 (3), the BLOSUM62 (4) and the VT160 (5) matrices. The numbers in those names stem from a framework wherein the matrix describes a particular evolutionary distance. For example, in the PAM and VT series of matrices the evolutionary distance is measured in multiples of the time during which on average 1 amino acid in 100 is mutated.

When a sequence contains one or several subsequences that resemble each other, one speaks of internal repeats. Sequence alignment may be a suboptimal tool for the delineation of repeats because one needs to inspect an unknown number of alternative alignments in order to spot all instances of a repeated region. This problem is better approached by inspecting a dot plot comparing

```
HBA_HUMAN      2 -VLSPADKTNVKAAGKVGAGHAGEYGAEALERMFSLFPTTKTYFPHF-----DLSHGSAQV
HBB_HUMAN      3 VHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKV
```

Sequence Comparison. Figure 1 Excerpt of the alignment of the amino acid sequences of human hemoglobin alpha and beta chains.



Sequence Comparison. Figure 2 Dot-plot depicting the comparison between the coding sequences of human hemoglobin alpha and beta chains.

a sequence to itself. In the presence of internal repeats the dot plot will display a characteristic pattern of parallel diagonals stemming from the subsequences that match up.

Multiple Alignment

Multiple alignment compares several sequences simultaneously and constitutes a computationally very hard problem. Practical programs rely on iterative execution of pairwise sequence alignment or implement heuristics to search the large space of possible multiple alignments. Scoring schemes are usually composed of the pairwise similarity scores, either of all protein pairs or of those protein pairs that are adjacent to each other in a phylogenetic tree. Widely used programs for multiple alignment are CLUSTALW (6) and DIALIGN (7) with the latter being more geared toward the identification of similar subsequences from and among the given ones. It should be noted, though, that no sequence alignment output by a program should be accepted without visual inspection, possibly leading to subsequent manual improvement. Alignment editors like seaview (8) are helpful in this context.

Applications

The applications of sequence alignment programs are manifold and range from evolutionary studies to structure prediction. In particular, a multiple alignment constitutes the typical starting point for a phylogenetic study. However, a certain circularity must be noted

there, since frequently a preliminary phylogenetic tree based on pairwise comparisons is used for the computation of the multiple alignment. Another prominent application of sequence comparison lies in function prediction. In particular, common [protein domains](#) can be detected that then lead to the hypothesis that both proteins share the function mediated by the particular domain. Structure prediction, and here in particular [homology modeling](#), also relies on a sequence alignment.

Database Searching

Sequences are collected in sequence databases. Given a new sequence the first question is whether it is similar to anything that is known already. This is answered by comparing the new sequence, the query, to all sequences in a database using the program BLAST (9). In effect, BLAST performs a similar operation to that of the Smith-Waterman alignment on the query against each sequence in the database, although much faster. BLAST can compare DNA with DNA and protein with protein, but it also allows comparison of for example, a protein sequence against a nucleotide database by implicitly translating in all reading frames. A more sensitive search is effected by the program PSI-BLAST (9), which uses newly identified sequences to initiate a new round of searching. Generally one faces a trade-off between speed and sensitivity here; the faster the search program the more unlikely it is to pick up distantly related sequences.

Correct judgment of the findings of a database search requires a statistical assessment of the scores that were achieved. Today, the random distribution of the similarity score is well understood (1) allowing us to determine the statistical significance of the hits that were obtained. A typical database search program will, for each sequence, output a so-called E-value. The E-value is an estimator of the number of sequences one would find by chance scoring better than this particular sequence. This figure not only reflects the quality of the similarity but also the size of the database, which constitutes a major factor in the judgment of significance.

Given a multiple alignment for a protein family or a protein domain, one often wishes to identify additional related sequences. This is achieved by converting the given alignment into an HMM, a hidden Markov model. An HMM can be thought of as an automaton that can generate sequences that look like the ones with which it was fed. A special algorithm can also compute the probability that another sequence could be the product of this generation process. HMMs are used for database searching by computing this probability for each entry in the database. Today, this process is the basis for the most sensitive database searching procedures (10).

Protein Families, Orthology, and Paralogy

Two homologous genes from two different organisms are called orthologous when on the evolutionary path from their most recent common ancestor to the present day genes no gene duplication event has occurred (11). Otherwise they are called paralogous. This definition is sometimes sharpened by also considering genomic order. Database searching serves as a basis for the delineation of protein families. Using each sequence in a database as a query for an individual search, one can compute an all-versus-all comparison. From this information, clustering algorithms are applied to determine groups of proteins that are likely to fall into families of paralogous genes. Determination of orthologous genes is more involved because one rarely has knowledge of the complete evolutionary tree of the genes under study in order to check for the absence of duplication events. Alternatively, the availability of complete sets of genes for two organisms allows the determination of orthology based on two “reverse” database searches. Two genes are likely to be orthologous when either one identifies the other one as most similar in a search among the genes of its organism. Linking together pairs of genes that have been determined as orthologous by this scheme leads to clusters of orthologous genes (12).

Genomic Comparison

Sequence comparison is not only used to identify homologous genes but has also become a major tool in the processing of genomic sequences. For example, alignment between the coding sequence of a gene and its genomic region is applied for the delineation of gene structure. When no coding information is available, comparison of genomic regions from say man and mouse may, by way of the conservation pattern, indicate the location of exons or possible regulatory elements. Emphasis in genomic comparisons is often on identifying near identical sequences to the query, where possible differences are due to sequencing errors or allelic differences. Software for this purpose is optimized for speed rather than sensitivity. A typical example is the program BLAT (13), which is used to locate a piece of sequence of an organism in a database holding this organism's genomic sequence. For the comparison of long stretches of sequence, a classical sequence alignment is not a practical output. The program BLASTZ (14) visualizes a comparison between two genomic regions by focusing the display on one organism or region and plotting over it the level of similarity one finds between the two sequences.

Pattern Finding

Complete genomic sequences are, in particular, supplying us with non-coding sequence in which to search for

regulatory signals. This, too, is approached using sequence comparison methods, albeit with an emphasis on subtle, short similarities. Input to such an analysis could come from for example a set of co-expressed genes as determined in a [▶microarray](#) experiment. The upstream regions of these genes are then expected to harbor common sequence patterns reflecting possible binding sites of transcription factors mediating co-expression of genes. Identifying such short patterns poses both an algorithmic and a statistical problem. Since the search space of all those subtle patterns is huge, not only deterministic algorithms but also stochastic algorithms have been developed for this purpose. A typical program to identify common motifs is MEME (16). The statistical issue stems from the simple observation that regulatory signals may be too short to be distinguishable from random patterns. A program can therefore only enumerate subtle patterns while the biological significance still needs to be established separately.

[▶Sequence Annotation in Evolution](#)

References

1. Waterman MS (1995) Introduction to Computational Biology. Chapman & Hall, London
2. Smith TF, Waterman MS (1981) The identification of molecular subsequences. *J Mol Biol* 147:195–197
3. Dayhoff MO, Schwartz RM, Orcutt BC (1978) A model of evolutionary change in proteins. In: Dayhoff MO (ed) *Atlas of protein sequence and structure* 5, pp 345–352
4. Henikoff S, Henikoff JG (1992) Amino acid substitution matrices from protein blocks. *Proc Nat Acad Sci USA* 89:10915–10919
5. Mueller T, Vingron M (2000) Modeling amino acid replacement. *J Computational Biol* 7:761–776
6. Thompson JD, Higgins DG, Gibson TJ (1994) CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Res* 22:4673–4680
7. Morgenstern B (1999) DIALIGN 2: Improvement of the segment-to-segment approach to multiple sequence alignment. *Bioinformatics* 15:211–218
8. Galtier N, Gouy M, Gautier C (1996) SEAVIEW and PHYLO_WIN: two graphic tools for sequence alignment and molecular phylogeny. *Comput Appl Biosci* 12: 543–548
9. Altschul SF, Madden TL, Schaffer AA et al (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic Acids Res* 25:3389–3402
10. Durbin R, Eddy S, Krogh A et al (1998) *Biological sequence analysis: probabilistic models of proteins and nucleic acids*. University press, Cambridge
11. Graur D, Li W-H (2000) *Fundamentals of molecular evolution*. Sinauer, Sunderland, MA
12. Tatusov RL, Natale DA, Garkavtsev IV et al (2001) The COG database: new developments in phylogenetic classification of proteins from complete genomes. *Nucleic Acids Res* 29:22–28

13. Kent WJ (2002) BLAT – the BLAST-like alignment tool. *Genome Res* 12:656–664
14. Schwartz S, Kent WJ, Smit A et al (2003) Human-mouse alignments with BLASTZ. *Genome Res* 13:103–107, (2004) Erratum in: *Genome Res* 14:786
15. Schwartz S, Zhang Z, Frazer KA et al (2000) PipMaker – a web server for aligning two genomic DNA sequences. *Genome Res* 10:577–586
16. Bailey TL, Elkan C (1994) Fitting a mixture model by expectation maximization to discover motifs in biopolymers. In: *Proceedings of the Second International Conference on Intelligent Systems for Molecular Biology*. AAAI Press, Menlo Park, California, pp 28–36

Sequence Databases

Definition

There is an ever-increasing amount of sequence data being generated from high-throughput methods such as large-scale genome sequencing projects.

Sequence databases contain DNA sequences, transcriptomics data (ESTs, ORESTES, full-length cDNAs) and/or proteomics data (protein sequence data, mutation databases and databases of regulatory regions and elements). Most of them provide additional annotated information such as gene expression, cellular role, protein family and taxonomic data for microbes, plants and humans. Each sequence can be identified by an Accession number.

Search and retrieval entry points are for example the ENTREZ portal (www.ncbi.nlm.nih.gov/entrez <<http://www.ncbi.nlm.nih.gov/entrez>>) and the sequence retrieval system (SRS) at the EMBL-IBI (<http://srs.ebi.ac.uk>). Classified protein family databases such as Prosite (www.expasy.org/prosite <<http://www.expasy.org/prosite>>) are helpful in grouping genes into protein families, and offer a valuable resource for integrating information on cellular and molecular function.

► [Protein Databases](#)

Sequence Polymorphism

Definition

Sequence polymorphism refers to variations of a genetic sequence that occur within a population.

► [Morpholino Oligonucleotides, Functional Genomics by Gene ‘Knock-down’](#)

Sequence-Specific Post-Transcriptional Gene Silencing by Double-Stranded RNA

► [RNA Interference in Mammalian Cells](#)

Sequencing

Definition

In molecular biology, sequencing usually means the determination of bases in DNA or RNA, although the sequence of amino acids in proteins can also be determined. For RNA and DNA different methods are used, some common techniques are chemical or enzymatic splitting by reverse transcriptase, or direct blotting electrophoresis. (see also ► [Sanger Sequencing Method](#))

► [Large-Scale ENU Mutagenesis in Mice](#)

► [Mouse Genomics](#)

► [Shotgun Libraries](#)

Ser

Definition

Ser stands for the Drosophila Serrate protein.

► [Notch Pathway](#)

Serial Analysis of Gene Expression

Definition

Serial Analysis of Gene Expression (SAGE) is based on the massive sequential analysis of short cDNA sequence tags. Each tag is derived from a defined position within a transcript. Its size (14 bp) is sufficient to identify the corresponding gene, and the number of times each tag is observed provides an accurate measurement of its expression level. Since tag populations can be widely amplified without altering their relative proportions, SAGE may be performed with minute amounts of biological extract. The most attractive feature of SAGE is its ability to evaluate the expression pattern of thousands of genes in a quantitative manner without prior sequence information.

- ES Cell Differentiation as a Model System for Functional
- Microarrays in Colorectal Cancer
- RNA Interference in Mammalian Cells

Seropositive

Definition

Rheumatoid arthritis patients with a positive ► [rheumatoid factor](#) test are designated as having a seropositive disease.

- [Rheumatoid Arthritis](#)

Sertoli Cells

Definition

Sertoli cells are cells in the seminiferous epithelium of the testis, which envelop male germ cells and support development of spermatogenic cells into mature spermatozoa. The Sertoli cells produce ► [antimüllerian hormone](#) during embryogenesis.

- [Mammalian Fertilization](#)
- [SRY – Sex Reversal](#)

SET

Definition

Solubility enhancement tag (SET) refers to a small and highly soluble protein domain attached to the protein of interest to improve the expression level, solubility and stability.

- [3D Structure by NMR](#)

SET Domain

Definition

SET domain is the short name* for a highly conserved 130 – to 140 amino acid motif of characterizing a group of proteins known to methylate histones on lysine. Histone methylation is important in the regulation of

chromatin and gene expression (*SET: Su(var)3-9, E(z) and Trithorax).

- [Hereditary Neuropathies, Motor and/or Sensory](#)

Seven-Transmembrane Receptors

Definition

Seven-Transmembrane Receptors (7-TMs) are a superfamily of transmembrane proteins that cross the cell membrane seven times (also referred to as seven-transmembrane receptors, or 7-TMs). These receptors bind peptide ligands and signals by activation of G-proteins.

- [G-Protein Coupled Receptor](#)
- [G-Proteins and G-Protein Mutations in Human Disease](#)
- [Transcription Factors and Regulation of Gene Expression](#)

Sex Chromosome Silencing

- [X-Chromosome Inactivation](#)

Sex Reversal

Definition

Sex reversal in mammals refers to a process whereby the sexual identity of an individual is changed from one sex to the other. Embryos with female chromosomes develop male organs, or embryos with male chromosomes develop female organs.

- [SRY – Sex Reversal](#)

Sex-Determining Gene

- [SRY – Sex Reversal](#)

Sex-Determining Region Y

► [SRY – Sex Reversal](#)

SH2/SH3 Domains

Definition

The Src Homology 2 (SH2) and Src Homology 3 (SH3) domains are small globular protein modules that mediate protein-protein interactions in signal transduction pathways that are activated by protein tyrosine kinases. SH2 domains bind to short phosphotyrosine-containing sequences in growth factor receptors and other phosphoproteins. SH3 domains bind to target proteins through sequences containing proline and hydrophobic amino acids.

- [Adherens Junctions](#)
- [Adhesion Molecules](#)
- [Cardiac Signaling: Cellular, Molecular and Clinical Aspects](#)
- [Methylation of Proteins](#)
- [Tight Junctions](#)

Shaker/Usher Syndrome

Definition

The Shaker/Usher syndrome is the result of a myosin VII mutation and leads to aberrant cochlear hair cells and deafness.

- [Actin Cytoskeleton](#)

Shc

Definition

Shc (src homology collagen) refers to SH2-containing, collagen-related proteins, a family of adaptor proteins that are recruited to activated tyrosine kinases in response to receptor ligation. These proteins are composed of three interaction domains: a carboxy-terminal Src homology 2 domain (SH2), a central collagen homology domain (CH), and an amino-terminal phosphotyrosine binding domain (PTB).

- [Signal Transduction: Integrin-Mediated Pathways](#)

SHH

Definition

Sonic hedgehog. The secreted protein sonic hedgehog (Shh) is essential for normal development of many organs. Targeted disruption of Shh in mouse leads to near complete absence of craniofacial skeletal elements at birth, and mutation of SHH gene in human causes ► [holoprosencephaly](#).

- [Hedgehog Signalling](#)

Shine-Dalgarno Site

Definition

Shine-Dalgarno site designates a bacterial mRNA sequence complementary to a sequence near the 3' end of 16S rRNA. Base pairing between the complementary sequences occurs during translation initiation, and directs the ribosome to a correct initiation codon. The same type of base pairing also occurs at many programmed frameshift sites in prokaryotes.

- [Translational Frameshifting, Non-standard Reading of the Genetic Code](#)

Shmoo

Definition

Shmoo is the polarization that yeast cells form towards a mating partner, named after a cartoon character which first appeared in 1948 and resembles the shape of an elongated yeast cell.

- [Two-Hybrid System](#)

Short Interfering RNA

- [Small Interfering RNA](#)

Short Interspersed Repeat

Definition

Short interspersed repeat denotes any non-autonomous non-LTR (long terminal repeat) retrotransposon

derived mostly from host genes containing a pol III (►[Polymerase gene](#)) internal promoter. SINE elements are retrotransposed by reverse transcriptase encoded by long interspersed repeat (LINE) elements.

►[Repetitive DNA](#)

Shotgun Libraries

MATTHEW C. JONES, KIRSTEN E. MCLAY

The Wellcome Trust Sanger Institute, Wellcome Trust
Genome Campus, Hinxton, Cambridgeshire, UK

mcj@sanger.ac.uk

kem@sanger.ac.uk

Definition

Shotgun libraries are generated to allow the sequence determination of specific ►[DNA](#) molecules. This strategy involves the random breakage of a particular DNA molecule into small pieces. Many copies of the DNA are broken, creating multiple starting points for DNA sequencing distributed throughout the DNA molecule. A series of overlapping DNA sequences can be used to reconstruct the entire sequence of the original DNA molecule. The scattered and random distribution of the DNA fragments bears some resemblance to the random distribution of pellets from a shotgun, hence the name shotgun libraries.

Characteristics

Introduction

The information that defines the nature of an organism (through the medium of polypeptides) is passed from generation to generation in the form of the sequence of four nucleotide bases in the molecule deoxyribose nucleic acid (DNA). It is possible to infer this sequence by the process of DNA sequencing. The majority of DNA sequencing utilises the Sanger method of incorporating chain terminating dideoxy ribonucleotides, whilst synthesising a copy of template DNA. This involves annealing a primer oligonucleotide to the template DNA, and allowing cycles of DNA synthesis (mediated by a thermostable DNA polymerase) to take place. The DNA polymerase generates a complementary strand of DNA to the template, but if it incorporates a dideoxynucleotide (instead of the deoxynucleotide), the complementary strand cannot be extended further. By manipulating the levels of deoxynucleotides and dideoxynucleotides available to the DNA polymerase, strands of a vast range of lengths will be produced. All strands of a specific length will terminate with the specific dideoxynucleotide that is

complementary to the template nucleotide base that distance from the oligonucleotide primer. If the DNA polymerase also incorporates some type of label (most often in the form of fluorescent dye conjugated dideoxy nucleotides) then electrophoresis of the sequencing reaction with a suitable detection system for the label will allow inference of the DNA sequence of the template DNA.

A number of factors influence the length of sequence that can be determined in a sequencing reaction (also called a read, hence the term ►[read length](#)). Current technology limits read length to the range of up to one thousand bases. This is orders of magnitude less than many molecules that are the subject of DNA sequencing projects. In order to determine the DNA sequence of molecules longer than the read-length obtained by current technology, a strategy for generating related smaller DNA sequences has to be employed. Shotgun libraries represent one such strategy, which has enormous potential for using high levels of automation to generate very large quantities of DNA sequence information.

In shotgun libraries, the DNA molecule that is to have its sequence determined is broken, randomly, into small pieces. Because many copies of the DNA molecule are broken and the breakage is random, the location of starting points for sequencing is distributed throughout the DNA molecule. Each randomly generated fragment of the DNA molecule (or a subset of the fragments separated by size) is ligated into a specific point of another DNA molecule (the sequencing vector). This generates a collection (or library) of new DNA molecules, part of each consisting of the same DNA (the vector) and the remainder consisting of different parts of the original DNA molecule. Each of these new molecules is a ►[subclone](#) of the original molecule. Once the sequence of many of these subclones has been determined, that data can be assembled to begin to reflect the DNA sequence of the original molecule. ►[Assembly](#) consists of comparing the DNA sequence of subclones, one with another, and considering those with identical parts of sequence to overlap. Sequence assembly requires a considerable amount of computing power for all but the smallest sequencing projects. Because the subclones are distributed randomly, the DNA sequence derived from them is scattered through the final assembly. This scattered and random distribution bears some resemblance to the random distribution of pellets from a shotgun, hence the name shotgun libraries.

The generation of shotgun libraries involves a number of techniques. The best technique to use at each stage will depend upon a number of factors. The first consideration is the starting DNA. For many large scale projects this will itself be a subclone (into a ►[cloning vector](#), for instance a bacterial artificial

chromosome or ►BAC) derived from genomic DNA. DNA preparations of such primary subclones are relatively easy to produce. Alternatively, the starting DNA may itself be genomic DNA (the whole genome shotgun approach), which may be hard to produce in large quantities, depending upon the organism. If the starting DNA is hard to produce in quantity, optimisation of the following steps should be done with other DNA.

Shearing

There are several techniques for the random (or near random) subdivision of DNA molecules. DNase I in the presence of manganese ions introduces double strand breaks in DNA (1), but this process can be difficult to optimise. Several physical breakage methods involve the passage of DNA solutions through tubing at high pressure; for instance needle shearing (using a 30 G hypodermic needle), ►HPLC tubing (also used in the Hydroshear device) and clinical nebulisers. A very popular method is to use cavitation produced by a sonicator. The sonicator methodology uses a cup-horn sonicator probe, which is functionally similar to an ultrasonic bath. Tubes containing the DNA to be disrupted are suspended in the bath during sonication. Because there is no direct contact between the sonicator and the DNA sample, there is no opportunity for sample-to-sample contamination.

End Repair

Once the DNA has been disrupted, it is necessary to ensure that all the newly generated ends of DNA molecules are in the same form. One technique is to make all the ends of the molecules blunt. This is to ensure that there are neither 3' nor 5' single strand extensions. Provided that starting DNA is not limiting, the easiest way to do this is to use the single strand specific nuclease, mung bean nuclease (EC 3.1.30.1) to digest any single stranded extensions. Where starting DNA is limiting, the use of a mixture of T4 DNA polymerase and ►polynucleotide kinase in the presence of dNTPs will fill-in single stranded ends (rather than nibbling back) and may give a higher yield of subclones per unit amount of starting material.

Ligation

The kinetics of the reaction in which end-repaired DNA fragments are ligated into sequencing vector are such that smaller inserts are favoured over longer inserts. As the costs of each sequencing reaction are about the same, regardless of length of the insert, smaller subclones may not provide a cost-effective amount of sequence. Some methods of disrupting DNA are reputed to produce a smaller range of fragment sizes than other methods. Alternatively, it is relatively easy to

remove small fragments by using preparative electrophoresis in agarose gels. Some staining and visualising methods for DNA electrophoresis can have a deleterious affect on subsequent cloning, but there are strategies to overcome this. One method is to cut off marker tracks, stain and view against a ruler, and then use this measurement to recover agarose slices containing DNA of the desired size, which will not have been exposed to any deleterious influences. The DNA fragments will need to be recovered from the agarose slices. A number of proprietary methods are available for this. It is possible to lose significant amounts of valuable DNA fragments during this part of the process; this should be taken into consideration when dealing with limiting amounts of starting DNA.

Sequencing Vector

A large number of plasmids exist which would be suitable sequencing vectors for shotgun libraries. All feature a selection method (usually antibiotic resistance) to allow the maintenance of the plasmid (and recombinant plasmids derived from it) in bacterial hosts grown in selective conditions. It is necessary to have a single site for the action of a restriction endonuclease, which will leave the resulting ends in a suitable state to accept the fragments of the DNA to be sequenced. In order to encourage the formation of recombinant plasmids (those that contain an insert) rather than re-circularising the plasmid, the linearized vector plasmid can be treated with a phosphatase enzyme such as calf intestinal alkaline phosphatase or shrimp alkaline phosphatase. These enzymes remove the 5'-phosphate group of the linearized vector molecule, preventing T4 DNA ligase from recircularising the plasmid. As the mung bean nuclease treated insert-DNAs maintain 5' phosphate groups, these are a substrate for T4 DNA ligase to create recombinant molecules with the linearized vector plasmid. It is possible to create inserts made up of two or more independent pieces of starting DNA, but this will be minimised by size selection of DNA fragments by preparative agarose gel electrophoresis. Although dephosphorylation of vector should inhibit the generation of recircularised (insert-less) molecules, some will still be produced. Most sequencing vectors in common use have some way of distinguishing between molecules with and without insert DNA. Many systems employ alpha-complementing fragments of the gene for the enzyme β -galactosidase. In suitable bacterial hosts, and plated on media containing IPTG (isopropyl- β -D-thiogalactopyranoside) and Xgal (5-bromo-4-chloro-3-indolyl- β -D-galactoside) bacterial colonies containing vectors without inserts are blue, whilst those with inserts have no colour and appear white or cream.

Template Preparation

Once a library of fragments has been created, the nucleotide sequence of the shotgun fragments can be determined by shotgun sequencing.

An effective template purification method is required to ensure good quality template for use in the sequencing reaction. An individual bacterial colony containing the insert DNA can be selected and placed into a suitable growth media, such as Luria Bertani (LB) broth. To ensure retention of the plasmid, the growth media should be supplemented with an appropriate antibiotic, such as ampicillin. The yield and quality of the plasmid is dependent on the conditions in which the bacterial culture is grown, so it is important to regulate the incubation time and temperature to the optimum.

Plasmid DNA can be prepared by a variety of modifications of the original alkaline lysis method of Birnboim and Doly (2). The alkaline lysis procedure essentially relies on bacterial lysis by sodium dodecyl sulphate and sodium hydroxide. Subsequent neutralization with potassium acetate results in selective precipitation of the bacterial chromosomal DNA along with other cell components of high molecular weight, leaving the desired DNA in suspension. The plasmid DNA suspension can easily be aspirated off and precipitated using isopropanol. The aim of this preparation method is to obtain plasmid DNA that is uncontaminated with bacterial chromosomal DNA. This is a relatively cheap and effective way of obtaining template of good sequencing quality, with many commercial kits now available for DNA purification based on this original method. For high throughput template preparation, microtitre plates and robotic workstations are regularly used where large numbers of subclones will be involved.

It would be prudent at this stage to assay the yield of plasmid DNA by agarose gel electrophoresis prior to sequencing. It is essential to have high quality, quantified DNA to ensure that the sequence data is also of high quality. Running a standard marker alongside the purified DNA ensures that the DNA is of adequate concentration and insert size to proceed with sequencing.

Sequencing

The two main methods of DNA sequencing described in 1977 are the dideoxy chain termination method (Sanger) (6) and the chemical degradation method (Maxam and Gilbert) (5).

Although both techniques are still used today, in large-scale genome sequencing the Sanger method is most commonly used. The Sanger method involves the synthesis of a DNA strand from a single-stranded template by a **DNA polymerase**. By using the correct ratio of the four conventional deoxynucleotides (dNTPs) and one of four dideoxynucleotides (ddNTPs)

in a reaction with DNA polymerase, polynucleotide chains of varying lengths are produced. The ddNTPs are incorporated into the growing strand in the same way as the conventional dNTPs. Due to ddNTPs lacking the 3'-hydroxyl group necessary for chain elongation, when a ddNTP is incorporated into the new strand, the chain elongation will terminate at this position.

By using the four different ddNTPs in four separate reactions the complete sequence information can be obtained. One of the dNTPs is usually radioactively labelled which allows visualization on an autoradiograph following the electrophoresis of the four samples on a polyacrylamide gel. There have been many improvements to this original method, with terminator sequencing now using fluorescent forms of each ddNTP. These can now be added together into a single reaction because each of the ddNTPs has an individual high sensitivity dye attached to it.

One of the major advances in sequencing technology has been the introduction of cycle sequencing which allows the series of events in the dideoxy sequencing method to occur many times under the control of a thermal cycler. Cycle sequencing improves the efficiency and reliability of sequencing double-stranded DNA and results in longer DNA sequences with improved resolution. Another major advance has been the introduction of automated DNA sequencing instruments. These instruments enable the fluorescence emitted from the ddNTPs to be detected and the nucleotide sequence generated.

Data Assembly

Once the DNA sequence has been obtained, the individual reads can then be re-assembled with the use of a computer-based assembly program. Two main assembly packages GAP4 (3) and Consed (4) are widely used to aid the re-construction of the original DNA. Sequencing from both ends of the subclone dramatically improves the efficiency of the final sequence assembly. In contrast to single sequence, (sequencing from only one end of the shotgun subclones), the pairs of sequence reads have known orientation and spacing, aiding the accurate re-assembly of the original DNA molecule.

The majority of large-scale sequencing projects have adopted the shotgun sequencing approach, due to both efficiency and the possibility of automating most of the DNA template preparation and sequencing protocols.

Clinical Relevance

Many clinical disorders have a genetic component. This could range from single gene disorders (such as **cystic fibrosis**) in which sufferers are homozygous for a defective recessive allele at a single locus, to complex

multifactorial conditions, which are influenced by alleles at many loci and by environmental conditions. The studies of these diseases and the discovery of model conditions in experimental organisms have been affected by the large amount of data generated by the human genome project and genome projects for model organisms. All of these projects are dependent on shotgun libraries during their sequencing phase.

References

1. Anderson S (1981) Shotgun DNA sequencing using cloned DNase-I-generated fragments. *Nucleic Acid Res* 10:3015–3027
2. Birnboim HC, Doly J (1979) A rapid alkaline procedure for screening recombinant plasmid DNA. *Nucleic Acids Res* 7:1513–1523
3. Bonfield JK, Smith KF, Staden R (1995) A new DNA sequence assembly program. *Nucleic Acids Res* 23: 4992–4999
4. Gordon D, Abajian C, Green P (1998) Consed: a graphical tool for sequence finishing. *Genome Research* 8:195–202
5. Maxam AM, Gilbert W (1977) A new method for sequencing DNA. *Proc Natl Acad Sci USA* 74:560–564
6. Sanger F, Nicklen S, Coulson AR (1977) DNA sequencing with chain terminating inhibitors. *Proc Natl Acad Sci USA* 74:5463–5467

Sialic Acid

Definition

Sialic acid (neuramic acid) represents the parent acid of a family of amino sugars containing 9 or more carbon atoms. It may be regarded as the aldol condensation product of pyruvic acid and N-acetyl-D-mannosamine. Sialic acid has not been isolated and characterized as such, but several N- and O-substituted derivatives are widely distributed in nature.

- [Biochemical Engineering of Glycoproteins](#)
- [Glycosylation of Proteins](#)

Sialidase

Definition

Sialidase (Neuraminidase) hydrolyses the terminal (alpha)-2 → 3, 2 → 6 and 2 → 8-linkages which bind N-acetylneuraminic acid to N-acetylhexosamines and N- or O-ac(et)ylated neuraminyl residues in oligosaccharides, glycolipids and glycoproteins.

- [Biochemical Engineering of Glycoproteins](#)

Signal Recognition Particle

Definition

Signal recognition particle (SRP) refers to a nucleoprotein complex that consists of a 7S RNA and 6 proteins. It prevents the release of secreted and transmembrane proteins into the cytosol. SRP binds the N-terminus of a nascent polypeptide chain emerging from a ribosome, transiently arrests translation, and targets the mRNA-ribosome-nascent polypeptide complex for endoplasmic reticulum where the translation is finished.

- [Repetitive DNA Transposons](#)

Signal Transduction (Pathway)

Definition

Signal transduction describes the sequential process initiated by the binding of an extracellular signal to a receptor converting an extracellular signal into a cellular response. Signal transduction pathway describes an intracellular response pathway of proteins that activate each other upon an extracellular stimulus (usually receptor activation), and which generally results in the modulation of the gene expression of target genes.

- [Cancer Invasion and Metastasis: Cellular, Molecular and Clinical Aspects](#)
- [G-Proteins and G-Protein Mutations in Human Disease](#)
- [Interleukins and Their Receptors, a Role in Regulating Hematopoietic Stem Cells](#)
- [Receptor Serine Threonine Kinases](#)
- [Signal Transduction: Integrin-Mediated Pathways](#)
- [Somitogenesis](#)
- [Splicing](#)

Signal Transduction: Integrin-Mediated Pathways

KERSTIN DANKER

Institute for Biochemistry und Molecular Biology,
Campus Benjamin Franklin, Charité-University
Medicine Berlin, Berlin, Germany
kerstin.danker@charite.de

Definition

Cell adhesion to the ► [extracellular matrix](#) (ECM) is a biologically important event, providing structural and signaling information for biological processes such as

cell proliferation, differentiation migration and apoptosis, but also for tumorigenesis and metastasis. These adhesive interactions are basically mediated by integrins, a large family of heterodimeric transmembrane cell surface receptors. Besides connecting the ECM with the ►cytoskeleton, ►integrins are also capable of specifically inducing various signaling pathways upon ligand binding. These include the activation of mitogen-activated protein kinases, small ►GTPases and protein kinase C, as well as the modulation of phosphoinositide levels, leading to cell- cycle control and cytoskeletal re-organization (outside-in signaling). Furthermore, signals from inside the cell can modulate the integrin affinity for its ligands, followed by changes in cell adhesion and migration (inside-out signaling). Engagement of integrins results in integrin clustering and the binding of signaling and cytoskeletal proteins to the integrin cytoplasmic domains, leading to the formation of multiprotein complexes called focal adhesions. ►Focal adhesions are heavily enriched in structural and signaling molecules, anchor actin fila-

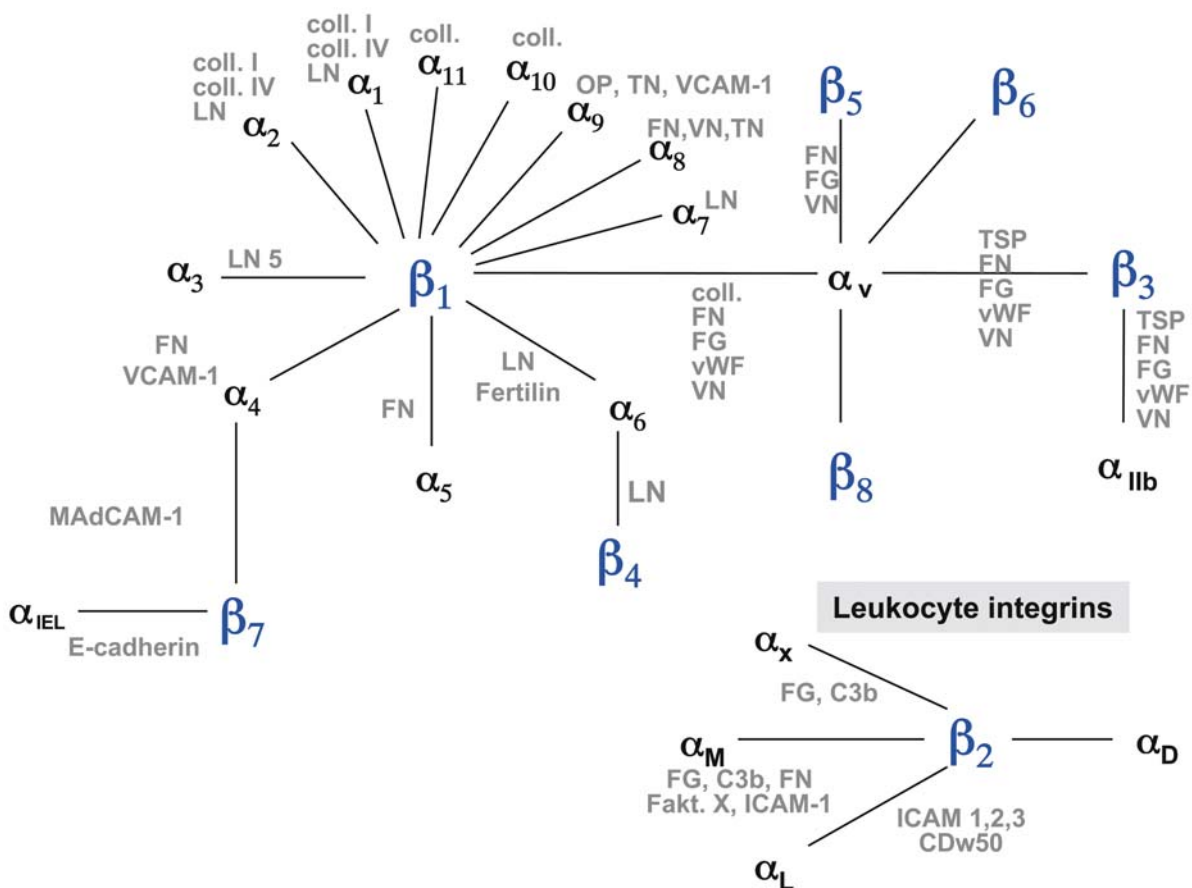
ments and are the starting points of integrin-dependent signaling processes.

The cytoplasmic domains of integrin subunits are essential for recruiting integrins to focal adhesions and for initiating integrin signaling cascades.

Characteristics

Integrins are a diverse family of glycoproteins that form heterodimeric receptors consisting of α and β subunits. These receptors possess a large extracellular domain responsible for ligand binding, a single transmembrane domain and a cytoplasmic domain that in most cases consists of 15–70 amino acid residues. The integrin receptor family in vertebrates includes at least 18 distinct α subunits and 8 β subunits, which can associate to form more than 20 distinct integrins.

The pairing of the two subunits determines ligand specificity. Integrin ligands are extracellular matrix proteins, counter-receptors on other cells such as ►ICAMs, soluble plasma proteins (e.g. fibrinogen, von-Willebrand factor; Fig. 1) or microorganisms.



Signal Transduction: Integrin-Mediated Pathways. Figure 1 Schematic overview of all known integrin receptors and their ligands. C3: complement component 3; Coll: collagen; FG: fibrinogen; FN: fibronectin; LN: laminin; OP: osteopontin; TSP: thrombospondin; VN: vitronectin; vWF: von-Willebrand factor.

Many of the integrins recognize the RGD (Arg-Gly-Asp) sequence within their ligands. Nevertheless, they are capable of distinguishing between different RGD-containing proteins, so that some bind primarily to fibronectin and others to vitronectin.

The strength of ligand binding is modulated by divalent cations, by receptor clustering and by the association of integrins with accessory molecules. Depending on the β subunit, integrins can be divided into subfamilies, the $\beta 1$ integrins being the largest subfamily. The $\beta 2$ integrins are exclusively expressed on leukocytes and are therefore also referred to as leukocyte integrins.

The name integrin was originally coined to denote the function of these proteins in linking the extracellular matrix (ECM) and the cytoskeleton. However, these receptors can regulate many aspects of cell behavior. Signaling enzymes, adaptor and cytoskeletal proteins regulated by integrin engagement control physiological processes through interactions with the cytoskeleton and with other receptors. The cytoplasmic domains of the integrin subunits are key nexuses of interaction between the extracellular environment and intracellular structures and signaling cascades. Upon activation by ligand binding, integrins cluster and recruit cytoplasmic proteins that form so-called focal adhesions. Focal adhesions can be considered as integrin-specific signal transduction entities that are starting points for F-Actin stress fibers and signaling cascades.

Both the α and the β subunit cytoplasmic sequences make important contributions to various aspects of overall integrin function including cytoskeletal reorganization, cell motility, signal transduction and modulation of integrin affinity for the respective ligand. Therefore a large number of cytoskeletal, adaptor and signaling proteins can interact with integrin cytoplasmic domains.

Regulatory Mechanisms

Protein-Interactions with the Cytoplasmic Domains

The cytoplasmic tails of integrins are generally short and always devoid of any enzymatic activity. Hence, integrins transduce signals by associating with adaptor proteins that connect the integrin to the cytoskeleton, cytoplasmic kinases and transmembrane growth factor receptors. Integrin signaling and assembly of the cytoskeleton are intimately linked. As integrins bind to ECM, they become clustered in the plane of the cell membrane and associate with a cytoskeletal and signaling complex that promotes the assembly of actin filaments. The reorganization of actin filaments into larger stress fibers, in turn causes more integrin clustering, thus enhancing the matrix binding and organization by integrins in a positive feedback system. For regulating these processes, the β cytoplasmic tails are required for correct subcellular localization of integrins, for activation of signaling pathways and for regulation of the affinity of integrins for their ligands.

The mechanisms by which integrin β tails function in both outside-in and inside-out signaling remain to be fully resolved. Nonetheless, these processes are probably mediated mainly through direct associations between integrin β tails and signaling and structural proteins. At least 21 proteins are known to bind to one or more integrin β tails. This diverse list of proteins includes actin-binding proteins, enzymes, adaptor proteins, a transcriptional co-activator and additional proteins of unknown function (1).

α -Integrin cytoplasmic tails are less well investigated for their role in signaling. The cytoplasmic tails and in particular the membrane proximal GFFKR motif, conserved among all α integrin subunits, might be critical for the regulation of the affinity states of the receptor.

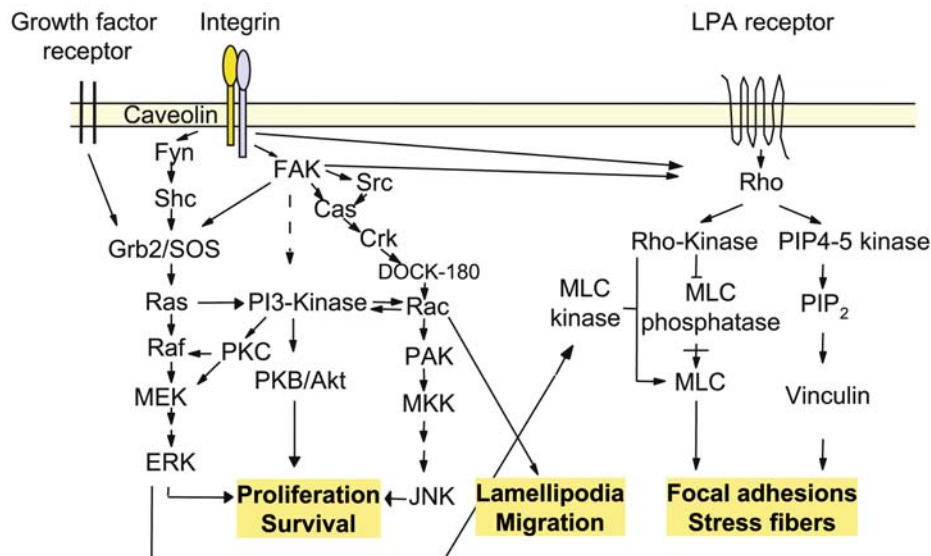
A few proteins have been identified so far that interact with sequences of the cytoplasmic tails of α subunits (2). One protein that binds to the highly conserved GFFKR motif is **calreticulin**, which seems to be a modulator both of integrin adhesive functions and integrin-initiated signaling. But adaptor proteins, like paxillin, and signaling molecules, like phospholipase $C\gamma$, can also bind to these short sequences.

Transmembrane Associations and Signaling

The association of integrins with other transmembrane proteins provides additional coordinate signals to cells that are also specific for individual integrins. **IAP**, an immunoglobulin superfamily transmembrane protein, cooperates with $\beta 3$ integrins in binding thrombospondin to cells and it also activates an inhibitory trimeric guanine nucleotide-binding protein. The $\alpha 3\beta 1$ and $\alpha 6\beta 1$ integrins associate with tetraspan (four transmembrane domain) proteins, which may link these integrins to phosphatidylinositol signaling pathways. When stimulated, some α subunits are able to recruit the adaptor protein **Shc**, via the membrane protein **caveolin-1** (Fig. 2).

Cooperations with Growth Factor Receptors and Other Membrane Proteins

Integrins not only signal on their own but are also necessary for optimal activation of growth factor receptors. The receptors for insulin, platelet-derived growth factor (PDGF), epidermal growth factor (EGF) and vascular endothelial growth factor (VEGF) are optimally activated by their ligands only under appropriate cell attachment conditions. Although a systematic analysis has not been conducted, certain integrins appear to be preferentially associated with specific growth factor receptors. Thus the $\alpha \nu \beta 3$ integrin can be immunoprecipitated in complexes with the insulin, PDGF and VEGF receptors, whereas some $\beta 1$ integrins associate with the EGF receptor. Integrin clustering and association with the cytoskeleton appear to give rise to integrin/growth factor receptor



Signal Transduction: Integrin-Mediated Pathways. Figure 2 Schematic overview of integrin-mediated signaling pathways. For details see the text. Cell proliferation and cell survival are primarily regulated by FAK and Shc, and cytoskeletal structures by the activities of Rho family GTPases.

complexes. Aggregation of the growth factor receptors results in their partial activation, possibly bringing growth factor signaling closer to a threshold of manifest activity and enabling cross talk between integrins and growth factor receptors.

Integrin-mediated Signal Transduction Network Outside-in-Signaling

Integrins regulate physiological processes like migration, not only because they directly mediate adhesion but also because they induce intracellular signaling pathways that control cytoskeletal organization, force generation and survival (3). Depending on the ligand(s), integrins activate one or more signaling pathways. These typically involve phosphorylation of focal adhesion kinase (▶FAK), recruitment of adaptor proteins, activation of small GTPases and subsequent activation of downstream effector molecules (Fig. 2). These signals, in concert with signals derived from growth factors, regulate cell behavior in a complex tissue microenvironment.

FAK-Pathway

Subsequent to integrin activation, the cytoplasmic protein kinase FAK is activated. Activated FAK binds to several signaling molecules, mediating integrin-induced activation of the ▶Ras▶Erk pathway, which promotes cell proliferation. Phosphorylated FAK can activate Erk by recruiting adaptor proteins such as ▶Grb2 and Cas, as well as by activating protein kinases such as the ▶Src-family kinases.

FAK activation has also been shown to promote cell survival and cell migration. *Fak*-null mice die before

birth and cells isolated from these embryos have migration defects.

FAK and Src both phosphorylate ▶Cas at multiple tyrosine residues, causing Cas to associate with another adaptor protein, ▶Crk. The Cas-Crk complex is a component of the molecular migration machinery, inducing cell motility by activating the protein ▶DOCK-180, leading to the activation of the small GTP binding protein Rac.

Shc-Pathway

Upon integrin activation Shc is recruited by the ▶Fyn/caveolin complex into focal adhesions in response to ligation of integrins α1β1, α6β4, α5β1 or αvβ3. Formation of this complex leads to Shc-dependent cell-cycle progression, migration, anti-apoptotic signals and Erk-activation.

In migration Shc is the “random motility component”, whereas FAK and Cas support migration in a persistent direction.

Small G Proteins

Highly motile cells display dramatic alterations in their cytoskeletal organization that facilitate their migratory behavior. The ▶Rho family of small GTPases, including Rho, Cdc42 and Rac, can regulate these changes (4).

Rac is activated by ligation of integrins and promotes membrane ruffling and migration, whereas Rho activation is dependent on integrins, as well as ▶syndecan-4 and additional cell surface receptors like the LPA receptor and leads to the formation of stress fibers and the assembly of focal adhesions.

Rho can activate a ser/thr protein kinase known as Rho kinase. This kinase phosphorylates the myosin light chain (►MLC) phosphatase, thereby suppressing the activity of the enzyme. This increases the myosin binding to actin filaments and subsequently increases actomyosin contractility and the formation of focal adhesions and stress fibers.

Furthermore, integrins can stimulate the production of the membrane compound phosphatidylinositol biphosphate (►PIP2) in a Rho-dependent manner. Rho is able to activate the phosphatidylinositol-4-phosphate-5-kinase (PIP4-5-K) that produces phosphatidylinositol-(4,5)-biphosphate. Phosphatidylinositol-(4,5)-biphosphate modulates the activity of several cytoskeletal proteins like ►vinculin, which in turn connect the focal adhesion complex to the F-actin cytoskeleton.

Rac activation also potentiates Erk signaling and increases cellular sensitivity to growth factors. Another mechanism of Erk activation is provided by the Rac-dependent activation of p21-activated kinase, ►PAK. This ser/thr kinase regulates the actin cytoskeleton as well as the ►JNK/SAP kinase pathway.

PI3-Kinase Pathway

►PI 3-kinase is an important lipid kinase, which has recently been demonstrated to play a role in integrin signal transduction. Some of its lipid products can act as second messengers in the regulation of protein kinase C (►PKC) family members as well of as the protein kinase B/Akt. Association of FAK with PI 3-kinase in response to integrin activation has been demonstrated in both platelets and fibroblasts, thereby regulating cell proliferation, apoptosis and migration.

In addition, FAK/PI 3-kinase association in fibroblasts is stimulated by cell treatment with platelet-derived growth factor (PDGF), suggesting a mechanism of cross-talk between integrin and the PDGF receptor signaling pathway.

PKC Pathway

The protein kinase C family is also important for regulating integrin function and signaling. PKC is required for focal adhesion formation, cell spreading, Shc-dependent Erk phosphorylation and migration. PKC interacts with $\beta 1$ integrins using tetraspanin-family proteins as bridging molecules. Furthermore, PKC regulates $\beta 1$ transport within the cell.

Inside-Out Signaling

Integrins not only send signals to the cell in response to the extracellular environment, but they also respond to intracellular cues and alter the way in which they interact with the extracellular environment. This process – termed “inside-out-signaling” – regulates integrin adhesiveness by modulating the affinity and avidity of integrins for their ligands, thereby regulating integrin-dependent processes like migration.

The signaling mechanisms that regulate integrin affinity and avidity are likely to vary for different integrin heterodimers, but small GTP-binding proteins like Rho, Rac, Cdc42 and Ras as well as the protein kinase C are known to be involved (5).

Therapeutic Applications

Investigation of integrin receptors and integrin-mediated signaling has also led to a number of therapeutic applications and targets for tumor therapy. An anti- $\beta 3$ -integrin antibody is used as an integrin-directed drug for prevention of arterial restenosis.

Several other integrin-based drugs, such as peptides containing the integrin-binding RGD sequence and mimics of such peptides that specifically block individual integrins and their functions are also under development. These drugs target thrombosis ($\alpha \text{IIb}\beta 3$ in platelets), osteoporosis ($\alpha \text{v}\beta 3$ in osteoclasts) and tumor-induced angiogenesis ($\alpha \text{v}\beta 3$ in neovascular endothelial cells).

►Integrin Signaling

References

1. Liu S, Calderwood DA, Ginsberg MH (2000) Integrin cytoplasmic domain-binding proteins. *J Cell Sci* 113:3563–3571
2. Dedhar S, Hannigan GE (1996) Integrin cytoplasmic interactions and bidirectional transmembrane signalling. *Curr Opin Cell Biol* 8:657–669
3. Giancotti FG, Ruoslahti E (1999) Integrin signaling. *Science* 285:1028–1032
4. Kumar CC (1998) Signaling by integrin receptors. *Oncogene* 17:1365–1373
5. Hood JD, Cheresch DA (2002) Role of integrins in cell invasion and migration. *Nature Reviews Cancer* 2:91–100

Signalling

Definition

Communication of cells with one another by means of biochemically diverse signalling molecules like proteins, small peptides, ions, nucleotides or fatty acid derivatives.

Signalling State

Definition

Signalling state defines a protein conformation which is recognized by the next downstream component of the signalling cascade.

►Photoreceptors

Signature

Definition

Signature means a combination of single genes into multi-gene expression patterns to describe the molecular differences between two disease entities.

► [Computational Diagnostics](#)

Silane Reagent

Definition

Silane reagent is an organometallic compound, traditionally applied as repelling reagent to coat a glass surfaces in order to prevent unspecific binding.

► [Protein Microarrays as Tool for Protein Profiling and Identification of Novel Protein-Protein Interactions](#)

Silencer

Definition

Silencer are regulatory elements with repressive functions on transcription found at different positions relative to the transcriptional start site. The DNA segment binds repressors/corepressors (negatively acting transcription factors) that can also act over long distance, with the effect that a remote transcription unit cannot be transcribed.

► [Enhancer](#)

► [Transcription Factors and Regulation of Gene Expression](#)

Silent Substitution

Definition

Single base substitutions in a codon are called silent if they do not affect the corresponding amino acid.

► [Sequence Annotation in Evolution](#)

Silver-Russell Syndrome

Definition

Silver-Russell syndrome is a congenital condition characterized by intra- and extrauterine growth retardation.

Main features of the Russell-Silver syndrome are lateral asymmetry and low-birth-weight dwarfism.

► [Prader Willi and Angelman Syndromes](#)

Simulated Annealing

Definition

Annealing refers to incubation of a mixture of DNAs in single-stranded form and to quantitative determination of the helical material formed. The method permits the detection of homologous regions in any two species of DNA. Simulated annealing is an optimization technique using simulations in which the temperature of the system is gradually decreased to a very low value.

► [Molecular Dynamics Simulation in Drug Design](#)

Simultaneous Slippage

Definition

Simultaneous slippage occurs during programmed -1 translational frameshifting. It defines the simultaneous disruption and re-pairing in the -1 frame of the codon/anticodon complexes of two ribosome-bound tRNAs.

► [Translational Frameshifting, Non-standard Reading of the Genetic Code](#)

SINE

► [Short Interspersed Repeat](#)

Single Gene Disorder

► [Monogenic Disorder](#)

Single Nucleotide Mismatch

Definition

Single nucleotide mismatch describes the pairing of a nucleotide with a non-complementary nucleotide (e.g. pairing of G with T instead of C nucleotide).

► [Hereditary Nonpolyposis Colorectal Cancer](#)

Single Nucleotide Polymorphisms

Definition

Single-nucleotide polymorphisms (SNPs) are the most common type of human DNA genetic sequence variation, occurring on average 1 per 1000 base pairs. They occur when a single nucleotide (A, T, C or G) in the genome sequence is altered. SNPs are mostly bi-allelic, hence less informative than ►microsatellites. Yet they are more frequent and more stable than microsatellites, as well as more amenable to automation and DNA-chip technology. SNPs have drawn attention as potential markers of variation because they are both common (occurring relatively frequently compared with other markers) and plentiful (there may be as many as 30 million in the human genome).

- Atopy Genetics
- Chromosome 21 Disorders
- Common Diseases Genetics
- COPD and Asthma Genetics
- DNA Chips
- Genomic Information and Cancer
- Genetic Predisposition to Multiple Sclerosis
- Large-Scale ENU Mutagenesis in Mice
- Microarrays in Pancreatic Cancer
- Mutagenesis Approaches in Medeka
- PNA Chips
- SNP Detection and Mass Spectrometry
- Splicing
- Thermodynamic Properties of DNA

Single Particle Tracking

- Fluorescence Microscopy: Single Particle Tracking

Single Particles

Definition

Single particles, in the context of electron microscopy, refers to macromolecular assemblies as isolated and unordered particles with identical structure.

- Cryo-Electron Microscopy: Single-Particle Reconstruction

Single Photon Emission Computer Tomography

Definition

Single photon emission (computer) tomography (SPECT or SPET) is a tomographic ►nuclear medicine imaging technique to acquire 3-dimensional images. Internal radiation is administered by means of a pharmaceutical that is labeled with a radioactive isotope (radiopharmaceutical; tracer). The radioactive isotope decays, resulting in the emission of gamma rays. These gamma rays are the source of information, rather than X-ray transmissions as used in conventional computed tomography. Gamma ray emissions SPECT data are acquired according to the original concept used in tomographic imaging: multiple views of the body part to be imaged acquired by rotating the camera head (s) around a craniocaudal axis.

- Structural Molecular Imaging

Single-Cell Gene Expression Profiling: Cell-Level Biology by Multiplexed Expression Fluorescence *in Situ* Hybridization

JEFFREY M. LEVSKY, ROBERT H. SINGER
Anatomy & Structural Biology, Albert Einstein College of Medicine, Bronx, New York, NY, USA
rhsinger@acom.yu.edu

Definition

Single-cell expression profiling is a means to take a 'snapshot' of the gene activity of a specimen at the individual cell level. The active sites of ►transcription are detected for many genes in their native cellular context using ►*in situ* hybridization. This technique preserves each cell's characteristic repertoire of expressed alleles and the location of expressed alleles in the sub-cellular and tissue-level organization. The reaction is performed *in situ*, at the proper location of the gene activation events, to yield a visual analysis of complex expression patterns. Hybridization detection of transcription sites measures new transcription specifically, without considering total cell levels of transcripts, ►translation to protein and all the other

levels of post-transcriptional expression regulation. This allows detection of small incremental changes in transcript levels and direct analysis of causes and effects upstream of transcription. To date, the assay has been used to analyze fixed cell samples from *in vitro* cultures, allowing high temporal and spatial resolution of expression (1). Many hypothetical uses exist for this approach, from experiments to improve the basic understanding of single cell biology and finely demarcated expression patterns to future clinical application in diagnostics.

Characteristics

Cell Level Biology

Undisturbed single-cell biology is largely inaccessible to most popular molecular assays because samples used are invariably derived by summation of many single cell responses or by artificial amplification steps such as ►[polymerase chain reaction](#) (PCR). These procedures are necessary in order to reach a detectable threshold of biomolecules due to the sensitivity constraints of the techniques. Summation obscures individual cell biology and amplification can be uneven, distorting the perceived composition of the cellular contents (2). Using *in situ* hybridization, DNA and RNA can be detected in their natural location, while preserving cell-to-cell heterogeneity (3). Summation of multiple cell contributions is not required as a single cell's signal is visualizable due to the intensely fluorescent probes and highly sensitive microscopy methods used. Amplification is also not necessary as hybridization is performed on the native targets only. Using probes of determined fluorescence intensity, signals can be quantified to measure absolute numbers of transcripts. The sensitivity of ►[fluorescence in situ hybridization](#) has been demonstrated to be sufficient for the detection of single nucleic acid molecules and parts thereof, allowing considerable freedom to measure low levels of transcripts (4). This is a potentially important feature as most messenger RNAs are present in very small amounts – less than ten copies per cell.

Multiplexed Detection

Applying fluorescence microscopy to the detection of several genes' transcription simultaneously yields a profile of gene activity for individual cells, or 'single-cell gene expression profiling' (1). Multiplexing, or concurrent assay of many different targets, is necessary to study certain aspects of cell physiology. For example, the observation of co-expression events at the single cell level and determination of relative locations of functioning genes are only possible by simultaneous assay of several different mRNAs. In general, the larger the number of genes accessible to analysis the greater is the depth of profile and

information content in the dataset. The number of detectable entities depends on the encoding scheme used to represent specific genes. Using fluorescent dyes, color combinations and color ratios have been employed to assay tens of specific nucleic acids at one time *in situ* (3), although there is potential to detect far more species of interest by more complete usage of the spectrum of dyes and digital processing and interpretation of the signals.

Data Analysis

Fluorescence images are interpreted by computer in order to objectively assign specific signals to gene identities based on the color-coding scheme. Then, signals are segmented, or separated, by nucleus in order to yield a dataset of individual cells. The simplest resultant datasets from single cell profiling consist of a catalogue of specific nuclei and the activated alleles they possess. From this information, the percentage of cells expressing a specific gene can be determined, as a measure of tissue-wide activation. Cells can be sorted on the basis of particular expressed gene combinations in order to highlight cell type-specific correlated activations. Using basic statistics, one can determine the likelihood with which a nucleus found to express one gene would express another. Spatial characteristics are incorporated to derive structural data from single cell expression sets. This can involve the precise measurement of distances between different sites of transcription and between sites and other nuclear features. Functional organization of the nucleus is thereby determined for several genes at once. When structural information is added to observations of expression, significant amounts of data can be derived for the classification of individual cells.

Clinical Relevance

Although there are no current clinical trials using single cell expression study as a diagnostic, the technique provides many avenues for pre-clinical investigation.

Expression Modulation

Many changes in physiology are known to be associated with modulation of gene activity. This is particularly important in the dysregulation of oncogenes and tumor suppressor genes during cancerous transformation. Presence or absence of expression can be assayed by single cell monitoring, to theoretically yield prognostic information regarding development of cancer. Expression levels can be changed in more complex ways than mere activation and deactivation. For example, many transcripts possess alternative splicing products, which can be selectively processed under specific conditions. These and

other co-transcriptional mechanisms can be measured using fluorescence color coding and single cell monitoring (5).

Small Populations of Cells

In many cases, diagnosis and prognosis can depend on a very small number of cells that have certain pathophysiological expression characteristics. Assays that do not preserve cellular heterogeneity naturally average many cellular responses and may obscure the few affected cells that have clinical importance. For example, the presence of HIV RNA levels in lymphocytes or activity of genes involved in cancer progression to metastasis in a small cell population within a tumor could be used as diagnostic information. Monitoring changes on this small level is an advantage of the single cell approach (2).

Interactions

Cell-to-cell interactions at the micro level are thought to influence many biological processes, including development and cancer. Often the population of cells involved in such signaling is quite small and difficult to micro-dissect from surrounding tissue types. Visualization of single-cell profiles is one solution to exploring this biology because it obviates the need to separate cells physically for analysis. As cells are measured with preserved organization, computer-based analysis can be used to measure the distances between aberrantly expressing cells and a tissue marker. This would allow the study of expression as influenced by proximity to a heterologous tissue type, as occurs in epithelial-mesenchymal signaling in physiology and disease.

Therapeutic Consequences

Therapy based on conclusions drawn from single cell expression information is yet hypothetical, as the technique has not yet been applied to clinical materials. Single cell expression data provided by this technique has not been determined to correlate with any particular physiological results. However, a closer understanding of single cell behavior has tremendous promise in therapeutics. Above all, cell-level analysis will allow a firmer factual basis for cell behavior upon which models of physiology and pathophysiology can be tested. The technique serves to further our understanding of any process of disease that depends on particular cellular expression events.

References

1. Levisky JM, Shenoy SM, Pezo RC et al (2002) Single-cell gene expression profiling. *Science* 297:836–840
2. Levisky JM, Singer RH (2003) Gene expression and the myth of the average cell. *Trends Cell Biol* 13:4–6
3. Levisky JM, Singer RH (2003) Fluorescence in situ hybridization: past, present and future. *J Cell Sci* 115:2833–2838
4. Femino AM, Fay FS, Fogarty K et al (1998) Visualization of single RNA transcripts *in situ*. *Science* 280:585–590
5. Wilson AJ, Velcich A, Arango D et al (2002) Novel detection and differential utilization of a c-myc transcriptional block in colon cancer chemoprevention. *Cancer Res* 2002 62:6006–6010

Single-Particle Reconstruction

Definition

This terminus refers to the 3D structure determination of single particles from ►2D (cryo-) electron microscopy images.

►Cryo-Electron Microscopy: Single-Particle Reconstruction

Single-Stranded Conformational Polymorphism

►SSCP

Single-Stranded DNA Binding Protein

Definition

Single-stranded DNA binding protein is a protein that protects and stabilizes the unwound single-stranded DNA (e.g. replication protein A).

►Replication Fork

Singlet State

Definition

Singlet state defines – in a multi-electron system like a molecule or an atom – a state of having a total electron spin quantum number equal to 0.

►FRET

Single-Wavelength Anomalous Diffraction

Definition

Single-wavelength anomalous diffraction (SAD) describes a method to determine the diffraction phases in protein crystallography, from a dataset measured at a wavelength for which the X-ray fluorescence, and the anomalous part f' of the scattering factor of a natural or artificially incorporated anomalous scatterer, is at or near a maximum. A tunable radiation source (synchrotron) is required for recording a high anomalous signal from the scatterer. By increasing the quality of the data measurement and the redundancy of the diffraction data set, even weak anomalous signals can be used for phase determination, e.g. when the diffraction experiment is performed at a wavelength far from the absorption edge of a scatterer (e.g. phasing based on anomalous scattering from sulfur with $\text{CuK}\alpha$ radiation from conventional rotating-anode X-ray generators).

► [MAD Phasing](#)

siRNA

► [Small Interfering RNA](#)

Short Interfering RNA

► [Small Interfering RNA](#)

Sister Chromatid

Definition

Sister chromatid designates the extra chromosome that is present in eukaryotic cells after replication and before cell division.

► [DNA Recombination](#)

Sister Chromatid Cohesion

Definition

Sister chromatid cohesion defines the linkage of newly synthesised DNA strands (sister chromatids) after DNA replication, which ensures the faithful inheritance of chromosomes by daughter cells.

► [DNA Polymerases](#)

Sister Chromatid Exchange

Definition

Sister chromatid exchange (SCE) refers to an observation through the light microscope of an exchange between the two sister chromatids in metaphase chromosomes. The basis is a differential staining of the sister chromatids; one chromatid is stained lightly and one is stained densely after they have completed two rounds of replication in the presence of the thymine analogue 5-bromo-deoxy-uridine (BrdU). Due to the semiconservative replication of DNA, one chromatid still has an original strand of DNA in the double helix opposite to a DNA strand that has incorporated BrdU. In the other chromatid both strands are substituted with BrdU. The latter stains less intensely than the former.

► [Bloom Syndrome](#)

Sister Kinetochores

Definition

The term sister kinetochores denotes each member of a pair of kinetochores facing opposite spindle poles.

► [Centromeres](#)

Site-Specific Protease

Definition

These proteases cleave at defined amino acid sequences. They are a useful tool to remove protein tags. Commonly used proteases are enterokinase, X_a , thrombin and TEV (tobacco etch virus) protease.

► [Protein Tags](#)

Site-Specific Recombination

► Cre/loxP Strategies

SIX5

Definition

SIX5 is a gene located in the DM1 locus encoding a homeodomain transcription factor. The instable (CTG)_n repeat, located almost in its promoter, causes transcriptional silencing of the flanking SIX5 allele.

► Myotonic Dystrophy

Size Exclusion Chromatography

Definition

Size exclusion chromatography refers to a chromatographic method used to separate molecules on the basis of their size.

► Protein Interaction Analysis: Chemical Cross-Linking

Skeletal Disorders

► Bone Disease and Skeletal Disorders, Genetics

Skeletonization

Definition

Skeletonization is an image processing method used to simplify complex features for better comprehension, and allows the retrieval of quantitative geometrical information about an object.

► Electron Tomography

Skewed X-Inactivation

Definition

In females, one of two X chromosomes is deactivated well before birth, in the early stages of embryonic development. This means that either the paternal X chromosome is active, or the maternal X chromosome is active. In the female population, one expects a 1:1 distribution of such cells; i.e. about 50% of cells show activation of the paternal X chromosome and 50% show activation of the maternal X chromosome. A distribution which differs from the normal 1:1 is called a skewed X inactivation.

► Duchenne Muscular Dystrophy

► X-Chromosome Inactivation

Skin Barrier

► Skin (Permeability) Barrier

Skin (Permeability) Barrier

Definition

The skin (permeability) barrier is formed by the epidermis and stratum corneum to protect the organism from water loss and physical, chemical, and mechanical insults.

► Heritable Skin Disorders

Skin and Hair

JOERG HUELSKEN

Epithelial Stem Cell Biology, ISREC, Epalinges, Switzerland

joerg.huelsken@isrec.unil.ch

Definition

The skin covers the outer surface of the body and forms the interface with the environment. Human skin has

a surface area of about 1.5–2.0 m² and accounts for 15–20% of our body weight. The skin serves a variety of functions: (1) protection: it builds a physical barrier to mechanical impact or penetration by microorganisms such as bacteria, viruses and fungi, it protects against ultraviolet light damage and prevents dehydration; (2) thermoregulation: it insulates the body *via* hair and adipose tissue and allows heat dissipation by sweat production and blood circulation; (3) sensation: it is the largest sensory organ which detects pressure, vibration, heat, cold, pain and itching; (4) metabolism: it stores energy in the adipose tissue and synthesizes vitamin D.

Characteristics

The skin consists of three layers, the epidermis, which is the most superficial layer, the dermis and the hypodermis or subcutis. The epidermis forms the external surface of the skin and is mainly composed of keratinocytes, which differentiate to form four layers, the ►stratum basale, ►stratum spinosum, ►stratum granulosum and ►stratum corneum (1). Keratinocytes are specialized epithelial cells responsible for epidermal cohesion and barrier function and are firmly interconnected with each other by desmosomes and ►keratin intermediate filaments (tonofibrils). Epidermal keratinocytes are generated in the stratum basale, migrate upwards while they differentiate terminally and are shed at the surface as cornified scales. This self-renewal process normally takes approximately four weeks. Interspersed in the stratified epithelium of the epidermis are other cell types such as melanocytes (UV protection), Merkel's cells (mechanosensory function) and Langerhans' cells and T-lymphocytes (immune defense). A complex basement membrane separates the epidermis from the underlying dermis, the connective tissue responsible for the mechanical resistance of the skin. This function is provided by dermal fibroblasts that synthesize fibers composed of collagens, reticulin and elastin interspersed glycosaminoglycans. The dermis contains sense organs for touch, pressure, pain and temperature (e.g. tactile corpuscles of Meissner, Pacinian corpuscles and free nerve endings) and a dense vascular network that supplies the avascular epidermis. A variety of cells of the immune system are found in the dermis such as perivascular macrophages, dendritic cells and mast cells. The subcutaneous layer below the dermis consists of loose connective tissue and many adipocytes that help to insulate the body.

The dermis also harbors epidermal appendages such as the pilosebaceous unit (hair follicle, sebaceous glands, arrector pili muscle) and sweat glands that extend deep into the dermis (2). Epidermal appendages, the epithelial parts of which are in continuity with the epidermis, develop during embryonic life through

complex epithelial-mesenchymal interactions by epithelial invagination. The hair follicle differentiates into three enclosed epithelial cylinders: the central most cylinder forms the hair shaft, the outer root sheath (ORS) represents the outermost cylinder that separates the whole structure from the dermis, and the middle cylinder, the inner root sheath (IRS), forms and guides the shaft in its passage outward. The shaft and the IRS move outward together. All mature follicles undergo a growth cycle consisting of phases of growth (anagen, 85–90% of hair cycle), regression and involution by ►apoptosis (catagen, 2–3% of cycle), rest (telogen, 10–15% of hair cycle) and shedding (exogen). A complete cycle lasts 2–6 years for human scalp hair during which it grows 0.4 mm/day. In addition to the cycling part, the follicle consists of an upper permanent, non-cycling portion. The lower end of this portion is defined by the hair bulge.

Molecular Interactions

Skin Neoplasms

Several benign and malignant neoplasms arise in the skin (3). These tumors often reflect the differentiation program of certain keratinocyte lineages, e.g. ►basal cell carcinomas (BCCs) mimic the relatively undifferentiated phenotype of the hair follicle outer root sheath, ►squamous cell carcinomas (SCCs) exhibit elements of interfollicular epidermal differentiation and ►pilomatricomas and trichofolliculomas contain cells undergoing differentiation along hair lineages.

BCC of the skin is the most common type of cancer in humans. Two ►tumor suppressor genes, *p53* and *ptch*, are major targets of mutation in sporadic BCC. P53 is important for the induction of cell-cycle arrest in response to cellular insults or DNA damage and can initiate cell death by apoptosis. Germline mutations of *ptch1* have been identified in Gorlin's syndrome, which is characterized, by multiple BCCs, keratocysts and skeletal defects. *Ptch* encodes a negative regulator of the ►Hedgehog signaling pathway. Hedgehog (Hh) signaling plays an important role in the patterning of many tissues and structures during embryogenesis. The response to the Hh signal is controlled by the tumor suppressor gene *ptch* and the proto-oncogene *smo* which both encode transmembrane proteins. In the absence of Hh, the activity of Smoothened (Smo) is suppressed by Patched (Ptc). Smo can signal to a cytoplasmic complex that includes the serine/threonine protein kinase Fused, Suppressor of Fused, and the transcription factor Cubitus interruptus (three genes in humans: *gli1*, *gli2* and *gli3*). Upon Hh stimulation, phosphorylation and subsequent processing of the transcription factor is blocked, and it is translocated to the nucleus as a full-length transcriptional activator. One of the general target genes is *ptch*, resulting in

feedback inhibition. In addition to *ptch1*, mutations in *ptch2* and activating mutations in *smoh* have been identified in sporadic BCC.

SCCs arise in areas of sun-exposed skin and typically involve ►UV signature mutations or ►LOH in key tumor suppressor genes involved in cell-cycle control, *p53*, *cdkn2a* and *rb*. Inactivating mutations in *p53* are an early event in carcinogenesis and have been found in precancerous lesions (actinic keratosis and SCC *in situ*). Through alternative splicing the *cdkn2a* locus encodes two proteins, INK4A and ARF, which positively regulates p53. INK4a is a cyclin-dependent kinase inhibitor that specifically blocks the activity of cyclin-dependent kinase 4 (CDK4) and CDK6 by preventing their interaction with D-type cyclins. These proteins act together to govern the phosphorylation status of the retinoblastoma protein (Rb). Hypophosphorylated Rb binds and inactivates E2F transcription factors responsible for cell-cycle progression. In addition, amplification of the proto-oncogene *ccnd1*, which encodes cyclin D1, has been identified in SCC. Activating mutations in the *ras* proto-oncogene occur in some spontaneous SCCs, and over-expression of *ras* can induce SCC in mouse models. However, expression of active Ras alone induces growth arrest in human epidermal cells by suppressing the cyclin-dependent kinase CDK4. This arrest can be overcome by inhibition of NFκB signaling which induces CDK4 expression. Moreover, expression of the α6β4 integrin is markedly increased in SCC but not in BCC and was shown to be essential for invasive growth of SCCs. This integrin is located to hemidesmosomes in the stratum basale where it attaches to laminin 5 in the basement membrane and can promote migration and survival.

Trichofolliculomas and pilomatricomas are well-differentiated hair follicle derived tumors. Mutations in *ctnnb1* have been found in these tumors with high frequency (1). *Ctnnb1* encodes for β-catenin, the central signal transmitter in the canonical ►Wnt signaling pathway (5). This pathway is crucial for pattern formation during embryonic development and for cell fate determination in the adult, e.g. in the skin and intestine. Many components of the Wnt signaling system have been identified. Extracellular Wnt/wingless ligands activate Frizzled receptors and through Dishevelled induce an increase in cytoplasmic β-catenin by preventing its degradation in proteasomes. Axin/Conductin, in cooperation with the tumor suppressor gene product ►adenomatous polyposis coli (APC), promote β-catenin degradation, which involves serine-threonine phosphorylation of the N-terminus of β-catenin by casein kinases and GSK3β and subsequent ubiquitination. Upon Wnt signaling, β-catenin

accumulates in the cytoplasm and is transported to the nucleus, where it interacts with members of the LEF/TCF family of transcription factors and activates gene expression. Mutations in a variety of human tumors have been identified in *apc* or in the N-terminal phosphorylation sites of β-catenin. These mutations result in accumulation of β-catenin, constitutive signaling and gene activation and are causally involved in human tumorigenesis. For example, germline mutations in *apc* have been linked to Gardner syndrome, which is characterized by early onset adenomatous polyps of the colon but also displays pilomatricomas and epidermoid cysts in the skin.

Human Keratin Diseases

A variety of epithelial fragility disorders was linked to mutations in epithelial *krt* genes. Most of these mutations are inherited as autosomal dominant traits and result in expression of a defective keratin protein, which tends to act in a dominant-negative fashion. For example, mutations in *krt14* and *krt5* have been identified in the skin blistering disorder epidermolysis bullosa simplex. These keratins are expressed in the stratum basale of the skin, and so patients suffer from blisters in response to mild mechanical trauma. Mutations in *krt1* and *krt10*, which are expressed in the stratum spinosum of the epidermis, lead to the skin thickening disorder bullous congenital ichthyosiform erythroderma. In this disorder, breakdown of suprabasal keratinocytes under mechanical stress leads to release of inflammatory cytokines. This stimulates uncontrolled proliferation of basal cells and induces hyperkeratosis.

Ectodermal Dysplasia

Ectodermal dysplasia (ED) represents a complex group of diseases characterized by impaired morphogenesis of epidermal appendages. Primary defects affect ectodermal structures including hair, skin, nails, teeth and sweat glands. EDs are rare diseases with an estimated incidence of 700 patients per million. Several genes that are involved in cell/cell communication and signaling have been identified in ED. The most frequent form of ED, X-linked anhidrotic ED, is caused by mutation in the *ed1* gene (*eda* in mice). The *ed1* gene encodes a transmembrane protein called ectodysplasin, which contains an essential tumor necrosis factor (TNF) domain. Ectodysplasin binds to receptors called XEDAR and EDAR (the human homologue of the mouse *downless* gene). Receptor activation is transmitted by the adapter protein EDARADD and signals via the NFκB and the JNK pathway. These pathways mediate positive signals for cell survival, growth and differentiation. Other genes,

which encode members of the EDA pathway such as EDAR, XEDAR, EDARADD and NEMO, have also been found to be mutated in ED. The NEMO gene encodes IKK γ , which is required for the activation of the IKK complex in the NF κ B pathway.

Regulatory Mechanisms

Multipotent stem cells are fundamental for tissue repair and tissue homeostasis in the adult and have been described in a variety of tissues such as bone marrow, gut and skin (4). Stem cells have been defined as a population of undifferentiated cells that are capable of self-renewing by symmetric divisions, production of a large number of differentiated progeny by asymmetric divisions, homing to an appropriate growth environment and tissue regeneration after injury. Tissue and lineage specification of stem cells is believed to be controlled in a self-organizing, dynamically regulated process based on cell-cell and cell-environment interactions. Within their tissues, stem cells are often localized to specific stem cell niches. In the skin, multipotent epithelial stem cells have been localized to a portion of the outer root sheath of the hair follicle, termed the hair bulge (2). These bulge stem cells can give rise not only to epidermal keratinocytes, but also to hair follicular keratinocytes and cells of the sebaceous glands. Stem-cell progeny were shown to exit the bulge and migrate either upwards into the stratum basale or downwards to populate the [▶hair matrix](#) of the follicle. These two progeny tissues contain the transit-amplifying (TA) cells of the skin that undergo a limited number of divisions before they terminally differentiate. Hair follicles periodically undergo cycles of destruction and reconstruction and entry into the hair cycle is believed to be induced by signals from the [▶dermal papilla](#). The exact nature of this signal is unknown, but it has been shown that Wnt signaling is essentially involved in anagen induction and controls lineage commitment of skin stem cells (1, 5).

References

1. Fuchs E, Raghavan S (2002) Getting under the skin of epidermal morphogenesis. *Nat Rev Genet* 3:199–209
2. Paus R, Cotsarelis G (1999) The biology of hair follicles. *N Engl J Med* 341:491–497
3. Owens DM, Watt FM (2002) Contribution of stem cells and differentiated cells to epidermal tumours. *Nat Rev Canc* 3:444–451
4. Gambardella L, Barrandon Y (2003) The multifaceted adult epidermal stem cell. *Curr Opin Cell Biol* 15: 771–777
5. Huelsken J, Vogel R, Erdmann B et al (2001) β -Catenin controls hair follicle morphogenesis and stem cell differentiation in the skin. *Cell* 105:533–45

SL1

Definition

SL1, or Selectivity factor 1, is a multisubunit complex of TBP (TATA binding protein) and TAFs (TBP associated factors), which are essential in RNA polymerase I transcription in mammalian cells (in the murine system it is also known as TIF-IB).

▶RNA Polymerase I

Slavotinek Syndrome

▶Monosomy 1p36

SLE Pathogenesis, Genetic Dissection

CHUN XIE, CHANDRA MOHAN

Simmons Arthritis Research Center and Department of Internal Medicine/Rheumatology, The University of Texas Southwestern Medical Center at Dallas, Dallas, TX, USA

chandra.mohan@utsouthwestern.edu

Synonyms

Systemic lupus erythematosus; lupus

Definition

Systemic lupus erythematosus (SLE) is a chronic, systemic autoimmune disease that involves multiple organ systems, hence manifesting a wide array of clinical symptoms. The etiology of SLE remains unclear. However, it is certain that both genetic and environmental factors contribute to the development of the disease.

Characteristics

SLE is a disease predominantly affecting women, with a female:male ratio of 9:1. The prevalence is higher in African American, Afro-Caribbeans and Asians than in Caucasians (6). The hallmark of SLE is the presence of a multitude of autoantibodies, especially anti-nuclear

autoantibodies (►ANA), which in turn cause the tissue damage seen in SLE. The pathogenesis of SLE is not fully understood. However, the past decade of research has greatly deepened our understanding of this disease.

Genetics

Genetic Studies in Human SLE

Epidemiological studies have established that genetic predisposition is an essential component in SLE susceptibility. Familial aggregation of SLE has been demonstrated in several investigations. Twin studies have shown approximately 10-fold higher concordance rates among monozygotic twins than among dizygotic twins, showing the strong genetic basis of SLE. Inheritance patterns of this disease indicate it to be a complex trait that does not follow classical Mendelian inheritance. Both association and linkage studies have been applied to the genetic analysis of human SLE. In association studies, several groups of genes have been identified as potential candidate genes, including ►major histocompatibility complex (MHC; in man: ►HLA) class II, complement, Fc receptor and the genes for different cytokines or their receptors. The *HLA-DRB1* types *DR2* and *DR3* have been shown to be consistently associated with SLE in Caucasians. The association of *DR* alleles with SLE in other ethnic groups has been less consistent. Interestingly, MHC class II alleles have been associated more strongly with particular autoantibody profiles than with the disease itself. Deficiencies in components of the early classical complement activation pathway have been noted to be associated with the development of SLE. The low affinity alleles of Fc receptor II and Fc receptor III (*FcRIIa*, *FcRIIIa*, *FcRIIIb*) have also been associated with susceptibility to lupus nephritis.

Genome-wide linkage analysis has been used in the genetic analysis of murine lupus, and more recently, in human SLE as well. The advent of microsatellite-based gene mapping techniques made it feasible to “scan” the entire genome for SLE susceptibility. More than six genome-wide screens of human SLE families have thus far been published, each identifying more than 10 susceptibility loci, as recently reviewed (7). Although some of these independently identified loci coincide or overlap, others are clearly peculiar to individual studies. Out of these, 6 genomic regions meet the criteria for ‘significant’ linkage (LOD score >3.3), with supporting evidence (LOD score ≥1.0) from at least one other independent study. These loci (and corresponding candidate genes) include *1q23* (bearing *FcRIIIa*), *1q41-42* (bearing *PARP*), *2q37* (bearing *INPP5D*), *4p15-16* (bearing *CD38*, *BST1*, *NZF36*), *6p11-12* (bearing *HLA class II genes*, *C4a*, *TNF*), and *16q12-13* (bearing *NOD2*). About a dozen additional loci have also been shown to be suggestively linked to SLE in at

least 2 independent studies. Thus, it is very clear that human SLE is richly polygenic. Current efforts in several laboratories are focused on trying to identify the culprit gene(s) within these loci of interest.

Genetic Studies in Murine SLE

Mouse models have been extremely useful in genetic dissection and analysis of lupus, because of the ready availability of large numbers of animals with homogeneous genetic backgrounds (and standardized environmental conditions), which allow further genetic manipulation and functional analyses. Indeed, we have learned a great deal about the genetic basis of lupus in murine models over the past decade, as recently reviewed (2, 7). These models can be broadly divided into two main groups: “►Forward Genetic” models (also referred to as “synthetic” or genetically-mutated models), and “►Reverse Genetic” models.

Forward Genetic Studies

In genetically mutated models, disruption of a variety of regulatory genes and over-expression of other genes have been found to promote the development of lupus-like disease, often unexpectedly. Genes identified by this approach can perhaps be classified into 4 functional categories. The first group of molecules includes the complement components *C1q*, *C3*, *C4*, serum amyloid P-component (*SAP*), *CRP*, *DNase I*, etc. It is believed that the deficiency of these molecules impedes the clearance of apoptotic debris, which then serves to breach immune tolerance. The second group of genes infringes lymphocyte apoptosis, potentially resulting in lymphoproliferation and other features of lupus. These genes include *Fas*, *FasL*, *Pten*, *FLIP*, etc. The third group includes a large array of molecules that augment or modulate T- and B-lymphocyte function. Thus, the aberrant expression of *CTLA4*, *PD-1*, *PI3 kinase*, and *p21* genes has been shown to facilitate heightened T-cell proliferation and lupus-like phenotypes. Similarly, the over-expression of *Bcl-2*, *Blys*, *CD19*, *CD21*, and deficiency of *CD22*, *Lyn*, *Fyn*, *Shp-1* and *Btk* genes lead to generalized B-cell hyperactivity and other features resembling lupus. The last group of genes directly impacts the onset and progression of end-organ disease. For example, molecules such as the FcγR, adhesion molecules (e.g., ICAM-1), and chemokines (e.g., MCP-1) may impact intrinsic susceptibility to renal damage following an immunological insult.

Reverse Genetic Studies

In spontaneous lupus-prone models, dozens of genome-wide scanning studies have been performed, resulting in the mapping of more than 40 chromosomal

SLE Pathogenesis, Genetic Dissection. Table 1 Susceptibility loci for SLE as mapped in mice

Chr ¹	Locus	Position ²	Segregating Genomes	Disease Allele	Phenotype ³	Reference ⁴
1	<i>Sle1</i>	88	B6/NZM	NZM	GN	Morel <i>et al.</i> 1994
	<i>Sle1</i>	88	B6/NZM	NZM	ANA, S	Morel <i>et al.</i> 1999
	<i>Nba2</i>	92	NZB/NZW/SMj	NZB	GN	Drake <i>et al.</i> 1995
		52	NZW/Balb/c	Balb/c	ANA	Vyse <i>et al.</i> 1996
	<i>Nba2</i>	94	NZB/B6/Balb/c	NZB	GN	Rozzo <i>et al.</i> 1996
	<i>Lbw7/Sbw1</i>	61	NZB/NZW	NZB	ANA, S	Kono <i>et al.</i> 1994
		49-100	BXSB/B10	BXSB	ANA, GN, S	Haywood <i>et al.</i> 2000
	<i>Bxs1, 2, 3</i>	32-100	BXSB/B10	BXSB	ANA, GN, S	Hogarth <i>et al.</i> 1998
	<i>Swrl-1</i>	86	SWR/NZB	SWR	ANA	Xie <i>et al.</i> 2001
3	<i>Sles3</i>	26	B6/NZW	NZW	GN	Morel <i>et al.</i> 1999
	<i>Bxs5</i>	31	BXSB/B10	B10	ANA, GN, S	Haywood <i>et al.</i> 2000
		63	BXSB/B10	BXSB	ANA, GN	Hogarth <i>et al.</i> 1998
4	<i>Sle2</i>	47	B6/NZM	NZM	ANA, GN	Morel <i>et al.</i> 1994
	<i>Sles2</i>	55	B6/NZW	NZW	ANA, GN	Morel <i>et al.</i> 1999
	<i>Nba1</i>	69	NZB/NZW/SMj	NZB	GN	Drake <i>et al.</i> 1995
	<i>Nba1</i>	72	NZB/NZW/SMj	NZB	GN	Vyse <i>et al.</i> 1996
		50	NZB/B6/Balb/c	NZB	GN	Rozzo <i>et al.</i> 1996
	<i>Lbw2/Sbw2</i>	53	NZB/NZW	NZB	GN, M, S	Kono <i>et al.</i> 1994
	<i>Lmb1</i>	54	MRL/B6	B6	ANA, S	Vidal <i>et al.</i> 1998
		47	BXSB/B10	BXSB	S	Hogarth <i>et al.</i> 1998
5	<i>Sle6</i>	15	B6/NZW	NZW	GN	Morel <i>et al.</i> 1999
	<i>Lbw3</i>	88	NZB/NZW	NZW	GN, M	Kono <i>et al.</i> 1994
	<i>Lmb2</i>	27	MRL/B6	MRL	ANA, S	Vidal <i>et al.</i> 1998
6		64	B6/NZW	B6	ANA	Morel <i>et al.</i> 1999
	<i>Lbw4</i>	60	NZB/NZW	NZB	M	Kono <i>et al.</i> 1994
		44	MRL/B6	B6	GN	Vidal <i>et al.</i> 1998
7	<i>Sle3</i>	4, 28	B6/NZM	NZM	GN	Morel <i>et al.</i> 1994
	<i>Sle3, 5</i>	4, 28	B6/NZM	NZM	ANA, GN	Morel <i>et al.</i> 1999
	<i>Sle3</i>	25	B6/NZW	NZW	ANA, GN	Morel <i>et al.</i> 1999
	<i>Nba3</i>	37	NZB/NZW/SMj	NZB	GN	Drake <i>et al.</i> 1995
		52	NZB/NZW/SMj	NZB	ANA	Vyse <i>et al.</i> 1996
	<i>Lbw5</i>	22	NZB/NZW	NZW	M	Kono <i>et al.</i> 1994
	<i>Lrdm1</i>	4	MRL/C3H	MRL	GN	Watson <i>et al.</i> 1992
	<i>Lmb3</i>	26	MRL/B6	MRL	ANA, S	Vidal <i>et al.</i> 1998
		25	NZW/B6.Yaa	NZW	ANA, GN	Santiago <i>et al.</i> 1998
9	<i>Baa1</i>	23	NZW/Balb/c	Balb/c	ANA	Vyse <i>et al.</i> 1996
10		69	B6/NZM	NZM	GN	Morel <i>et al.</i> 1999

SLE Pathogenesis, Genetic Dissection. Table 1 Susceptibility loci for SLE as mapped in mice (Continued)

Chr ¹	Locus	Position ²	Segregating Genomes	Disease Allele	Phenotype ³	Reference ⁴
	<i>Sles4</i>	2	B6/NZW	NZW	GN	Morel <i>et al.</i> 1999
		25	NZB/NZW/SMj	NZB	GN	Drake <i>et al.</i> 1995
	<i>Lmb4</i>	50	MRL/B6	MRL	S	Vidal <i>et al.</i> 1998
		30	BXSB/B10	B10	ANA	Hogarth <i>et al.</i> 1998
11	<i>Swrl-4</i>	25	SWR/NZB	SWR	ANA	Xie <i>et al.</i> 2001
		20	B6/NZM	NZM	ANA, GN	Morel <i>et al.</i> 1999
		25	NZW/Balb/c	NZW	ANA	Vyse <i>et al.</i> 1996
		17, 43	NZB/B6/Balb/c	NZB	ANA	Rozzo <i>et al.</i> 1996
12		38	NZB/NZW	NZB	ANA	Kono <i>et al.</i> 1994
		25	NZW/Balb/c	NZW	ANA	Vyse <i>et al.</i> 1996
	<i>Lrdm2</i>	27	MRL/C3h	MRL	GN	Watson <i>et al.</i> 1992
		50	NZB/NZW/SMj	NZB	GN	Drake <i>et al.</i> 1995
13		29	BXSB/B10	BXSB	ANA	Haywood <i>et al.</i> 2000
	<i>Swrl-2</i>	36	SWR/NZB	SWR	GN	Xie <i>et al.</i> 2001
14		32	NZB/NZW/SMj	NZW	ANA, GN	Vyse <i>et al.</i> 1996
16	<i>Sle4</i>	19	B6/NZM	NZM	GN	Morel <i>et al.</i> 1994
	<i>Sles1</i>	19	B6/NZW	NZW	ANA, GN	Morel <i>et al.</i> 1999
		18	NZB/NZW/SMj	NZB	GN	Vyse <i>et al.</i> 1996
		19	NZW/Balb/c	Balb/c	ANA	Vyse <i>et al.</i> 1996
17	<i>Lbw1</i>	19	NZB/NZW	NZB/W	ANA, GN, M	Kono <i>et al.</i> 1994
		6	NZW/B6.Yaa	NZW	ANA	Santiago <i>et al.</i> 1998
		19	SWR/NZB	SWR	ANA, GN, M	Xie <i>et al.</i> 2001
		49	NZB/NZW	NZW	GN, M	Kono <i>et al.</i> 1994
18	<i>Swrl-3</i>	14	SWR/NZB	SWR	ANA	Xie <i>et al.</i> 2001
19		38	NZN/NZW/SMj	NZB	ANA	Vyse <i>et al.</i> 1996
		38	NZW/Balb/c	NZW	ANA	Vyse <i>et al.</i> 1996
	<i>FAS</i>	44	MRL/C3H		GN	Watson <i>et al.</i> 1992

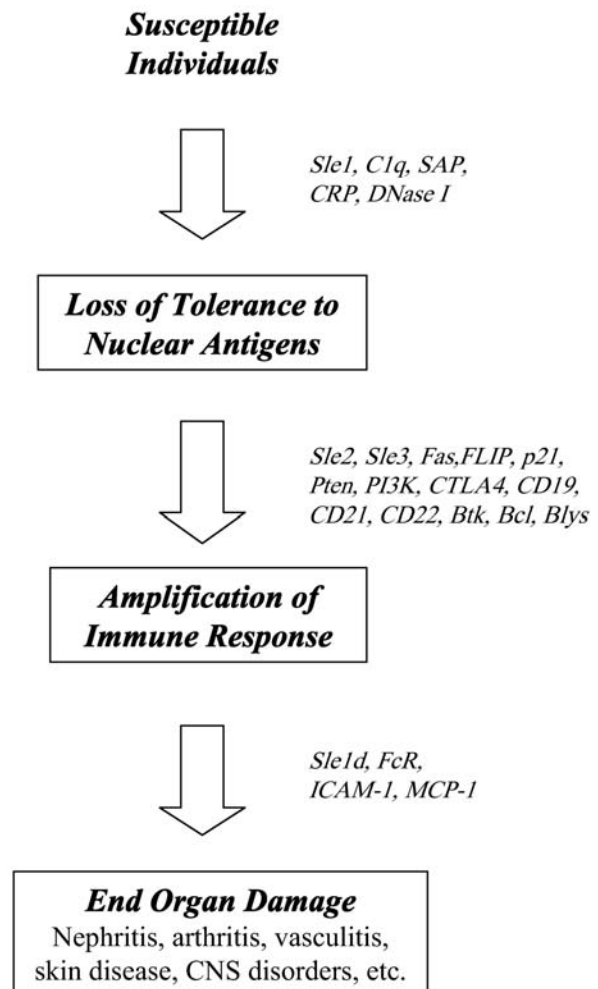
¹Chromosomal location of locus²Chromosomal position (in cM) from the acromere³Linked phenotypes: ANA, antinuclear autoantibodies; GN, glomerulonephritis; M, early mortality; S, splenomegaly⁴Please refer to reference #2 for complete citations of these studies

intervals associated with lupus susceptibility, as summarized in Table 1, and reviewed elsewhere (2, 7). These data have originated from studies using the NZB/NZW and the related NZM2410 strains, as well as the MRL/*lpr*, BXSB, and SWR/NZB strains. These lupus susceptibility loci appear to be randomly

distributed over all 19 chromosomes, indicating that murine lupus is also richly polygenic. Although some of these susceptibility loci have been repeatedly mapped to certain “hotspots” on chromosomes 1, 4, 7, and 17, the majority of detected loci appear to be strain-specific. Whereas the majority of the susceptibility

loci have originated from lupus-prone strains, a few loci have arisen from “normal” (non-autoimmunity prone) genomes. These studies have also demonstrated a “many to many” relationship between the different component lupus phenotypes and the susceptibility loci. Thus, for instance, *Sle1* on chromosome 1 has been demonstrated to be associated with production of autoantibodies, development of GN, as well as splenomegaly; on the other hand, *Nba2* on chromosome 1, *Sle3* on chromosome 7 and *Swrl-2* on chromosome 14 are all associated with the development of GN (2). In addition to these disease ▶susceptibility loci, several lupus resistance (or “suppressor”) loci have also been identified, including *Sles1*, *Sles2*, *Sles3*, and *Sles4*, all of NZM2410/NZW origin (3). It thus appears that the eventual disease outcome in any given strain or individual is the consequence of the interplay between a suite of lupus-accelerating or lupus-attenuating genes, impacting the immune system in apparently opposite fashions.

One of the strengths of using mouse models of lupus is the ability to perform “▶congenic dissection”. Essentially, in this approach, the researchers genetically transfer one or more of the susceptibility loci detected in the original mapping onto “normal”, non-autoimmune mouse backgrounds, in order to transform the polygenic model system into a series of monogenic disease models. This strategy has allowed researchers to analyze the function of each individual susceptibility locus thoroughly by phenotyping the corresponding mono▶congenic strains, and to further narrow the genetic intervals (to an interval that is small enough to allow physical cloning of the culprit genes) through the breeding and analysis of congenic recombinants. Congenic dissection of murine lupus has been extensively used in the analysis of disease in the NZM2410 model (2, 7). Thus, the three major susceptibility loci, *Sle1*, *Sle2*, and *Sle3*, have been successfully bred onto the non-autoimmune C57BL/6 (B6) background, and the phenotypes of each congenic strain have been thoroughly analyzed. B6 mice congenic for *Sle1* exhibit autoantibodies to H2A/H2B/DNA subnucleosomes, with minimal renal disease (1, 2). B6 mice bearing *Sle2*, on the other hand, exhibit B-cell hyperactivity with polyreactive IgM antibodies and an expansion of B1 cells, but without renal involvement (2). In contrast, *Sle3* leads to T-cell hyperactivation, modest sero-reactivity to several nuclear antigens, decreased activation-induced cell death, and moderate nephritis (2). Recently, fine mapping of the *Sle1* interval using congenic recombinants has revealed that *Sle1* consists of a cluster of several distinct loci, including *Sle1a*, *Sle1b*, and *Sle1c*, that independently breach tolerance to chromatin and



SLE Pathogenesis, Genetic Dissection. Figure 1 A model of SLE pathogenesis demonstrating how the epistatic interaction of several different lupus susceptibility genes might orchestrate disease.

Sle1d that facilitates GN (4). It is also intriguing to note that some of the lupus susceptibility loci identified, such as the loci on mid-chromosome 3, telomeric chromosome 6, proximal chromosome 7, mid-chromosome 10, and the *H2* locus on chromosome 17, appear to co-cluster with susceptibility loci for other autoimmune diseases, including murine diabetes, experimental autoimmune encephalomyelitis and collagen-induced arthritis (2). Importantly, some of the identified murine loci appear to be syntenic with some of the human intervals linked to SLE (7). Thus, NZM2410-derived *Sle1a* and *Sle1b* are syntenic with *1q21-23*, and *Sle1d* with *1q41*, on human chromosome 1. ▶Epistasis of different disease susceptibility loci plays a very important role in the genesis of lupus. This has

been well illustrated in the B6.*Sle* congenic models. As mentioned above, none of the B6.*Sle* monocongenic strains develop severe pathology. In contrast, when both *Sle1* and *Sle3* were transferred onto the B6 background, the resulting bicongenic strain developed high titers of nephrophilic anti-dsDNA antibodies, severe GN, and splenomegaly (2). Likewise, the triple congenic B6.*Sle* strain bearing *Sle1*, *Sle2*, and *Sle3* also developed severe systemic autoimmunity and fully penetrant, fatal GN (5). Thus, the “fully developed” version of lupus may require the combinatorial participation of several different genes, each impacting a different pathogenic step. Figure 1 illustrates how different genes may impact several of the key steps leading to the development of SLE. In the first step, genes such as *Sle1*, *C1q*, *SAP* and *DNase I* may act to initiate autoimmunity by breaching tolerance to nuclear antigens, leading to the formation of non-pathogenic autoantibodies. *Sle2*, *Sle3* and yet other genes may then serve to amplify the ongoing autoimmune response, leading to generalized T- and B-cell hyperactivation and the production of pathogenic, nephrophilic autoantibodies. The final step may involve genes such as *Sle1d*, *FcR*, *ICAM-1* (and possibly other adhesion molecules), *MCP-1* (and possibly other chemokines), that facilitate immune-mediated end organ injury. Recent advances in the fields of gene expression microarrays and proteomics are likely to add further dimensions to our understanding of lupus in the coming decade.

Clinical Relevance

Currently, the diagnosis of SLE is mainly based on clinical manifestations of the disease and immunological evaluation. Molecular diagnostics (and gene/molecule-targeted therapy) will be possible in the future as we uncover the genetic and molecular footprints of lupus.

References

1. Mohan C, Alas E, Morel L et al (1998) Genetic dissection of SLE pathogenesis. *Sle1* on murine chromosome 1 leads to a selective loss of tolerance to H2A/H2B/DNA subnucleosomes. *J Clin Invest* 101:1362–1372
2. Mohan C (2001) Murine Lupus Genetics: Lessons Learned. *Curr Opin Rheumatology* 13:352–360
3. Morel L, Tian XH, Croke BP et al (1999) Epistatic modifiers of autoimmunity in a murine model of lupus nephritis. *Immunity* 11:131–139
4. Morel L, Blenman KR, Croke BP et al (2000) The major murine systemic lupus erythematosus locus, *Sle1*, is a cluster of functionally related genes. *Proc Natl Acad Sci USA* 98:1787–1792

5. Morel L, Croker BP, Blenman KR et al (2000) Genetic reconstitution of systemic lupus erythematosus immunopathology with polycongenic murine strains. *Proc Natl Acad Sci USA* 97:6670–6675
6. Rus V, Hochberg MC (2001) The epidemiology of systemic lupus erythematosus. In: Wallace DJ, Hahn BH (eds) *Dubois's lupus erythematosus*. Williams and Wilkins, Baltimore, p 65–83
7. Wakeland EK, Liu K, Graham RR et al (2001) Delineating the genetic basis of systemic lupus erythematosus. *Immunity* 15:397–408

Sleep Apnea

Definition

Sleep apnea is a breathing disorder characterized by brief interruptions of breathing during sleep.

► [Prader Willi and Angelman Syndromes](#)

Sleep Paralysis

Definition

Sleep paralysis is a phenomenon occurring rarely in healthy individuals and frequently in narcoleptic patients. Sleep paralysis is characterized by an inability to move, which either occurs at the transition from waking to sleeping or (more frequently) at the transition from sleep to wakefulness. In this state consciousness is preserved and the inability to move is felt as extremely unpleasant. Episodes might last several minutes.

► [Narcolepsy](#)

Slippery Heptamer

Definition

Slippery heptamer denotes a seven nucleotide sequence whose primary sequence allows base paired tRNAs to slip into an alternative reading frame (either +1 or –1).

► [Translational Frameshifting, Non-standard Reading of the Genetic Code](#)

SLOS

- Smith-Lemli-Opitz Syndrome

SMA

- Spinal Muscular Atrophy

Slow and Fast Muscle Fibres

Definition

Slow and fast muscle fibres are fibres that perform contractions of different speed and are innervated by nerves firing with different frequencies. Slow fibres are fuelled by oxidative phosphorylation, fast fibers by glycolysis.

- Muscle Development

Smac

Definition

Smac is an acronym for “second mitochondria derived activator of caspase”. It blocks IAP (inhibitor of apoptosis protein) activity, allowing activation of caspases by the apoptosome.

- Apoptosis, Regulation and Clinical Implications

Slow Exchange Limit

Definition

Slow exchange limit describes an interaction where the life-time of a ligand in a molecular complex is large, compared to the time-scale defined by the chemical shift difference of a nuclear resonance in the free and in the bound state.

- Protein-Ligand-Interaction by NMR

Smad

Definition

The term Smad is an amalgamation of the *C. elegans* small worm phenotype (Sma) with the *D. melanogaster* mother against decapentaplegic (Mad) phenotype. Smads are intracellular signal transducing proteins and participate in regulation of gene expression as transcription factors.

- Receptor Serine/Threonine Kinases
- Transcription Factors and Regulation of Gene Expression

Slow Muscle Fibres

- Slow and Fast Muscle Fibres

Smad Anchor for Receptor Activation

- SARA

SLT

- Specific Locus Test

Smad Ubiquitin Regulatory Factor

- Smurf

Small Dendritic RNA

Definition

Small dendritic RNA refers to a cytoplasmic RNA expressed in a highly defined neural and subcellular distribution in the brain. This RNA is also known as BC1, BC200 and BCYRN1. It regulates protein synthesis in the neuronal dendrites without being translated into protein.

► [Fragile X Syndrome](#)

Gene Function in Mammalian Cells

- [Micro RNA's](#)
- [Morpholino Oligonucleotides, Functional Genomics by Gene 'Knock-down'](#)
- [RNA Interference in Mammalian Cells](#)

Small GTPases

Definition

Small GTPases is the name for a family of small guanine nucleotide-binding proteins including ras and rho GTPases, which cycle between an inactive GDP-bound state and an active GTP-bound state, through the activity of specific ► [guanine nucleotide exchange factors](#) (GEFs) or GTPase activating proteins (GAPs). Small GTPases are involved in the regulation of several cellular activities such as: ► [cell division](#), motility, migration, proliferation and invasion.

► [Cancer Invasion and Metastasis: Cellular, Molecular and Clinical Aspects](#)

► [Focal Complexes/Focal Contacts](#)

► [Microvilli](#)

► [Rho, Rac, Cdc42](#)

► [RNA Interference in Mammalian Cells](#)

Small Intragenic Mutations

Definition

Small intragenic (subtle) mutations are base substitutions and small deletions/insertions in a given gene. Small intragenic mutations in the SMN 1 (survival motor neuron) gene have been associated with spinal muscular atrophy.

► [Spinal Muscular Atrophy](#)

Small Interfering RNA

Definition

Small interfering RNA (siRNA; short interfering RNA) refers to double stranded RNAs of 21–23 nucleotides (nts), which are generated from long dsRNA with ~2 nt 3' overhangs. They can mediate sequence-specific inhibition of gene expression in mammalian cells via a posttranscriptional gene silencing mechanism, and act as guide molecules to specify the cleavage of mRNAs during ► [RNA interference](#) (RNAi).

► [Catalytic RNA](#)

► [Functional Assays](#)

► [High-Throughput Approaches to the Analysis of](#)

Definition

Small molecule inhibitors are chemical compounds that bind the active site of a protein, and inhibit either enzyme activity or protein-protein interactions. Much effort is going into screening large libraries of compounds as leads for drug development.

► [Methylation of Proteins](#)

Small Vessel Vasculitis

► [Morbus Wegener](#)

SMART Database

Definition

SMART database stands for Simple Modular Architecture Research Tool. This database allows the identification and annotation of genetically mobile domains, and the analysis of domain architectures.

- [Functional Assays](#)
- [Protein Databases](#)

SMase

- [Sphingomyelinase](#)

SMC Proteins

Definition

Abbreviation for Structural Maintenance of Chromosomes. The SMCs are known to be present in species from mycoplasma to mammals, and they constitute a family of conserved and ubiquitous proteins called the SMC family (with 6 subfamilies). The proteins have two extended coiled-coil domains, which are separated by a central globular hinge region. These proteins, when dimerised, have ATPase activity by bringing together the N- (ATP binding) and C- (ATP hydrolysis) termini.

- [Chromosome Condensation](#)

Smith-Lemli-Opitz Syndrome

Definition

Smith-Lemli-Opitz syndrome (SLOS) is an autosomal recessive disorder characterised by multiple congenital anomalies including mental retardation, facial and brain abnormalities, and limb, heart, renal, genital and skeletal defects. SLOS is due to mutations in the gene encoding 7-dehydrocholesterol reductase, the last enzyme of cholesterol biosynthesis.

- [Hedgehog Signaling](#)

Smith-Magenis Syndrome

Definition

Smith-Magenis syndrome (SMS) is a contiguous gene syndrome (1:25000 newborns) that is caused by hemizygous deletion at human chromosome 17p11.2–p12, and characterized by behavioural abnormalities such as inverse sleep pattern (long nocturnal wake periods creating profound intrafamilial problems, and short daytime naps), autoaggression and self-injurious behaviors (presumably due to decreased sensitivity to pain). Facial dysmorphism is mild and growth is normal.

- [Microdeletion Syndromes](#)

SMN-Complex

Definition

The Survival of Motor Neurons (SMN) protein, the product of the spinal muscular atrophy-determining gene, is part of a large macromolecular complex (SMN complex). The SMN complex plays a critical role in the assembly of spliceosomal small nuclear ribonucleoproteins (snRNPs) and possibly other RNPs, in pre-mRNA ► [splicing](#), and in the assembly of mRNA factories (transcriptosomes).

Smoothened

Definition

Smoothened characterises a component of the hedgehog receptor signaling complex and is similar in structure to ► [Frizzled](#).

- [Wnt/Beta-Catenin Signaling Pathway](#)

SMS

- [Smith-Magenis Syndrome](#)

Smurf

Definition

Smurf is a ►**Smad** ubiquitinylation regulatory factor that functions as an E3 ligase, which contains a HECT domain that specifically interacts with Smads and mediates their ubiquitinylation and proteasomal degradation. Two isoforms, Smurf1 and Smurf2, have been identified, and in addition to Smads they ubiquitinylate the TGF- β receptor complex.

►**Receptor Serine/Threonine Kinases**

SNARE Proteins

Definition

SNARE (soluble *N*-ethylmaleimide-sensitive factor (NSF) attachment protein receptors) proteins constitute an evolutionarily conserved super-family of small, coiled-coil membrane proteins (21 members in yeast and over 35 in mammals) with compartment-specific distribution patterns. SNAREs assemble as cognate pairs into specific complexes that drive fusion between appropriate lipid bilayers and exocytosis of intracellular vesicles. SNAREs constitute the essential minimal membrane-fusion machinery in eukaryotic cells. Most of them are attached to the membrane via a single, carboxy terminal transmembrane domain or posttranslational modification. The amino terminal cytosolic domain contains a conserved SNARE motif, a stretch of 60–70 amino acids organized in eight heptad repeats, which form α -helices and engage in the formation of coiled-coil structures.

►**Cell Polarity**

►**Epithelial Cells**

►**Protein and Membrane Transport in Eukaryotic Cells**

►**Vesicular Traffic**

SNP

►**Single Nucleotide Polymorphism**

►**SNP Detection and Mass Spectrometry**

SNP Detection and Mass Spectrometry

SASCHA SAUER

Max-Planck-Institute for Molecular Genetics, Berlin, Germany

sauer@molgen.mpg.de

Definition

Single Nucleotide Polymorphisms

Single nucleotide ►**polymorphisms** (►**SNPs**) are single-base changes that occur at a specific position in a genome. Very often this definition also comprises single-base insertions and deletions (so called “indels”). SNPs are the most abundant polymorphisms in the human genome, accounting for more than 90% of all sequence variations. In ►**diploid** species, SNPs are usually ►**biallelic** polymorphisms but have the potential to be tri- or even tetra-allelic. If in a general population the abundance of the less frequent allele of a base-change is not lower than 1% it is defined to be an SNP. Otherwise it is termed ►**point mutation**.

SNPs feature low mutation rates and have been estimated to occur in the human genome at a frequency of one in every 300–1000 nucleotides. In contrast to other ►**DNA markers** such as ►**STRs** (short tandem repeats), SNPs are more stable over generations. Due to the binary nature of SNPs, they can be more easily analysed than complex STRs. In order to achieve a maximum of information from DNA markers such as SNPs, the ►**haplotype** of an individual – i.e. the phase of the ►**alleles** of several SNPs on a single chromosome – is considered to be important for genetic studies.

Most SNPs occur in non-coding regions of the human genome. The frequency of SNPs in these regions is about fourfold higher than in ►**coding regions**. About 3% of the human genome refers to coding DNA. Given the size of the haploid human genome with 3×10^9 base pairs and one SNP on average per 300 nucleotides, there are approximately ten million SNPs, of which ca. 333,333 are expected to occur in coding sequences. Only half of these SNPs result in an amino acid change, which possibly represents 166,666 amino acid variations between two individuals.

Over half of the SNPs occurring in coding regions are ►**synonymous codon changes**; thus they do not affect protein biosynthesis. In humans, the C \rightarrow T/G \rightarrow A transition is the most common variant, representing about two-thirds of all SNPs. This striking abundance reflects the well-known ►**deamination** of 5-methylcytosine to thymine.

SNP Discovery

About 1.4 million human SNPs stored in public databases have been discovered in the course of the human genome sequencing effort. In this project several individuals were sequenced in parallel. Due to natural diversity the respective sequences differ slightly from one individual to the other. As a result of possible experimental errors and due to frequency changes of DNA markers from one population to the other, SNPs have always to be validated in the population to be studied. Furthermore, it quite often happens that no SNPs are available in the genomic region of interest. In this case novel SNPs have to be identified. In general, screening for SNPs comprises experimental approaches and ►*in-silico* procedures. Usually, SNP discovery is performed experimentally by DNA sequencing using a number of DNA samples in different populations. Although considered to be the “gold standard” for sequence determination, the ►Sanger sequencing method is prone to error production (as was observed in sequencing projects). Therefore experimental results must be confirmed, which is commonly done by screening polymorphisms in genomic sequences from different individuals using software analysis. This includes likelihood estimation of a particular locus on a genome being polymorphic. In this way true polymorphisms and sequencing errors can be distinguished.

SNP ► Genotyping

Genotyping refers to the assignment of different variants in an otherwise conserved DNA region. Gene variants can be associated with phenotypes such as ►complex diseases. ►Association studies were proposed as a suitable approach in those cases, where traditional family-based ►linkage analyses fail presumably because of many contributing loci. In practice, many SNPs from selected populations of affected and non-affected (control) individuals are scored. This is called case-control study design. A significant frequency change of typed SNPs indicates an association between these SNPs and the complex trait under investigation.

Mass Spectrometry

In principle, ►mass spectrometry embodies a very attractive solution for the detection of nucleic acids as it provides direct and fast analysis. The main advantage of mass spectrometry over other DNA characterisation techniques is its signal accuracy. Signals obtained reflect the molecular weight, a physical and intrinsic property, of an analyte. Two main inventions have paved the way for mass spectrometry in life sciences, matrix-assisted laser ►desorption/►ionisation time-of-flight mass spectrometry (►MALDI-TOF-MS) and electrospray ionisation mass spectrometry (►ESI-MS).

By means of either technology femtomolar amounts of nucleic acids can be detected. Besides these two methodologies alternative mass spectrometric procedures have been applied for the detection of SNPs, such as atmospheric pressure chemical ionisation (►APCI).

Characteristics

Matrix-Assisted Laser Desorption/Ionisation Mass Spectrometry

Franz Hillenkamp and Michael Karas discovered that irradiation of crystals formed by suitable small, acidic organic molecules, termed “matrix”, with a pulsed laser at a wavelength close to a resonant absorption band of these molecules induces an energy transfer and desorption process, evaporating matrix ions into the gas-phase. By embedding biomolecules such as peptides, proteins or short nucleic acids into the crystalline structure of matrix molecules, the biomolecules are co-desorbed into the gas-phase by fast heating. Ionisation occurs upon irradiation with the laser. However, the ionisation process in MALDI remains to be elucidated. Predominantly either positively or negatively single charged molecules are detected by MALDI. These ions are produced by a proton-transfer reaction between matrix and analyte molecules in the gas-phase. Ions are accelerated in an electric field. Usually, MALDI-MS is performed with simple and fast time-of-flight separation (MALDI-TOF-MS). By means of ion optics, molecules with (more or less) the same kinetic energy are guided into a flight-tube where they are separated according to their mass to charge ratio (m/z) before they finally reach the detector. The velocity of molecules is inversely proportional to the square root of their m/z . Molecules with a higher m/z ratio reach the detector later than molecules with a lower m/z .

Initially, MALDI was predominantly applied in protein chemistry. More recently it has also been employed for nucleic acid research. For the analysis of nucleic acids the main problem is the negatively charged sugar-phosphate backbone. In solution phosphate residues constitute a site of negative charge. Each oligonucleotide carries as many negative charges as phosphate residues. The affinity of phosphate residues for counterions such as sodium and potassium is high, but unfortunately not high enough to result in complete saturation. Metal ions cause adduct formation with phosphate linkages thereby significantly reducing the signal quality of nucleic acids. As a consequence purification strategies have been included in sample preparation protocols. Most popular procedures, including reversed-phase column purification, magnetic bead technology and gel filtration, are expensive and cumbersome for automation.

The acid instability of DNA poses another problem for MALDI analysis. Efficient MALDI matrices are

usually acidic. However in the gas-phase, long DNA molecules readily fragment in the presence of such matrices.

Several DNA characteristics have been observed and exploited for improved MALDI detection. ▶**Depurination** has been found for larger DNA oligomers. Replacing natural purines by 7-deaza-analogues is an efficient approach to overcome DNA degradation. Furthermore, ribonucleotides containing 2'-OH groups are more stable than DNA in the gas-phase. A third approach using chemical modification to improve MALDI detection of nucleic acids is named ▶**charge tagging**. Charge-tagged DNA carries either a single positive or a single negative charge. In this way MALDI detection sensitivity is increased about 100-fold compared to unmodified DNA and the susceptibility to ion adduct formation is significantly decreased.

The optimisation of the MALDI process also includes the identification of the most suitable matrix and sample preparation method. Three different matrix preparation methods are generally applied in MALDI depending on the matrix used; thin-layer (using for example α -cyano-4-hydroxy-cinnamic acid or derivatives as matrix), dried droplet preparations (using for example 3-hydroxypicolinic acid (HPA) as matrix) and Sequenom's SpectroChips™ (▶www.sequenom.com). For thin-layer preparations, matrix is spread over the MALDI target plate in a volatile solvent such as acetone. The solvent evaporates immediately, leaving behind a thin layer of small matrix crystals. Subsequently sample is dispensed onto the thin-layer in a solvent that does not dissolve the matrix crystals. Analyte molecules co-crystallise into the surface of matrix. Hence during laser desorption analytes are desorbed approximately equally all over the spot. Due to the limited thickness of the matrix layer, a better mass accuracy can be achieved than by dried-droplet preparations. However, efficient matrices for DNA analysis such as HPA can only be used for dried-droplet preparation because of their crystallisation behaviour. For dried droplet preparations, matrix and analyte solution are mixed and spotted onto the MALDI target plate. The mass accuracy for dried droplet preparations tends to be less stable than that for thin-layer preparations because of the uneven height of the matrix layer. As MALDI analysis is based on the determination of the time-of-flight of an ion, a variation of matrix layer height results in a shift in the starting point of desorption. Furthermore, certain positions on a dried droplet preparation tend to give rise to better results than others (so-called "sweet spots"). The disadvantages of dried droplet preparation can be overcome by using SpectroChips. Nanoliter quantities of analyte are transferred onto tiny spots of dried matrix (usually 3-hydroxypicolinic acid) deposited on silicon chips. As the entire preparation is volatilised with a few laser shots,

spot-to-spot reproducibility increases significantly compared to classical dried droplet preparations.

Electrospray Ionisation Mass Spectrometry

Electrospray ionisation mass spectrometry (ESI-MS) was mainly developed by John Fenn. It has become an important method in life sciences. Like MALDI-MS it has been used successfully for the analysis of biomolecules. ▶**Chromatographic procedures** are a preferred option for sample purification. ESI-MS can be coupled with ▶**capillary electrophoresis** or ▶**HPLC**. ESI is currently one of the gentlest methods for ionising molecules used in mass spectrometry, even gentler than conventional MALDI. In the process of ESI, charged droplets are generated at atmospheric pressure by spraying sample solution into an electric field. Analyte molecules are ionised in these droplets leaving the ion source largely intact. In addition to the time-of-flight separation also commonly used for MALDI, for ESI, ▶**quadrupoles** and/or ▶**ion traps** are used to guide the analyte molecules to the detector efficiently. In combination with ▶**tandem mass spectrometry** (MS/MS), which is based on the gas-phase collision induced fragmentation of biomolecules, structural and sequence information can be obtained.

ESI-MS can be used to detect intact ▶**PCR** products of up to 120 bases. Particularly in this mass range, its accuracy is better than that of MALDI-MS because of the multiple charge states of the analyte molecules used to determine their molecular mass. A number of procedures for the analysis of SNPs have been described in the past few years but most of them have not (yet) become mature technologies. An inherent drawback of ESI is that due to analyte injection throughput is severely limited. With one ESI mass spectrometer, several hundreds of samples could be analysed per day as routine, which is far too few for efficient screening or genotyping of SNPs. Therefore at present it is inconceivable that ESI will replace MALDI for DNA analysis in SNP discovery or in SNP typing applications.

SNP Discovery

Standardized DNA re-sequencing that is based on the procedure developed by Sanger and co-workers is currently the most straightforward approach to detect novel polymorphisms. Mass spectrometry was proposed as an alternative to gel- or capillary-based analysis of large DNA sequencing products (500–1000 bp). Among the procedures for sequencing DNA by mass spectrometry, most combined Sanger sequencing with MALDI-MS. However, only short sequences of up to 50 bases could be routinely detected, which impedes applications such as genome sequencing and efficient SNP screening.

Alternative (re-)sequencing approaches were introduced making use of base-specific fragmentation

patterns of genomic sequences of interest. For example workers from Methexis Genomics and Sequenom invented a method in which single-stranded RNA transcripts are produced from both strands of a PCR product and the resulting two RNAs are fragmented by the use of ►**RNAases**, for instance by G-specific RNase T1. Fragments are analysed by MALDI and signals are compared to values calculated from *in silico* digestions of reference sequences. Differences in expected masses can be identified and used for the detection of unknown SNPs and unambiguous sequence assignment.

SNP Genotyping

With the increased popularity of using SNPs in genetic analysis, the possibility of employing mass spectrometry for SNP typing has been extensively explored. After an initial step of complexity reduction and amplification of the genomic DNA, most (mass spectrometric) methods for SNP genotyping apply primer extension for allele querying. Other methodologies for the generation of allele-specific products include oligonucleotide hybridisation, oligonucleotide ligation and allele-specific cleavage of oligonucleotides.

The primer extension reaction is the most widely applied molecular biological procedure for SNP analysis. Primer extension is robust and flexible. What is more important for mass spectrometric analysis, it generates fairly small products. In a primer extension reaction, a ►**DNA polymerase** extends a primer upstream of the SNP of interest with a set of ►**dNTPs** and/or ►**ddNTPs** on a PCR product, resulting in allele-specific products. The DNA polymerase extends the 3'-end of the primer by specifically incorporating nucleotides that are complementary to the DNA template. The extension reaction terminates at the first ►**nucleobase** in the template where a nucleotide occurs that is complementary to one of the ddNTPs in the reaction mix. Generally, a thermostable DNA polymerase is used in a temperature-cycled reaction resulting in a linear amplification. Products of the primer extension reaction are detected by mass spectrometry. The different masses of the DNA products detected represent the different alleles. Most methods use MALDI detection to allow high-throughput analysis. Prominent assays are the MassEXTEND assay (formerly called PROBE assay), the PinPoint assay, the VSET assay, the GenoSNIP assay and the GOOD assay.

Clinical Relevance

The use of SNPs has attracted most attention to two major applications. Firstly, SNPs can help geneticists to trace sequences associated with susceptibility to

complex diseases. Secondly, candidate gene association studies can be a powerful tool for the development of personalised medicine. This area is called ►**pharmacogenomics**. Very often, drugs interact with proteins such as metabolising enzymes or receptors, not taking into account sequence variations within populations. Patients can therefore respond to different extents to the same drug. In the future, based on the results of pharmacogenomic research, physicians could routinely screen for particular SNPs and thereby tailor drug treatments to each patient's individual genetic background. Mass spectrometry is a powerful method to detect SNPs for clinical studies.

References

1. Kirpekar F, Nordhoff E, Larsen LK et al (1998) DNA sequence analysis by MALDI mass spectrometry. *Nucleic Acids Res* 26:2554–2559
2. Sauer S, Lechner D, Berlin K et al (2000) A novel procedure for efficient genotyping of single nucleotide polymorphisms. *Nucleic Acids Res* 28:e13
3. Gut IG (2001) Automation in genotyping of single nucleotide polymorphisms. *Hum Mutat* 17:475–492
4. Huber CG, Oberacher H (2001) Analysis of nucleic acids by on-line liquid chromatography-mass spectrometry. *Mass Spectrom Rev* 20:310–343
5. Sauer S, Gut IG (2002) Genotyping SNPs by MALDI Mass Spectrometry. *J Chromatogr B Analyt Technol Biomed Life Sci* 782:73–87

SOC

►Store-Operated Calcium Entry

SOD

►Superoxide Dismutase

Solid- or Split-Pin

Definition

Solid- or split-pin refers to two types of rigid metallic pins that are used in arraying to deliver in a contact-fashion a

droplet to the surface. Split pins contain a sample reservoir which is filled by capillary forces when the pin is dipped into the sample.

► [Protein Microarrays as Tool for Protein Profiling and Identification of Novel Protein-Protein Interactions](#)

Solubilization

Definition

A biological membrane is solubilized by interaction with a detergent. Ultimately, individual components of the membrane, i.e. proteins and lipids, are incorporated into individual detergent micelles.

► [Two-Dimensional Crystallization of Membrane Proteins](#)

Solvation Potential

Definition

Solvation potential describes the ability of a compound to dissolve in water. This term is related to hydrophobicity

► [Protein-Protein Interaction](#)
 ► [Two Hybrid System](#)

Soma

Definition

The soma is the main compartment (cell body) of a neuron containing the nucleus, Golgi apparatus, the rough and smooth endoplasmic reticulum, ribosomes and mitochondria. The soma may also receive synaptic input.

► [Neurons](#)

Somatic Hypermutation

Definition

Somatic hypermutation is the process by which affinity maturation of antibodies occurs. Here, specific segments of the expressed immunoglobulin genes undergo

nearly a million-fold increase in mutation rates compared to the spontaneous background mutation rate.

► [DNA Polymerases](#)

Somatic Mosaicism

Definition

Somatic mosaicism describes that cells of a tissue consist of two different populations, due to a mutation that has arisen in one cell at the early stages of embryonic development, and the subsequent proliferation and migration of this cell. This can give rise to two different cell lines in a single individual, each of which has a different phenotype.

► [Heritable Skin Disorders](#)

Somatic Mutation

Definition

Somatic mutation refers to a non-hereditary mutation in a somatic cell, i.e. not in a germ cell. A somatic mutation is not passed onto (inherited by) the offspring. Some ► [neoplasia](#) is due to somatic mutation. Somatic mutation is probably also important in generating diversity in the variable (V) regions of ► [immunoglobulins](#).

► [Bloom Syndrome](#)
 ► [Heritable Skin Disorders](#)
 ► [Microarrays in Pancreatic Carcinomas](#)
 ► [Mouse Genomics](#)
 ► [Neurofibromatosis Type 1 \(NF1\), Genetics](#)

Somite

Definition

Somites are transient organizational structures of the embryo located on both sides of the neural tube. The total number of somites is constant and specific to each species. In avian embryos, each somite pair individualizes every 90 minutes until a total of about 50 somite pairs have been formed; in mice the number of

segments attains 65. The somites are initially epithelial and then dissociate into sclerotome that gives rise to the vertebrae, and the dermomyotome which evolves into dorsal dermis and vertebral muscles.

- Bone Disease and Skeletal Disorders, Genetics
- Neural Crest Cells and their Derivatives
- Somitogenesis

Somitogenesis

PHILIPPE DAUBAS, MARGARET E. BUCKINGHAM
Department of Developmental Biology, CNRS, Pasteur
Institute, Paris, France
margab@pasteur.fr

Definition

Somites are transient segments of ►paraxial mesoderm that are present in developing ►cephalochordates and vertebrates. In many vertebrate species, somites form as blocks of cells, which bud off in a highly coordinated fashion from the anterior end of the unsegmented ►presomitic mesoderm (PSM), following an anterior/posterior gradient. The strict temporal and spatial regulation of somitogenesis is of crucial developmental importance because segmental structures (peripheral spinal nerves, vertebrae, axial muscles and early blood vessels) develop according to the periodicity of somite segmentation. Until recently, the mechanisms that govern somitogenesis were largely unknown. However, the identification of ►cycling genes, the expression of which turns on and off, has provided evidence for an intrinsic oscillatory mechanism in PSM cells. The newly formed somite, as it separates from the PSM (initial segmentation) has an epithelial structure, which matures in such a way that an epithelial ►dermomyotome is retained dorsally, while ventrally mesenchymal cells form the ►sclerotome. Multipotent mesodermal cells in the immature somite acquire a differentiated cell fate in response to signals from surrounding tissues. Subsequently sclerotomal cells differentiate into the cartilage and bone of the vertebral column and ribs while the dermomyotome gives rise to skeletal muscle and dorsal dermis.

Characteristics

Two main features characterise somitogenesis:

1. the formation of the somite and the molecular events leading to successive segmentation. The precise spatiotemporal regulation of somite segmentation

requires the activity of a biochemical oscillator known as the somite-segmentation clock. The molecular identity of the clock itself is still unclear, but genetic and embryological data indicate how the periodicity of somite formation is generated and how the position of segment borders is determined.

2. the developmental regulation of somite derivatives. The specification of somite derivatives such as bone and muscle depends on patterning signals secreted from adjacent tissues that lead to the activation, in particular somitic compartments, of genes promoting ►cell lineage specification.

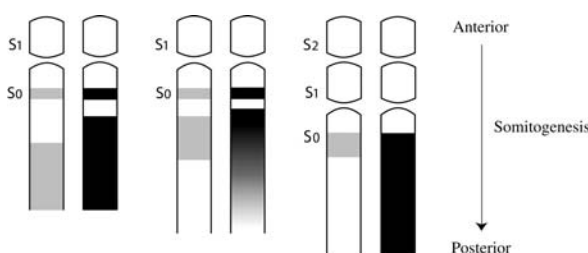
Molecular Interactions and Regulatory Mechanisms

The Making of the Somite

Key transcription factors and molecules of different ►signal transduction pathways are involved in vertebrate segmentation, and have now begun to be identified.

Cycling Genes – see Review (1) and Fig. 1

The first evidence for such genes was provided by the observation of regular pulses of expression in PSM cells of the mRNA coding for the basic helix-loop-helix (b-HLH) transcription factor *c-hairy 1*. *c-hairy* mRNA is expressed as a wave sweeping across the whole PSM once during the formation of each somite. The existence of such an oscillator or clock, whose role is to generate a temporal periodicity which can be translated spatially into the periodic boundaries of the somites, was proposed originally as the “clock and wavefront” model. It involves a set of genes involved in Notch signalling (2). Such genes encode transcription factors of the Hairy and Enhancer of Split (HES) family, acting downstream of Notch signalling, as well as the glycosyl transferase, Lunatic Fringe, and the Notch ligand, Delta C. Their cycling behaviour in the PSM is



Somitogenesis. Figure 1 Oscillation of *Lfng/c-hairy1* and *Axin2* during Somitogenesis. Comparison of the three phases of *Lfng/c-hairy1* (in light grey, left side of each schema) and *Axin2* (in black, right side) expression during each oscillation of the segmental clock. *Axin2* is out of phase with *Lfng/c-hairy1* (S0, next somite to form; S1, most recently formed somite; S2, somite formed prior to S1)

regulated at the transcriptional level. All these genes oscillate in synchrony in the PSM, suggesting that they are downstream of a common cycling activator. The mechanism driving the oscillations of the cycling genes has been extensively studied in zebrafish, chick and mouse. In all species examined thus far, Notch activation lies at the heart of the oscillator. Notch plays a critical role in the control of the oscillations by directly activating the cycling genes, thus accounting for their synchronous expression. Notch activation drives the expression of transcriptional repressors of the HES family, which in turn negatively regulate their own expression and that of the other genes. This model, based on a negative feed-back loop, requires that the HER (Hairy and Enhancer of Split-Related) proteins be highly unstable and that their repressive effect be dominant over Notch activation. A second group of cycling genes linked to the Wnt signalling pathway has recently been uncovered. One cycling gene in this class has been identified which encodes the inhibitor of Wnt signalling, *Axin2*. In the mouse, *Axin2* is expressed in a dynamic sequence similar to, but out of phase with, that of the Notch-related cycling genes. *Axin2* is directly regulated by Wnt signalling and may participate in the establishment of an autoregulatory negative feedback loop involved in its periodic expression. Results with mouse *Wnt* mutants indicate that Wnt signalling acts upstream of the Notch-regulated cycling genes. Therefore, in the mouse, the segmentation clock appears to be composed of a Wnt-based regulatory loop that generates a series of Notch-based loops.

The Determination Front and Segment Boundary Formation

The secreted growth factor FGF8 (fibroblast growth factor 8) is implicated in converting the clock pulsation into the periodic arrangement of segment boundaries. *Fgf8* mRNA is strongly expressed in PSM precursors in the ►primitive streak and tail bud as well as in the posteriormost PSM, and its expression progressively decreases in more anterior cells, establishing a gradient. *Fgf8* mRNA is progressively degraded in newly formed tissues and this is translated into a gradient of FGF8 protein. This mechanism provides a novel model for morphogen gradient formation (3). It is proposed that, because of the progressive decrease of *Fgf8* expression during maturation of the PSM, when cells become located in the anterior PSM, they reach a threshold of FGF signalling allowing them to activate the segmentation programme. This threshold level is termed the “determination front”. Wnt3a has also been proposed to assume a role similar to that of FGF8 by establishing a gradient controlling segmentation in the PSM (2). At the level of the determination front, a gene coding for the transcription factor *Mesp2/c-mesol* becomes periodically activated in a one-somite wide

domain (2). It is tempting to speculate that the periodic activation of the genes of the *Mesp* family is controlled by the segmentation clock. This would provide a link between the segmentation clock, the determination front and the boundary-formation process. How the segmentation clock itself is activated remains to be established.

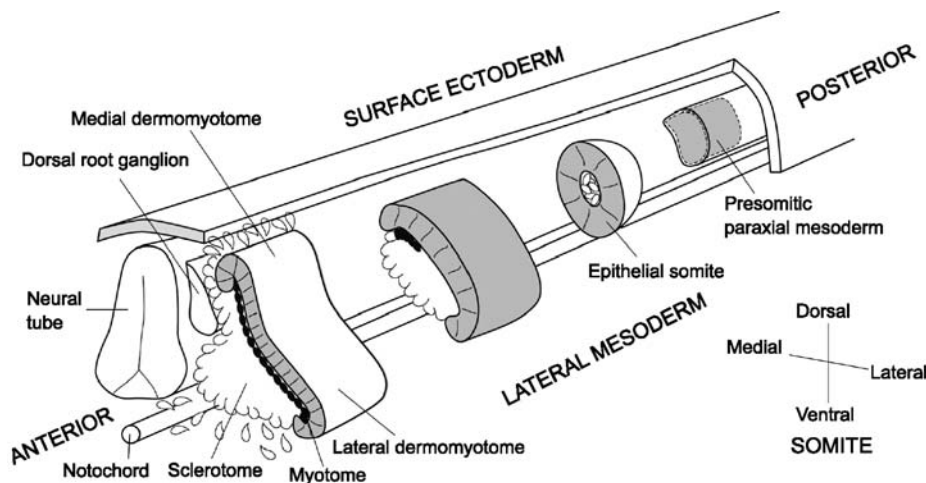
Hox Genes

Positional information along the body axis, defining for example the particular shape of the vertebrae, is controlled by the *Hox* genes in the somites. The segmentation clock also controls aspects of spatiotemporal activation of Hox genes providing a co-ordination between the positioning of segment boundaries and the future positional identity of the somite derivatives (4).

Maturation of the Somite and its Tissue Derivatives

The somite buds off from the PSM as an epithelial structure. Epithelialisation depends on a number of genes and notably the b-HLH transcription factor Paraxis. In its absence, somite derivatives form, but are disorganised (5). Patterning of the somite is characterised morphologically by the retention of a dorsal epithelium, the dermomyotome, while the ventral somite acquires the loose mesenchymal aspect of the sclerotome (Fig. 2). At the molecular level the homeo/paired domain protein Pax3, present in the PSM, continues to be expressed in the dermomyotome while the sclerotomal cells now express Pax1 (and Pax9). *Pax* genes are regulators of the myogenic and cartilaginous progenitor cells derived from the dorsal and ventral somite compartments respectively (6).

Further patterning of the somite occurs on the medial/lateral axis, such that the medial lip of the dermomyotome gives rise to the early myogenic progenitor cells in which the myogenic determination gene, *Myf5*, is activated (7). These cells are the first to contribute to the skeletal muscle of the ►myotome, a third somite compartment located between the dermomyotome and the sclerotome. As cells continue to enter the myotome, an intercalated (dermo)myotome domain can be distinguished, characterized by *Engrailed 1* expression. Myogenic cells derived from the medial dermomyotome later contribute to epaxial musculature, notably the deep back muscles. Subsequently muscle progenitor cells, specified in the lateral dermomyotome will also contribute to the myotome (Fig. 2). Entry of these cells into the myogenic programme also depends on the activation of *Myf5* and/or *Mrf4* and, later, *MyoD*, encoding other members of this b-HLH family of myogenic regulatory factors. Hypaxial muscles, which constitute most of the trunk musculature, originate from the lateral dermomyotome, which initially extends into lateral mesoderm as an epithelial



Somitogenesis. Figure 2 Schematic representation of Somitogenesis.

structure, the somitic bud. At certain axial levels – at limb level for example – the somitic bud disintegrates to give rise to migratory muscle progenitor cells which will activate myogenic regulatory factors once they have arrived at the sites where limb and other muscles will form. Genes coding for the transcription factors *Pax3* and *Lbx1* and for *Met*, a tyrosine kinase receptor, are implicated in the migration of these cells from the somite. A number of genes encoding homeo-domain factors are also involved in the specification of somitic cell-type derivatives. *Meox1* and *Meox2* genes function together and upstream of several genetic hierarchies that are required for the development of somites. Double mutant animals lack an axial skeleton and skeletal muscles are severely deficient. The genes encoding the homeoproteins *Six 1.4* and *5* have been shown to be expressed in somites during development. The *Six* proteins, acting in collaboration with an *Eya* co-activator, can directly transactivate some skeletal muscle target genes. The central dermomyotome also gives rise to cells which will form the dorsal dermis in which the homeoprotein *Msx1* is expressed, as well as contributing muscle precursor cells to the older myotome as the epithelial structure of the dermomyotome disintegrates. It has been proposed that a subpopulation of endothelial cells is also derived from the dorsal somite. The sclerotomal cells of the ventral somite give rise to the axial skeleton, both the vertebral column and the ribs. The sclerotome is also patterned on the medio-lateral axis. *Pax1* is more rapidly down-regulated in the lateral domain where rib primordia will develop, in close proximity to the myogenic cells of the somitic bud. A further sclerotomal compartment (the ►syndetome), lying under the myotome, is characterized by the expression of the b-HLH transcription factor, scleraxis. Cells expressing this factor will

contribute to tendon formation, at the expense of skeletogenic *Pax1* expressing cells (8).

Signalling from Surrounding Tissues Regulates the Tissue Specification of Somitic Cells

The dorso-ventral and medio-lateral axes of the newly formed somite become patterned in response to signals from surrounding tissues (6, 7). These also play a role in activating the genes that determine the entry of cells into a differentiation programme. The notochord and floor plate of the neural tube are sources of Sonic hedgehog (Shh), which plays a key role in sclerotome formation. Shh, with other Hh homologues, are required for the activation of *Pax1*. In the dorso-medial sclerotome *Pax1* probably acts transiently to render cells competent to respond to ►BMP signalling which promotes chondrogenesis. *Bapx1/Nkx3.2*, a homeodomain transcriptional repressor, expressed initially in the same domain as *Pax1*, also lies downstream of Shh signalling and probably plays a role in repressing transcription factors that inhibit the chondrogenic-promoting activities of BMPs. The recently identified sub-compartment of the sclerotome, that contributes to tendon formation, is subject to FGF signalling from the adjacent myotome.

The dorsal somite is patterned by positive and negative signalling from axial structures, and from surface ectoderm and lateral mesoderm. Shh probably has a long-range effect on the medial dermomyotome, although its direct effect on *Myf5* activation in this domain is controversial. Wnt signalling from the dorsal neural tube and surface ectoderm is critical for *Myf5* and *MyoD* activation, and indeed *Pax3* expression also depends on Wnt signals from surface ectoderm. The early activation of *Myf5* in the medial dermomyotome depends on TCF/βcatenin mediated Wnt

signalling, whereas later activation of *MyoD* in the lateral domain is influenced by Wnt7a which acts through a non-canonical Wnt pathway. Negative regulators of Wnt signalling, such as the Frzbs show complex expression patterns in the somite, where they probably contribute to the fine modulation of dorsal somite maturation. BMPs from the surrounding tissues inhibit the onset of myogenesis. The negative influence of BMP signalling from the dorsal neural tube is counteracted by Noggin, produced in the medial dermomyotome and later in the lateral domain where it neutralises the effect of BMP4 signalling from lateral mesoderm. Other signalling systems, such as those mediated by IGF and PDGF, have also been implicated in the onset of myogenesis in the dorsal somite. FGF signalling within the somite is also involved in later phases of myotome development. FGF inhibitors, the Sproutys, show complex expression patterns as the somite matures. Notch signalling in the myotome is also a negative regulator of myogenesis, affecting *MyoD* expression. In conclusion the dorsal somite, which is the source of all skeletal muscle apart from that in the head, is subject to highly complex positive and negative signalling which regulates the spatiotemporal specification of myogenic cells (7). Tissue determination genes, such as *Myf5*, are controlled by a correspondingly complex series of regulatory elements (9) which integrate these incoming signals, leading to the activation of the gene and hence the entry of a particular sub-population of cells into the myogenic programme.

Acknowledgements

Philippe Daubas and Margaret Buckingham are supported by the CNRS and the Pasteur Institute.

References

1. Pourquié O (2003) The segmentation clock: converting embryonic time into spatial pattern. *Science* 301:328–330
2. Rida PCG, Le Minh N, Jiang Y-L (2004) A Notch feeling of somite segmentation and beyond. *Dev Biol* 265:2–22
3. Dubrulle J, Pourquié O (2004) *fgf8* mRNA decay establishes a gradient that couples axial elongation to patterning in the vertebrate embryo. *Nature* 427:419–422
4. Kmita M, Duboule D (2003) Organizing axes in time and space; 25 years of colinear tinkering. *Science* 301:331–333
5. Buckingham M (2001) Skeletal muscle formation in vertebrates. *Curr Opin Genet Dev* 11:440–448
6. Brent AE, Tabin CJ (2002) Developmental regulation of somite derivatives: muscle, cartilage and tendon. *Curr Opin Genet Dev* 12:548–557
7. Tajbakhsh S, Buckingham M (2000) The Birth of Muscle Progenitor Cells in the Mouse: Spatiotemporal Considerations. *Curr Top Dev Biol* 48:225–267
8. Brent AE, Schweitzer R, Tabin C (2003) A somitic Compartment of Tendon Progenitors Cell 113:235–248
9. Hadchouel J, Carvajal J, Daubas P et al (2003) Analysis of a key regulatory region upstream of the *Myf5* gene reveals multiple phases of myogenesis, orchestrated at each site by elements dispersed throughout the locus. *Development* 130:3415–3426

SOP

► Standard Operating Procedure

Sorting Signals

Definition

Sorting signals are sequence or structural motifs that determine the subcellular localization or the trafficking of newly synthesised cellular proteins by mediating the association with specific recognition proteins (targeting).

► Cell Polarity

► Protein and Membrane Transport in Eukaryotic Cells

SOS

Definition

SOS stands for son of sevenless. It promotes the exchange of Ras-bound GDP by GTP.

► Signal Transduction: Integrin-Mediated Pathways

Southern Blot

Definition

Southern blot describes a method in molecular biology, pioneered by E. M. Southern, where DNA fragments are first separated according to molecular weight on a gel, then transferred to an absorbent sheet of material, such as Nylon membrane, and finally analyzed with a (non-) radioactive probe that only sticks to the DNA sequence(s) of interest. Southern blotting is very appropriate to detect the presence/absence or size changes of a certain stretch of DNA.

- Large-Scale Homologous Recombination Approaches in Mice
- Mitochondrial Myopathies
- YAC and PAC Maps

Spastic Paralysis

Definition

Spastic paralysis refers to a type of muscle weakness associated with muscular stiffness and exaggerated deep tendon reflexes.

- Hereditary Spastic Paraplegias

Specific Locus Test

Definition

Specific locus test (SLT) designates a specific genetic test for mice to assess the mutagenicity of chemicals or radiation. For this purpose, a Tester stock is used containing seven recessive loci for coat colour and other visible phenotypes, e.g. mutation rate of a mutagen can be determined by mating a treated male mouse to a Tester stock female mouse, and subsequently counting the number of altered phenotypes in the offspring generation.

- Large-Scale ENU Mutagenesis in Mice

Specific Rotation $[\alpha]^{25^\circ}_D$

Definition

A quantitative expression of the optical activity of a given substance. $[\alpha]^{25^\circ}_D$ is the rotation (in degrees) measured by using plane-polarized light of D line of sodium (wavelength of 5.461 nm) at 25°C, an optical path length of 1 dm and a concentration of 1 g/100 ml.

- Amino Acids: Physicochemical Properties

Specificity

Definition

Specificity is the likelihood of a negative test result in patients without disease. It measures how well the test

excludes disease. It is the complement of the false-positive rate.

- Mass Spectrometry: SELDI

SPECT

- Single Photon Emission Computer Tomography

Spectral Imaging

- Fluorescence Microscopy: Spectral Imaging vs. Filter-based Imaging

Sperm

Definition

Sperm is a highly specialized, haploid, male germ cell that fertilizes the egg, thereby restoring a diploid state to the zygote.

- Mammalian Fertilization

Spermatogenesis

Definition

Spermatogenesis designates a sequential process in males involving mitosis, meiosis, and cellular differentiation, whereby spermatogonial stem cells give rise to mature spermatozoa.

- Mammalian Fertilization

Spermatogonia

Definition

Spermatogonia are the male germ cells in the testis. They contain 46 chromosomes, and permanently divide

by mitosis to maintain a stem cell population. A special type of spermatogonia undergoes meiosis and differentiates into sperm.

► [Fragile X Syndrome](#)

Sphingomyelinase

Definition

Sphingomyelin (SM), and in particular its metabolite, ceramide, is involved in a variety of cellular processes including differentiation, cellular senescence, apoptosis and proliferation. Ceramide is produced by hydrolysis of SM by acid or neutral sphingomyelinase, and can act as a second messenger. Some members of the TNF receptor family stimulate sphingomyelinases. Ceramide produced by neutral SMase activated by TNF-R1 seems to be involved in the regulation of MAP kinases and PLA2, whereas ceramide generated by acid SMase, which can be activated by Fas and TNF-R1, is involved in NF- κ B activation and apoptosis induction.

► [TNF Receptor/Fas Signaling Pathways](#)

Spinal Muscular Atrophy

SHUJI OGINO

Department of Pathology, Brigham and Women's Hospital; Department of Medical Oncology, Dana-Farber Cancer Institute; and Harvard Medical School, Boston, MD, USA
sogino@partners.org

Synonyms

Autosomal recessive spinal muscular atrophy; proximal spinal muscular atrophy; infantile spinal muscular atrophy; spinal muscular atrophy, infantile acute form, Werdnig-Hoffmann disease; spinal muscular atrophy, infantile chronic form; juvenile spinal muscular atrophy; spinal muscular atrophy, childhood and adolescent form; Kugelberg-Welander disease

Definition

Spinal muscular atrophy (SMA) is characterized by symmetric proximal muscle weakness and is caused by degeneration of the ► [anterior horn cells](#) of the spinal cord. SMA is clinically subdivided into three types

based on age of onset and clinical severity. Type I SMA (Werdnig-Hoffmann disease; OMIM #253300) is characterized by the onset of severe muscle weakness and hypotonia in the first few months of life, and the inability to sit or walk. Fatal respiratory failure usually occurs before the age of 2 years. Type II SMA (intermediate type; OMIM #253550) is characterized by the onset of proximal muscle weakness before 18 months of age, the ability to sit but not to walk unaided, and survival beyond 4 years of age. Type III SMA (Kugelberg-Welander disease; OMIM #253400) is characterized by the onset of proximal muscle weakness after the age of 2 years, the ability to walk independently until the disease progresses and survival into adulthood.

Since the discovery of the disease gene, *SMN1* (OMIM #600354), there has been considerable progress in understanding of molecular genetics and pathogenesis of SMA. It is now understood that SMA is caused by deficiency of the SMN protein. There are diseases that present symptoms resembling those of SMN-deficient SMA, but are caused by mutations in genes other than *SMN1* (1).

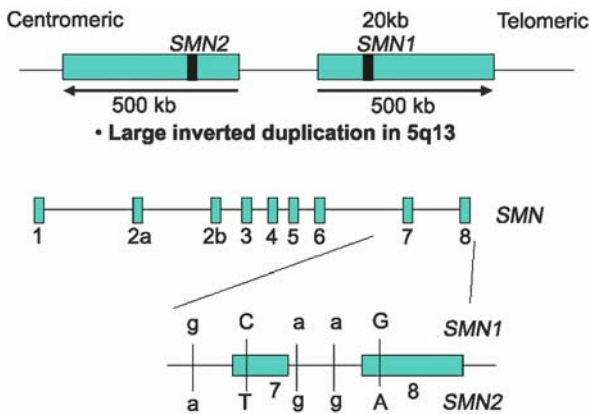
Characteristics

SMN1, the SMA Disease Gene and *SMN2*

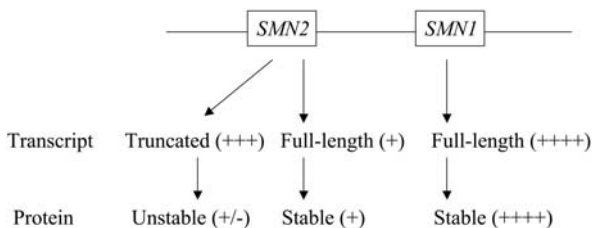
By gene deletion or ► [gene conversion](#), *SMN1* exon 7 is homozygously absent in 94% of patients with clinically typical SMA (2). *SMN1* and its centromeric homologue, *SMN2* (OMIM #601627), lie within the telomeric and centromeric halves, respectively, of a large inverted repeat in chromosome region 5q13 (Fig. 1). *SMN* consists of 9 exons (exons 1, 2a, 2b, 3 through 8), with the stop codon present near the end of exon 7. Although *SMN2* typically differs from *SMN1* in nucleotides in intron 6, exon 7, intron 7, and exon 8, the coding sequence of *SMN2* differs from that of *SMN1* only by a single nucleotide (c.840C > T) in exon 7 (Fig. 1). In contrast to *SMN1*, at least one copy of which is indispensable for the normal survival of motor neurons, *SMN2* is dispensable since approximately 5–10% of normal individuals lack both copies. The failure of *SMN2* to compensate fully for the loss of *SMN1* is due to the sequence difference (c.840C > T) in exon 7. Although translationally silent, the c.840C > T transition in *SMN2* decreases the activity of an ► [exonic splicing enhancer](#) so that less full-length transcript is generated, leading to deficiency of stable SMN protein (Fig. 2). The *SMN* gene duplication occurred more than five million years ago, before the separation of human and chimpanzee lineages, but the *SMN2* gene sequence *per se* appeared for the first time only in *Homo sapiens*.

SMN1 Deletion Mutations and Conversion Mutations

Approximately 94% of individuals with clinically typical SMA lack both copies of *SMN1* exon 7 (2).



Spinal Muscular Atrophy. Figure 1 Genomic Organization of the *SMN1* and *SMN2* genes. *SMN2* typically differs in nucleotides in intron 6, exon 7, intron 7 and exon 8 from *SMN1*.



Spinal Muscular Atrophy. Figure 2 Transcript and protein productions from the *SMN1* and *SMN2* genes. Most transcripts from *SMN2* are truncated forms, and produce unstable proteins.

Loss of *SMN1* can occur by deletion (typically a large deletion that includes the whole gene) or by gene conversion to *SMN2*. A large majority of genes converted from *SMN1* to *SMN2* had *SMN2* sequences in most of the polymorphic nucleotides except for the *SMN1* sequence in exon 8. Minor variants of hybrid *SMN* genes have been described. These include *SMN2* intron 6 and exon 7 juxtaposed to *SMN1* intron 7 and exon 8, *SMN2* exon 7 flanked by *SMN1* intron 6 and exon 8 and *SMN1* sequences in all of the polymorphic nucleotides except for the *SMN2* sequence in exon 8; the latter is not a disease allele. A number of hypotheses have been proposed for the generation of these hybrid genes, including unequal recombination, intrachromosomal deletion and gene conversion. The gene conversion hypothesis is generally accepted.

The absence of *SMN1* exon 7 sequence correlates with either a functional or a *de facto* absence of the *SMN1* gene. At least one of the factors underlying the variable phenotype of SMA is the *SMN2* copy number. Individuals with type III SMA have, on average, more

copies of *SMN2* than individuals with type II or type I SMA, and probably express more full-length transcript, and thus more stable SMN protein, from *SMN2*. *SMN2* copy number also correlates directly with length of survival of SMA patients.

SMN1 Small Intragenic Mutations and Polymorphisms

The SMA in many of the approximately 6% of SMA patients who do not lack both copies of *SMN1* is probably unrelated to *SMN1*. However, in a substantial number of the SMA patients who do not lack both copies of *SMN1*, small intragenic mutations (also referred to as “subtle” or “non-deletion” mutations) in *SMN1* have been identified (2). These intragenic mutations provide solid evidence that *SMN1* is indeed the SMA gene, since intragenic *SMN1* mutations are associated with the SMA phenotype regardless of the status of other candidate genes. Twenty-nine small intragenic mutations have been reported [Table 1, modified from (1)]. The most common locations for these mutations are exon 6 and exon 3. The two most frequently reported small intragenic mutations, both in exon 6, are the p.Y272C missense mutation (c.815A>G) and the c.770_780dup frameshift mutation.

SMN polymorphisms have been reviewed (2). Most polymorphisms are present in the promoter region and intron 6.

SMN1 ▶ De Novo Mutations

SMN1 has a high rate of *de novo* deletion mutations due to a high frequency of recombination and crossover events. The high rate of *de novo* mutations in *SMN1* probably underlies the high carrier frequency for SMA in the general population despite the high genetic lethality of the disease. Most reported *de novo* deletion mutations were paternal in origin, and 1.7% of affected individuals received *de novo* deletion mutations mostly from their fathers. The large inverted repeat at chromosome 5q13, as well as smaller repeat sequences around the *SMN1* and *SMN2* loci, probably predispose to unequal crossing over and other recombination events, and therefore probably underlie the high *de novo* mutation rate of *SMN1*. However, the reason why *de novo* mutations of paternal origin are much more frequent than those of maternal origin is unknown. The paternal *de novo* mutation rates have been calculated as 1.6×10^{-4} (1). However, this figure is based on very small numbers of patients who had *de novo* mutations. Further studies are necessary to calculate *de novo* mutation rates more precisely.

Cellular and Molecular Regulation

The SMN Protein and the SMN Complex

SMN is a ubiquitously expressed protein (38 kD) consisting of 294 amino acids. Most of the SMN

Spinal Muscular Atrophy. Table 1 List of Small Intragenic *SMN1* Mutations {modified from (1)}. Mutations are listed in order of nucleotide number starting from A in the initiation codon ATG

Standard nomenclature of nucleotide change	Single amino acid change	Other nomenclature used in the literature	Site of mutation	Type of mutation	Type of SMA	No. of reported patients
c.5C>G	p.A2G	38C>G	Exon 1	Missense	II, III	3
c.22dupA			Exon 1	Frameshift	I or II	1
c.43C>T	p.Q15X	78C>T	Exon 1	Nonsense	I, III	2
c.81dupG		Q27insG	Exon 1/ Intron 1	Frameshift	II	2
c.91dupT		124insT	Exon 2a	Frameshift	I	1
c.208_209insGTGT		241-242in4	Exon 2b	Frameshift	III	1
c.305G>A	p.W102X		Exon 3	Nonsense	II, III	2
c.399_402delAGAG		430del4	Exon 3	Frameshift	I, II, III	5
c.400G>A	p.E134K	433G>A	Exon 3	Missense	I	1
c.439_443delGAAGT		425del5, 472del5	Exon 3	Frameshift	I	2
c.509_510delGT		542delGT	Exon 4	Frameshift	I, II, III	3
c.558delA		591delA	Exon 4	Frameshift	II	1
c.585dupT		618insT	Exon 4	Frameshift	I	1
IVS4_IVS6del			Intron 4 to intron 6	Deletion	I	2
c.683T>A	p.L228X		Exon 5	Nonsense	I or II	1
c.734C>T	p.P245L	767C>T	Exon 6	Missense	III	1
c.740dupC		773insC	Exon 6	Frameshift	III	1
c.770_780dupCTGATGCTTTG		813ins/dup11, 800ins11	Exon 6	Frameshift	I, II	8
c.785G>T	p.S262I	818G>T	Exon 6	Missense	III	2
c.815A>G	p.Y272C	848A>G	Exon 6	Missense	I, II, III	9
c.821C>T	p.T274I	854C>T	Exon 6	Missense	II, III	4
c.823G>A	p.G275S		Exon 6	Missense	III	1
c.834+2T>G		c.867+2T>G	Intron 6	Splice site	I	1
c.835_18_835-12delCCTTTAT		c.868-11del7	Intron 6	Splice site	I	1
c.835G>T	p.G279C	868G>T	Exon 7	Missense	II, III	2
c.836G>T	p.G279V	869G>T	Exon 7	Missense	I	2
IVS7+4_IVS7+7delAGTC		c.922+3del4	Intron 7	Splice site	II	1
IVS7+6T>G		c.922+6T>G	Intron 7	Splice site	III	1
EX8del ^b			Exon 8	Deletion	II, III	2

^bExact extent of deletion not determined

protein in a normal individual is derived from *SMN1*, whereas the scanty amount of the SMN protein in an affected individual with homozygous absence of *SMN1* is derived from *SMN2* (Fig. 2). SMA patients with homozygous absence of *SMN1* always have at least one copy of *SMN2*. Homozygous deletions of the mouse *SMN* cause embryonic lethality, and human *SMN2* can rescue mice from embryonic lethality, and alleviate symptoms resembling human SMA in a dose dependent manner (3). Thus, the SMN protein is essential and indispensable for cell survival. It still remains a question why relative deficiency of such a ubiquitous and essential protein, SMN, causes loss of neurons, but not other cell types. It is probably because requirements of the SMN protein may differ among various cell types.

SMN is involved with biogenesis of uridine-rich small nuclear [▶ribonucleoproteins](#) (UsnRNPs) and facilitates cytoplasmic assembly of UsnRNPs into the spliceosome, a large snRNA-protein complex that catalyzes pre-mRNA splicing (3, 4). In the nucleus, SMN appears to be directly involved in pre-mRNA splicing. SMN associates with multiple other proteins to form a complex (4, 5). Core SMN-interacting proteins in the complex include [▶Gemin2](#), Gemin3, Gemin4, Gemin5, Gemin6, and Gemin7. Substrates and substoichiometric components include Sm and Sm-like (LSm) proteins, uridine-rich snRNAs (UsnRNAs), fibrillarin, GAR1, coilin, profilin, ZPR1, OSF, nucleolin, B23, hsc70, RNA helicase A, RNA polymerase II, snurportin, importin β , galectin1/3, and heterogeneous nuclear RNP-Q and RNP-R (3, 4, 5). UsnRNP assembly is mediated by the SMN complex, which subsequently facilitates hypermethylation of the methyl-G caps of UsnRNAs (4). The hypermethylated cap (m_3 G-cap) structure is specifically bound by snurportin, which then interacts with the import receptor importin β and, together with a not-yet-known import receptor that recognizes the Sm core, mediates UsnRNA import into the nucleus (4). After each round of assembly, the SMN complex must be reloaded with fresh Sm proteins to regain its activity, and this loading reaction may be regulated by the PRMT5 complex, composed of methylosome [pICln, WD45, and PRMT5 (JBP1)] and Sm proteins (4, 5). Sm proteins B, D1 and D3 are methylated by methylosome before being loaded into the SMN complex (4).

Modifying Factors for SMA

SMN2 is a well-established disease-modifying gene for SMA, and has been shown to decrease severity of SMA in a dose-dependent manner. The copy number of *SMN2* also correlates with longer survival of individuals affected with type I SMA. The dosage effect of human *SMN2* has also been observed in animal model of SMA (3, 6). However, the *SMN2* copy numbers in

type I, type II, and type III SMA overlap. Therefore, there must be other modifying factors.

Other candidate modifier genes include *NAIP*, *H4F5* (*SERF1*), and *Htra2-1* [reviewed (1)]. It remains a possibility that SMA phenotypes are modified by genes responsible for other neuromuscular disorders, including *IGHMBP2*, *HEXA*, *SCO2*, and others [reviewed (1)]. It is also possible that some of the components of the SMN and PRMT5 complexes and other components of the spliceosome may be modifiers for SMA.

Clinical Relevance

SMN1 Normal and Disease Allele Frequencies

SMA is the second most common severe autosomal recessive disease, affecting 1/10,000 live births, with a disease allele frequency of 1/100 and a carrier frequency of 1/50 (1). Approximately 60% of SMA cases are type I, and the remaining 40% type II or III with more type II cases (1).

Various *SMN1* normal and disease allele frequencies for each SMA type as well as in the general population have been estimated (1). Most disease alleles in SMA are deletions (or conversions) of at least exon 7 of *SMN1*, or what is referred to as “0-copy” alleles (frequency, 9.8×10^{-3}). A normal chromosome 5 usually has one copy of *SMN1* (“1-copy” allele; frequency, 9.6×10^{-1}). However, two copies of *SMN1* are often present on the same chromosome 5, for which the term “2-copy allele” (frequency, 3.3×10^{-2}) is used. We refer to disease alleles with small intragenic mutations in *SMN1* as “1^D” (“1-copy-Disease”) alleles (frequency, 1.8×10^{-4}). The 1^D alleles are indistinguishable from normal, 1-copy alleles by standard molecular analysis.

Molecular Genetic Testing

Since the discovery of the disease gene *SMN1*, various methods of molecular genetic testing for SMA have been developed [reviewed in (1)], including linkage analysis, PCR-[▶restriction fragment length polymorphism](#) (RFLP), PCR-[▶denaturing high-performance liquid chromatography](#) (DHPLC), quantitative *SMN* [▶gene dosage analysis](#), *SMN1* small intragenic mutation analysis, and [▶haploid analysis](#). A reasonably sensitive PCR-RFLP assay is the most commonly used method to detect SMA patients who lack both copies of *SMN1* exon 7. *SMN* gene dosage analysis may utilize quantitative competitive PCR-RFLP or quantitative real-time PCR. *SMN* gene dosage analysis is performed to identify a carrier individual with only one copy of *SMN1*, or to identify an SMA patient who has a small intragenic mutation in the patient's only copy of *SMN1*. Such a patient cannot be detected by the standard PCR-RFLP assay. The haploid analysis technique can be used in combination with *SMN* gene dosage analysis to determine *SMN1* copy number in each chromosome 5. This method will be particularly

useful to detect a carrier individual who has two copies of *SMN1* on one chromosome 5 and a deletion/conversion mutation on the other chromosome 5. Such a carrier cannot be distinguished from a normal individual by *SMN* gene dosage analysis alone.

Genetic Counseling and Risk Assessment

Methods of SMA genetic risk assessment in various clinical settings have been described in great depth (1). Risk assessment is an integral part of genetic testing and counseling. The purpose of genetic risk assessment is to provide family members with risk estimates as accurately as possible for their decision-making. Risk assessment for SMA is particularly important, because of the complexity of SMA genetics. The disease allele frequency is relatively high, small intragenic mutations are undetectable by the standard assays including *SMN* gene dosage analysis, the paternal *de novo* mutation rate is high and the presence of two copies of *SMN1* on one chromosome 5 can mask a deletion/conversion mutation on the other chromosome 5 in *SMN* gene dosage analysis.

Potential Therapeutic Approach

Currently there is no therapy that cures SMA. However active investigations are underway, and new therapeutic approaches will probably emerge in the near future. Potential approaches include: 1) to increase full-length mRNA from *SMN2* by an *in vivo* alteration in the splicing pattern of *SMN2* exon 7 through activation of trans-acting factors or ►antisense oligonucleotides (7), 2) to increase the overall transcription level of *SMN* mRNA by activating the *SMN* promoter, 3) to stabilize the *SMN* protein, 4) to repair degenerated motor neurons, 5) to replace degenerated motor neurons by ►stem cell therapy and 6) to target modifying factors other than *SMN2* (6). A high throughput drug screening for the identification of compounds that up-regulate full length *SMN* protein is underway (6). Further understanding of the molecular genetics and pathogenesis of SMA will facilitate the development of new therapies that may cure SMA in the future.

►Spinobulbar Muscular Atrophy (SBMA) (Kennedy Disease)

►Splicing

References

1. Ogino S, Wilson RB (2002) Genetic testing and risk assessment for spinal muscular atrophy (SMA). *Hum Genet* 111:477–500
2. Wirth B (2000) An update of the mutation spectrum of the survival motor neuron gene (*SMN1*) in autosomal recessive spinal muscular atrophy (SMA). *Hum Mutat* 15:228–237
3. Frugier T, Nicole S, Cifuentes-Diaz C et al (2002) The molecular bases of spinal muscular atrophy. *Curr Opin Genet Dev* 12:294–298

4. Paushkin S, Gubitz AK, Massenet S et al (2002) The *SMN* complex, an assemblysome of ribonucleoproteins. *Curr Opin Cell Biol* 14:305–312
5. Meister G, Eggert C, Fischer U (2002) *SMN*-mediated assembly of RNPs: a complex story. *Trends Cell Biol* 12:472–478
6. Wirth B (2002) Spinal muscular atrophy: state-of-the-art and therapeutic perspectives. *Amyotroph Lateral Scler Other Motor Neuron Disord* 3:87–95
7. Skordis LA, Duncley MG, Yue B et al (2003) Bifunctional antisense oligonucleotides provide a trans-acting splicing enhancer that stimulates *SMN2* gene expression in patient fibroblasts. *Proc Natl Acad Sci USA* 100:4114–4119

Spindle Check-Point

Definition

The spindle check-point serves to sense if all kinetochores are properly attached. It is believed that unattached ►kinetochores are able to emit a negative signal. If not all the chromosomes are attached to spindles, ►anaphase will not begin. Failure to execute the spindle check-point may lead to a change of the chromosomal content called aneuploidy.

►Cell Division

►Cell Cycle – Overview

►Mitotic Spindle

Spindle Microtubules

Definition

Spindle microtubules are long, tubular structures that make up the mitotic spindle. The plus end of the spindle microtubule attaches to the kinetochore, and the minus end attaches to the spindle pole.

►Centromeres

►Mitotic Spindle

Spines

Definition

Spines are specialised structural compartments of a dendrite of various shapes and sizes, which receive exclusively excitatory synaptic input. Spines may receive input from a single or multiple presynaptic

boutons, and may contain an internal spine apparatus comprised of internal stores and filaments.

► [Neurons](#)

Spinobulbar Muscular Atrophy

Definition

Spinal and bulbar muscular atrophy (SBMA), also known as Kennedy's disease, is an X-chromosomal recessively transmitted adult-onset neurodegenerative disorder that affects males. It is caused by degeneration of the spinal and bulbar motor neurons and dorsal root ganglia due to the expansion of a polyglutamine (polyQ) tract in the androgen receptor gene (sometimes combined with signs of partial androgen insensitivity).

► [Huntington's Disease](#)

► [Repeat Expansion Diseases](#)

► [SCA](#)

Spinocerebellar Ataxia

► [SCA](#)

Spinocerebellar Degeneration

Definition

Spinocerebellar degeneration defines a group of Mendelian disorders affecting the structure and function of nuclei and tracts, normally forming functional circuits within the cerebellum and spinal cord typically manifesting as impaired co-ordination and balance.

► [Genetic Predisposition to Multiple Sclerosis](#)

Splanchnic (Visceral)

Definition

Splanchnic (Visceral) is related to, situated in or affecting the soft internal organs of the body, especially those contained within the abdominal and thoracic cavities.

► [Lung](#)

Splice (Junction) Mutation

Definition

Splice (junction) mutation is a mutation that alters the junction between an intron and an exon so that it no longer functions properly, and often leads to exon skipping or premature translation termination.

► [Hereditary Spastic Paraplegias](#)

► [Heritable Skin Disorders](#)

Splice Acceptor

Definition

Some gene-trap vectors contain splice acceptors to ensure the production of a fusion transcript if the trapping vector inserts into the intron of a gene. In the case of such an event, the splice acceptor allows a fusion transcript with the upstream coding sequence and the reporter gene to be formed.

► [Mouse Genomics](#)

Splice Enhancer

Definition

A splice enhancer is an RNA segment in the primary transcript that binds regulatory protein(s) and thus facilitates usage of a nearby splice site.

► [Enhancer](#)

3' Splice Site

Definition

3' splice site defines a sequence at the 3' end of an intron, which is the 5' end of an exon. The human 3' splice site is described by the consensus yyyyyyyyn-cagG. Only the **ag** is invariable, all other positions can deviate from the consensus sequence. Capital letters are located in the intron, small letters in the exon.

► [Alternative Splicing](#)

5' Splice Site

Definition

5' splice site defines a sequence at the 5' end of an intron, which is the 3' end of an exon. The human 5'

splice site is described by the consensus CAGgtaagt. Only the **gt** is invariable, all other positions can deviate from the consensus sequence. Capital letters are located in the intron, small letters in the exon.

► Alternative Splicing

Splice Mutation

► Splice (Junction) Mutation

Splice Site Mutation

► Splice (Junction) Mutation

Spliceosome

Definition

Spliceosome is the large RNA-protein complex that removes introns from eukaryotic nuclear mRNA precursors.

► Splicing

Splicing

CHIN LI, WOAN-YUH TARN
Institute of Biomedical Sciences, Academia Sinica,
Taipei, Taiwan
wtarn@ibms.sinica.edu.tw

Synonyms

Removal of introns and joining of exons in RNA

Definition

Most eukaryotic protein coding genes are discontinuous, with the ► **exons** (coding sequences) interrupted by stretches of the ► **introns** (non-coding sequences). Splicing is required for excision of the introns with

concurrent joining of the exons from the primary transcripts. Precise splicing of precursor mRNA (pre-mRNA) is necessary to place the coding sequences in frame for protein ► **translation** and thus is critical for genetic expression in eukaryotes. Through regulated splicing, a single gene is capable of generating multiple transcripts encoding proteins with different functions. Thus, splicing is a post-transcriptional mechanism by which the diversity of higher eukaryotic genomes can be greatly expanded.

Characteristics

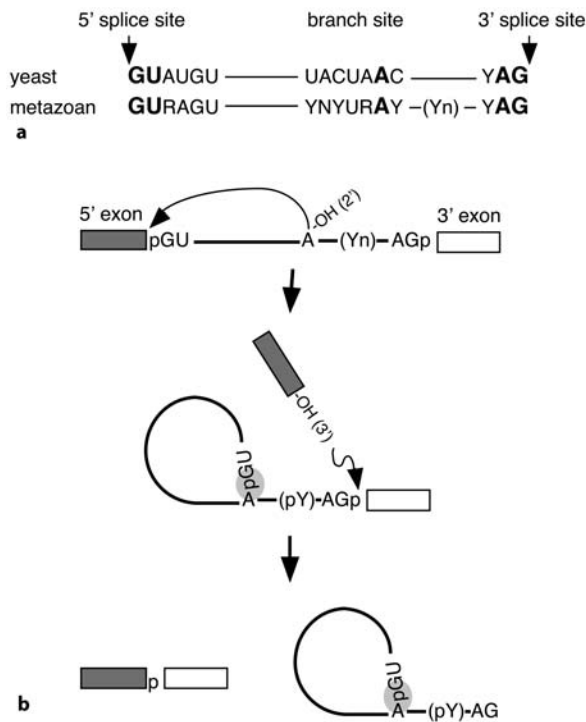
Introns in Nuclear Pre-mRNA

Splicing of nuclear pre-mRNA is directed by sequences at the splice junctions of the intron and a region near the 3' splice site that includes the branch point. These elements are well conserved in yeast but much less so in higher eukaryotes (Fig. 1a). Nevertheless, intron sequences almost invariantly start with the dinucleotide GU and end with AG. In some higher organisms, a minor class of pre-mRNA introns exists, which exhibits highly conserved but not conventional consensus sequences at the splice sites. Indeed, the two types of pre-mRNA introns are spliced by distinct, but analogous, splicing apparatuses.

Mechanism of Nuclear Pre-mRNA Splicing

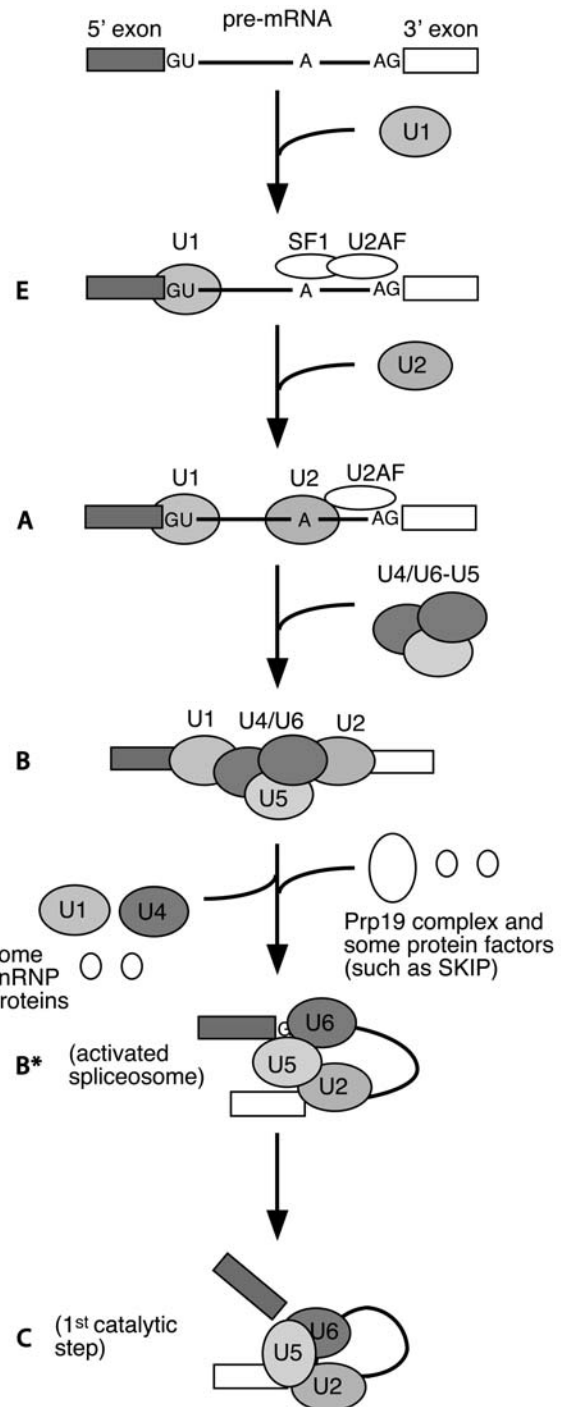
Pre-mRNA splicing occurs *via* a two-step ► **transesterification** pathway (Fig. 1b). In the first step, the branch point adenosine initiates nucleophilic attack by its 2' hydroxyl at the 5' splice junction, generating 5' exon and intron-3' exon RNA species as intermediates. The latter, referred to as the lariat intermediate, contains a characteristic 2'-5' ► **phosphodiester** linkage. The second step involves nucleophilic attack at the 3' splice junction by the 3' hydroxyl of the liberated 5' exon. This results in joining of the two exons and release of the intron in a lariat form.

Splicing of pre-mRNA is catalyzed by a dynamic complex of 3 MD, termed the ► **spliceosome**. The spliceosome is composed of four small nuclear ► **ribonucleoproteins** (U1, U2, U4/U6 and U5 snRNP), each of which contains one or two snRNA and a number of proteins. As illustrated in Fig. 2, *de novo* assembly of the spliceosome on each pre-mRNA is an ordered process with several distinct intermediates, namely E (early)/commitment (CC), A, B, and C complexes. These subcomplexes can be distinguished by their snRNP and protein factor composition. In complex E, the U1 snRNP recognizes the 5' splice site *via* a base-pairing interaction and protein factors SF1 and U2 snRNP auxiliary factor (U2AF) bind to the elements near the 3' splice site. Complex A is subsequently formed when the U2 snRNP binds *via* a base-pairing interaction with the branch site sequence.

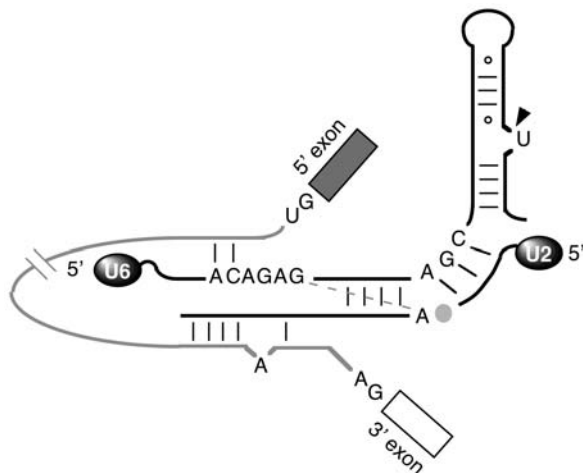


Splicing. Figure 1 (a) Canonical splicing signals in yeast and metazoan pre-mRNA introns. Highly conserved residues are shown in large bold letters. (b) The chemical reaction of pre-mRNA splicing. Splicing proceeds via two consecutive transesterification steps (see text). The 2'-5' phosphodiester bonds are shadowed.

This complex is then joined by the preassembled U4/U6-U5 tri-snRNP to form complex B. Complex B promotes a major structural change and is poised to perform the first chemical reaction. This reorganization of the spliceosome results in dissociation of the U1 and U4 snRNPs, U6 base-pairing with both the U2 snRNA and the 5' end of the intron and formation of a U6 intramolecular stem-loop structure. Concurrently, a dramatic change of protein components occurs during spliceosome activation, including dissociation of several snRNP proteins and recruitment of some protein factors that are less stably associated with snRNPs. Subsequently, the activated spliceosome catalyzes the first transesterification reaction, leading to the formation of complex C. After the second catalytic step, the mRNA is released and the disassembled snRNPs are then recycled for new rounds of splicing. Although current dogma asserts that the spliceosome is built in a stepwise manner, recent identification of a splicing-competent penta-snRNP complex from *S. cerevisiae* suggests the possibility that *in vivo* intron removal is orchestrated by a pre-assembled spliceosome.



Splicing. Figure 2 The spliceosome assembly pathway. This diagram emphasizes snRNP binding and remodeling events in each intermediate complex. The current model holds that U1, U2 and U4/U6-U5 snRNPs bind sequentially to the pre-mRNA. A major conformational change in the spliceosome occurs to form complex B* and creates the catalytic site for the first step. The first transesterification takes place in complex C, whereby the 5' exon is released and the lariat intermediate is formed.



Splicing. Figure 3 Model of the spliceosomal catalytic core. The core comprises the U2-U6 and intramolecular U6 helices. U6 and U2 base pair with the 5' splice site and branch site sequences of the pre-mRNA, respectively. Conserved residues of intron and snRNAs are highlighted. The phosphate 5' (arrowhead) of the bulged U residue in the U6 stem may coordinate a Mg^{2+} ion for catalysis. The dashed line indicates a proposed long-range tertiary interaction. Open circles in the U6 stem are putative non-Watson-Crick interactions. The gray sphere represents a non-conserved residue.

There is compelling evidence that the chemical reaction of pre-mRNA splicing is catalyzed by an RNA structure that is provided by a network of RNA-RNA interactions in the spliceosome. The catalytic core structure is mainly contributed by a U2-U6 helix and an intramolecular U6 stem-loop (Fig. 3). The former juxtaposes the reactive sites for the first catalytic step by the combined interactions of U2 with the branch site and U6 with the 5' splice site. In the intramolecular stem of U6, an evolutionarily invariant U residue situated at the bulge region coordinates an Mg^{2+} ion that may stabilize the RNA tertiary structure or even participate directly in the catalysis of the phosphoryl-transfer steps. Accordingly, selected pieces of U2 and U6 snRNAs can form a high-order structure, analogous to that of the spliceosome core, in the presence of Mg^{2+} . This protein-free U2-U6 RNA complex is capable of catalyzing a reaction closely related to the first chemical reaction of splicing. Thus, the spliceosome acts as a metal-ion affiliated RNA enzyme. In addition to RNA components, protein factors also contribute to the catalytic steps of splicing. A number of splicing factors possess the DExD/H box, characteristic of RNA-dependent \blacktriangleright ATPases. These proteins are thought to participate in ATP-dependent structural transitions during splicing. Among these factors, U5-associated protein Brr2 promotes unwinding of the U4/U6 snRNA duplex, which is crucial for

spliceosome activation. After U4 dissociation, the Prp19 protein complex joins to stabilize the remaining snRNPs and promote a structural change in the U6 snRNP that leads to activation of the spliceosome. More directly, the U5-associated Prp8/p220 protein is in close proximity to the splice sites and the snRNAs in the active site and may act as a cofactor to stabilize tertiary interactions in the catalytic core or to facilitate its formation.

Ensuring the fidelity of splice site recognition is a challenging objective for splicing in higher eukaryotes because a majority of metazoan pre-mRNAs contain poorly conserved splice sites and multiple introns. Many splicing factors containing a domain rich in arginine-serine (RS) dipeptides (thus called \blacktriangleright SR proteins) can promote spliceosome assembly and pairing of the 5' and 3' splice sites across introns. It is noteworthy that vertebrate pre-mRNA exons are often short (<300 nucleotides) compared to introns that vary considerably in size (10^2 – 10^5 nucleotides), invoking the exon definition model. SR protein-mediated splice site recognition must also occur across exons and joining of neighboring exons is then achieved by interactions between the SR proteins bound to individual exons. In cells, pre-mRNA splicing occurs while \blacktriangleright transcription proceeds. \blacktriangleright RNA polymerase II transcription complexes specifically recruit mRNA processing activities to nascent transcripts. RNA Pol II can facilitate exon recognition and juxtaposition *via* its interaction with splicing factors during transcriptional elongation, integrating a role of transcription into exon definition. Moreover, the Pol II \blacktriangleright promoter content can predetermine differential splice site selection, further supporting an intimate relationship between transcription and pre-mRNA splicing *in vivo*.

Alternative Pre-mRNA Splicing

Another level of complexity in pre-mRNA splicing among higher eukaryotes comes from alternative splicing. \blacktriangleright Alternative splicing is a phenomenon by which the splicing machinery assembles various combinations of exons to produce different mRNA isoforms from a single pre-mRNA, potentially generating multiple proteins with distinct functions. Alternative splicing may occur in a tissue- or development-specific manner or in response to specific cellular signals. Nearly 60% of human genes are estimated to undergo alternative splicing, allowing a much higher diversity within the proteome.

Alternative splicing patterns basically result from alterations in splice site utilization, including the use of alternative 5' or 3' splice sites, inclusion or exclusion of exons and retention of introns. Determination of splice site selection is generally initiated by binding of a *trans*-acting RNA binding protein with sequence specificity to a corresponding *cis*-element on the

pre-mRNA. Upon binding to the ►exonic splicing enhancer (ESE), the *trans*-acting factors may stabilize snRNP association with the regulated intron through protein-protein interactions and thus enhance splice site utilization. Binding of a splicing stimulation factor can be excluded by antagonizing factors that recognize neighboring ►exonic splicing silencers (ESS), resulting in lower usage of the splice site. Splicing enhancers or silencers may reside within introns as intronic elements. Most *trans*-acting splicing regulators have a modular structure. They carry either an RNA recognition motif (RRM) or K-homology (KH) domain for RNA binding. In addition, they possess one or more domains that mediate protein-protein interactions or determine their subcellular localization. SR proteins constitute a large family of splicing regulatory factors and play crucial roles in alternative splicing. Notably, differential ►phosphorylation of RS domains is a mechanism that controls the activity, localization or stability of SR proteins. Developmental or ►signal transduction pathways may therefore modulate the functions of splicing regulators, providing a mechanism for a wide range of splicing regulation.

A well-characterized example of splicing regulation is the control of sex determination in *Drosophila* (Fig. 4a). Sex-lethal protein (Sxl) is an RRM-containing protein expressed from an early promoter during early development of female fruit flies. Sxl blocks access of U2AF to the 3' splice site preceding a default exon of its own mRNA and thereby represses exon inclusion. This allows stable expression of Sxl protein and initiates a positive autoregulatory feedback loop in females. By a similar mechanism, Sxl protein directs 3' splice site utilization of the *transformer* (*tra*) pre-mRNA, leading to production of functional Tra protein in females. Tra protein, a member of the SR protein family, functions as a positive splicing regulator. Tra binds cooperatively with constitutive proteins Tra2 and RBP1 to an exonic enhancer sequence of the *doublesex* (*dsx*) pre-mRNA, thereby promoting inclusion of the female-specific exon and leading to production of the sex-specific transcriptional regulator Dsx^F. Numerous studies have so far indicated that interplay between regulatory elements within pre-mRNAs and *trans*-acting factors (SR and ►heteronuclear RNP proteins) is key in regulation of alternative splicing.

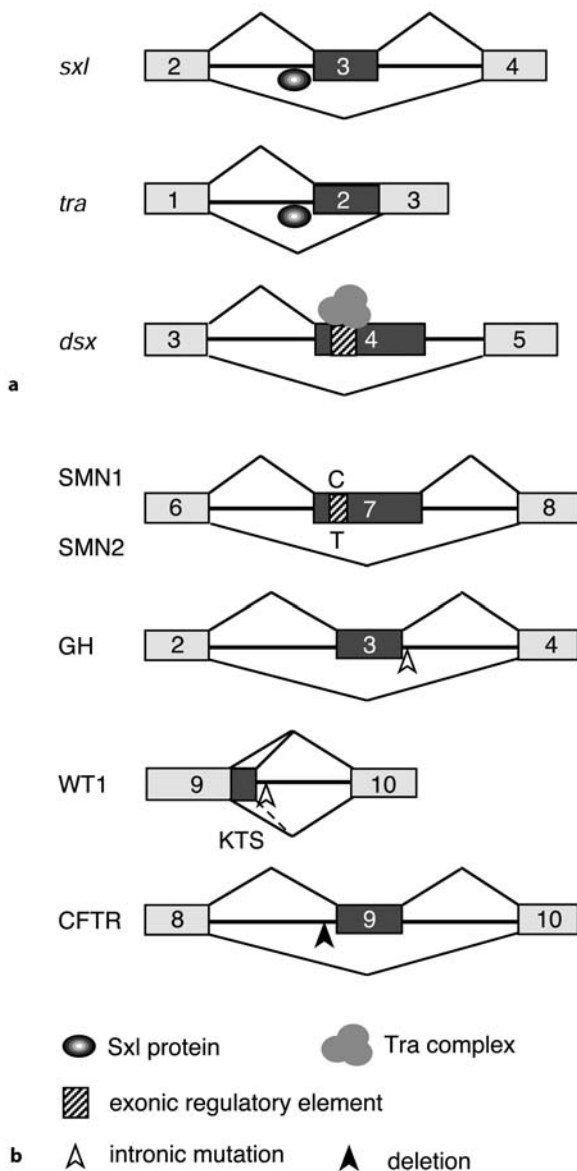
Clinical Relevance

Genetic mutations in protein coding regions may generate functionally defective and sometimes harmful proteins that are detrimental to cell functions. However, accumulating evidence has indicated that some ►translationally silent mutations have a significant impact on the yield or quality of gene products. Conceivably, mutations that disturb the splice site sequences or splicing regulatory elements may

influence the efficiency or patterns of mRNA splicing, which can in turn cause diseases or vary ►penetrance of mutations on phenotypes.

Defects in Pre-mRNA Splicing Can Cause or Predispose Individuals to a Disease

The notion that exonic nucleotide changes cause altered splicing is clearly exemplified by the study of ►spinal muscular atrophy (SMA) (Fig. 4b). SMA is caused by mutations of the *survival of motor neuron 1* (*SMN1*) gene, which encodes a protein required for the biogenesis of splicing snRNPs. The duplicated gene *SMN2* is remarkably similar to *SMN1* with only a few nucleotide substitutions. Although expressed, *SMN2* cannot completely compensate for the loss of *SMN1* function in SMA. Compelling evidence indicates that a silent C to T ►transition in exon 7 of *SMN2* results in inefficient inclusion of exon 7 by either disruption of an ESE or activation of an ESS. The level of functional SMN protein is thus reduced in all tissues, but deficiency of SMN inflicts particularly severe damage on motor neurons, leading to SMA. It is also evident that aberrant splicing analogously results from ►missense or ►nonsense single-base substitutions that disrupt regulatory *cis*-elements prevalent within exons. Intronic mutations near the splice sites may perturb association of the spliceosome and subsequently result in aberrant splicing (Fig. 4b). ►Growth hormone deficiency type II is an inherited condition resulting in patients of short-stature. In some alleles, mutations in the *GH* gene disrupt a 5' splice site, leading to exon skipping. The truncated GH produced exerts a dominant-negative effect that causes the symptom. Similarly, individuals with ►Frasier syndrome (FS) acquire mutations that inactivate an alternative 5' splice site downstream from exon 9 of the *Wilms tumor suppressor* (*WT1*) gene. Notably, in the wild-type allele, two functionally distinct WT1 isoforms can be generated by the use of different 5' splice sites for exon 9. Thus, the 5' splice site mutations detected in FS may result in an imbalance between the two isoforms that leads to pathogenic effects. Nucleotide alterations near the 3' end of an intron may also have detrimental effects on splicing. An example is the *cystic fibrosis transmembrane conductance regulator* (►CFTR) gene, in which an abbreviated polypyrimidine tract near the 3' end of an intron is associated with an increased frequency of exon skipping and severity of ►cystic fibrosis phenotypes. Furthermore, *trans*-acting mutations that disturb the function of the basal splicing machinery or splicing regulatory factors can potentially affect splicing on a broad scale. Alterations in the abundance or activities of splicing regulators have been implicated in ►neoplasia and ►malignancy by inducing misregulated splicing of mRNAs associated with ►cell transformation and cell-cell and cell-matrix interactions. Finally, it is likely that more pathogenic splicing abnormalities



Splicing. Figure 4 (a) Alternative splicing events in the sex determination pathway of *Drosophila*. The binding of Sxl protein to its own and *tra* pre-mRNAs blocks access of U2AF, leading to exon skipping or the use of a downstream 3' splice site. In *dsx* pre-mRNA, Tra with other splicing regulators binds to the enhancer of exon 4, thereby promoting the use of the preceding 3' splice site. (b) Simplified schemes of the disease-causing mutations described in the text. In SMN, the C to T substitution in a putative splicing regulatory element (hatched box) leads to exon skipping. In three other cases, intronic mutations or deletions also cause exon skipping or a shift of splice site utilization.

have yet to be unveiled. Mapping the functional significance of ▶[single-nucleotide polymorphisms](#) (SNPs) has been a powerful tool to identify potential

disease-causing mutations. With completion of the human genome project, it is now possible to search for SNPs on a genomic scale and identify their potential effects on splicing.

Acknowledgement

The authors thank Dr. Yi-Tao Yu (Rochester, NY) for critical reading of the manuscript.

- ▶[Cap-Independent Translational Control](#)
- ▶[Full-length cDNA Sequencing](#)
- ▶[Large-Scale ENU Mutagenesis in Mice](#)
- ▶[Methylation of Proteins](#)
- ▶[RNA Polymerase II Transcription](#)

References

1. Burge C, Tuschl T, Sharp PA (1999) Splicing of precursors to mRNAs by the spliceosomes. In: Gesteland RF, Cech TR, Atkins JF (eds) *The RNA World*, 2nd edn. Cold Spring Harbor Laboratory, Cold Spring Harbor, pp 525–560
2. Villa T, Pleiss JA, Guthrie C (2002) Spliceosomal snRNAs: Mg²⁺-dependent chemistry at the catalytic core? *Cell* 109:149–152
3. Jurica MS, Moore MJ (2003) Pre-mRNA splicing: awash in a sea of proteins. *Mol Cell* 12:5–14
4. Black DL (2003) Mechanisms of alternative pre-messenger RNA splicing. *Annu Rev Biochem* 72:291–336
5. Faustino NA, Cooper TA (2003) Pre-mRNA splicing and human disease. *Genes Dev* 17:419–437
6. Hastings ML, Krainer AR (2001) Pre-mRNA splicing in the new millennium. *Curr Opin Cell Biol* 13:302–309

Split-Hand-Foot

Definition

Split-hand-foot refers to a malformation of the hands/feet with an extremely variable phenotype, mainly characterized by missing central rays. Two main types are described; in the “lobster claw deformity” the hand or foot is divided by a cleft – the two parts can be opposed like lobster claws. Monodactyly is characterized by deficiency of the central and radial digits

- ▶[Bone Disease and Skeletal Disorders, Genetics](#)

Spo11

Definition

Spo11 is a Type IIB topoisomerase homolog present in eukaryotes. It is expressed during ▶[meiosis](#), where it

initiates meiotic recombination by essentially introducing double strand DNA breaks through mechanisms resembling DNA cleavage by Type II topoisomerases.

► [DNA Topoisomerases](#)

Sporadic (Cases of Disease)

Definition

Sporadic cases (of disease) are incidents of a disease with no discernable involvement of a genetic cause. Sporadic tumors, for instance, refers to tumors arising in individuals without an inherited mutation in cancer genes.

► [Microarrays in Pancreatic Cancer](#)

► [Tumor Suppressor Genes](#)

Sporadic Mutation

► [De Novo \(Mutation\)](#)

Spot Quantitation Matrix

Definition

Spot quantitation matrix refers to a table where each row corresponds to a spot on the array, and each column represents a particular characterisation of each spot, such as mean or median pixel intensity in the spot, or mean or median pixel intensity in the local background.

► [Microarray Data Analysis](#)

SPR

► [Surface Plasmon Resonance](#)

Spreading

Definition

Spreading describes a phenomenon that is observed with a cell in culture that switches from the round,

loosely adherent stage (often the growth phase) to the tightly substrate attached phase. The cell thereby modulates to a large, flat shape with a prominent cytoskeleton.

► [Focal Complexes/Focal Contacts](#)

SP-RING

Definition

SP-RING designates a motif ($SX_2CX_{15}CX_1HX_2C/SX_{17}CX_2C$) that is found in Siz/PIAS proteins that has a predicted ► [RING finger](#) structure.

► [Sumoylation](#)

Squamous Cell Carcinoma

Definition

SCCs are the most common tumors arising from sun-damaged skin, and manifest at the ages of 60–70. The lesion appears as a thick, indurated plaque with cells which tend to mimic the squamous differentiation of mid-spinous keratinocytes. SCCs have the potential to metastasize to regional lymph nodes.

► [Skin and Hair](#)

SR Proteins

Definition

SR proteins comprise of a highly conserved family of splicing factors in higher organisms, characteristic of a domain enriched in arginine/serine dipeptide repeats (known as RS domain). Some SR proteins shuttle between the nucleus and cytoplasm and appear to have a role in the export of some mRNAs.

► [RNA Export](#)

► [Splicing](#)

Src

Definition

v-Src (synonyms v-Src, c-Src) is a 60 kD protein encoded by the oncogenic retrovirus, rous sarcoma

virus. The protein derives its name from its ability to induce sarcomas in experimental animals and malignant transformed cells in tissue culture. c-Src, or cellular Src, is the normal cellular progenitor of v-Src. Both v-Src and c-Src belong to the family of Src kinases (other members are for example lck, lyn, fyn), which comprise cytoplasmic tyrosine kinases that transfer phosphate from ATP to tyrosine residues within specific protein substrates. The resulting phosphate acts as a docking site for other molecules that transmit growth signals to the nucleus on a chain of events involving multiple phosphorylation and binding reactions. c-Src contains a carboxy-terminal region that maintains the molecule in a mostly inactive state. In v-Src, this 12 amino acid region is deleted, rendering the molecule constitutively active. The Src kinase family consists of several members, which play important roles in numerous receptor signaling pathways.

► [Interleukins and Their Receptors, a Role in Regulating Hematopoietic Stem Cells](#)

► [Signal Transduction: Integrin-Mediated Pathways](#)

SRP

► [Signal Recognition Particle](#)

Sry

Definition

Sex-determining region Y (Sry) encodes a testis-determining factor in mammals, which is a member of the high mobility group (HMG)-box family of DNA binding proteins.

► [Mutagenesis Approaches in Medeka](#)

SRY – Sex Reversal

STEPHEN S. WACHTEL
Reproductive Genetics, University of Tennessee,
Memphis, TN, USA
swachtel@utmem.edu

Synonyms

Sex-determining region Y; Sex-determining gene; Testis-determining gene; XX male syndrome; XY

gonadal dysgenesis; True hermaphroditism; Mixed gonadal dysgenesis

Definition

The *SRY* (*sex-determining region Y*) gene was discovered in 1990 by Andrew Sinclair and colleagues, working in the laboratory of Peter Goodfellow (3). *SRY* is the switch that determines the sex of the mammalian embryo. In the presence of a functioning *SRY* gene, the ► [indifferent embryonic gonad](#) becomes a testis; the testis secretes ► [testosterone](#) and ► [anti-müllerian hormone](#) (AMH) and the embryo becomes a male. In the absence of the *SRY* gene, the gonad becomes an ovary; testosterone and AMH are not secreted and the embryo becomes a female. Since *SRY* is on the Y chromosome, XY embryos become males and XX embryos become females – under normal circumstances. But errors in the expression and regulation of *SRY* may lead to development of testes in XX embryos and failure of testicular development in XY embryos. These errors of gonadal or primary sex determination constitute the various forms of ► [sex reversal](#) discussed here.

Characteristics

According to Alfred Jost (1), sex determination may be viewed as a series of orderly processes: (i) establishment of genetic sex at fertilization (XY or XX), (ii) translation of genetic sex into gonadal sex (testis or ovary) and (iii) translation of gonadal sex into body sex (male or female). Step (ii) is sometimes referred to as gonadal or primary sex determination and step (iii) as secondary sex differentiation. We are here concerned with errors occurring at step (ii) – i.e. gonadal sex reversal resulting from aberrant translation of the Y-chromosome-mediated sex-determining message. Examples of gonadal sex reversal in the human are XX male syndrome, XY gonadal dysgenesis, true hermaphroditism and mixed gonadal dysgenesis.

XX Male Syndrome

Crossover of X and Y chromosomes during male meiosis may result in ► [translocation](#) of the *SRY* gene from the Y to the X. When XX embryos carry the translocated *SRY* gene, they become “XX males.” Human XX males are outwardly unremarkable in appearance, with normal intelligence and standard psychosexual orientation. But their gonads are small and sterile, their body hair may exhibit female patterns of distribution and abundance and about a third develop female-like breasts. The *SRY* gene has not been detected in about 20% of cases of XX male syndrome. Absence of *SRY* in these cases is often associated with abnormal differentiation of the external male genitalia (► [hypospadias](#) etc.). Evidently, sex reversal in these cases is due to constitutive activation of the

testis-determining pathway – i.e. to the spontaneous ►[transcription](#) of sex-determining genes downstream from *SRY*. About 1/20,000 newborn human males has the XX sex chromosome constitution.

XY Gonadal Dysgenesis

When the *SRY* gene is lost or mutated in XY embryos, the indifferent gonad becomes an ovary. But in humans, ovaries lacking two X chromosomes are unstable. There is a progressive loss of germ cells and a breakdown of follicular integrity. Although ovarian ►[follicles](#) may occasionally persist into adulthood, the XY ovary is usually represented at or around the time of birth by a dysgenetic streak gonad devoid of endocrinological activity. The resulting syndrome, occurring in about 1/20,000 women, is characterized by amenorrhea, eunuchoid habitus, tall stature and elevated ►[gonadotropins](#). In general, the external genitalia are feminine, but breast development is minimal. Pubic and axillary hairs are sparse and uterus and tubes are most often hypoplastic, although successful pregnancy has been reported in rare cases after hormonal intervention and ►[embryo transfer](#). Some 25–30% of women with this condition develop gonadal tumors. Geneticists were surprised to learn that the *SRY* gene is intact in most women with XY gonadal dysgenesis. In these cases, testicular differentiation is blocked by the action of other genes, which negate or antagonize the function of *SRY*.

True Hermaphroditism

Simultaneous occurrence of testicular and ovarian tissue, in separate gonads, or in a single gonad called the “ovotestis,” can lead to profound ambiguity of the secondary sex traits. This condition, called “true hermaphroditism” has been recognized for thousands of years, but less than 500 cases have been reported in the medical literature. The clinical diagnosis requires presence of testicular tubules and ovarian follicles. In adults the most striking features are ambiguous genitalia and well-developed breasts, but there is a bewildering variety of whole body ►[phenotypes](#), ranging from nearly normal female to nearly normal male. After surgery, some human true hermaphrodites have borne children. About 60% of human true hermaphrodites have the XX chromosome constitution. Some 12% are XY, and some have mosaic chromosome constitutions such as XX/XY. The testicular tissue of XX true hermaphrodites resembles that in XX males; they produce no sperm. *SRY* is not detected in most XX true hermaphrodites, indicating that testicular differentiation in these cases is induced by constitutive activation of downstream genes, as in XX males who lack *SRY*.

Mixed Gonadal Dysgenesis

In this variant of true hermaphroditism, there is a degenerative ovary – a streak gonad – on one side of

the body and a testis on the other. Affected persons have an uterus and at least one fallopian tube. The external genitalia are ambiguous, and the body type may range from masculine to feminine. Common karyotypes are 46,XY and 45,X/46,XY. Cryptic mosaicism – i.e. hidden 45,X cell lines in the dysgenetic ovarian gonad – may be suspected in patients with the “46,XY” karyotype. *SRY* would be expected to be present in 46,XY cell lines and absent in 45,X cell lines. This would explain the asymmetrical development of gonads in this condition. Development of testis *versus* contralateral ovary could also be explained by mosaic expression in XY gonads of a mutant testis-determining gene.

Cellular and Molecular Regulation

SRY

Each of these sex-reversed conditions is due to aberrant expression or absence of *SRY* or other genes in the sex-determining cascade. The *SRY* gene is situated on the short arm of the human Y chromosome in band Yp11.3, near the ►[pseudoautosomal region](#). It consists of a single ►[exon](#), which contains a conservative motif common among the high mobility group (HMG) proteins. The conservative motif, called an “►[HMG box](#),” spans 79 amino acids (►[Amino Acids: Physicochemical Properties](#)) associated with DNA binding. The *SRY* protein binds to sequences AACAAAG and AACAAAT and can bind to DNA 4-way junctions regardless of sequence. *SRY* intercalates in the minor groove of the DNA double helix and can bend DNA as much as 85°, thereby facilitating contact between remote ►[nucleotide](#) sequences. This would support the presumptive role of *SRY* as a transcription factor. *SRY* protein is expressed in the nuclei of the primordial ►[Sertoli cells](#), which organize the seminiferous tubules of the differentiating testis.

The Y chromosome mutations that cause XY gonadal dysgenesis tend to cluster in the *SRY* HMG box. But in the 1990's, just after the discovery of *SRY*, some 80% of women with XY gonadal dysgenesis were found to have intact *SRY* genes. This underscored the existence of a testis-determining pathway and other downstream genes, perturbation of which could block the male determining dictate of the Y chromosome. Other genes that have been found to impact differentiation of the testis are *SOX9*, *WT1*, *SFI DAX1* and *WNT-4*. The *AMH* gene, which plays a significant role in secondary sex differentiation of the male, is a downstream target of *SOX9*.

SOX9

The *SRY* gene is one member of a family of genes that code for HMG proteins. Collectively, these are known as *SOX* (*SRY-like HMG box*) genes. In one, *SOX9*, located on chromosome 17q24.1-q25.1, mutations are associated with campomelic dysplasia, a skeletal

malformation syndrome notable for bowing of the long bones and other abnormalities. Affected children with the 46,XY karyotype often exhibit XY gonadal dysgenesis, indicating that *SOX9* is part of the testis-determining pathway. In addition, there is evidence that duplication of *SOX9* can cause XX sex reversal – i.e. can induce the formation of XX testes – in the absence of *SRY*. Since *SOX9* is the first marker to appear in newly differentiated Sertoli cells, the *SOX9* gene may be an early downstream target of *SRY*. In the mouse embryo, *Sox9* is expressed on day 10.5 in the indifferent gonads of presumptive male and female. But when *Sry* is expressed on day 11.5 (the time of differentiation of the mouse testis), transcription of *Sox9* is increased in the male and decreased in the female.

AMH

The antimüllerian hormone, encoded by the *AMH* gene on chromosome 19p13.3-p13.2, blocks formation of the uterus and fallopian tubes in male embryos. There are indications that transcription of *AMH* is activated by *SOX9* (reviewed in 2). For example, the *AMH* ▶promoter contains an HMG binding site reactive with *SOX9* *in vitro*, and some XY women with *SOX9* mutations may have persistent uterus and fallopian tubes. This would place *AMH* in the male sex-determining pathway, although AMH itself is not involved in the formation of the testis.

WT1

Wilms tumor of the kidney, accounting for 15% of childhood cancers, may be found in association with aniridia, urogenital malformations and mental retardation (“WAGR syndrome”). The tumor develops as a consequence of mutation or loss of the *WT1* tumor suppressor gene on chromosome 11p13. Affected XY children exhibit gonadal dysgenesis, implying a role for *WT1* in testicular differentiation. The gene contains 10 exons, which generate 24 isoforms – i.e. functionally similar proteins having minor differences in amino acid sequence. *In vitro*, certain WT1 isoforms activate the *SRY* promoter and the *DAX1* promoter (see below).

SF1

The *SF1* (*steroidogenic factor*) gene plays a critical role in development of the adrenal gland and the gonad in humans and in mammals generally. Thus mutation in *SF1* results in adrenal failure and XY gonadal dysgenesis. In mice with homozygous *Sf1* deletions, the gonads and adrenals fail to develop and the pups die shortly after birth. Since these organs begin but do not complete development, *SF1* probably participates in formation of the indifferent gonad, but not directly in testicular organogenesis. The human *SF1* gene is on human chromosome 9p33 and contains a conservative DNA binding motif consisting of two zinc fingers.

DAX1

Human XY embryos carrying duplications in Xp21 are sex reversed (2, 4). This indicates occurrence of an X-linked gene, duplication of which can block the function of *SRY*. The gene is called *DAX1* for *dosage sensitive sex reversal-adrenal hypoplasia congenita-critical region of the X chromosome, gene 1*. It is a member of the nuclear hormone receptor superfamily, and encodes a nuclear protein that binds DNA and regulates transcription mediated by the retinoic acid receptor. Mutations in *DAX1* cause adrenal hypoplasia congenita but do not affect sex determination in human males. In mice, XX embryos lacking both copies of the *Dax1* gene become females. In mouse embryos, *Dax1* is expressed in the somatic cells of the indifferent gonad regardless of karyotype, but it is down-regulated in the developing testis. These observations suggest that, in addition to its endocrinological function, *DAX1* acts as an antagonist to *SRY*.

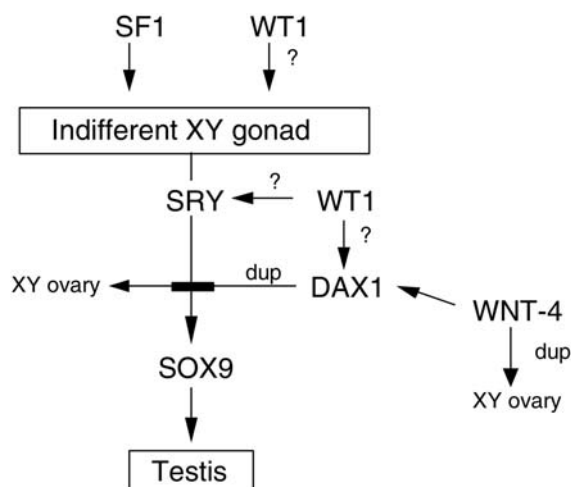
WNT-4

The product of the *WNT-4* gene on human chromosome 1p35, is one of a family of locally acting cellular growth factors. In both sexes, *WNT-4* is expressed in the ▶mesonephros and in the adrenal gland. In females, *WNT-4* signaling is required for development of the müllerian ducts and differentiation of oocytes and interstitial cells. Although the gene is expressed in the gonads of both sexes, it is down-regulated in the developing testis. The following observations indicate a key role for Wnt-4 signaling in gonadal sex determination. In mice, deletion of *Wnt-4* causes XX sex reversal. In humans, duplication of the WNT-4 region on chromosome 1 causes XY sex reversal. Moreover, *Wnt-4* up-regulates *Dax1*, so duplication of *Wnt-4* would be expected to suppress expression of *Sry*.

Summary

On the basis of the preceeding observations, the mammalian sex-determining pathway can be perceived as a complicated network of genes on several different chromosomes (Fig. 1). Although the particular steps of the pathway remain to be clarified, several of the individual genes have been identified. The pathway is initiated by *SRY*, a transcriptional activator with downstream targets including *SOX9*. The mechanism by which *SRY* is activated is unknown.

For the present, the various etiologies of gonadal sex reversal can be characterized as follows. XX sex reversal (XX male syndrome), resulting generally from translocation of the *SRY* gene to the X chromosome during meiosis, may also result from deletion of the *WNT-4* gene or from duplication of *SOX9* – in the absence of *SRY*. Ambiguity of the external genitalia in *SRY*(–) XX males and *SRY*(–) XX true hermaphrodites



SRY – Sex Reversal. Figure 1 Hypothetical cascade for differentiation of the XY testis. *SF1* and *WT1* genes may play a role in the development of the indifferent embryonic gonad, and *WT1* may also bind the *SRY* and *DAX1* promoters. According to this scheme, *SRY*, turned on in the indifferent XY gonad, activates *SOX9* leading to differentiation of the testis. *DAX1* antagonizes *SRY*, and duplication (dup) of *DAX1* leads to XY gonadal dysgenesis. *WNT-4* up-regulates *DAX1*. Duplication of *WNT-4* also leads to XY gonadal dysgenesis.

can be attributed to failure of activation of other genes downstream of *SRY*. Differences of gonadal phenotype in XX male syndrome and XX true hermaphroditism could be attributed to ▶lyonization involving X-linked sex-determining genes as yet undiscovered. XY sex reversal (XY gonadal dysgenesis) may result from loss or mutation of the *SRY* gene. It may also result from mutation within *WT1* or *SF1* or from duplication in the *DAX1* or *WNT-4* genes. XY gonadal dysgenesis in association with campomelic dysplasia is due to mutation within the *SOX9* gene. Sex reversal in a single gonad in an XY true hermaphrodite could be attributed to mosaic expression of the mutated or duplicated gene. As noted above, this could also explain nonmosaic cases of XY mixed gonadal dysgenesis.

Clinical Relevance

The manner in which the embryonic gonad is induced to become testis or ovary is a prime question in developmental biology and may have special relevance to the practice of medicine – particularly in the areas of pediatrics and endocrinology. Elucidation of the sex-determining cascade and identification of the individual sex-determining genes will provide insights into the nature and development of the various sex reversed syndromes, with implications for diagnosis and therapy.

References

1. Jost A (1970) Hormonal factors in the sex differentiation of the mammalian foetus. *Phil Transact Roy Soc Lond B* 259:119–130
2. Morrish BC, Sinclair AH (2002) Vertebrate sex determination: many means to an end. *Reproduction* 124:447–457
3. Sinclair AH, Berta P, Palmer MS et al (1990) A gene from the human sex-determining regions encodes a protein with homology to a conserved DNA-binding motif. *Nature* 346:240–244
4. Wachtel SS (1998) X-linked sex-reversing genes. *Cytogenet Cell Genet* 80:222–225

SSCP

Definition

Single-stranded conformational polymorphism (SSCP) detection is a method aimed at ▶mutation and polymorphism scanning, which identifies abnormally migrating single-stranded DNA segments on gel electrophoresis.

▶Hereditary Neuropathies, Motor and/or Sensory

SSH

▶Suppressive Subtractive Hybridization

Stable Isotope

Definition

Non-radioactive isotopes, often with a spin 1/2 magnetic moment such as ^{13}C , ^{15}N , ^{19}F , and ^{113}Cd , are used in NMR spectroscopy for labeling atoms, where the natural abundant nuclei are too weak to be detected.

▶3D Structure by NMR

STAGA

Definition

STAGA stands for the human histone acetyltransferases complex SPT3-TAF-GCN5. The human

STAGA complex is a chromatin-acetylating transcription coactivator that interacts with pre-mRNA splicing and DNA damage-binding factors *in vivo*.

►Polyglutamine Disease, the Emerging Role of Transcription Interference

Standard Operating Procedure

Definition

Standard operating procedure (SOP) describes a formal process established for methods that are to be followed routinely for the performance of designated operations; used in industrial settings.

►Large-Scale Homologous Recombination Approaches in Mice

Standard Proteomics Technologies

Definition

Standard proteomics technologies comprise for example technologies such as the combination of 2-D-SDS PAGE (2-D gel electrophoresis) and mass spectrometry, multidimensional chromatography and mass spectrometry, isotope coded affinity tagging methods (ICAT) and mass spectrometry.

►Protein Microarrays as Tool for Protein Profiling and Identification of Novel Protein-Protein Interactions

Staphylococcal Scalded Skin Syndrome

Definition

The Staphylococcal Scalded Skin Syndrome designates a syndrome in which the patient exhibits extensive exfoliation or shedding of skin, caused by staphylococcal secreted exfoliative toxin.

►Desmosomes

State Function

Definition

State function defines a thermodynamic function that only depends on the state of a system and not on the path by which this state was reached.

►Molecular Dynamics Simulation in Drug Design
►Thermodynamic Properties of DNA

Statistical Thermodynamics

Definition

Statistical thermodynamics refers to a branch of physics and physical chemistry that studies macroscopic systems (in equilibrium) from a microscopic or molecular point of view. It forms a bridge between classical thermodynamics and molecular physics.

►Molecular Dynamics Simulation in Drug Design

STATs

Definition

STATs (Signal Transducers and Activators of Transcription) constitute a highly conserved family of proteins with the dual function of transducing signals from the cell surface into the nucleus, as well as activating transcription of target genes. They convert extracellular stimuli into a wide range of appropriate cellular processes, such as immune response, antiviral protection and proliferation.

►JAK/STAT

►Polycystic Kidney Disease Autosomal Dominant

Steatocystoma Multiplex

Definition

Steatocystoma multiplex is a rare autosomal dominant disorder characterized by widespread development of sebum-containing dermal cysts. Development of steatocystomas has been hypothesized to be due to alterations in the structure of keratin 17.

►Heritable Skin Disorders

Steinerts Disease

►Myotonic Dystrophy Type 1

Stem Cell Plasticity or Stem Cell Transdifferentiation

Definition

The term reflects the theory that differentiated cell types can change into another cell type. Hence, in some cases it was demonstrated that adult stem cells such as hematopoietic stem cells transdifferentiate into endothelial cells. However, there are other examples where transdifferentiation is completely questioned.

► [Stem Cells: an Overview](#)

Stem Cell Therapy

Definition

Transplantation of normal stem cells that can differentiate into cells of a particular function, in order to replace the cells that die or are lost and restore the particular function.

► [Spinal Muscular Atrophy](#)

► [Stem Cells: an Overview](#)

Stem Cell Transdifferentiation

► [Stem Cell Plasticity or Stem Cell Transdifferentiation](#)

Stem Cells: an Overview

JÜRGEN HESCHELER¹, BERND K. FLEISCHMANN²

¹Institute of Neurophysiology, University of Cologne, Cologne, Germany

²Institute of Physiology I, University of Bonn, Bonn, Germany

j.hescheler@uni-koeln.de

Definition

In general, stem cells are defined by their capacity to keep (i) the potential to proliferate *in vitro* in an

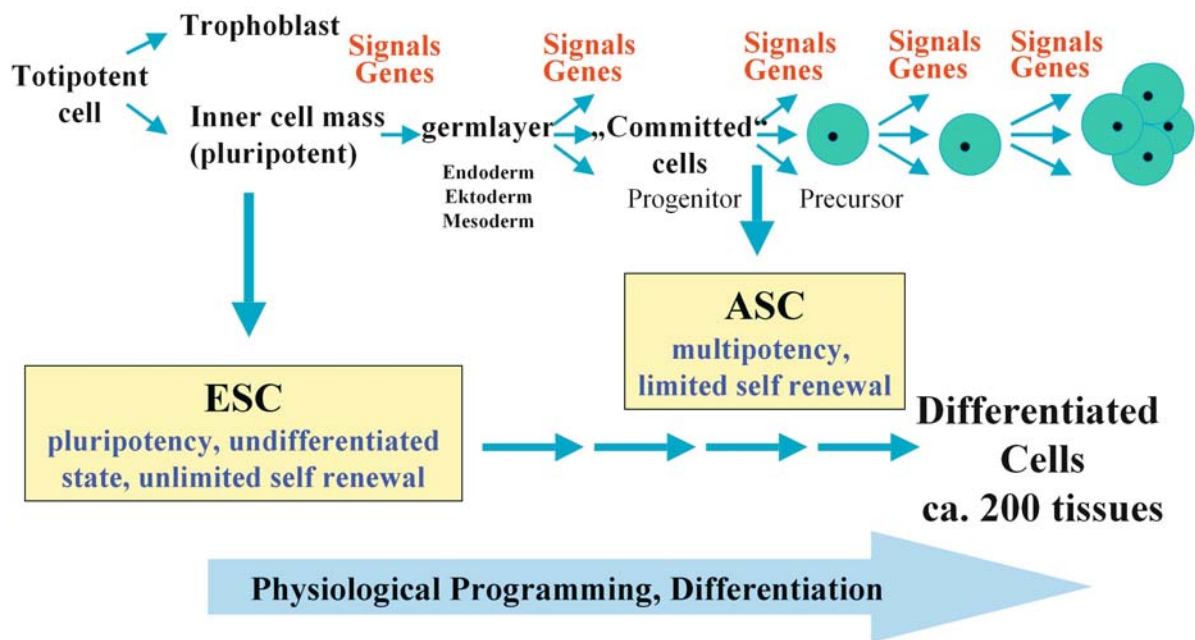
(ii) undifferentiated and (iii) ► [pluripotent](#) or ► [multipotent](#) state, thereby maintaining a normal and stable ► [karyotype](#) even with continual passaging (► [self-renewal](#)). The most prominent stem cells are of course the earliest formed during development, named ► [embryonic stem \(ES\) cells](#), which occur within the inner cell mass of blastocysts and are the origin of all the cells within our body. Obviously during their further development, mammalian organisms keep a pool of undifferentiated cells, which hold a certain regenerative potential for the organs. It is not yet unequivocally clarified what the properties of these ► [adult stem cells](#) are, but it is mostly assumed that they are in the genetic state of committed cells, i.e. progenitor cells which are already genetically determined to become a certain organotypic cell. It is still unproven whether adult stem cells have the potential to transdifferentiate, i.e. to cross boundaries and assume a different fate from that of the organ they are derived from (stem cell plasticity, see Fig. 1).

Characteristics

The term “blood stem cell” had already been used in the scientific literature as a key word in the 1960s. It appeared intermittently but gained attention when the first publication came out in the middle of the 1980s on the clinical use of blood stem cells to restore hematopoiesis in patients rendered aplastic as a consequence of extensive radiation and/or chemotherapy. Twenty years later, stem cells became even more interesting, since there was growing evidence that this population of undifferentiated cells was much more diverse. In the meantime stem cells (i.e. neural, ► [mesenchymal stem cells](#)) displaying the principal properties of self-renewal and differentiation were identified in various organs. Nowadays it is generally agreed that adult as well as embryonic stem cells provide one of the most promising tools to develop new therapeutic strategies in regenerative medicine, including the treatment of cardiac infarction and heart failure, diabetes type I, Parkinson's disease, spinal cord lesions and all other degenerative diseases. The concept of cell replacement therapy is based on the differentiation of stem cells, either *in vitro* or *in vivo*, and their implantation into the damaged tissue. Stem cells also provide an extraordinarily interesting tool for developmental biology as the consecutive expression of genes and proteins can be investigated during organotypic development under controlled conditions. Stem cells also open a new field of high turnover systems for screenings of new drugs or toxic agents under physiological conditions.

Adult Stem Cells

Stem cells are defined as clonogenic self-renewing progenitor cells that can generate one or more



Stem Cells: an Overview. Figure 1 Stem cell plasticity.

specialized cell types. Hence, the first group consists only of embryonic stem (ES) cells. These pluripotent stem cells are capable of generating all the differentiated cell types in the body. ES cells in turn generate the second group, which are called organ- or tissue-specific stem cells. Such multipotent stem cells generate cell types, comprising a particular tissue in embryos and in some cases in adults. The prototypic example of this second group is the hematopoietic or blood stem cell, which generates all the cell types of the blood and immune system. In addition, in tissues such as the intestine and skin, ongoing cellular turnover provides a rationale for the persistence of stem cells. In other organs such as the brain and heart, the presence of stem cells is proven by their isolation from these tissues and their subsequent growth in culture and differentiation, either *in vitro* or after transplantation *in vivo*. It is not clear however, whether these cells are actually used by the body to replace diseased or damaged cells. Adult stem cells can be autologous and are ethically less sensitive, but evidence for their multipotency, particularly cross-lineage is much weaker and it is not clear whether transplantation to damaged tissue is sufficient to induce differentiation to a functional cell *in vivo*, or whether *in vitro* differentiation protocols analogous to those being developed for embryonic stem cells would be a more effective approach. Mesenchymal stem cell populations retain the capacity to differentiate into the osteogenic lineage and produce cartilage or bone tissue when implanted within porous hydroxyapatite biomaterials in immunocompromised mice.

► Hematopoietic Stem Cells

The small population of multipotent hematopoietic stem cells (PHSCs) in the bone marrow is classified into short-term reconstituting cells (STRCs) and long-term reconstituting cells (LTRCs), based on how quickly the transplanted cells can produce progeny in an irradiated recipient. They can be purified using a combination of cell size, density, fluorescent dye uptake, resistance to cytotoxic chemicals and cell-surface markers including Thy1.1, Sca-1, c-kit, lineage markers, CD 38 and CD34. Experimental transplantation studies indicate that the best reconstitution occurs when both cell populations are present, the more mature cells activating the immature cells after myeloablation, whereas the mature cells provide negative control in normal animals. Functionally the type of assay that has been most widely used for the quantitation of mouse stem cells is the *in vivo* repopulating assay. Typically hematopoietic stem cells are derived from the bone marrow, which is collected from the femur and tibia of mice. Cells are suspended in PBS and incubated with rat anti-mouse monoclonal antibodies specific for the following hematopoietic lineages: CD4 and CD8 (T lymphocytes), B-220 (B lymphocytes), Mac-1 (macrophages), GR-1 (granulocytes) and TER-119 (erythrocytes). Cells are then rinsed in PBS and incubated for 30 min with magnetic beads coated with goat anti-rat immunoglobulin. Lineage-positive cells are removed by a biomagnet and the 10% remaining lineage-negative cells are then stained with ACK-4 biotin followed by streptavidin-conjugated phycoerythrin and sorted by FACS. The Lin-cells are

selected as c-kit-positive and c-kit-negative with a 1–2 log difference in staining intensity.

Mesenchymal Stem Cells

The term mesenchymal stem cells (MSC) refers to the adult mesenchymal progenitor cells with the potential to produce progeny that differentiate to produce a variety of mesenchymal cell types including fibroblasts, muscle, bone, tendon, ligament and adipose tissue. It is not known if these cells actually have the capacity to self-renew, which is a property of stem cells. MSC may be found in muscle, skin and adipose tissue as well as in bone marrow. MSC in bone marrow may be identified by colony-forming units that produce fibroblasts and make up a very small percentage of the total marrow population. MSCs or their progeny in the bone marrow provide a stromal microenvironment for hematopoiesis. During development, MSCs in the bone marrow may derive from the developing vessels or from circulating precursors.

MSCs may be isolated from bone marrow, peripheral blood, fat, skin, vasculature and muscle, where they are most probably responsible for normal tissue renewal as well as for response to injury. MSCs are negative for primitive hematopoietic cell markers but express antibody-defined markers, SH2, SH3 and SH4 and STRO-1. Systemic transplantation of MSCs has not always led to functional results in tissue. MSCs are obtained by the Caplan protocol, i.e. under sterile conditions the femur and tibia of 2-month-old female rats are excised and the cells are dispersed. The mesenchymal population is isolated on the basis of its ability to adhere to the culture plate.

►Neural Stem Cells

Fetal tissue has long been used as a source of multipotent neural progenitors. However, until recently, the adult CNS was considered incapable of self-renewal or repair. The discovery of stem cells within the adult CNS rendered this concept at least in part, obsolete. These cells are capable of self-renewal, and can give rise to highly differentiated progeny (neurons and glia) thus conforming to the standard definition of a stem cell. Adult neural precursors were initially obtained from the forebrain subventricular zone and have since also been isolated from the hippocampus, spinal cord, striatum, septum, cortex and optic nerve. Both the adult stem cells and those of fetal origin are isolated *via* dissection.

►Stem Cell Plasticity or Stem Cell Transdifferentiation

Orlic and coworkers (4) suggested pronounced transdifferentiation of hematopoietic precursors into cardiomyocytes. However this was questioned by three independent groups very recently. These studies appear to indicate that hematopoietic precursors do not

transdifferentiate into cardiomyocytes after injection into the injured mouse heart (3).

►Embryonic Stem Cells

Embryonic Stem (ES) cells originate from the inner cell mass of the pre-implantation embryo, the blastocyst. They are pluri-potent, completely undifferentiated and keep the potential for unlimited expansion, the unlimited self-renewal capacity. Stem cells with similar properties can be also obtained from embryonic carcinoma (teratocarcinomas) and embryonic germ cells. The latter are derived from primordial germ cells isolated primarily from the genital ridge. ES cells have the potential to proliferate infinitely *in vitro* in an undifferentiated and pluripotent state thereby maintaining a normal and stable karyotype even with continual passaging. Remarkably, *in vivo* they can be reincorporated into normal embryonic development by transfer into a host blastocyst or aggregation with blastomere stage embryos. They can contribute to all tissues in the resulting chimeras including gametes. When cultivated *in vitro*, ES cells differentiate under appropriate cell culture conditions, i.e. in the absence of leukemia inhibitory factor (LIF) into cell types of all three germ layers, endoderm, ectoderm and mesoderm. However, the differentiation process occurs only when ES cells are cultivated under culture conditions where they can form multicellular spheroidal tissues, termed embryoid bodies. In recent years a large number of different cell types have been described to differentiate within embryoid bodies. These include hematopoietic and endothelial cells, cartilage and neurons as well as smooth, skeletal and cardiac (6) muscle cells. The capacity of ES cells to differentiate into cell types of the mesodermal lineage has been extensively used by us and others to investigate the molecular and physiological events occurring during development into the cardiovascular system, i.e. cardiac and endothelial cells. We have observed cardiac development from mouse ES cells as early as 7 days after formation of the aggregates, which correlates well with the murine embryo, where the first beating is seen on day E8.5–E9.5 (i.e. 8.5–9.5 days post coitum). Cardiomyogenesis in embryoid bodies is paralleled by the development of vascular structures, which starts on day 5 of differentiation and results in the formation of hollow capillary-like tubules within 3–4 days. These improve the supply of nutrients and oxygen to the embryoid bodies as well as the export of catabolic end products (5). The close relationship of ES cells and primordial germ cells has recently been demonstrated by the group of Schöler who succeeded in differentiating premitotic oocytes from a pluripotent ES cell line (1). Furthermore, it could be shown that ES cells give rise to primordial germline cells and haploid male gametes.

Before clinical use of human ES cells for therapeutic trials, two major prerequisites must be fulfilled: (i) they must be safe, i.e. the development of tumors because of the high proliferative potential must not occur and (ii) the rejection of the transplanted cells must be prevented. For criterion (i) the technology of “lineage selection” has been developed in order to allow only the needed cell to differentiate from human ES cells, whilst survival of all other cells is prevented. This can be achieved by modifying the culture conditions and/or adding a combination of various growth factors and signaling molecules, which preferentially support the growth of a specific cell type, but prevent the development of the unwanted cell types. Up to now, the transgenic drug selection approach has proven to be the most specific and effective for reliable purification of ES cell-derived organotypic cells. Application of antibiotics will eliminate all other cells except those containing the resistance gene (purification of over 99.5%).

In respect of criterion (ii) several suggestions have been made. The simplest one is the establishment of a stem cell bank containing a large number of different human ES cell lines with various human leucocyte antigen (HLA) determinants. An alternative suggestion is to use “therapeutic cloning”, i.e. the exchange of the haploid nucleus of a donor oocyte by that of the patient. It has been demonstrated that the oocytes with the new nucleus develop to the blastocyst stage from which a new ES cell line can be generated, which is allogenic to the patient (2). In order to reduce the large number of donor oocytes, it may also be possible to differentiate oocytes by the method introduced by Schöler (1).

► [ES Cell Differentiation as a Model System for Functional Genomics](#)

References

1. Hubner K, Fuhrmann G, Christenson LK et al (2003) Derivation of oocytes from mouse embryonic stem cells. *Science* 300:1251–1256
2. Hwang WS, Ryu YJ, Park JH et al (2004) Evidence of a pluripotent human embryonic stem cell line derived from a cloned blastocyst. *Science* 303:1669–1674
3. Nygren JM, Jovinge S, Breitbach M et al (2004) Bone marrow-derived hematopoietic cells generate cardiomyocytes at a low frequency through cell fusion, but not transdifferentiation. *Nat Med* 10:494–501
4. Orlic D, Kajstura J, Chimenti S et al (2001) Bone marrow cells regenerate infarcted myocardium. *Nature* 410:701–705
5. Wartenberg M, Donmez F, Ling FC et al (2001) Tumor-induced angiogenesis studied in confrontation cultures of multicellular tumor spheroids and embryoid bodies grown from pluripotent embryonic stem cells. *FASEB J* 15:995–1005
6. Wobus AM, Wallukat G, Hescheler J (1991) Pluripotent mouse embryonic stem cells are able to differentiate into cardiomyocytes expressing chronotropic responses to adrenergic and cholinergic agents and Ca^{2+} channel blockers. *Differentiation* 48:173–182

Stem Loop

Definition

Stem loop defines an RNA secondary structure consisting of a stem that is formed by base pairing between adjacent complementary sequences, and an intervening single-stranded loop.

► [Translational Frameshifting, Non-standard Reading of the Genetic Code](#)

Stereocilia

Definition

Stereocilia are extremely elongated microvilli (also called stereovilli) originating from the apical surface of the mechanosensory hair cells of the inner ear. Stereocilia of the inner ear are implicated in the transduction of mechanical force into a neuronal signal. Stereovilli have also been described on epithelial cells of the epididymal duct.

► [Microvilli](#)

Sterile Mutations

Definition

Sterile mutations are mutations in genes that result in the absence or non-function of germ cells. Such mutations are propagated by picking heterozygous siblings.

► [C. elegans Genome, Comparative Sequencing](#)

Steroid Hormone Receptor Defects, Molecular Basis

TOMOSHIGE KINO¹, GEORGE P. CHROUSOS^{1,2}

¹National Institute of Child Health and Human Development, Pediatric and Reproductive Endocrinology Branch, Bethesda, MD, USA

²University of Athens, Athens, Greece

kinot@mail.nih.gov

chrousog@mail.nih.gov

Definition

Two adrenal corticosteroids, the mineralocorticoid aldosterone and the glucocorticoid cortisol, exert profound influences on many physiological functions by virtue of their diverse roles in growth, development and maintenance of homeostasis. Their actions are mediated by intracellular receptor proteins, the **glucocorticoid (GR) and mineralocorticoid (MR) receptors**, which function as hormone-activated transcription factors that regulate the expression of, respectively, the glucocorticoid and mineralocorticoid target genes.

Characteristics

GR is ubiquitously expressed in almost all human tissues and organs. The presence of glucocorticoids is crucial for the integrity of CNS function and for maintenance of cardiovascular, metabolic and immune homeostasis. Increased glucocorticoid secretion during stress alters CNS function, assists with adjustments in energy expenditures and modulates the inflammatory/immune response. Since glucocorticoids possess a broad array of life-sustaining functions, only partial or incomplete glucocorticoid resistance has been reported so far, suggesting that complete inability of glucocorticoids to exert effects on their target tissues is incompatible with human life.

MR mediates the sodium-retaining effects of aldosterone in the kidney, salivary and sweat glands, as well as colon. In addition, the MR located in the central nervous system (CNS) – also called glucocorticoid receptor type I – appears to have a role in the regulation of stress response and feedback control in the hypothalamic-pituitary-adrenal (HPA) axis. MR has a high affinity for both aldosterone and cortisol. The circulating levels of cortisol are over 100× higher than those of aldosterone. The MR of the distal convoluted tubule and possibly other mineralocorticoid target tissues are protected from the actions of cortisol by expression of 11β-hydroxysteroid dehydrogenase type 2, which converts cortisol into the inactive cortisone.

Structure and Actions of the Glucocorticoid and Mineralocorticoid Receptors

GR and MR are members of the steroid/sterol/thyroid/retinoid/orphan receptor superfamily of nuclear transcription factors, with over 150 members currently cloned and characterized across species. The human GR cDNA was isolated by expression cloning in 1985. Located on chromosome 5, the gene of the GR is comprised of 9 exons (Fig. 1a). It encodes two 3' splicing variants, GRα and GRβ, as a result of alternatively using either terminal exon 9α or 9β. The GRα encodes a 777 amino acid protein and is the classic GR that mediates actions of glucocorticoids,

while the shorter GRβ cannot bind the ligand and its physiological and pathological roles are not well understood. The cDNA for human MR was isolated by low-stringency hybridization using human GR cDNA as a probe in 1987. The gene coding for MR also consists of 9 exons (Fig. 1b) and its locus is on human chromosome 4.

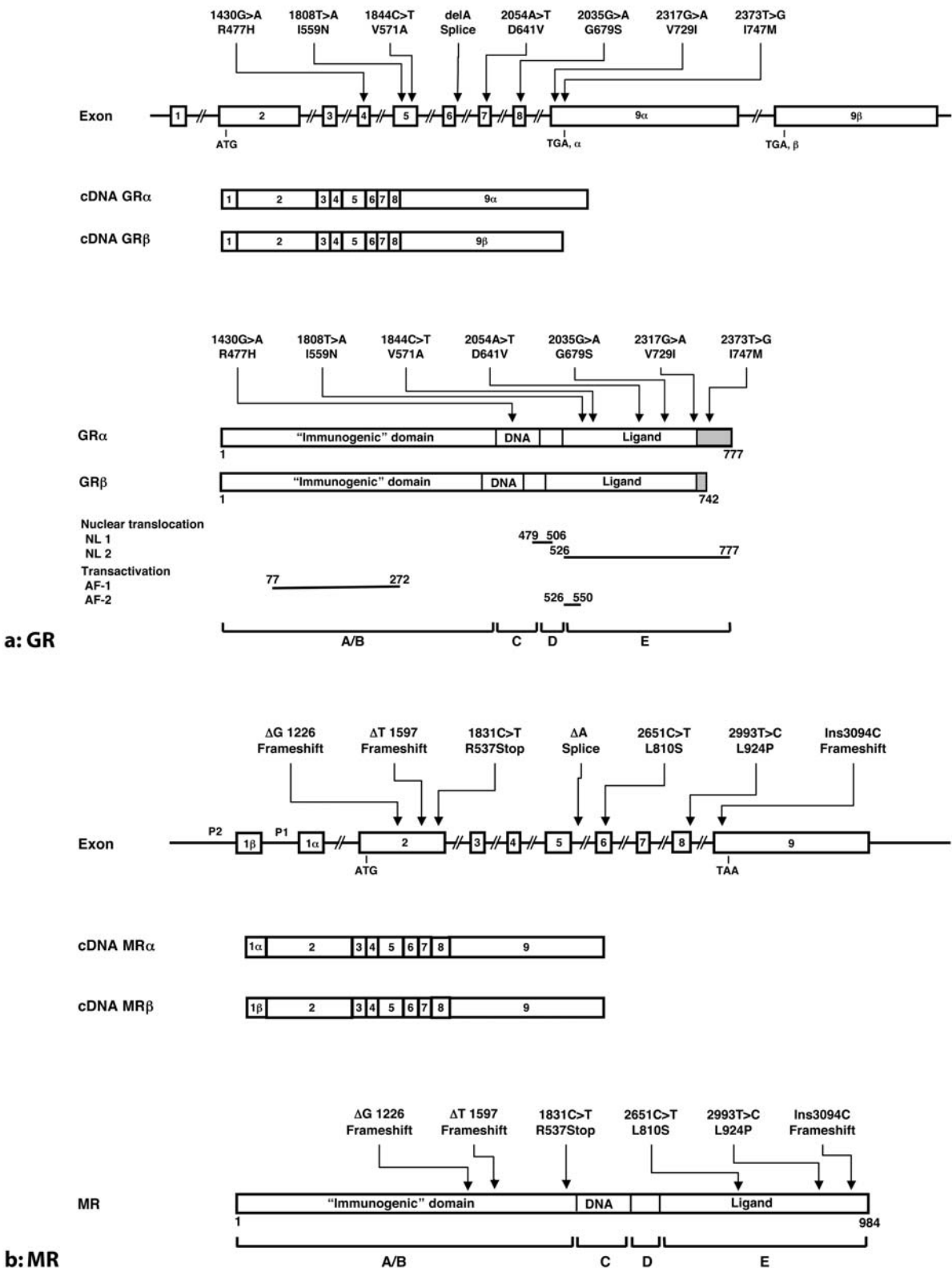
GR and MR in the unliganded state are located primarily in the cytoplasm, as part of hetero-oligomeric complexes including heat shock proteins 90, 70 and 50. After binding to their agonist ligand, they undergo conformational changes, dissociate from the heat shock proteins, homodimerize and translocate into the nucleus. There, the ligand-activated GR and MR directly interact with DNA sequences in the promoter regions of target genes.

The promoter-bound GR and MR then modulate the transcription rate of responsive genes by facilitating the formation of transcription initiation complexes including **RNA polymerase II** and its ancillary factors. In addition to these molecules, GR and MR, *via* their two transactivation domains, attract several proteins and protein complexes, namely co-activators, which help to transform the receptor signal to the transcription initiation complex as well as contain intrinsic histone acetyltransferase (HAT) and other enzymatic activities, through which they loosen the chromatin structure and facilitate access and/or binding of transcription machinery components to DNA. They include molecules known as acetyltransferase co-activators, such as the homologous p300 and cAMP-responsive element binding protein (CREB)-binding protein (CBP), the p160 family of coactivators and p/CAF, as well as huge protein complexes including SWI/SNF and thyroid hormone receptor-associated protein (TRAP)/vitamin D receptor-interacting protein DRIP complexes.

Clinical Relevance

Glucocorticoid Receptor Mutations

Familial/sporadic **glucocorticoid resistance syndrome** was first described in 1976, as a disorder characterized by hypercorticism without Cushingoid features. Since then, over 10 kindreds and sporadic cases with abnormalities of the GR number, affinity for glucocorticoids, stability and translocation into the nucleus have been reported (1). However, the molecular defects have been elucidated in only 5 kindreds and three sporadic cases (Fig. 1A). The proband of the original kindred was homozygous for a single non-conservative point mutation, replacing aspartic acid with valine at amino acid 641 in the GR LBD. This mutation reduced binding affinity for dexamethasone three-fold and caused concomitant loss of transactivation activity (1). The proband of the second family had a 4-base deletion at the 3'-boundary of exon 6, removing a donor splice site (1). This



Steroid Hormone Receptor Defects, Molecular Basis. Figure 1 Genomic, complementary DNAs, protein structures and localization of known pathologic mutations of the human glucocorticoid (a) and mineralocorticoid (b) receptors. The human glucocorticoid or mineralocorticoid receptor genes consist of 10 exons. Exon 1 is not

mutation resulted in complete ablation of one of the GR alleles in affected members of the family. The proband of the third kindred had a single point mutation in a homozygous state at amino acid 729 (valine to isoleucine) in the LBD, which reduced both the affinity and transactivation activity of GR (1). There was also an interesting sporadic case of a heterozygous man carrying a *de novo* germ-line mutation at amino acid 559 (isoleucine to asparagine) also in the LBD close to the NL1 (1). This mutant GR bound no ligand but exerted dominant negative activity on the wild type receptor. This was caused by inhibition of the translocation of the wild type receptor into the nucleus (1).

Study of a fifth case/kindred with glucocorticoid resistance and a GR mutation in the LBD (amino acid 747, replacing isoleucine with methionine) in a heterozygous state revealed that the mutant receptor had mildly reduced affinity for dexamethasone and markedly decreased transactivation activity, interestingly associated with dominant negative activity on the wild type receptor. The mutation is located several amino acids more amino terminal to helix 12 of the GR LBD and causes defective formation of the activation function (AF) 2 surface that interacts with several co-activators. The sixth and seventh sporadic cases also had mutations in a heterozygous state, replacing arginine to histidine at amino acid 477 and glycine to serine at amino acid 679, respectively (2). The former is located in the second zinc finger in the DBD. This mutant receptor has no transactivation activity due to impaired binding to glucocorticoid-response elements (GREs). The latter mutation is in the LBD, outside the ligand-binding pocket. This mutation causes 50% reduction of ligand binding affinity with comparable reduction in transactivation activity.

The proband of the eighth familial case had a point mutation in a homozygous state replacing valine to alanine at amino acid 571 in the LBD (3). The mutant receptor had a 6-fold reduction in binding affinity to dexamethasone and 10–50 fold less transactivation activity than the wild type receptor. Interestingly, the proband also suffered from a 21-hydroxylase deficiency, suggesting that association of this congenital

disease exacerbated the hyperandrogenism and virilization potential of the glucocorticoid resistance syndrome.

A complex negative feedback system exists in the human CNS that regulates glucocorticoid homeostasis. Glucocorticoids exert negative feedback effects on several higher regulatory centers; thus, the negative loop is activated in patients with loss-of-function GR mutations, resulting in compensatory increases in ACTH and cortisol secretion (Fig. 2). Although adequate compensation is apparently achieved by elevated cortisol concentrations in the great majority of the patients described, excess ACTH secretion also results in increased production of adrenal steroids with mineralocorticoid activity and enhanced secretion of adrenal androgens. The former, together with cortisol, is responsible for causing symptoms and signs of mineralocorticoid excess, such as ▶hypertension and/or hypokalemic alkalosis, whereas the latter cause varying manifestations of hyperandrogenism, such as acne, hirsutism, male pattern baldness and menstrual irregularities and infertility in women. Precocious puberty has been seen in a child due to early and excessive prepubertal adrenal androgen secretion.

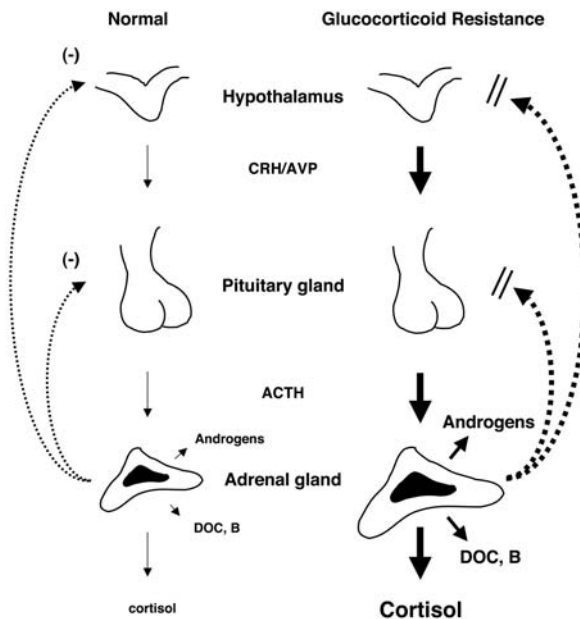
Patients are treated with high doses of mineralocorticoid-sparing synthetic glucocorticoids. The goal is to suppress the increased levels of ACTH, which cause overproduction of mineralocorticoids and androgens (Fig. 2) (1). As all cases described thus far have had partial inactivation of GR activity, the use of potent synthetic glucocorticoids with minimal intrinsic mineralocorticoid activity, such as dexamethasone, is a rational approach.

Mineralocorticoid Receptor Mutations

Inactivating Mutations of the Mineralocorticoid Receptor

The mechanism by which aldosterone stimulates sodium transport in its target tissues may involve the synthesis of a protein associated with the function of the amiloride-sensitive sodium channel (ASSC). The latter is located in the apical membrane of epithelial cells of the renal distal convoluted tubule and in the plasma membranes of cells in other tissues involved with salt conservation. The phenotype of patients with

translated, exon 2 codes for the immunogenic domain (A/B), exons 3 and 4 for the DNA-binding domain (C), exons 5–9 for the hinge region (D) and the ligand-binding domain (E). The glucocorticoid receptor gene contains either exon 9 α or 9 β alternatively spliced to produce the classic GR α or the nonligand-binding GR β . C-terminally gray colored domains in GR α and GR β compose specific portions. Functional domains of the glucocorticoid receptor are indicated. The mineralocorticoid receptor gene contains either exon 1 α or 1 β alternatively spliced to produce the MR α or MR β cDNAs, respectively, but produces a single MR protein. Locations of the known mutations in the glucocorticoid (A) or mineralocorticoid (B) receptor genes are indicated in their genomic and protein structures. GR, glucocorticoid receptor; MR, mineralocorticoid receptor; HR, hinge region; DBD, DNA-binding domain; LBD, ligand-binding domain; NL1 and 2, nuclear localization signal 1 and 2; AF-1 and -2, activation function-1 and -2.



Steroid Hormone Receptor Defects, Molecular Basis. Figure 2 The increased activity of the hypothalamic-pituitary-adrenal (HPA) axis in patients with glucocorticoid resistance results in hypersecretion of CRH, ACTH, cortisol, deoxycorticosterone (DOC), corticosterone (B) and adrenal androgens causing salt retentions and hyperandrogenism.

loss-of-function mutations of the MR mimics that of patients with defects in the subunits of the ASSC, who, however, represent the bulk of patients with ►**pseudo-hypoaldosteronism** type 1 (PHA1) or mineralocorticoid resistance.

In kindreds with PHA1, both autosomal dominantly and recessively transmitted forms have been observed. The autosomal recessive form was associated with severe disease, with manifestations persisting into adulthood and was found to be caused by loss-of-function mutations in genes encoding subunits of the ASSC. However, recently, Geller et al. identified heterozygotic MR gene loss-of-function mutations in one sporadic case and four autosomal dominant cases of PHA1 (Fig. 1b) (4). These included two frameshift mutations, each deleting a single base pair in exon 2; the resultant frame shifts led to a gene product lacking the entire DNA- and hormone-binding domains, as well as a dimerization motif. Two families had an identical mutation, introducing a premature termination codon in exon 2 at position 537. One case showed a single base-pair deletion in the intron-5 splice donor site. Subsequently, Tajima et al. reported a fifth family with a pathological mutation in the MR gene from patients with autosomal dominant PHA1 (5). The proband had a single point mutation on one chromosome at codon 924

(leucine to proline) in the LBD of MR. A sixth inactivating MR mutation was found in a sporadic case of PHA1, harboring a frame shift mutation with insertion of a cytosine at position 3094 of exon 9, which resulted in a nonsense protein with a first stop codon at position 1012 (6).

As indicated above, all reported MR mutations causing the PHA1 phenotype were in the heterozygous state affecting only one of two MR alleles, indicating that PHA1 can be caused by haploinsufficiency. Patients develop PHA1 at a very early stage of their life and their clinical condition remits with age, although plasma aldosterone concentrations and renin activity remain high. These findings suggest that the intact MR gene is required for the reabsorption of salt in infancy, while by some other as yet unknown mechanism, the defect of MR function may be overcome partially with age.

Treatment of PHA1 caused by MR mutations is by supplementation with sodium chloride. The amounts of salt may be different from patient to patient depending on their degree of salt waste. The salt administration may be lessened or become unnecessary later in their life.

Activating Mutation of the Mineralocorticoid Receptor

The first activating MR mutation was found as an early-onset hypertension markedly exacerbated in pregnancy (Fig. 1b) (7). The probanda had a point mutation replacing serine at amino acid 810 by leucine in a heterozygous state. The mutation is localized in helix 5 of the MR LBD; the leucine side chain projects into the ligand-binding pocket, potentially forming additional van der Waals interactions with alanine 773 of helix 3 and the C19 methyl group of steroid hormones. Therefore, the mutation changes the ligand specificity of the mutant receptor, conferring receptor binding to progestins in addition to mineralocorticoids and glucocorticoids. Thus, the patient developed hypertension particularly in pregnancy due to the activation of the mutant receptor by progestins secreted from the placenta.

References

1. Kino T, De Martino MU, Charmandari E et al (2003) Tissue glucocorticoid resistance/hypersensitivity syndromes. *J Steroid Biochem Mol Biol* 85:457–467
2. Ruiz M, Lind U, Gafvels M et al (2001) Characterization of two novel mutations in the glucocorticoid receptor gene in patients with primary cortisol resistance. *Clin Endocrinol (Oxf)* 55:363–371
3. Chrousos GP, Olefsky JM, Samols E (2002) Hormone Resistance and Hypersensitivity States. In: *Modern Endocrinology Series*. Lippincott, Williams and Wilkins Publ. Co., Philadelphia, PA

4. Geller DS, Rodriguez-Soriano J, Vallo Boado A et al (1998) Mutations in the mineralocorticoid receptor gene cause autosomal dominant pseudohypoaldosteronism type I. *Nat Genet* 19:279–281
5. Tajima T, Kitagawa H, Yokoya S et al (2000) A novel missense mutation of mineralocorticoid receptor gene in one Japanese family with a renal form of pseudohypoaldosteronism type I. *J Clin Endocrinol Metab* 85:4690–4694
6. Viemann M, Peter M, Lopez-Siguero JP et al (2001) Evidence for genetic heterogeneity of pseudohypoaldosteronism type I: identification of a novel mutation in the human mineralocorticoid receptor in one sporadic case and no mutations in two autosomal dominant kindreds. *J Clin Endocrinol Metab* 86:2056–2059
7. Geller DS, Farhi A, Pinkerton N et al (2000) Activating mineralocorticoid receptor mutation in hypertension exacerbated by pregnancy. *Science* 289:119–123

Steroid Hormones

Definition

Steroid hormones like sex hormones or cortisol are synthesized from cholesterol and have similar chemical skeletons. These and other lipophilic hormones bind specifically to intracellular or nuclear receptors, i.e. steroid hormone receptors.

► [Steroid Hormone Receptor Defects, Molecular Basis](#)
 ► [Transcription Factors and Regulation of Gene Expression](#)

Steroids

Definition

Steroids are a class of organic molecules characterized by a cyclopentanophenanthrene ring structure with the atomic numbering structure of cholesterol.

► [Nuclear Hormone Receptors: Regulators of Gene Transcription and Cellular Signaling](#)
 ► [Steroid Hormones](#)

Stickler Syndrome

Definition

Stickler syndrome is a connective tissue disorder caused by mutations in the collagen Type II gene (COL2A1). The disease can include ocular symptoms

like myopia, cataract, and retinal detachment; hearing loss; midfacial underdevelopment and cleft palate.

► [Bone Disease and Skeletal Disorders, Genetics](#)

STM

► [Scanning Tunneling Microscope](#)

Stochastic

Definition

Stochastic refers to a term of analytical statistics based on the theory of probability.

► [X-Chromosome Inactivation](#)

Stokes Radius

Definition

The Stokes radius of a molecule is the radius of a sphere that would have the same diffusion coefficient.

► [Analytical Ultracentrifugation](#)

Stokes Shift

Definition

Stokes shift describes the difference in wavelength between the maximum of the excitation spectrum (shorter wavelength, higher energy), and the maximum of the emission spectrum (longer wavelength, lower energy).

► [Fluorescence Microscopy: Spectral Imaging vs. Filter-Based Imaging](#)

Stop Codon

Definition

Translation elongation stops when the translation machinery encounters one of 3 stop codons: UAG, UAA or UGA, which do not code for an amino acid.

► [Translational Control in Eukaryotes](#)

Storage Polysaccharide

Definition

Starch and/or glycogen are commonly referred to as storage polysaccharide. They are linked to a core protein occurring in the cytosol, and are primarily used for energy metabolism.

► [Glycosylation of Proteins](#)

Store-Operated Calcium Entry

Definition

The store-operated calcium entry (SOC) refers to a flux of extracellular calcium ions into the cell. The SOC is triggered by reduction in the stored calcium within the cell. This calcium entry accounts for the second, more prolonged phase of the intracellular calcium response after metabotropic receptor stimulation.

► [Polycystic Kidney Disease, Autosomal Dominant](#)

STR

Definition

STR stands for Short Tandem Repeat. Often called

► [microsatellites](#), which in most cases consist of (CA)_n repeats of different length.

► [SNP Detection and Mass Spectrometry](#)

Strand Transfer

Definition

Strand transfer describes the transfer of a strand of DNA from one molecule to another.

► [DNA Recombination](#)

Stratified Epithelium

Definition

Stratified epithelium is a type of tissue in which there are several layers of epithelial cells, e.g. epidermis of skin.

► [Desmosomes](#)

Stratum Basale

Definition

The basal layer of the epidermis consists of a single cell layer of cuboidal cells, attached by hemidesmosomes to a basement membrane which separates it from the underlying dermis. It is the highly proliferative, amplifying compartment of the epidermis with cells interconnected by gap junctions.

► [Skin and Hair](#)

Stratum Corneum

Definition

Stratum corneum refers to the outer, cornified layer of the epidermis that consists of flattened, enucleated and fused cells that are filled with keratin and bound together by lipids. Most of the barrier function of the skin is provided by this layer, which is continually shed and regenerated.

► [Skin and Hair](#)

Stratum Granulosum

Definition

Stratum granulosum is the granular layer of the epidermis which contains keratinocytes with basophilic keratohyalin granules. The protein filaggrin is a major component of these granules and further complexes the keratin filaments. Lipid lamellae (Odland bodies) are extruded from lamellar granules into the intercellular spaces of the upper granular layer and serve as “intercellular cement”.

► [Skin and Hair](#)

Stratum Spinosum

Definition

Stratum spinosum is the spinous or prickle cell layer of the epidermis which is composed of differentiating keratinocytes. These are highly interconnected by

intercellular bridges (or prickles) containing desmosomes. Fibrillar ►[keratin](#) is produced and aggregates to form tonofibrils which are attached to the ►[desmosomes](#)

►[Skin and Hair](#)

Streptavidin

Definition

Streptavidin is a protein that is found in *Streptomyces avidinii*. It has a very high affinity for biotin, and is therefore often used in coupling procedures for bioanalytical purposes.

►[Surface Plasmon Resonance](#)

Stress Fibres

Definition

Stress fibres are contractile bundles of actin and myosin filaments that span across the cell attaching to the plasma membrane at ►[focal adhesions](#).

►[Rho](#), [Rac](#), [Cdc42](#)

Stress Response

Definition

The stress response is the set of physiological changes in response to a threat/pathogen or stress. In humans, it is sometimes called the “fight-or-flight” response. In a stressful situation, the body releases chemicals that trigger increased heart rate and blood pressure, rapid breathing, increased sweating, a sudden rush of strength, slowed digestion and dilated pupils. On a cellular level, stress response defines physiological changes in reaction to a variety of stressful (intracellular or extracellular) stimuli that are typically detrimental to the organism. Stress response refers to a genetic program in response to extracellular cues, which includes DNA repair, induction of survival functions, or initiation of the apoptotic program. Stress response usually involves the increased synthesis of stress proteins (►[heat shock response](#)), which have a protective effect against cellular stress.

►[Jun/Fos](#)

►[Proteomics in Human-Pathogen Interactions](#)

Stress-Induced Premature Senescence

Definition

Long-term appearance of the biomarkers of replicative senescence after exposure of human diploid cells to subcytotoxic stress.

►[Proteomics in Ageing](#)

Striae Atrophicae

Definition

Striae atrophicae are stretch marks of the skin resulting from atrophy of the ►[dermis](#) and overextension of the skin; they are not associated with pregnancy or marked weight changes.

►[Marfan Syndrome](#)

Stroma

Definition

Stroma designates a connective tissue made up of cells, such as fibroblasts, and matrix, such as collagen.

►[Kidney](#)

Structural Databases

The earliest protein structures were solved in the 1970s by X-ray crystallography, and later by NMR-spectroscopy, and deposited in the Protein Databank (PDB). This resource, which is now based in the Research Collaboratory of Structural Biology (RCSB) Rutgers University, contains around 30,000 structures of macromolecules (these are mainly protein structures but also include some nucleic acid- and carbohydrate structures), comprising of more than 60,000 individual

protein domains (► www.rcsb.org/pdb). Since the early 1990s, there have been sufficient structures to cluster evolutionary relatives into protein families and super-families. This has given rise to comprehensive hierarchical databases of protein structures, such as CATH and SCOP (structural classification of proteins, ► <http://scop.mrc-lmb.cam.ac.uk/scop/>), which rely on a combination of manual expert classification and structural comparison methods.

► [Protein Databases](#)

Structural Genomics

Definition

Structural genomics refers to a large-scale project to determine the three-dimensional shapes of all proteins and other important biomolecules encoded by the genomes of key organisms, using both experimental and computational techniques. A rational, genome-driven target selection, high-throughput innovations, and international coordination are key elements of structural genomics. Structural genomics is sometimes also called ► [structural proteomics](#).

► [Protein Databases](#)

► [Structure-Based Drug Design](#)

► [Structural Genomics: Structure-to-Function Approaches](#)

Structural Genomics: Structure-to-Function Approaches

UDO HEINEMANN

Max Delbrück Center for Molecular Medicine, Berlin, Germany

heinemann@mdc-berlin.de

Definition

According to international consensus, the term ► [structural genomics](#) (also sometimes referred to as ► [structural proteomics](#)) describes a large-scale project to determine the three-dimensional shapes of all proteins and other important biomolecules encoded by the genomes of key organisms. To reach this goal, structural genomics initiatives select protein targets for structure analysis following bioinformatics-driven

strategies based on known genome sequences. In order to complete a large number of structure analyses by X-ray crystallography or ► [NMR spectroscopy](#), high-throughput innovations in several preparative, analytical and computational aspects of protein structure determination are required and a degree of international coordination of the common effort is called for. Very often, little is known about the function of a protein after its three-dimensional structure has been elucidated. Although the structure often allows the formulation of hypotheses regarding the biochemical function of a protein molecule, additional computational and experimental analyses are usually required to derive function from structure.

Description

Structural genomics initiatives were formed in the late 1990s when it became clear that, after the completion of the sequence analysis of many genomes, the attention of researchers needed to focus on the gene products in order to gain an understanding of gene function. As a part of a broad research program in ► [functional genomics](#), the systematic, genome-driven and high-throughput crystal and NMR structure analysis of proteins was initiated. Gene products other than proteins have so far remained outside the focus of structural genomics. The new approach meant a paradigm shift for many structural biologists. Whereas, in the past, each laboratory invested much effort in the production and structural analysis of a few, biochemically well characterized protein molecules, now they were dealing with a multitude of different proteins, many with partially or completely unknown function. Herein lies a practical problem and intellectual challenge, which will be addressed in this paper. How may functional insight be derived from three-dimensional protein structure in the initial absence of relevant biochemical data? The concepts, scope and achievements of structural genomics will be briefly summarized before the main issue of deriving function from structure is addressed. With regard to this, alternative approaches, either relying on bioinformatics or based on experiment, will be discussed.

Structural Genomics

Structural genomics aspires “to determine the three-dimensional shapes of all proteins and other important biomolecules encoded by the genomes of key organisms” as stated at the Second International Structural Genomics Meeting held at Airlie House, VA, USA, in April 2001. This is an ambitious research agenda, indeed. It does not, however, imply that all these structures have to be determined experimentally. Accepting the assertion that a reasonably accurate protein structure can be modeled by a computer based on a template structure sharing at least 30% sequence

identity, one may conclude that approximately 16,000 carefully chosen new protein structures will be required to permit the ▶**homology modeling** of the large majority of all protein structures (1). For structure analysis by X-ray crystallography or NMR spectroscopy, these proteins must be selected by bioinformatics tools to represent sequence families of yet unknown three-dimensional structure. In addition to the protein structures already available in the ▶**Protein Data Bank**, these structures would represent the large majority of all ▶**domain folds** present in proteins.

In order to reach the goal of 16,000 or more new protein structures within a reasonable time, high-throughput innovations are needed to speed up the sample preparation and structure analysis steps. Laboratory automation, robotics and improved software have been developed over the past few years for expression cloning of protein-coding genes, for protein purification and biochemical and biophysical characterization, for X-ray diffraction experiments, for the acquisition of NMR spectra and the processing of the experimental data and for structure modeling and validation. These innovations permit the structure analysis of many proteins in parallel, but without loss of accuracy. Especially, but not exclusively for those structural genomics centers dealing with eukaryotic proteins, the yield of crystallizable soluble protein samples from recombinant sources has proven rate-limiting in the structure analysis. Most target selection schemes focus on the coverage of new sequence families and place little priority on prior knowledge of protein function. For example, schemes have been developed that select proteins for structure analysis in part by crystallizability. The structures of these proteins may hold very valuable functional clues which can give the analysis added scientific impact, but often the protein function will not be immediately revealed without further analysis.

In a similar way to genomic sequencing, structural genomics stands to benefit from international coordination. Under the guidance of the International Structural Genomics Organization (ISGO, ▶<http://www.isgo.org/>) modes of international cooperation in structural genomics have been established. The most visible aspect of this cooperation is the construction and maintenance of a ▶**Target Registration Data Base** (Target DB) at the Protein Data Bank. 20 structural genomics consortia from around the world enter their target proteins into the Target DB and allow the database to automatically monitor the progress of the structure analysis of these proteins (Table 1). The 20 projects contributing to the Target DB are the core of the worldwide structural genomics effort.

Through structural genomics, a renewed emphasis has been placed on methods developments in all aspects of protein structure analysis from which the structural

biology sector as a whole stands to benefit. For the US structural genomics pilot centers forming the Protein Structure Initiative (PSI), these innovations have allowed a significant increase in the output of new protein structures over recent years and a significant reduction in the cost per structure. As of August 10, 2004, world-wide 36,603 (32,531) protein-coding genes had been cloned, 7138 (5783) proteins had been purified, 924 (683) crystal and 343 (64) NMR structures had been determined and 940 (624) structure entries had been added to the Protein Data Bank according to the summary status report from the Target DB. The numbers in parentheses denote the contribution of the nine PSI pilot centers. Compared to the number of 25,348 structures (proteins, peptides, viruses, and protein-nucleic acid complexes) currently archived in the Protein Data Bank, the 940 structures contributed by structural genomics projects may still seem little. It must be pointed out, however, that the content of the Protein Data Bank is highly redundant with multiple copies of the same protein molecule or variants present, so that the contribution of structural genomics to the number of domain folds or sequence families covered is probably much higher. Unfortunately, reliable statistics regarding the source of protein structures representing new sequence families in the Protein Data Bank are not available.

In addition to assigning a three-dimensional fold to a sequence family, a structure determined in structural genomics projects, as any other protein structure, may provide a wealth of insight into the function of the molecule. In view of the accelerated rate at which these structures can now be determined and the typical scarcity of biochemical information accompanying the proteins, assigning function at a rate matching that of structure determination is now a major challenge. This is akin to moving “from words to literature in structural proteomics” (2). We are mainly interested in the type of functional insight that can be deduced from the three-dimensional structure, but not from the sequence.

In the following discussion of functional assignments from structure, function is always taken to mean “biochemical function”. How a protein molecule performs its enzymatic role, how it physically interacts with large and small ligands and how it cooperates to form larger cellular structures can often be illuminated by the three-dimensional structure. Conversely, crystal or NMR structures have very limited direct use for explaining biological function. They cannot tell us whether a protein arrests the cell-cycle, in which signaling pathway it is active or how it affects the phenotype of an organism.

Function from Structure: Bioinformatics

Bioinformatics lends itself to high throughput and may thus be employed for functional assignments based on

Structural Genomics: Structure-to-Function Approaches. Table 1 Compilation of worldwide structural genomics initiatives. The consortia listed here report their work progress to the Target Registration Data Base of the Protein Data Bank at Rutgers University, USA. The Pilot Centers forming the Protein Structure Initiative (PSI) funded by the National Institutes of General Medical Science of the National Institutes of Health, USA, are marked (*)

Project	Acronym	PI	Focus
Bacterial Targets IGS	BIGS	J.-M. Claverie, CNRS Institut de Biologie Structurale et Microbiologie, Marseilles, France	Proteins from <i>Rickettsia</i> , <i>Escherichia coli</i>
Berkeley Structural Genomics Center*	BSGC	S.-H. Kim, University of California at Berkeley, USA	Proteins from <i>Mycoplasma genitalium</i> and <i>M. pneumoniae</i>
Montreal-Kingston Bacterial Structural Genomics Initiative	BSGI	M. Cygler, Biotechnology Research Institute, Montreal, Canada	Proteins from <i>Escherichia coli</i>
Center for Eukaryotic Structural Genomics*	CESG	J. Markley, University of Wisconsin, Madison, USA	Proteins from <i>Arabidopsis thaliana</i>
Israel Structural Proteomics Center	ISPC	J.L. Sussman, Weizmann Institute of Science, Rehovot, Israel	Bacterial proteins, disease-related human proteins
Joint Center for Structural Genomics*	JCSG	I. Wilson, Scripps Research Institute, La Jolla, CA, USA	Proteins from <i>Thermotoga maritima</i> and <i>Caenorhabditis elegans</i>
Midwest Center for Structural Genomics*	MCSG	A. Joachimiak, Argonne National Laboratory, Argonne, IL, USA	Bacterial and archaeal proteins (mostly)
Marseille Structural Genomics Platform	MSGP	C. Cambillau AFMB UMR 6098, CNRS, Marseille, France	Proteins from <i>Escherichia coli</i>
Northeast Structural Genomics Consortium*	NESG	G.T. Montelione, Rutgers University, NJ, USA	Proteins from <i>Saccharomyces cerevisiae</i> , <i>Caenorhabditis elegans</i> and <i>Drosophila melanogaster</i>
New York Structural Genomics Research Consortium*	NYSGRC	S.K. Burley, Structural Genomics, Inc., San Diego, CA, USA	Yeast, bacterial, archaeal and other proteins
Oxford Protein Production Facility	OPPF	D. Stuart, Wellcome Trust Centre for Human Genetics, Oxford, UK	Human proteins and proteins from human pathogens
Protein Structure Factory	PSF	U. Heinemann, Max-Delbrück Center for Molecular Medicine, Berlin, Germany	Human proteins of presumed medical relevance
RIKEN Structural Genomics/Proteomics Initiative	RIKEN	S. Yokoyama, University of Tokyo, Japan	Proteins from mouse, <i>Arabidopsis thaliana</i> and <i>Thermus thermophilus</i>
Structure 2 Function Project	S2F	O. Herzberg, Center for Advanced Research in Biotechnology, Rockville, MD, USA	Proteins from <i>Haemophilus influenzae</i> and bacterial homologs
Southeast Collaboratory for Structural Genomics*	SECSG	B.-C. Wang, University of Georgia, Athens, USA	Proteins from <i>Homo sapiens</i> , <i>C. elegans</i> and <i>Pyrococcus furiosus</i>
Structural Genomics of Pathogenic Protozoa*	SGPP	W.G.J. Hol, University of Washington, Seattle, USA	Proteins from <i>Trypanosoma cruzi</i> , <i>Trypanosoma brucei</i> , <i>Leishmania</i> spp., <i>Plasmodium falciparum</i> and <i>Plasmodium vivax</i>
Structural Proteomics in Europe	SPINE	D. Stuart, Wellcome Trust Centre for Human Genetics, Oxford, UK	Bacterial, viral and human proteins

Structural Genomics: Structure-to-Function Approaches. Table 1 Compilation of worldwide structural genomics initiatives. The consortia listed here report their work progress to the Target Registration Data Base of the Protein Data Bank at Rutgers University, USA. The Pilot Centers forming the Protein Structure Initiative (PSI) funded by the National Institutes of General Medical Science of the National Institutes of Health, USA, are marked (*) (Continued)

Project	Acronym	PI	Focus
<i>Mycobacterium tuberculosis</i> Structural Genomics Consortium*	TB	T.C. Terwilliger, Los Alamos National Laboratory, NM, USA	Proteins from <i>Mycobacterium tuberculosis</i>
German Mycobacterium Tuberculosis Structural Genomics Consortium	XMTB	M. Wilmanns, EMBL Outstation at DESY, Hamburg, Germany	Proteins from <i>Mycobacterium tuberculosis</i>
Yeast Structural Genomics	YSG	J. Janin, Université Paris-Sud 11, Gif-sur-Yvette, France	Proteins from <i>Saccharomyces cerevisiae</i>

structure. The first check a crystallographer or NMR spectroscopist will perform once a new protein structure has been unveiled is to look for structural homologs already present in the Protein Data Bank, using standard bioinformatics tools such as DALI, VAST or DEJAVU (3, 4, 5). Even if his structure represents a new sequence family, more often than not will he find a similar fold in a protein of known function. From this, the hypothesis may be generated that the protein of interest shares a similar function, since it is known that the vast majority of known [protein folds](#) is associated with one type of function only. The hypothesis thus generated may subsequently be proven or disproven by experiment. Obviously, this approach will fail for some of the very common folds such as the TIM barrel, the P-loop NTP hydrolases or the Rossmann fold, which all support numerous biochemical functions, but it allows the formulation of reasonable hypotheses for many other proteins. Likewise, heterologous interactions of our protein of interest with other folds, which may be of biological relevance, may be predicted based on the structure, since, again, the majority of folds is associated with one or very few interacting folds.

In general, computer-based approaches have been of limited use in assigning functions to proteins adopting a fold not observed before. Enzymes may be an exception to this rule, since their active sites tend to be defined by strict spatial relationships between few amino acid side chains or backbone segments. If this precise arrangement of structural elements is present in a new protein structure, there is a good possibility that the enzymatic function is retained. At least this observation, which may be derived from checking the protein structure against an atlas of active-site geometries (6), will give rise to a highly detailed hypothesis (that the protein is an enzyme, which chemical reaction it catalyzes, which substrate it accepts) that can be

experimentally tested. Alternative approaches for identifying enzymes based on three-dimensional structure and not on sequence and for predicting their functions have recently been described. One uses a vector machine-learning algorithm based on protein secondary structure, amino acid propensities and surface properties to discriminate enzymes from non-enzymes (7). Another analyzes protein surface charges to locate potential active sites in proteins (8).

Function from Structure: Experimental Approaches

Even without sophisticated bioinformatics analysis it is often possible to find clues for function in a novel protein structure. A decade ago, we determined the crystal structure of the major cold shock protein, CspB, from *Bacillus subtilis* (9). At the time of the structure analysis, proteins of this family had been reported to be involved in gene regulation, but there were also ideas that they might act as antifreeze proteins. By revealing one basic face on the surface of an acidic protein with a number of unusually exposed aromatic side chains anchored in sequence motifs known to be involved in RNA binding, the crystal structure of CspB clearly had the looks of a single-strand nucleic acid-binding protein. This *ad hoc* hypothesis could be corroborated by demonstrating that CspB indeed bound preferentially to one strand of a 35-bp promoter fragment, but not to the double strand or to the complementary single strand (9). Only recently, however, have we been able to show by crystallographic analysis that indeed exactly those amino acids that had been predicted from the structure of the isolated CspB (Max et al., unpublished) are involved in binding of a 6-nt ssDNA fragment.

Another well-documented example of protein function deduced, *via* experiment, from structure concerns the tubby proteins. These are found in various multicellular eukaryotes and are associated with at least three human

disease phenotypes. As in the case of CspB, the molecular architecture, revealed by crystal structure analysis, was considered ideal for DNA binding (10). This finding, along with the location in the protein of disease-related point mutations, prompted experiments that proved the binding of tubby to double-stranded DNA, its nuclear localization and transcription activation by tubby.

In a structural genomics context, it is imperative that functional assays can follow the structure analysis swiftly in order to avoid undue delay in making the structural data available to the public. There are a number of examples on record to show that this is possible. The protein YrbI from *Haemophilus influenzae* (HI1679) was shown to adopt an α/β -hydrolase fold resembling haloacid dehalogenases. Based on this observation, enzymatic assays on a number of potential substrates were performed that showed YrbI to be a Mg^{2+} -dependent phosphatase. A new class of truncated thioredoxins was defined following the crystal structure analysis of the archaeal protein MTH895 and structure-prompted experiments proving the thioredoxin activity. Likewise, the archaeal protein TM1643 could be shown to be an L-aspartate dehydrogenase by structure analysis followed by enzymatic screening. A combination of crystal structure analysis, bioinformatics and biochemistry was employed to reveal the structure and function of the *E. coli* protein BioH, which was known to be involved in biotin biosynthesis but had no assigned biochemical function. Screening the protein structure against a library of enzyme active sites (6) led to the discovery of an enzyme active site harboring a Ser/His/Asp catalytic triad characteristic of hydrolases. A series of hydrolase assays then identified short-chain carboxylesters as preferred substrates and marked BioH as a carboxylesterase.

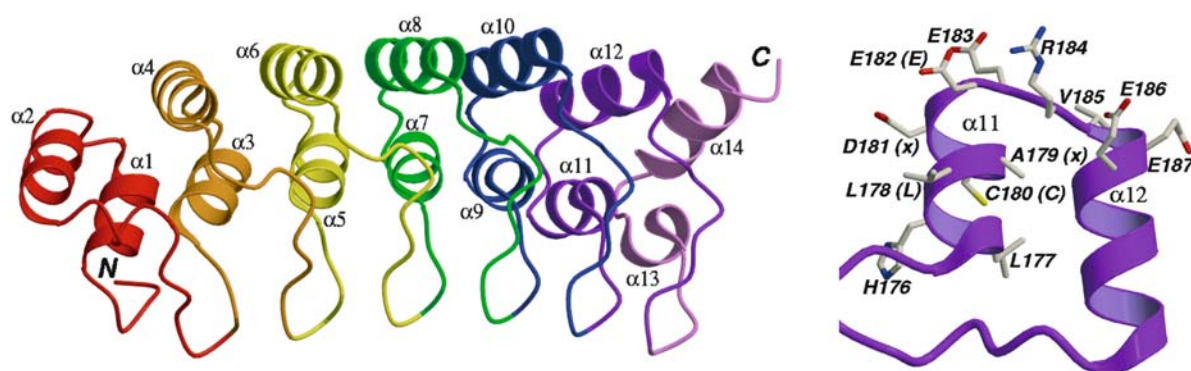
Metal centers in proteins often mark active sites and can guide a functional annotation, either directly from the structure or *via* experiment. The *Methanococcus jannaschii* protein MJ0936 has a three-dimensional structure reminiscent of nucleases, phosphatases or nucleotidases and possesses a binuclear metal cluster when crystallized from solutions containing either Ni^{2+} or Mn^{2+} . Enzymatic activity was found against indicator substances for phosphatases, clearly annotating this protein as a metal-dependent phosphatase. The human protein FLJ36880 of unknown function was shown by crystal structure analysis to be dimeric and to have one active site per subunit in which a Mg^{2+} ion is coordinated by three carboxylate groups. Structure and sequence comparisons indicated that this protein is a fumarylacetoacetate hydrolase family member with yet uncharacterized substrate specificity.

Occasionally, structure analysis can serve to correct misannotations in databases as in the case of the protein product of the open reading frame Rv3853 from

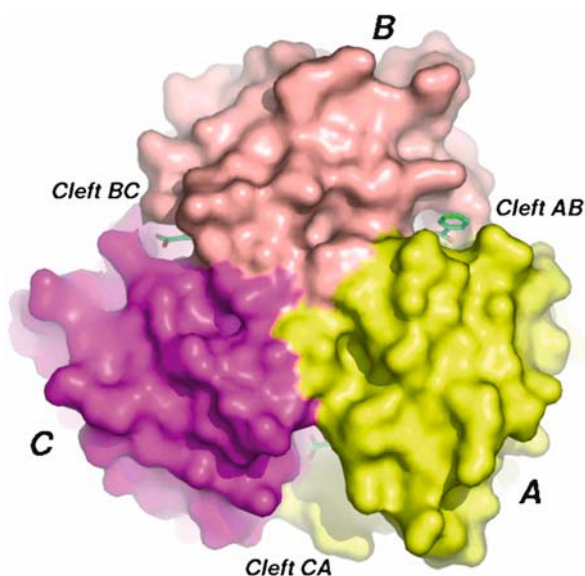
Mycobacterium tuberculosis, which had been annotated as an S-adenosylmethionine (SAM)-dependent methyltransferase and assumed to catalyze the final step in menaquinone biosynthesis (gene denoted *menG*). The crystal structure clearly showed this protein to have a structure completely unlike all known SAM-dependent methyltransferases and instead suggests a role as subunit of a larger assembly binding nucleic acids or proteins. A similar question mark is set by the crystal structure of the human oncoprotein gankyrin linked to hepatocellular carcinoma (11). Gankyrin has been described to interact with and modify the phosphorylation status of the retinoblastoma protein (Rb), and an LXCXE sequence motif present in gankyrin was implicated in this process. However, the crystal structure showed this sequence motif in gankyrin to adopt an α -helical conformation (Fig. 1), whereas in other Rb-binding peptides the LXCXE sequence is extended, thus casting doubt on the proposed gankyrin-Rb interaction.

Small molecules bound to proteins can be another source of functional clues from crystal and NMR structures. For example, the human protein hp14.5, a member of the large YjgF/YER057c/UK114 protein family was found to adopt a trimeric structure in crystals and to bind benzoate molecules picked up from the crystallization buffer at its inter-subunit clefts (Fig. 2). The clefts are lined by residues conserved in this protein family (12). Both hp14.5 and its rat homolog, L-PSP, have been described to act as translational inhibitors, the latter protein also being ascribed ribonuclease activity against single-stranded RNA. Therefore, it seems likely that the benzoate molecules mark a hydrolytic active site on the surface of hp14.5, even though the physiological substrates remain to be identified (12). This may be achieved by NMR-based ligand screening as shown for the hp14.5 homolog HI0719 from *Haemophilus influenzae*. By this technique, HI0719 was shown to bind several small α -keto acids or α - β -unsaturated acids, i.e. molecules with chemical similarity to benzoate. The same residues in the inter-subunits clefts were involved as in hp14.5, corroborating the assignment of this cleft as the site of enzymatic activity in YjgF/YER057c/UK114 proteins. NMR or co-crystallization screening for ligands using appropriate libraries is in principle automatable and could develop into a valuable technique for structure-based functional annotations of proteins.

Post-translational modifications of proteins can easily be detected in crystal and NMR structures of proteins. They also can be sources of functional information, since they are characteristic of the subcellular location of proteins in eukaryotic cells and indicative of the functional state of a protein. Unfortunately, proteins produced recombinantly in *E. coli* will only carry a subset of the post-translational modifications observed



Structural Genomics: Structure-to-Function Approaches. Figure 1 Crystal structure of the human oncoprotein gankyrin, the product of the first gene shown to be associated with hepatocellular carcinoma. The structure of gankyrin (11) is organized into five complete and two incomplete ankyrin repeats (left). An LXCXE sequence motif (¹⁷⁸LACDE¹⁸² in gankyrin) usually implicated in binding to the retinoblastoma (Rb) tumor suppressor protein adopts an α -helical conformation in gankyrin (right). LXCXE peptides have previously been found in an extended conformation when bound to the pocket domain of Rb (11). Figure taken and modified from (11).



Structural Genomics: Structure-to-Function Approaches. Figure 2 Crystal structure of the trimeric human protein hp14.5 (12). The inter-subunit clefts are lined with residues highly conserved in members of the large YjgF/YER057c/UK114 protein family spanning all kingdoms of life. Benzoate molecules picked up from the crystallization buffer bind to the clefts, probably marking them as active sites of hydrolytic activity. The identity of the substrates turned over by hp14.5 remains to be determined. Figure courtesy of B.A.Manjasetty.

in eukaryotic proteins. This is one reason why it might be desirable to invest work in developing high-throughput techniques for protein production in recombinant yeast, insect, mammalian or other eukaryotic cells.

In conclusion, it seems that structural genomics has much to gain from techniques allowing the assignment of function to initially uncharacterized proteins on the basis of their three-dimensional structure. There are examples demonstrating how it can be achieved, but the development of automatable methods for high-throughput enzymatic assays or ligand-binding screens is only just starting.

References

1. Vitkup D, Melamud E, Moult J et al (2001) Completeness in structural genomics. *Nat Struct Biol* 8:559–566
2. Sali A, Glaeser R, Earnest T et al (2003) From words to literature in structural proteomics. *Nature* 422:216–225
3. Holm L, Sander C (1993) Protein structure comparison by alignment of distance matrices. *J Mol Biol* 233:123–138
4. Madej J, Gibrat JF, Bryant SH (1995) Threading a database of protein cores. *Proteins Struct Funct Genet* 23:356–369
5. Kleywegt GJ, Jones TA (1997) Detecting folding motifs and similarities in protein structures. *Methods Enzymol* 277:525–545
6. Porter CT, Bartlett GJ, Thornton JM (2004) The Catalytic Site Atlas: A resource of catalytic sites and residues identified in enzymes using structural data. *Nucleic Acids Res* 32:D129–D133
7. Dobson PD, Doig AJ (2003) Distinguishing enzyme structures from non-enzymes without alignments. *J Mol Biol* 330:771–783
8. Bate P, Warwicker J (2004) Enzyme/non-enzyme discrimination and prediction of enzyme active site location using charge-based methods. *J Mol Biol* 340:263–276
9. Schindelin H, Marahiel MA, Heinemann U (1993) Universal nucleic acid-binding domain revealed by crystal structure of the *B. subtilis* major cold-shock protein. *Nature* 364:164–168

10. Boggon TJ, Shan WS, Santagata S et al (1999) Implication of tubby proteins as transcription factors by structure-based functional analysis. *Science* 286: 2119–2125
11. Manjasetty BA, Quedenau C, Sievert V et al (2004) X-ray structure of human gankyrin, the product of a gene linked to hepatocellular carcinoma. *Proteins Struct Funct Bioinf* 55:214–217
12. Manjasetty BA, Delbrück H, Pham D-T et al (2004) Crystal structure of *Homo sapiens* protein hp14.5. *Proteins Struct Funct Bioinf* 54:797–800

Structural Proteomics

Definition

Structural proteomics is a term sometimes used for ► [structural genomics](#).

► [Structural Genomics: Structure-to-Function Approaches](#)

Structure-based Drug Design

GERHARD KLEBE

Institute of Pharmaceutical Chemistry, Philipps-University, Marburg, Germany
klebe@mail.uni-marburg.de

Definition

Structure-based design starts from the assumption that a drug molecule exerts its biological activity through specific binding to a macromolecular target receptor, usually a protein. In consequence, the biological function of this target protein is modulated and hopefully this process leads to the cure of the disease. Prerequisite for strong and selective target binding is a perfect structural and chemical complementarity of ligand and receptor. The strategy of “structure-based design” comprises a variety of methods to discover and optimize small molecule ligands that fit precisely into a depression at the surface of a protein where the actual function of the protein is usually carried out. Through perfect and high affinity binding the ligand modulates the function of the protein. The design process starts with the information that can be extracted from the shape and composition of the binding pocket, estimates the properties a putative ligand must exhibit to be accommodated and tries to translate these properties into molecular skeletons of possible drug molecules. After synthesis and biological testing, the structure of the designed ligand together with the protein is determined (Fig. 1). This complex is the starting point for a new design cycle (Fig. 1) that

hopefully results after several cycles in a prospective drug candidate for clinical trials.

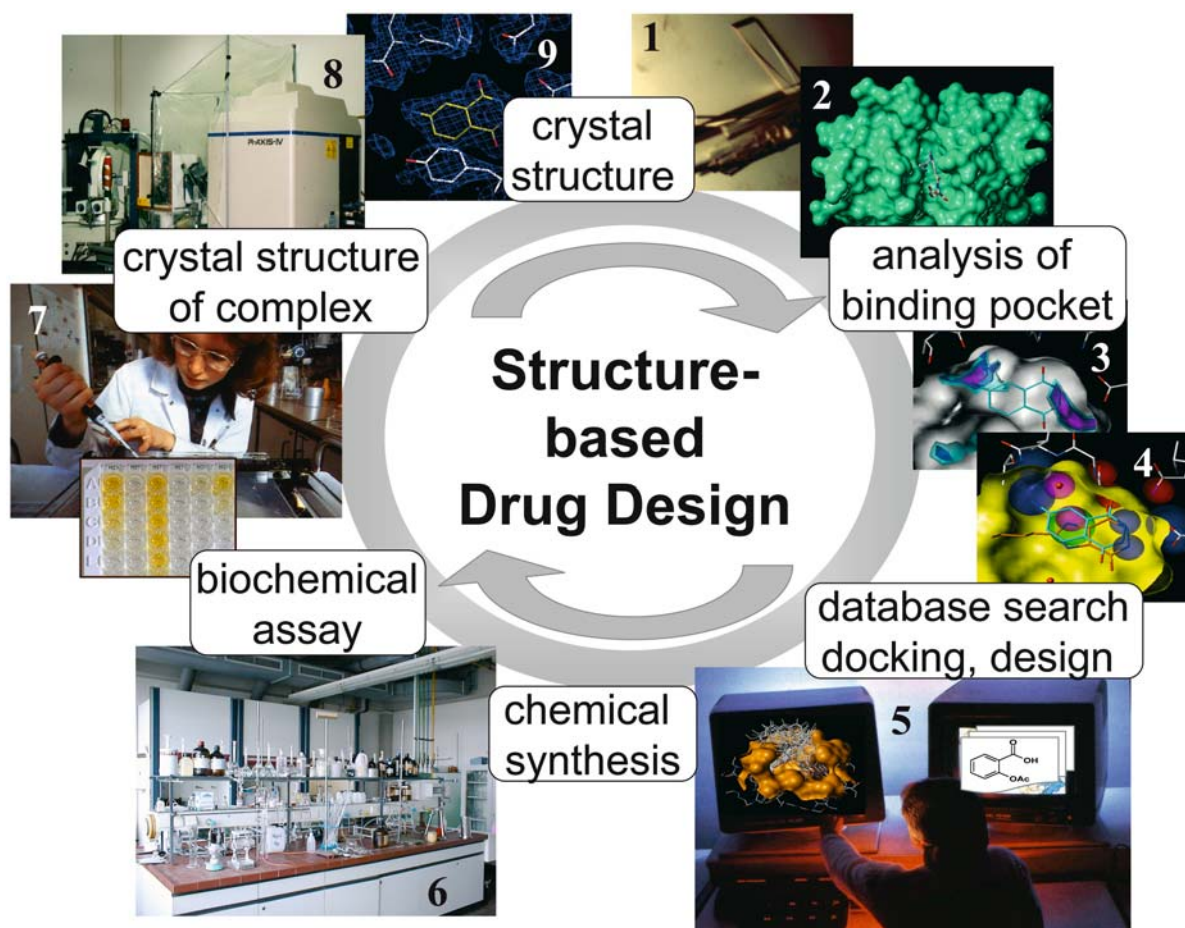
Characteristics

Structure-based Drug Design Needs Structures

The number of proteins of known 3D structure has increased exponentially over the last few years. Whereas in 1988 only 200 structures had been published, the number has now increased to nearly 30,000 examples. Of these, many are proteins with bound ligands. This dramatic growth is mainly due to technological and methodological improvements in gene technology, protein chemistry, protein purification and structure determination techniques. It is predominantly ► [protein crystallography](#) and multidimensional high-resolution ► [NMR spectroscopy](#) besides ► [cryo electron crystallography](#) that give us experimental access to the spatial arrangement of atoms in molecules. While many of the structures of globular water-soluble proteins have been determined, the number of structure determinations of membrane-bound proteins is steadily increasing. After the human genome project has been accomplished, advances in ► [proteomics](#) and ► [bioinformatics](#) have opened the floodgates to thousands of new proteins that might serve as putative targets for drug discovery. Worldwide efforts in ► [structural genomics](#) now aim to obtain a comprehensive view of the world of proteins. In addition, this impressive increase in experimentally resolved protein structures will be accompanied by an even larger number of protein models computed by ► [homology modelling](#). Thus, in the near future we will be increasingly faced with the situation that the 3D structure of a therapeutically relevant protein is available at an early stage of a drug development project. However, suitable ligands ready to interfere with the biological function are missing.

What is Required: The Complement of the Binding Site

Structure-based drug design starts from a detailed analysis of the binding pocket (Fig. 1). Usually such a binding pocket is the site in a protein where an endogenous ligand such as an ► [agonist](#), ► [antagonist](#), ► [effector](#) or ► [substrate](#) is recognized and in the case of enzymes chemically modified. This recognition of small molecule ligands or molecular portions of substrate molecules occurs in well characterized clefts or cavities located at the protein surface or accessible from the outside by some mechanism involving globular protein movement. Recognition between molecules is effected by the formation of noncovalent contacts among ligand and protein ► [functional groups](#). These groups and the overall shapes of ligand and binding pocket must be mutually complementary, similar to a lock and its key, to efficiently establish these interactions. Thus, analysing the recognition



Structure-based Drug Design. Figure 1 Structure-based drug design starts with the detailed analysis of the binding pocket of the target protein (1). Using different probes the binding pocket is analyzed in terms of “hot-spots” of binding (2, 3). The centers of such areas, indicated as beneficial for the occupation with particular ligand functional groups, are contoured and translated into a protein-based pharmacophore (4). The latter is then used to constrain a database screening on the computer (5). Docking is applied in the last step of a sequential filtering strategy and the generated solutions are ranked according to the predicted affinities. After synthesis (6) and biological testing (7) the discovered hits are soaked or co-crystallized with the protein to determine their binding modes (8). Their structures serve as a starting point for the following drug design cycle (9).

properties exposed by the protein in the binding pocket allows one to estimate what the prerequisites on the side of the ligand will be. The composite picture of all these properties required by a ligand is called “the pharmacophore”. If this ligand-defining [▶pharmacophore](#) is established using binding-site features, it is also called a “protein-based pharmacophore hypothesis”.

A Quantitative Picture: Mapping “Hot Spots” of Binding

To find quantifiable measures for such a hypothesis several tools have been developed that map the binding-site features fast and efficiently. The archetypal method in this area is Peter Goodford's GRID program. It is based on a [▶molecular mechanics force field](#). The

entire area of the binding pocket is analyzed and special, more local regions such as sub-pockets are contoured in terms of interaction energies that various types of functional groups experience at the intersections of a regularly spaced grid. Other methods use information about frequently observed contact geometries to identify the “hot spots” of binding in the binding pocket (Fig. 1).

As a subsequent step, this information about the hot spots of binding that define the type of functional groups required by the protein to be present in a putative ligand has to be translated into real molecules. This step is not trivial, since the information about the type of functional group is rather generic and only indicates the kind of interaction required, e.g. “a hydrogen-bond donor or acceptor” or “a bulky hydrophobic group”.

Docking: Fitting of Molecular Skeletons into the Binding Pocket

Most approaches in structure-based drug design try to place actual molecules into the binding pocket (Fig. 1). A force field together with molecular dynamics can be used to guide the placement of the ligand into the binding pocket. Much faster methods exist that exploit the results of the hot-spot analysis, classify the ligand in terms of its functional groups and test for complementarity in the placement or docking step. After such a first small molecule fragment (a "seed") has been placed, two strategies can be followed. In a *de novo* design approach, further fragments are retrieved from a library and subsequently attached to the initial seed to finally result in a molecule fitting as well as possible. The consecutive addition of new fragments should consider some rules about the synthesis of molecules so that the ligands finally suggested are easy to make. As an alternative, the complete molecules to be docked into the binding pocket can be taken from a library. *Via* computational simulation they are decomposed into fragments and, starting with the placement of a seed or anchor fragment, the entire molecule is constructed stepwise in an incremental build-up procedure. These docking methods have been accelerated so that nowadays libraries of up to several million compounds can be screened on the computer with respect to their putative protein binding. However, the different suggestions produced by docking have to be compared and ranked in terms of the expected **binding affinities**. The problem of predicting binding affinity based solely on the generated ligand binding geometry is a very difficult task and substantial effort is at present being given to this essential step in computer screening (also termed virtual screening). The libraries used for docking runs are either compiled from existing molecules that are in stock or from those that can be purchased after being identified as potential hits. Alternatively, libraries of virtual compounds can be screened and synthesis postponed to a later stage considering only the most likely hits. In recent years, much effort has been spent in the optimization of such screening libraries to focus the search on molecules that are predicted to possess satisfactory bioavailability properties.

Iterative Design: A Stepwise Optimization from Lead to Drug Candidate

As a next step, the hits detected by computer methods are confirmed experimentally in a functional assay and subjected to a structure determination together with the protein. This elucidates the actually adopted binding geometry of the first hits in the binding pocket. In a subsequent step, the information provided by their binding geometry is used to either improve the protein-based pharmacophore hypothesis for a further screening cycle or, through analysis of the observed binding mode of the first hits, a hypothesis can be established as to how

to improve the interactions of these ligands with the protein. Usually this newly designed compound is the starting point for a further design cycle involving docking, computer modelling, chemical synthesis, biological testing and structure determination (2, 3).

Today structure-based drug design is used as a routine tool in the development of new drugs. Many projects have been reported where this technique has significantly contributed to the discovery and development of clinical candidates and marketed drugs (4, 5, 6). However, the process of ligand-protein binding is very complex and at present only poorly understood. A particular complication arises from the fact that the conformations of the ligand and the protein mutually adapt to each other. It can be expected that ongoing research will contribute to a more profound understanding of the basic principles that determine selective and tight ligand-receptor recognition. This understanding will be a prerequisite for improving the concepts and strategies in structure-based drug design. In many systems protein function is triggered by the formation of protein-protein complexes. Evidence has been collected that the formation of such protein-protein complexes, resulting in large contact interfaces of several hundred square-angstroms, can also be modulated by small molecules. However, today it is not understood how this modulation is performed in structural and energetic terms. Nevertheless, it can be expected that, once the principles of such interference are known, a structure-based route to novel and innovative drugs will be opened.

References

1. Goodford P (1984) Drug Design by the Method of Receptor Fit. *J Med Chem* 27:557–564
2. Greer J, Erickson JW, Baldwin JJ et al (1994) Application of the Three-Dimensional Structures of Protein Target Molecules in Structure-Based Drug Design. *J Med Chem* 37:1035–1054
3. Kuntz ID (1992) Structure-Based Strategies for Drug Design and Discovery. *Science* 257:1078–1082
4. Babine RE, Bender SL (1997) Molecular Recognition of Protein-Ligand Complexes. *Chem Rev* 97:1359–1472
5. Böhm HJ, Klebe G, Kubinyi H (1996) *Wirkstoffdesign*, Spektrum Akad. Verlag, Heidelberg

3D Structure Determination by NMR

ZHEN-YU J. SUN, GERHARD WAGNER

Department of Biological Chemistry and Molecular Pharmacology, Harvard Medical School, Boston, MD, USA

zhen-yu_sun@hms.harvard.edu

gerhard_wagner@hms.harvard.edu

Definition

► **Nuclear magnetic resonance (NMR)** is a technique used for observing signals from isotopes with magnetic moments. For biological applications these are primarily ^1H , ^{13}C , ^{15}N , ^{19}F , ^{32}P and occasionally ^{113}Cd , which can sometimes be used to substitute for protein-bound metals, such as zinc or calcium. Between 0.2 and 0.5 ml of dissolved sample in concentrations of 0.2 mM or higher are needed for the experiments and placed in a glass tube in the center of a high-field magnet. Irradiation of the samples with trains of radio frequency pulses, termed ► **pulse sequences**, produces response signals that allow detection of ► **correlations** between different nuclei. This can reveal information about the three-dimensional structure and dynamic state of biological macromolecules, such as proteins or polynucleotides. Proteins can adopt well-defined 3D structures or form flexible chains, each exhibiting characteristic NMR signals. In the case of folded protein, the protein structure can be determined using a sequence of procedures that have evolved over the past two decades. The routines used for structure determination depend on whether proteins can be enriched with NMR-active ► **stable isotopes**. If this is not possible, one has to rely on spectroscopy with ^1H only, since this is the only NMR-active stable isotope found at high abundance in proteins. This limits structure determination to proteins of less than 15 kD. In order to study larger proteins and to facilitate structural analysis, most proteins studied today are labeled with ^{13}C , ^{15}N and occasionally ^2H , using bacterial expression systems. This enables structure determinations of proteins in the 40 kD range and beyond.

Description

Preparation of Protein Samples Suitable for NMR Structure Determination

The most important and time-consuming step is the production of a sample that is suitable for the process of protein solution structure determination. A typical NMR sample has a protein concentration of 0.2–1.0 mM in 0.2–0.5 ml buffer at a pH of less than 8 and should ideally be stable at room temperature for one week or longer. Often, the protein solubility and stability can be improved by increasing the salt concentration (up to 300 mM NaCl) and by variation of pH, temperature and buffer, as well as by adding small amounts of detergents. In some cases, new protein constructs are required in which a few surface-exposed non-essential hydrophobic residues are substituted with hydrophilic ones. Occasionally, a small ► **SET** (solubility enhancement tag) is attached, such as the highly soluble 6 kD B1 domain of streptococcal protein G (GB1), to make the total construct more soluble and inhibit irreversible precipitation. The procedures for isotopically enriching proteins for

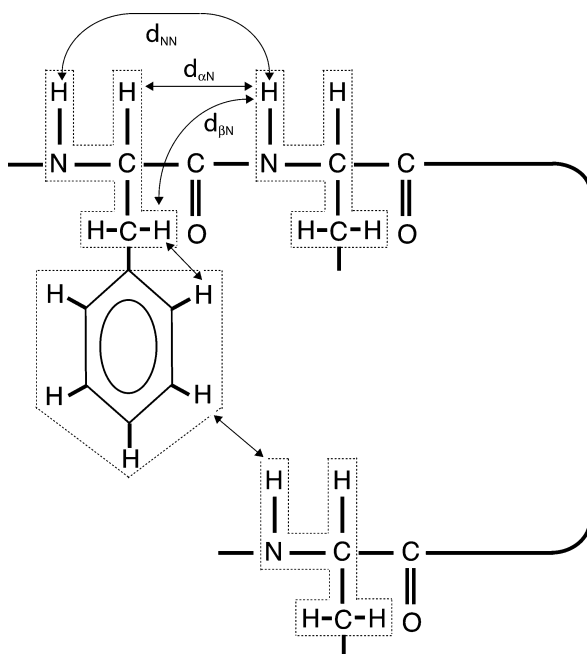
NMR studies sometimes require serious efforts to optimize expression strategies and develop purification protocols. Although different expression systems for labeling have been developed, including insect cell, yeast, slime mold and mammalian cell lines, the expense of the stable isotopes limits the expression of most proteins to the old-fashioned *E. coli* as a host. Different protein expression vectors can result in varying amounts of total, soluble and insoluble (inclusion body) proteins being produced. To facilitate purification, soluble proteins are often expressed with a glutathione S transferase (GST), protein G B1 (GB1) or hexa-histidine tag that can be removed subsequently by enzymatic cleavage. Insoluble proteins are easy to purify but must be painstakingly refolded into their native conformations for structural studies. Impurities remaining in the NMR sample after purification are not of serious concern if they constitute less than 5% of the solute and do not include endogenous proteases.

Sequence Specific Assignments for Unlabeled Proteins

For short peptides and small proteins, it is possible to complete the ^1H resonance assignments without isotope labeling. The strategy used in this case relies only on two-dimensional experiments as shown in Fig. 1 (1). The first step is to assign the amino acid spin system of each residue by through-bond ^1H - ^1H correlations with ► **COSY**-type experiments. The experiments used can correlate through the scalar coupling spins that are separated by no more than three chemical bonds. The second step is to find sequential through-space connectivities between the backbone amide protons and protons of the preceding residue. This step uses the dipole-dipole interaction, which gives rise to cross peaks in two-dimensional ► **NOE** (nuclear Overhauser effect) spectroscopy (► **NOESY**). In the third and final step, the sequentially connected amino acid spin systems are matched to the protein sequence. At this point, the spin-system topologies are checked for consistency with those expected from the primary sequence.

Sequential Backbone Resonance Assignments for Labeled Proteins

Labeling of proteins with ^{13}C and ^{15}N dramatically facilitates structure determination. In this case, backbone and side-chain ^1H , ^{13}C and ^{15}N resonances can be assigned efficiently and unambiguously with a set of so-called triple-resonance ($^{15}\text{N}/^{13}\text{C}/^1\text{H}$) experiments (2, 3). This set typically includes three pairs of experiments termed HNCA and HN(CO)CA, HNCB and HN(CA)CO and HNCACB and HN(CO)CACB. The letters indicate the nuclei that are connected by the particular pulse sequence (Fig. 2). The experiments are recorded in pairs where the N-H group is correlated with either the CA, the CO or the CB signals of the



3D Structure Determination by NMR. Figure 1
Schematic illustration of the sequential assignment strategy for unlabeled proteins. The intra-residue spin systems identified from COSY-type experiments are shown as enclosed by dotted lines. The sequential NOE connectivities are shown by solid lines with arrows.

same or the preceding residue. The intra-residue correlations between the carbon and the backbone amide resonances are compared with the sequential inter-residue correlations and the matching pairs are linked to a thread of spin systems assigned to the protein amino acid sequence. This set of six triple-resonance experiments run for obtaining the complete backbone assignments typically takes a total of 6–12 days, depending on the protein concentration and molecular weight. By using ▶non-uniform sampling techniques and cryogenic probes (4), the total experimental time for recording the set of six triple resonance experiments can be shortened to one day. The spectral data can be analyzed with the aid of computer automation programs to establish sequential assignments within a few hours after the spectra have been recorded and processed.

Side-Chain Assignments for Labeled Proteins

Side-chain assignments for labeled proteins are typically pursued with the HCCH-TOCSY experiment, which connects all ^{13}C - ^1H resonances that belong to the same amino acid spin system (2). The aliphatic ^{13}C and ^1H resonances are then correlated with the backbone NH *via* the C(CO)NH and H(CCO)NH experiments. Incidentally, the correlations identified from the C(CO)NH experiment are also particularly

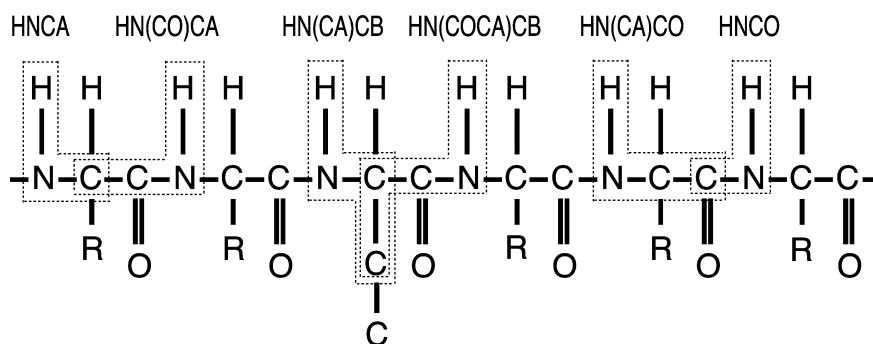
useful in assisting the sequential backbone assignments by providing the amino acid type information. Precise chemical shift assignments, crucial for unambiguous assignment of the NOE cross peaks, can be obtained with a recently developed method that combines simultaneously-evolved nuclei (SEN) HC(CO)NH experiments and non-uniform sampling techniques. In this way, two 3D experiments plus the C(CO)NH and H(CCO)NH experiments can yield the same information as a four-dimensional HC(CO)NH experiment, but at much higher resolution and in as little as three days. This method has the benefit of correlating all side-chain ^{13}C and ^1H resonances with the backbone amide groups in a single data set and the assignment procedure can be easily automated.

Collection of Structural Parameters Initial Fold Determination

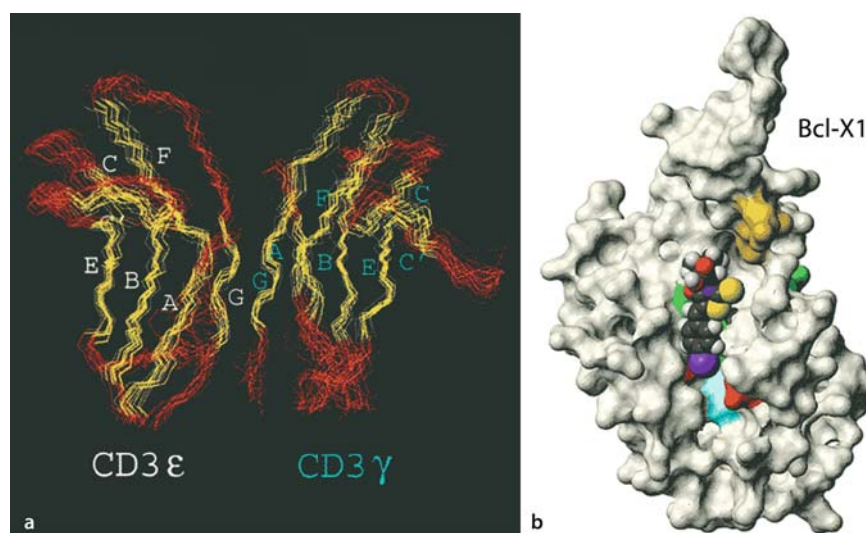
The protein secondary structure topology can be first identified from the deviation of the backbone atoms and beta carbon chemical shifts relative to the random coil values for specific amino acids. Two-dimensional NOESY and three-dimensional heteronuclear ^{15}N - or ^{13}C -edited NOESY experimental data are acquired in order to extract the inter-nuclei distance information. With 3 to 4 NOE distance constraints assigned per residue on average, the initial protein 3D fold can be determined with software programs such as CYANA, XPLOR and CNS. The computation methods include distance geometrical algorithms that use NOE constraints and simulated annealing methods that minimize a target function derived from the violations of NOE constraints (inconsistencies between the input distance constraints and the distances obtained in the computed structural models). At this stage the distance constraints related to the backbone amide, side-chain aromatic and methyl groups are most important for defining the regular secondary structure elements and the hydrophobic core of the protein. Several software programs such as ARIA and CANDID have been developed to assign NOE peaks automatically and to calculate the initial protein fold.

Structure Refinement

Based on the initial protein model, more NOE peaks can be assigned and more distance constraints included in the structure refinement stage. During this iterative process, it is also critical to eliminate wrong NOE assignments that would cause distorted structures and prevent convergence of structure calculations. The erroneous NOE assignments leading to serious constraint violations can be tracked by calculating structures where NOE constraints are eliminated for a single residue, di-peptides, tri-peptides, etc. while sliding the window of “constraint-silencing” along the protein sequence. Disappearance of constraint



3D Structure Determination by NMR. Figure 2 Schematic illustration of the triple-resonance experiment strategy. HNCA and HN(CO)CA, HNCO and HN(CA)CO and HN(CA)CB and HN(COCA)CB for backbone sequential assignments of labeled proteins.



3D Structure Determination by NMR. Figure 3 (a) The backbone traces of several superimposed structures for the extra-cellular part of the CD3γ heterodimer. (b) A ligand docked to the BH3-binding site of the Bcl-xL oncogene product.

violations and improvement of the target function in one of these cases would reveal the presence of the incorrect NOE assignment within the “silenced” segment. On average, 10–20 constraints per residue are required for calculating good NMR structures. Hydrogen-bond constraints, backbone dihedral angle constraints, orientation constraints from residual dipolar couplings and stereo-specific assignments of methylene and methyl groups as well as side-chain rotamer libraries can significantly improve the convergence and precision of the NMR structure ensemble.

Structure Validation

The quality of the final ensemble of NMR structures is routinely checked with validation software, such as ▶[ProcheckNMR](#). Coordinates and restraint files are

usually submitted to the protein data bank (PDB), where the staff of the data bank perform additional validation procedures. As an example of an NMR-derived protein structure, Fig. 3a shows the backbone traces of several superimposed structures for the extra-cellular part of the CD3γ heterodimer, which is a crucial element of the invariant portion of the T-cell receptor complex. Some rules of thumb for assessing the quality of NMR structures include: (1) the average root mean square deviation of backbone atoms relative to the mean structure should be smaller than 1.0, (2) the number of residues with backbone phi-psi angles in the disallowed region in the Ramachandran plot (obtained from ProcheckNMR) should be less than 1%, (3) there should be no NOE distance violations greater than 0.5 and no dihedral angular constraint violations greater than 5°.

Clinical Applications

NMR structural studies of proteins are helpful for understanding the basic health-related cell functions and the underlying mechanisms of some diseases by studying protein complexes and protein conformational changes in solution that are at or close to physiological conditions. Examples include studies of the synergistic co-folding of many cancer-related transcriptional activator complexes and the conformational changes and folding intermediates of prion proteins.

Therapeutic Consequences

Protein structures determined by NMR can accelerate the drug development process. First, the structural studies of protein-protein and protein-ligand interactions can be used to identify the target sites for drug intervention and provide the basis for rational drug design (5). A model of the oncogene protein Bcl-xL with an inhibitor docked to its BH3-domain binding site is shown in Fig. 3b. Second, NMR resonance assignments acquired during structural determination can also be used in rapid drug screening by monitoring specific protein signal changes or signal transfer to interacting chemical compounds (6). ►**SAR (structure-activity-relationship) by NMR**, a good example of both applications, involves first identifying weak interacting compounds and then covalently linking two fragments with neighboring binding sites to form a tighter-binding drug target.

References

1. Kurt Wuthrich (1986) *NMR of Proteins and Nucleic Acids*. John Wiley, pp 130–161
2. Cavanagh J, Fairbrother W, Palmer A et al (1995) *Protein NMR spectroscopy – Principles and Practice*. Academic Press, pp 411–556
3. Ven de F Van (1995) *Multidimensional NMR in Liquids – Basic Principles and Experimental Methods*. Wiley-VCH
4. Hoch J, Stern A (1996) *NMR Data Processing*. Wiley-Liss
5. Craik D (1995) *NMR in Drug Design*. CRC Press
6. James T, Dötsch V, Schmitz U (2001) *Methods in Enzymology*, vol 338. *Nuclear Magnetic Resonance of Biological Macromolecules Part A*. Academic Press, pp 202–230

3D Structure Determination by X-Ray

RICHARD A. ENGH

Max Planck Institute of Biochemistry, Martinsried,
Germany
engh@biochem.mpg.de

Definition

3D structure determination using X-ray radiation refers to the calculation of the spatial distribution of electrons in a crystallized molecule – the crystal structure – based on the distribution of intensities of X-rays diffracted by the crystal. The macromolecules or molecular assemblies that are relevant to genomics and proteomics, here generally referred to as proteins, require data collection from single crystals. The degree of order of the crystal, dependent on the structural homogeneity (rigidity, purity) of the molecule and on the quality of the crystal, determines the spatial resolution to which the electron distribution may be estimated. Additional sources of information and assumptions about the properties of the electron density distribution simplify the structure solution and greatly enhance the accuracy of the final model, the crystal structure.

The process by which the three dimensional structures are determined depends on a variety of factors, including molecule size, the types of atoms in the molecule, knowledge of similar structures, quality of crystals, the available X-ray sources and the information desired from the 3D model. Structural genomics proteomics programs often aim to determine structures of a wide variety of structures in a highly automated, high throughput mode and generate thereby a reference database of 3D structures. Other programs may aim to characterize particular macromolecules more fully by solving their structures in a variety of crystallization environments; still others may focus on very high-resolution chemical phenomena.

Description

Purpose of Structural Models

The goal in solving a crystal structure determines how the problem is approached throughout the entire process. For ►**structural genomics** (see related articles), the goal may be to determine structures for the largest number of genomic or protein sequences, giving priority to new folds in order to enable at least homology modeling for most sequences. Large-scale automation may accelerate the search for suitable conditions for protein expression, purification and crystallization for a wide range of proteins. Studies of protein-protein interactions require the crystallization of the relevant protein complexes; the requirements for the crystal structure determination depend on the structural information available for the individual components. Structures of extremely large protein complexes have special requirements for experimental equipment as well as theoretical techniques. Studies of enzyme catalysis mechanisms may require very high resolution and careful structure determination to avoid the introduction of chemical bias into the model and its analysis. Similarly, structures provide the greatest value for drug design purposes when they have a ►**resolution**

that enables the analysis of solvent interactions and that shows details of chemical geometries.

Protein Preparation

Although many proteins have been purified from their natural sources for crystallization, recombinant methods are now standard as they typically provide the purest protein in the highest amounts. The structure determination of a protein for which no approximate three dimensional model is available will require the presence of heavy, electron rich atoms in one or more crystals of the protein in order to solve the ►phase problem. These can be introduced intrinsically into the protein by the incorporation of non-natural amino acids. Most common is the substitution of selenomethionine for methionine using auxotrophic organisms.

The protein sequence must also be chosen with respect to the expected crystallizability of the protein. Crystal formation requires ordering crystal contacts between ordered protein domains. Thus, crystallization chances are enhanced if the inherent flexibility of the protein is minimized and if surface properties of the protein may be varied in order to increase the variety of possible ordering crystal contacts. As a result, it may be critical to prepare a variety of constructs with varying lengths and varying point mutations at the surface to produce a “crystallizable” protein. These constructs need to be chosen with respect to the desired use of the structure and with an eye toward monitoring possible artifacts introduced by the variations.

Crystallizability is also enhanced by maximizing the structural homogeneity of the protein and by minimizing flexibility and also heterogeneous ►post-translational modifications, such as phosphorylation, deamidation, mixed disulfide bond formation and so on. Autocatalytic functions of a protein may require inactivation of the protein in order to prevent the presence of an equilibrium mixture of states. Finally, affinity tags appended to the protein may assist purification, but may also increase structural heterogeneity and inhibit crystal growth.

Crystallization (See ►X-Ray Crystallography – Basic Principles)

A protein may crystallize when it tends to form ordered interactions with identical molecules in an arrangement compatible with a three dimensional periodic lattice. The conditions under which this occurs are not predictable for a given protein, so that the crystallization strategy is generally to screen ►precipitant conditions and monitor for the appearance of crystals. The most common method for protein crystallization is to mix a protein solution with a precipitant solution in vapor phase contact with a larger reservoir of precipitant solution and allow volatile components to equalize their chemical potentials (or conceptually their concentrations) *via* vapor phase

diffusion. This typically increases the protein concentration (if water is the principal volatile constituent) or the precipitant concentration (for volatile precipitants such as alcohols) or both, in order to create a supersaturated state. Mixing protein with precipitant may already suffice to produce a suitable supersaturated state (batch crystallization). Diffusions through membranes, gels or channels are alternative mechanisms for mixing precipitant with protein. The initial crystallization conditions found may have produced crystals suitable for data collection, if not, many variables such as temperature, additives, protein buffers, pH and so on may be varied in order to optimize crystal quality.

Crystal Preparation for Data Collection

The crystals may need or benefit from additional modification prior to measurement. Soaking the crystal in a solution of ligands may be required for co-crystal structure determination. Soaking the crystal in a solution of heavy atom derivatives may be necessary if the structure is to be solved by MIR or MAD techniques (see below). The crystal quality may be improved by soaking in solutions that alter the solvent content of the crystal. An alternative to these soaking experiments may be found in controlled humidity gas phase experiments, where the addition of ligand or heavy metals may be performed *via* ink jet type devices. Crystals can be harvested into thin walled glass capillaries, adjacent to mother liquor solution to maintain crystal quality during data measurement. These crystals decay during measurement, usually due to the formation of chemically reactive free radicals that diffuse through the crystal. Freezing the crystals in liquid nitrogen greatly retards crystal decay. With the widespread availability of ►synchrotron radiation, freezing with liquid nitrogen has become virtually essential because of the high level of radiation damage and heat deposition due to high intensity synchrotron radiation.

Data Collection and Evaluation (See ►X-Ray Crystallography – Basic Principles)

Data is collected by illuminating the crystal with intense X-ray radiation and measuring the positions and intensities of the diffraction spots as a function of crystal rotation. The rotation range that provides a complete dataset depends on the symmetry of the crystal and on whether ►anomalous diffraction data may be measured. The strategy of data collection is also determined by the strategy of solving the “phase problem” (see related article). If the structure can be solved by molecular replacement, a single complete dataset without anomalous diffraction data is sufficient. If multiple isomorphous replacement (MIR) methods are used, multiple datasets of multiple crystals soaked with different heavy atoms must be measured. For multi-wavelength anomalous dispersion (MAD, see

related article) methods, a single crystal containing anomalous scatterers may be measured at multiple wavelengths. Because of the high intensities and tunable wavelengths of synchrotron X-ray radiation, data collection at synchrotron sources is preferred because of the shorter data collection times or is required because of tunable wavelengths. Advances in X-ray generation technologies do however make laboratory generators cost-effective for many applications.

Initial Phasing

“Phasing” is the critical step in solving the structure from the diffraction measurements, equivalent to solving the “phase problem” of crystallography (see related article). Without phases, the measured diffraction intensities can be converted into the ►[Patterson function](#), a three dimensional function of interatomic vectors (the electron density function is a three dimensional function of atomic positions). As listed above, molecular replacement, multiple isomorphous replacement and multiple anomalous dispersion are three methods used to estimate the phases of the diffraction intensities. (A fourth, direct method, can be used in special cases, see related article.) Molecular replacement relies on the availability of a closely related structure for phasing. Phases can be estimated from the model if its position and orientation in the unit cell are known. The problem in molecular replacement is thus to determine the six parameters that determine the position and orientation, usually by first fitting the three rotational parameters to the Patterson function or its equivalent, then the three positional parameters. The method can fail if the model is not sufficiently similar to the molecule in the crystal or if the model represents too small a subset of the contents of the asymmetric unit of the crystal.

Multiple isomorphous replacement (MIR) relies on the use of one or more strongly scattering atoms bound at different fixed positions in the crystal. Because of the strength of their scattering, their positions in the unit cell can be determined directly. This enables calculation of the phases and amplitudes of the X-rays scattered by the heavy atoms, which combine with the phases and amplitudes of the X-rays scattered by the native structure by ►[vector addition](#). Estimation of the phases of the X-rays scattered by the native crystal can then be estimated by analyzing the intensity shifts caused by the addition of the heavy atoms. This method can fail if addition of the heavy atom scatterers alters the crystal packing, so that the different crystals are no longer isomorphous.

Multi-wavelength anomalous dispersion (MAD) methods are in principle similar to MIR methods, except that multiple wavelengths are used instead of multiple crystals. The requirement is the presence of atoms that have significant ►[anomalous scattering](#). This type of scattering is wavelength dependent, because it is strongest at wavelengths near ►[electronic transition](#)

[energies](#). Measurement at multiple wavelengths and calculations of differences allows identification of the positions of the anomalous scatterers, their relative vector contributions to the total scattered radiation and, again by analysis of the vector addition properties, the phases of the total scattered waves.

Structure Refinement

Once the phases have been estimated well enough to provide an interpretable electron density, the structure may be built. For molecular replacement, the initial model needs revision; for the other methods, an initial model needs to be generated. This is a straightforward task for an easily interpretable electron density, either automatically or manually, using suitable computer graphics and model building programs. In either case, the test of the interpretability of the electron density is that the phases, re-estimated using the new model, should produce an electron density still clearer than the first and from which further corrections, if any, are also clearer.

This iterative procedure is a process of fitting the model parameters, in this case the spatial coordinates and parameters that model atomic motion (usually the ►[temperature factor](#) or B-factor), to best fit the measured data, the diffraction intensities. When the temperature factor is a single value per atom, the number of parameters to be fitted for an N-atom model is 4N. For an asymmetric unit consisting of one single domain protein, 4N can be 15,000 or higher; larger asymmetric units and/or multimeric proteins are much larger still. Therefore, the diffraction data require additional information in order to maintain a reasonable parameter to observation ratio. This additional information also allows considerable improvement in the accuracy of the model.

The most useful information derives from the requirement for a reasonable chemical geometry of the final structure. Average chemical bond lengths are known to accuracies of hundredths of ►[ångströms](#); average bond angles to within tenths of degrees. Bonds and angles in protein structures can deviate from these averages, but the deviations will be small and restraints can be included in the iterative model refinement procedure to improve model accuracy. Because however the restraints introduce bias into the model, restraints must be selected carefully if the structural information is to be used to analyze details of chemical geometries in the structure. This can be the case if, for example, the structure is to clarify mechanisms of enzyme catalysis. The final crystal structure is thus usually derived from experimental diffraction data supplemented by knowledge of chemical geometries in the macromolecule. Sources of errors in the final model include experimental noise, crystal defects, false assumptions regarding chemical identities in the protein, such as ►[protonation states](#) or post-translational modifications,

false identification of solvent molecules and oversimplification of motion models. Proper interpretation of the structure requires consideration of these effects, crystal packing and crystallization condition artifacts and that a single crystal structure fails to show a full range of protein flexibility. Multiple structure determinations in a variety of conditions and crystal packing arrangements are required to adequately assess these effects. Alternate methods such as NMR (See Related Article) can assist the interpretation of crystal structures by characterizing flexibilities in solution.

References

1. Drenth J (1994) Principles of Protein X-ray Crystallography. Springer-Verlag, New York
2. Rhodes G (2006) Crystallography Made Crystal Clear, 3rd edn. Academic Press Inc. (London) Ltd.

Strümpell Disease

► Hereditary Spastic Paraplegia

SU (Surface)

Definition

SU (surface) designates the retrovirus protein (encoded by the env gene) that forms the external part of the Env trimer, and is responsible for interacting with the cell surface receptor.

► Retroviruses

SU (Surface) TAR

Definition

SU (surface) TAR describes the region on HIV RNA to which ► Tat binds.

► Retroviruses

Subcellular Compartment (or Organelle)

Definition

Each cell is comprised of numerous elements, referred to as organelles or subcellular compartments, each involved in highly specific functions. For example, the nucleus contains one copy of the genomic DNA for the corresponding organism, and is the site of gene transcription to RNA; the ► endoplasmic reticulum is involved in protein processing and trafficking; and the ► mitochondria are responsible for energy production via oxidative phosphorylation.

► Immunochemical Methods, Localization

Subcellular Localization

► Immunochemical Methods, Localization

Subclone

Definition

Subclone defines a fragment of DNA that is derived from previously cloned material (grown in bacterial or yeast hosts) and ligated into a cloning or sequencing vector. It is capable of being propagated in suitable host strains.

► Shotgun Libraries

Subependymal Giant Cell Astrocytoma

Definition

Subependymal giant cell astrocytoma is a tumor typically seen in patients with tuberous sclerosis. The tumor is considered to be a variant of astrocytoma and contains large fusiform or pyramidal cells. The tumor is located in the lateral ventricular wall at the foramen of

Monro, where, if large, it may cause obstruction to the flow of cerebrospinal fluid causing hydrocephalus.

► [Tuberous Sclerosis](#)

Subependymal Nodule

Definition

Subependymal nodule refers to benign tumors commonly seen in tuberous sclerosis. Subependymal nodule lesions are composed of calcified proliferating glial and vascular elements, typically found in the anterior portions of the lateral ventricles. The nodules may undergo malignant degeneration into ► [subependymal giant cell astrocytomas](#).

► [Tuberous Sclerosis](#)

Subjective Day/Subjective Night

Definition

Subjective day/subjective night refers to the portion of a circadian cycle in constant conditions which correspond to day or night in a light-dark cycle.

► [Circadian Clocks](#)

Sub-Proteome

Definition

Sub-proteome denotes a collection of proteins that are part of a proteome and share a common characteristic. For example, proteins that reside on the cell membrane of a cell could be a subproteome of cell surface proteins.

► [Proteomics in Human-Pathogen Interactions](#)

Substantia Nigra

Definition

Substantia nigra describes a region of the midbrain that is pigmented black due to its melanin content, and in which the neurons produce dopamine.

► [Parkinson's Disease: Insights from Genetic Causes](#)

Substrate

Definition

Substrate is a ligand that binds to an enzyme and is chemically modified in an enzyme reaction.

► [Structure-Based Drug Design](#)

Subtelomeric Repeats

Definition

Subtelomeric repeats refer to satellite-like sequences that are found in highly repetitive regions between telomeres and the rest of chromosomes, and are usually derived from TTAGGG telomeric repeats and similar motifs.

► [Repetitive DNA](#)

Subtractive Analysis

Definition

Subtractive analysis is the comparison of proteomes of at least two different biological conditions.

► [Two-Dimensional Gel Electrophoresis](#)

Succinate Dehydrogenase

► [SDH](#)

Sumo

Definition

Small Ubiquitin Related Modifier

► [Sumoylation](#)

Sumoylation

GUILLAUME BOSSIS, FRAUKE MELCHIOR
University of Göttingen, Göttingen, Germany
f.melchior@medizin.uni-goettingen.de

Definition

The term SUMO, which stands for small ubiquitin related modifier, refers to a family of eukaryotic proteins that are structurally and functionally related to ubiquitin. Alternate names for SUMO proteins are sentrin or Smt3. Like ►ubiquitin, SUMO proteins can be covalently coupled to specific nuclear or cytoplasmic proteins in the cell. SUMO attachment, also called “sumoylation”, is achieved through an enzymatic cascade that involves E1, E2 and E3 enzymes and ATP. The resulting ►isopeptide bond between SUMO and its targets can be cleaved by the function of specific proteases (isopeptidases). Reversible sumoylation serves to regulate protein functions by, e.g., changing protein/protein or ►protein/DNA interactions, localization or stability.

Characteristics

SUMO Genes and Proteins

The number of SUMO proteins that are expressed within an organism varies significantly. *S. cerevisiae* contains a single SUMO gene, *Smt3*, which is essential for progression through the cell-cycle. Eight SUMO proteins are known to exist in *Arabidopsis*, but only one each in *C. elegans* and *D. melanogaster*. In mammals, the SUMO family comprises three members, SUMO-1, -2, and -3. Human SUMO-1 is a 101 amino acid protein with 18% sequence identity to ubiquitin. hSUMO-2 and its twin SUMO-3 are very similar (95% identical), but are each only approximately 50% identical to SUMO-1. Different SUMO proteins within one organism appear to serve distinct functions and/or are subject to distinct regulation. For example, under normal growth conditions SUMO-1 is mainly found in conjugates, while SUMO-2 and -3 exist predominantly in free form. Interestingly, SUMO-2/-3 become rapidly conjugated upon certain stress stimuli. The mechanisms underlying these differences are not currently understood.

SUMO Structure

Despite low sequence conservation between SUMO and ubiquitin, their three dimensional fold is very similar. It is characterized by the characteristic $\beta\beta\alpha\beta\beta\alpha\beta$ ubiquitin-folding motif. Nascent SUMO proteins contain a short stretch of C-terminal amino acids that needs to be removed before SUMO can be

conjugated to other proteins. Processing of the C-terminus, which is accomplished by SUMO proteases of the Ulp family (see below), exposes a glycine-glycine motif required for conjugation. A hallmark for SUMO family members that is absent from ubiquitin and other ubiquitin-related proteins is a flexible N-terminal extension that varies in length between 10 and 20 amino acids. The precise function of this N-terminus is currently unknown, but *in vitro* and *in vivo* evidence suggests that it can serve in SUMO-SUMO chain formation.

The Mechanism of Sumoylation

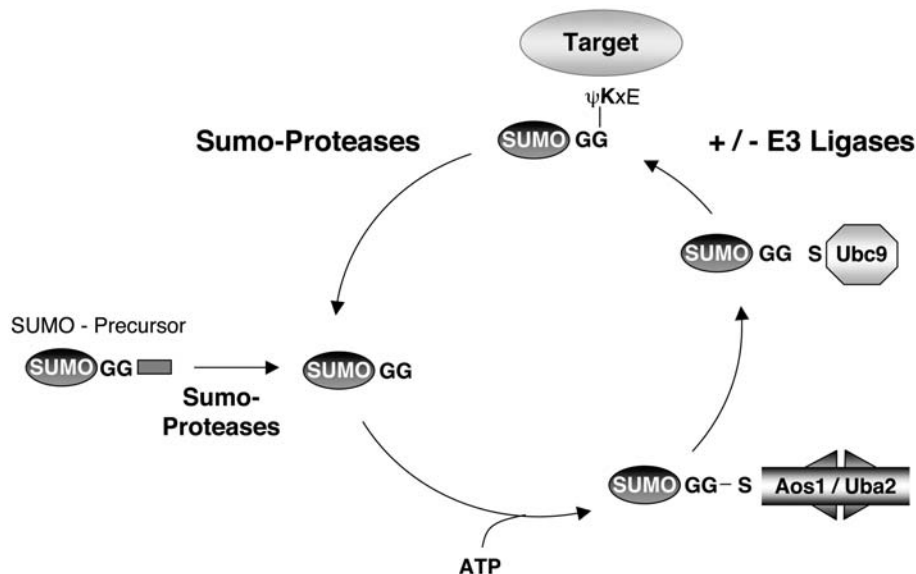
SUMO proteins serve as small protein tags that are covalently coupled to many distinct cellular proteins (targets). Coupling is accomplished by isopeptide bond formation between the carboxy terminal glycine in mature SUMO and an -amino group in a lysine side chain of the target. Considering the structural homology between SUMO and ubiquitin, it is not surprising that the conjugation mechanisms are related. Both processes require ATP and involve three distinct types of enzymes (E1 activating, E2 conjugating and E3 ligating enzymes). While ubiquitination often results in ►ubiquitin chain formation, sumoylation usually leads to attachment of one or several single SUMO entities. Evidence for chain formation exists as well, but whether SUMO chains are physiologically relevant is still an open question.

SUMO Activating Enzyme Aos1/Uba2

Activation of the C-terminal glycine of SUMO is the first step towards isopeptide bond formation. This is achieved *via* an ATP dependent mechanism that results in ►thioester bond formation between SUMO and the SUMO specific E1 enzyme. The SUMO activating enzyme consists of two subunits, Aos1/Uba2 (SAE1/SAE2), that resemble the N- and C-terminal halves of the single ubiquitin E1 enzyme, respectively. Aos1/Uba2 is strongly enriched in the nucleoplasm, but present also in the cytoplasm.

SUMO Conjugating Enzyme Ubc9

Once activated, SUMO is transferred to the single E2 conjugation enzyme Ubc9 (again, a thioester bond is formed). It is localized mainly in the nucleus, but is also present in the cytoplasm and at ►nuclear pore complexes. Ubc9 shares significant homology with the large number of ubiquitin conjugating enzymes (Ubc's), but is specific for SUMO proteins. An important difference between Ubc9 and Ubc's is Ubc9's ability to directly and specifically recognize SUMO targets *via* their SUMO attachment site motifs (see below). Consequently, the last step in forming an isopeptide bond – transfer of activated SUMO to an -amino group in lysines of the acceptor protein – can in principle be



Sumoylation. Figure 1 Sumoylation and desumoylation cycle. SUMO is synthesized as a precursor, whose C-terminal extremity is cleaved by SUMO specific proteases (C-terminal hydrolase activity). Mature SUMO is activated by the SUMO E1 activating enzyme (composed of the two sub-units Aos1 and Uba2) in an ATP dependant manner, before it is transferred to the E2 conjugating enzyme Ubc9. Isopeptide bond formation with a target protein is accomplished with (or in some cases without) the help of an E3 ligase. Three types of SUMO Ligases are known so far, the PIAS family, RanBP2 and PC2. SUMO is often conjugated to lysine residues that are located within the consensus motif ψ KxE (ψ is a bulky hydrophobic residue). SUMO can be removed from the target by SUMO proteases (isopeptidase activity). Of note, several known SUMO proteases contain C-terminal hydrolase as well as isopeptidase activity.

accomplished by Ubc9 alone. However, modification of most targets is extremely inefficient and slow in the absence of additional factors (E3 ligases).

SUMO E3 Ligases

Three types of factors have been identified that strongly accelerate sumoylation of specific targets *in vitro* and/or *in vivo*. These are referred to as SUMO E3 ligases or E3 ligase like factors. The first type is encoded by the protein inhibitor of activated STAT (PIAS) family. PIAS proteins were originally identified as co-regulators for **STAT** – and **nuclear hormone receptor** – mediated transcription. PIAS proteins are characterized by a predicted RING domain (the **SP-RING**), and are thought to function like ubiquitin RING finger E3 ligases as adaptors between the thioester charged E2 and specific targets. PIAS proteins reside predominantly within the nucleus, where they are enriched in diverse subnuclear domains, but are also found in the cytoplasm. Mammals express at least 5 different PIAS proteins, PIAS1, PIAS3, PIAS α , PIAS β and PIAS γ , all of which have been shown to function as SUMO E3 ligases. Two PIAS homologs, Siz1 and Siz2, exist in yeast. They appear to be responsible for sumoylating a

large number of proteins, as simultaneous disruption of both genes significantly reduces overall levels of sumoylation.

The second type of SUMO E3 ligases is currently represented by a single vertebrate specific member, the nuclear pore complex protein RanBP2/Nup358 (Ran binding protein/nucleoporin of 358 kD). Its SUMO ligase activity is restricted to a 30 kD domain, which shows no homology to PIAS proteins or ubiquitin ligases. RanBP2 is a major component of cytoplasmic filaments of nuclear pore complexes, where it serves as a docking site for complexes that travel into or out of the nucleus. This allows speculation that RanBP2 may sumoylate specific proteins on their way into or out of the nucleus.

The third SUMO E3 ligase type that has been described is the human Polycomb member Pc2. Polycomb group proteins form large protein complexes that are involved in gene silencing. Pc2 is unrelated in sequence to PIAS or RanBP2, and its mode of function awaits further investigation.

Comparison of different SUMO E3 ligases demonstrates that they have some target preferences *in vitro*. In combination with their distinct localizations *in vivo*,

this may contribute significantly to selective modification of target proteins within the cell. Whether additional SUMO E3 ligases exist, and how the distinct SUMO E3 ligases function on a molecular level to accelerate modification of specific targets is currently unknown.

Deconjugation

A key feature of posttranslational modification with SUMO is its reversible and at times highly dynamic nature. Cleavage of the isopeptide bond between SUMO and its targets is accomplished by SUMO-specific proteases (SUSP) of the Ulp1 family (also referred to as SENP for sentrin specific proteases). These ►cysteine proteases, which contain a C-terminal 200 amino acid catalytic core domain with similarity to adenovirus proteases, also serve in maturation of nascent SUMO.

Two members of the Ulp family exist in yeast (Ulp1 and Ulp2). The essential protein Ulp1 localizes to nuclear pore complexes, the non-essential Ulp2 is nucleoplasmic. Mammals express a much larger family of Ulp proteins, in part due to the presence of several distinct genes (7 in humans), in part due to extensive alternative splicing. The exact number of distinct SUMO-specific proteases in mammals is currently unknown. Analysis is complicated by the recent discovery that not all Ulp family members are specific for SUMO. At least one of the Ulp genes (*SENP8*) encodes a protease with specificity for the ubiquitin related protein Nedd8. Several full length and splice variants encoded by the genes *SENP1*, *SENP2* and *SENP6* have been characterized in some detail. All the proteins contain the catalytic core, but vary in the length and sequence of their N-terminal domains. The most prominent difference between these different proteases is their localization to different subcellular compartments (the nucleoplasm, nuclear bodies, the nuclear pore complex, the nucleolus and the cytoplasm). Differences in catalytic properties, e.g. preferences for one or the other SUMO protein, have also been described, but a systematic analysis is still lacking.

Molecular Interactions

With the exception of non-covalent SUMO – enzyme interactions, most proteins (targets) bind to SUMO covalently. These targets have no apparent affinity for SUMO proteins themselves. Instead, binding is accomplished through interaction with conjugating and ligating enzymes and subsequent isopeptide bond formation with SUMO. One determinant for covalent SUMO interaction is the presence of a SUMO attachment site in the target. In many proteins this motif consists of just four amino acids, ψ KxE/D (ψ stands

for a bulky hydrophobic residue, K is the lysine to which SUMO is attached). Interestingly, short peptides containing this minimal motif can be sumoylated *in vitro*. Obviously, the sequence ψ KxE/D exists in many non-sumoylated proteins, and additional features such as accessibility and secondary structure of the motif will be important for recognition by Ubc9. In the efficient SUMO target RanGAP1 this motif is part of an accessible loop structure. Alternative SUMO attachment sites are beginning to emerge, e.g. in PCNA, but it is not yet clear how they are recognized by Ubc9.

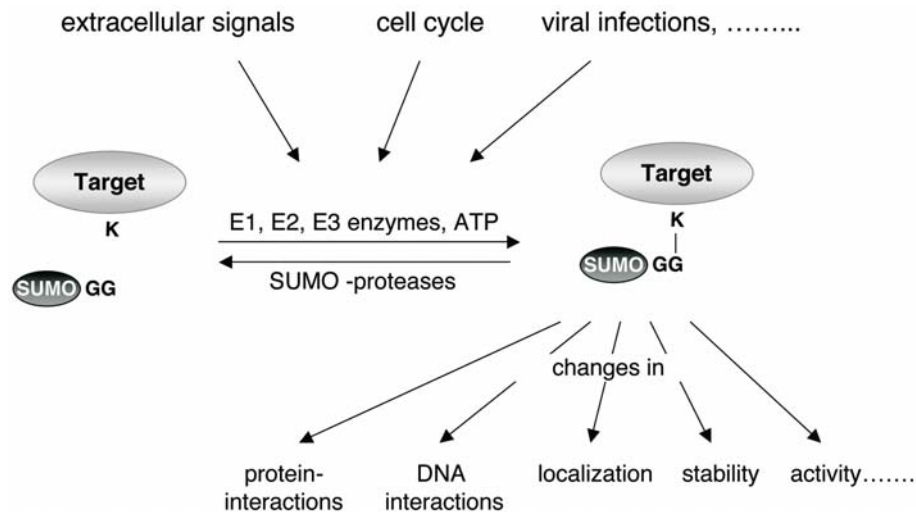
The number of known covalent interactors (SUMO targets) has reached more than 100 and is growing rapidly. Characterized SUMO targets fall into just a handful of protein groups, most of which have functions within the nucleus. These include transcription factors (e.g., p53, c-jun, ►HSF, Sp3, Lef1, Elk1, nuclear hormone receptors), regulators of transcription and/or chromatin remodeling (e.g., nuclear hormone receptor co-regulators, histones deacetylases, DNA methyl transferases) or proteins involved in replication and DNA repair (topoisomerases, PCNA, thymine-DNA glycosylase). Sumoylated proteins are however not restricted to the nucleus. Examples are yeast septins, which reside at the ►bud neck and are modified specifically in mitosis, *Dictyostelium* MAP kinase kinase MEK1, whose sumoylated form is found in the cytoplasm and at the cell cortex and the Ran GTPase activating protein RanGAP1, whose sumoylated form is found at cytoplasmic filaments of nuclear pore complexes.

Regulatory Mechanisms

Two aspects of regulation will be discussed here. First, reversible attachment of SUMO to specific targets can be constitutive or a highly regulated mechanism. Second, SUMO attachment serves to regulate proteins in their function or stability. Selected examples will be discussed to highlight both aspects.

SUMO Attachment Can Be Regulated

While some proteins appear to be modified constitutively, many proteins obtain SUMO only transiently, for example at a specific time of the cell-cycle, upon heat shock or DNA damage, or due to extracellular signals. The precise mechanisms underlying this dynamic modification are unknown for most targets. From the available data it is however clear that regulation can be accomplished by several distinct means. First, upstream events may result in phosphorylation or dephosphorylation of a specific target, which in turn allows or inhibits interactions of the target with the sumoylation machinery. For example, the heat shock factor HSF1 is sumoylated upon phosphorylation



Sumoylation. Figure 2 Regulation and function of sumoylation. SUMO conjugation is a reversible process that affects many different nuclear and cytoplasmic proteins (more than 100 targets are currently known). Modification of a specific target can be constitutive or regulated, e.g., by extracellular signals, position in the cell-cycle, viral infections, DNA damage. Consequences of modification are target specific, and include, amongst others, changes in protein-protein and protein-DNA interactions, localization, stability or activity.

in close proximity to its SUMO acceptor site. That this is due to better recognition by the modification machinery was demonstrated by the finding that a phosphomimetic mutant of HSF1 is more efficiently modified *in vitro* than its wild type counterpart. Down-regulation by target phosphorylation has for example been reported for the transcription factors c-Jun and Elk1. Second, upstream events can also act at the level of the modifying or demodifying enzymes. Cell-cycle dependent changes in localization have, e.g., been reported for the E3 ligase Siz1 in *S. cerevisiae*. Siz1 is predominantly intranuclear in interphase, but it partially relocates to the bud neck in mitosis. Concomitant with this, yeast septins, which localize to this site, are sumoylated. The fission yeast isopeptidase Ulp1 also undergoes cell-cycle dependent changes in localization (it moves from the nuclear periphery to the nucleoplasm during mitosis), but this has not yet been linked to changes in the sumoylation pattern of specific proteins.

SUMO Attachment Regulates Protein Function

Considering that addition of even a single phosphate can have dramatic effects on proteins, it is not surprising that attachment of the 11 kD SUMO protein may have significant consequences for its targets. Due to its size it may mask large binding interfaces, add new interaction sites or influence the folding of a protein. Consequently, a plethora of distinct target specific

functions have been reported for sumoylation. These include, but are not limited to, roles in nuclear import and export, enhancement or repression of transcription, changes in activity, increase or decrease in protein stability and inhibition of mono- or poly-ubiquitination. On a molecular level, the only common denominator that one can extract from available data is that SUMO alters target interactions. These may be with proteins or with DNA (RNA interactions have not yet been reported) and they may either be positively or negatively influenced by SUMO.

In conclusion, posttranslational modification with SUMO is an essential and widely used regulatory mechanism. Like phosphorylation, its effect on protein functions is not predictable, but has to be analyzed for each target individually.

Further Reading

1. Melchior F (2000) SUMO – nonclassical ubiquitin. *Annu Rev Cell Genet Biol* 16:591–626
2. Wilson VG, Rangasamy D (2001) Viral interaction with the host cell sumoylation system. *Virus Res* 81:17–27
3. Schwartz D, Hochstrasser M (2003) A superfamily of protein tags: ubiquitin, SUMO and related modifiers. *Trends Biochem Sci* 28:321–328
4. Seeler J, Dejean A (2003) Nuclear and unclear functions of SUMO. *Nat Rev Mol Cell Biol* 4:690–699
5. Melchior F, Schergaut M, Pichler A (2003) SUMO: ligases, isopeptidases and nuclear pores. *Trends Biochem Sci* 28:612–618

6. Muller S, Ledl A, Schmidt D (2004) SUMO: a regulator of gene expression and genome integrity. *Oncogene* 23:1998–2008
7. Johnson ES (2004) Protein modification by SUMO. *Annu Rev Biochem* 73:355–382

Superantigens

Definition

Superantigens are usually bacterial antigens that possess potent immunostimulatory properties and the ability to crosslink MHC class II molecules with T-cell receptors. Thereby, they polyclonally activate large fractions (2–20%) of the T-cell population at picomolar concentrations. Studies have implicated superantigens in the pathogenetic aspect of such diseases as food poisoning, toxic shock syndrome, and autoimmune phenomena.

The best characterized superantigens are the microbial toxins from *Staphylococcus aureus* and *Streptococcus pyogenes*. Superantigens associated with *Streptococcus pyogenes* are also hypothesized in part to be responsible for psoriasis. Antigens associated with *Mycobacterium tuberculosis*, the rabies virus, and possibly HIV may also function as superantigens.

- [Steroid Hormone Receptor Defects, Molecular Basis](#)
- [Psoriasis, Molecular Basis](#)

Supercoiled DNA

Definition

Supercoiled DNA defines a molecule of DNA or a region of DNA in which the double helix is further twisted on itself.

- [DNA Ligases](#)
- [Nucleosomes](#)

Superoxide Dismutase

Definition

SOD is any of a member of three distinct protein families that dismutate the superoxide anion (O_2^-) to molecular oxygen (O_2) and hydrogen peroxide (H_2O_2). The three SOD families are unrelated in sequence, and also differ in their catalytic transition metal content (Cu, Mn or Fe).

- [Free Radicals](#)

Superresolution

Definition

The light microscopic optical resolution is given by $d = 0.61 \lambda / NA$, where λ designates the wavelength of the employed light, and NA the numerical aperture of the utilized objective lens. Objects with sizes smaller than d are always imaged as diffraction-limited light spots (see also ► [Localization Precision](#)). The details of such objects cannot be recognized, and further magnification produces only an enlargement of the light spot, but does not reveal further detail. Usually one matches the half-width of the diffraction-limited spot to the size of a single picture element (pixel) of the detector. In contrast to this, superresolution imaging means that the diffraction-limited spot is mapped with a higher magnification onto the detector, such that a single spot extends over a greater number of pixels (at minimum 3 3, but up to 10 10).

- [Fluorescence Microscopy: Single Particle Tracking](#)
- [Single Particle Tracking](#)

Supervised Analysis

Definition

Supervised analysis refers to a data analysis where a priori information about the samples or genes is used from the beginning of the analysis.

- [Microarray Data Analysis](#)

Supervised Learning

Definition

Supervised learning comprises of generalizing a structure from given examples to future data.

- [Computational Diagnostics](#)

Suppressive Subtractive Hybridization

Definition

SSH is a molecular screening technique that is used to identify unknown genes or genes with unknown function, which are differentially expressed in two samples. The underlying principle of this technique is

the hybridization of molecules that are expressed in both samples. Molecules without a hybridization partner are differentially expressed and can be sequenced.

► [Rheumatism Related Genes, Identification](#)

Suppressor Hunting

► [Protein:Protein Analysis, Suppressor Hunting](#)

Surface Plasmon Resonance

KARL-WILHELM KOCH

Division of Biochemistry, Institute of Biology and Environmental Sciences, Carl von Ossietzky University, Oldenburg, Germany
karl.w.koch@uni-oldenburg.de

Definition

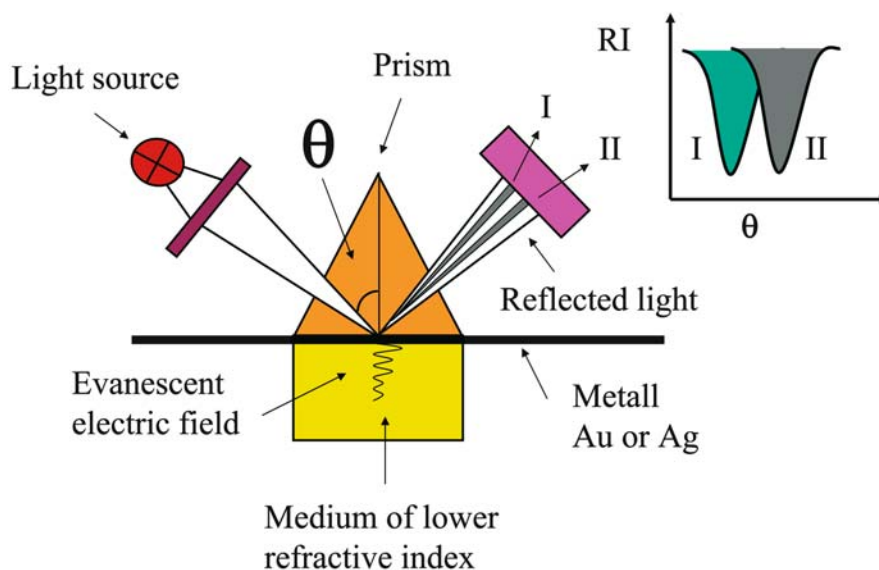
► [Surface plasmon resonance](#) (SPR) is an optical phenomenon that is observed on ultrathin surfaces (50 nm) of specific metals like gold, silver and copper

(1, 2, 3). The metal layer is located at the interface between two media of different refractive indices (a prism and a compartment containing an aqueous buffer). When light is shone on to the prism under the angle of total internal reflection, the electric field of the photons extends about one quarter of the wavelength (► [evanescent electric field](#)) beyond the reflecting interface into the medium of lower refractive index (Fig. 1). A resonance phenomenon occurs when a thin metal film is present at the reflecting surface. Free electron clouds or certain electron densities of the metal interact with the evanescent electric field and the incident light energy is converted into electron density waves in the metal film under specific conditions. As a result the intensity of the reflected light beam on the side of the prism is extinguished. This reduction of reflected light intensity is maximal at a specific angle of the incident light beam called the resonance angle (Fig. 1). Changes in mass at the surface layer are related to changes in resonance angle. This allows the detection of binding events on the surface in real time and thereby the determination of rate and binding constants.

Description

Configuration of System

SPR technology is applied in optical biosensors to monitor interactions of biomolecules and is available in commercial systems like e.g. Biacore®. A typical



Surface Plasmon Resonance. Figure 1 Principle of surface plasmon resonance. A light source emits monochromatic light that traverses a prism at a certain angle θ and is totally internally reflected at the interface of two media. A thin metal layer is at the boundary between the prism and a medium of lower refractive index (e.g. aqueous buffer). Interaction of the evanescent electric field and the electron constellations of the metal causes a reduction in the reflected light intensity (RI). This is illustrated as a gray shadow in the reflected light beam and as a decrease of RI in the graph RI versus θ . Changes in the dielectric properties in the vicinity of the metal surface change the resonance conditions and lead to a shift of θ (I→II).

configuration consists of a light-emitting diode as light source (near infra red), an optical unit with a glass prism, an array of light-sensitive diodes as detector unit and a microfluidic cartridge system with channels and valves to allow delivery of solvents and samples into the system (1). Interaction of molecules occurs on the surface of special sensor chips that are introduced into the system. The sensor chips are docked to the microfluidic cartridge with their metal (gold) coated surface and are in contact with the glass prism on the other side. The only variable in this configuration is the refractive index of the medium on the side of the gold-coated sensor chip surface and the buffer flow. Thus, any binding event on the gold-coated surface or any transient change of buffer close to the surface is seen as a change in the resonance angle. Changes in resonance angle are directly correlated to changes in mass at the surface layer. Any given change of mass results in the same change of refractive index for all proteins and peptides. Thus, the detection principle is independent of the amino acid sequence of a given protein. A shift in the resonance angle of 10^{-4} is defined as 1 resonance unit (RU) and corresponds to an increase of surface mass of 1 pg protein/mm². The whole dynamic range of the system is 30 kRU. The use of the system is not restricted to proteins and peptides, but can also be applied to the study of glycoproteins, oligosaccharides, lipids and nucleic acids. The relation of mass changes to changes in the refractive index is the same as for proteins.

Immobilisation Strategies

In order to use SPR to study interactions of biomolecules in a protein-protein or protein-lipid system etc., it is necessary to immobilize one binding partner termed the ligand on the surface of a sensor chip. The other binding partner termed the analyte is supplied in the mobile phase and injected into the buffer stream flowing over the sensor chip surface. Several immobilization strategies employing a variety of chemical or biochemical procedures have been used so far (3). These can be summarized as in the following.

Immobilization on ► Carboxymethylated Dextran Surfaces

Sensor chips with a gold surface are coated with carboxymethylated dextran of approx. 100 nm thickness. Carboxy groups on this matrix can be modified to enable covalent coupling to primary amines (-NH₂), sulfhydryl groups (-SH), aldehyde (-CHO) or carboxyl (-COOH) groups. First, the dextran matrix is activated by derivatization with N-hydroxysuccinimide (NHS) to yield a reactive N-hydroxysuccinimide ester. The activated carboxy group then can react with substituents carrying an uncharged primary amino group. This allows direct covalent linkage of proteins and peptides (amine coupling) by lysine or free NH₂-termini.

Alternatively, activated carboxy groups are further derivatised by 2-(2-pyridinyldithio)ethaneamine hydrochloride (PDEA) yielding a modified surface with a reactive disulfide group. Proteins and peptides can be selectively coupled by cysteines *via* a disulfide exchange reaction (ligand thiol coupling). A second way to utilize thiol chemistry for coupling is to introduce a reactive disulfide group into the ligand before immobilization, which then allows direct binding to a thiol group on the surface (surface thiol coupling).

The aldehyde coupling procedure is a method to immobilize glycoproteins, glycoconjugates and polysaccharides. The dextran surface of the sensor chip is derivatised by NHS (see above) and can subsequently react with hydrazine to yield a hydrazide surface. Glycoconjugates with *cis*-diols are oxidized by periodate to an aldehyde function. Aldehydes then readily react with hydrazide on the sensor chip surface.

Immobilization by Antigen-Antibody Interaction

Antibody coated surfaces allow a monospecific immobilization of the ligand. The antibody can be directly coupled to activated carboxy groups *via* amine coupling or, in a less random process, bound to immobilized ►protein A or ►protein G on the sensor chip surface.

Immobilization of Tagged Proteins

Fusion proteins containing a glutathione-S-transferase or a maltose-binding protein can be coupled in a site directed way on a surface with the corresponding antibody. Specific sensor chips harbouring a preprepared NTA (nitriloacetic acid)-matrix can be loaded with Ni²⁺ and are designed for the immobilization of poly-histidine-tagged proteins. The high affinity interaction between biotin and ►streptavidin or ►avidin allows the immobilization of biotinylated proteins.

Immobilization of Lipids and Membrane Proteins

A special challenge of SPR applications is the immobilization of lipid bilayers on the sensor chip surface to study membrane proteins and membrane bound processes. Commercially available hydrophobic sensor chips contain hydrophobic groups within a dextran matrix, which capture ►liposomes from the mobile phase and anchor the lipids in the form of a bilayer on the surface. Liposomes with integrated membrane proteins can be immobilized in the same way. Alternative approaches to immobilizing membrane proteins are to utilize an immobilized antibody (within the dextran layer) directed against the membrane protein or to first immobilize a lectin that binds with high affinity to the glycosyl moiety present in most membrane proteins.

Blank gold surfaces of sensor chips without any additional coating also offer the possibility of using a

►self-assembled monolayer (SAM) as a basis for further immobilization steps. When ►alkane thiols are used to build up a SAM *via* a sulphur-gold contact, lipids can be fused to this layer and thereby form a heterobilayer. When alkane thiols are mixed with other thiols, e.g. biotin thiol or alkane thiols modified with COOH- or OH-groups, further derivatization is possible and allows building up of sandwich-like structures on the chip surface.

Data Acquisition

Interaction analysis by SPR is performed by injection of one binding partner into the buffer stream that is carried through the flow cell over the surface coated with the corresponding immobilized binding partner. Binding events on the surface of the chip are sensed and recorded as a change in resonance units resulting in a graph called a sensorgram. A typical sensorgram consist of an association phase and a dissociation phase (Fig. 2). Dissociation of a formed complex on the sensor surface is usually initiated by the end of analyte injection when only the running buffer is in the mobile phase. Association and dissociation rate constants can be obtained from a sensorgram by applying suitable curve fitting procedures, which in addition give equilibrium binding constants (see below) (4, 5). Further, SPR is a tool to investigate an interaction process by qualitative and semiquantitative analysis. These applications include, but are not restricted to

- screening for a possible interaction partner out of a heterogeneous mixture of biomolecules
- identification of an interaction site within a protein by screening of peptide libraries or testing of overlapping protein constructs
- investigation of mutations in proteins
- small molecule screening to facilitate drug design

Quantitative Analysis of Sensorgrams

A sensorgram is a change of resonance units RU with time and reflects the formation of a complex



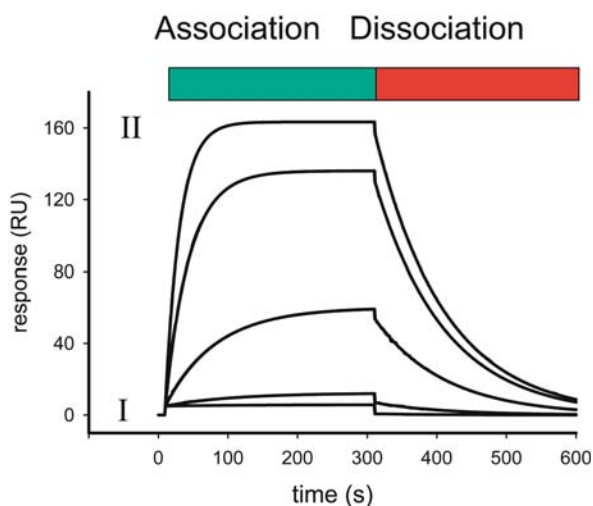
The rate of formation of complex AB with time t is

$$d[AB]/dt = k_a[A]([B]_0 - [AB]) - k_d[AB] \quad (2)$$

or as a change of resonance units with time

$$dRU/dt = k_a C(RU_{\max} - RU_t) - k_d RU_t \quad (3)$$

dRU/dt is the rate at which a complex of two components A and B is formed. RU_{\max} represents the total binding capacity on the surface of the sensor chip and is equivalent to the number of immobilized binding sites B at $t = 0$ ($[B]_0$).



Surface Plasmon Resonance. Figure 2 Example of SPR Sensorgrams. Traces represent simulated SPR recordings showing an increase in resonance units (RU) when two binding partners associate and a decrease in resonance units when they dissociate. The two states I and II correspond to different angles θ under which surface plasmon resonance is maximal and therefore the reflected light intensity RI reaches a minimum (Fig. 1). The concentration of the binding partner in the mobile phase (analyte) was varied from bottom to top as follows: 0.1 μ M, 1 μ M, 10 μ M, and 100 μ M. Dissociation constant $K_D = 26.4 \mu$ M.

RU_t is the number of occupied binding sites at time t (equivalent to $[AB]$). $RU_{\max} - RU_t$ is the number of free binding sites on the chip surface at time t (or $[B]_0 - [AB]$); k_a and k_d are association and dissociation rate constants, respectively. C is the concentration of the binding partner in the mobile phase ($[A]$).

At equilibrium $dRU/dt = 0$, the equation

$$k_a C(RU_{\max} - RU_t) - k_d RU_t = 0 \quad (4)$$

can be arranged to

$$k_a C RU_{\max} - k_a C RU_t = k_d RU_t \quad (5)$$

$$C RU_{\max} - C RU_t = k_d / k_a RU_t \quad (6)$$

$$RU_t / C = K_A RU_{\max} - K_A RU_t \quad (7)$$

At equilibrium RU_t is expressed as R_{eq} . A Scatchard plot analysis (R_{eq}/C against R_{eq}) can be used to obtain K_A (or K_D) from the slope of the plot.

Integration of the rate equation (3) yields:

$$RU_t = \frac{C k_a RU_{\max}}{C k_a + k_d} (1 - e^{-[(C k_a + k_d)(t - t_0)]}) \quad (8)$$

The dissociation phase of a sensorgram can be described as

$$dRU/dt = -k_d RU \quad (9)$$

or in integrated form

$$RU_t = RU_0 e^{-k_d(t-t_0)} \quad (10)$$

Primary data from a sensorgram RU *versus* t during the association phase can be directly fitted to equations (8) and (10) assuming a 1:1 interaction model.

Common Problems in SPR Experiments

Recorded sensorgrams very often do not fit to a simple 1:1 binding model as outlined above. In most cases deviations from this bimolecular interaction model originate from experimental artefacts that can influence a correct analysis of sensorgrams (2, 3, 5). These include:

Non-Specific Binding

An analyte molecule could interact with the coating of the sensor chip surface or could interact with the immobilized ligand in a non-specific manner. The problem can be addressed by running the experiment in parallel on a control surface and subtracting the control signal.

Heterogeneity of Chip Surface

Immobilization of the ligand can occur at different sites on the molecule. For example a protein might have several exposed lysines that react with the activated dextran surface. The problem can be avoided by using a more specific site-directed immobilization, for example by selective coupling to an introduced cysteine. High immobilization densities of the ligand can also lead to a heterogeneous surface, where binding conditions within the dextran matrix are different from the conditions on the surface of the dextran layer. This problem can be avoided by using low surface densities on the chip surface (see also mass transfer limitations).

Mass Transfer Limitations

Several factors determine the transport of the analyte to the immobilized ligand on the sensor chip surface. It has to pass the unstirred layer on the boundary layer between chip surface and buffer flow, it has to partition into the matrix and diffuse through the dextran matrix to reach the immobilized ligand within the matrix. During dissociation, rebinding of the analyte can occur due to a hindered transport of the analyte away from the ligand-binding site. These diffusion processes can affect the observed chemical rate constants and lead to a misinterpretation of the kinetic evaluation of sensorgrams. The main ways to avoid problems arising from mass transfer limitation are the use of very low surface densities (only several fmol/mm²) and high buffer flow rates.

Other Problems

A variety of other problems can influence an SPR analysis, e.g. multivalent interactions, steric hindrance, crowding or matrix effects like swelling or shrinking of the dextran layer by changing the composition of the running buffer. Most problems can be avoided by using low surface densities, appropriate control injections and reversing the geometry of ligand immobilization, i.e. to immobilizing the other binding partner.

Clinical Applications

SPR is a tool of basic biomedical research and is applied in academia and industry. It is widely used for the characterization of antibodies and in particular for epitope mapping of antibodies. In combination with high-throughput screening in the area of drug discovery, it serves to characterize small molecules as well as antibodies. In addition it is a tool in the quality control of products in the biomedical industry.

Therapeutic Consequences

As a typical tool of basic research SPR has only indirect therapeutic consequences, which might originate from characterization of ligand-analyte interactions.

► Immunological Methods, the Use of Antibodies to Discover the Function of Novel Gene Products

References

1. Jönsson U, Fägerstam L, Ivarsson B et al (1991) Real-time biospecific interaction analysis using surface plasmon resonance and a sensor chip technology, *BioFeature* 11:620–627
2. Schuck P (1997) Use of surface plasmon resonance to probe the equilibrium and dynamic aspects of interactions between biological macromolecules, *Annual Reviews of Biophysics and Biomolecular Structure* 26:541–566
3. Herberg FW and Zimmermann B (1999) Analysis of protein kinase interactions using biomolecular interaction analysis. In: Hardie DG (ed) *Protein Phosphorylation – A practical approach*, 2nd edn. Oxford University Press, Oxford 335–371
4. O'Shannessy DJ, Brigham-Burke M, Sonesson KK et al (1994) Determination of rate and equilibrium binding constants for macromolecular interaction by surface plasmon resonance. *Methods in Enzymology* Vol. 240:323–349
5. Myszkka DG (1999) Improving biosensor analysis. *Journal of Molecular Recognition* 12:279–284

Surfactant Proteins

Definition

Surfactant proteins are proteins which enhance the spreading, adsorption and stability of surfactant lipids that are required for the reduction of surface tension in the alveolus.

► Lung

Susceptibility Gene

► [Susceptibility Locus/Susceptibility Gene](#)

Susceptibility Locus/Susceptibility Gene

Definition

A susceptibility locus/gene is a specific region of any given chromosome that is linked to a particular phenotype or disease.

► [SLE Pathogenesis Genetic Dissection](#)

SV

► [Synaptic Vesicle](#)

Swiss Institute of Bioinformatics

Definition

Swiss Institute of Bioinformatics (SIB) is an academic institution established March 30, 1998 as a non-profit foundation. The goals of this institution are to promote the development of software tools and databases in the field of Bioinformatics (<http://www.isb-sib.ch/>).

► [Protein Databases](#)

Swiss-PdbViewer

Definition

Swiss-PdbViewer is a 3-D structure viewer providing a user-friendly interface, allowing analysis of several proteins at the same time. (<http://us.expasy.org/spdbv/>)

► [Protein Databases](#)

Symmetry Operations

Definition

Transformations of an object that create an indistinguishable object. Transformations of a unit cell (rotations, screw dislocations, mirror planes, inversions) that create an indistinguishable unit cell (by mapping the asymmetric unit onto its symmetry image) represent crystal symmetry operations.

► [X-Ray Crystallography—Basic Principles](#)

Synapse

Definition

Synapse is a specialized zone of contact between neurons, or a neuron and another excitable cell (e.g. a muscle cell), across which impulses are transmitted via synaptic transmitter molecules. Synapses consist of the presynaptic axon, and the postsynaptic target (axon, soma, dendritic shaft or spine, muscle cell, etc.). Some cells in the nervous system have as many as 200,000 synaptic connections.

► [Fragile X Syndrome](#)

► [Neurons](#)

► [Neutrophilic Factors](#)

Synaptic Efficacy

Definition

The synaptic efficacy E is dependent on three parameters: The number of release sites n , the probability of neurotransmitter release P_r , and the amplitude of the signal per synaptic contact, the so-called quantal amplitude q . The synaptic efficacy is a measure of the strength of a given synaptic connection.

► [Neurons](#)

Synaptic Vesicle

Definition

Synaptic Vesicles (SV) are small, non-peptide neurotransmitter-filled vesicles that are concentrated at specialized zones of the presynaptic plasma membrane, described as active zone. Upon Ca^{2+} -influx they fuse

with the plasma membrane and release their content into the synaptic cleft. Thereafter, constituents of the SVs are retrieved from the plasma membrane by endocytosis.

► [Neurons](#)

► [Protein and Membrane Transport in Eukaryotic Cells](#)

Synaptogenesis

Definition

Synaptogenesis designates the building of the synapse, the fundamental structure that is used for neuron-neuron and neuron-muscular communication via neurotransmitters.

► [C. Elegans as a Model Organism for Functional Genomics](#)

► [Neurons](#)

Synaptonemal Complex

Definition

Synaptonemal complex designates a tripartite, proteinaceous structure that forms between the axes of intimately associated (synapsed) homologous chromosomes.

► [Meiosis and Meiotic Recombination](#)

Synaptotagmin I

Definition

Synaptotagmin I is an integral membrane protein of synaptic vesicles. It contains an amino terminal transmembrane domain and two cytoplasmic C2 domains that bind to Ca^{2+} and phospholipids. Among its binding partners are ► [SNARE proteins](#), but also the adaptor protein AP2. Therefore, synaptotagmin I has been implicated in both exo- and endocytosis.

► [Protein and Membrane Transport in Eukaryotic Cells](#)

Synchrotron

Definition

Synchrotron is a circular particle accelerator that generates X-rays as a by-product of particle acceleration. Many modern synchrotrons are designed to

maximize X-ray radiation as the principle benefit of the particle acceleration.

► [3D Structure Determination by X-Ray](#)

Syndecans

Definition

Syndecans are a small family of ► [heparan sulfate proteoglycans](#) of the cell surface. They act as co-receptors for growth factor receptors, but can also bind extracellular matrix glycoproteins such as ► [fibronectins](#) and ► [laminins](#). Intracellularly they are connected to the cytoskeleton and signalling pathways.

► [Extracellular Matrix](#)

► [Signal Transduction: Integrin-Mediated Pathways](#)

Syndetome

Definition

Syndetome denotes a sub-compartment of the ► [sclerotomes](#), lying under the myotome, which will contribute to tendon formation.

► [Somitogenesis](#)

Syndrome

Definition

The term syndrome describes a group of symptoms of congenital anomalies that occur together and define a particular disorder.

► [Chromosomal Instability Syndromes](#)

► [Cleft Lip Palate](#)

Synergism

Definition

Synergism is an effect achieved by two agents that is greater than the combined effect of the individual agents.

► [Neurotrophic Factors](#)

Synergistic Interaction

Definition

Synergistic interaction refers to the arithmetic sum of two phenotypes due to the malfunction of two (or more) genes, which is less than what is observed, or in reverse case, more than what is observed.

► [Drosophila Model of Cardiac Disease](#)

one species is observed to be similarly linked in a second species through evolution.

► [Chromosome 21, Disorders](#)

► [Common Diseases, Genetics](#)

Synonymous Codon Changes

Definition

Nucleobase(s) of a coding triplet are changed in such a way that – due to the degenerated code – the triplet still encodes the same amino acid.

► [SNP Detection and Mass Spectrometry](#)

Syringomyelia

Definition

Syringomyelia describes a spinal lesion with fluid filled cavities.

► [High-HDL Syndrome](#)

Systemic Lupus Erythematosus

► [SLE Pathogenesis, Genetic Dissection](#)

Synpolydactyly

Definition

Synpolydactyly is an autosomal dominant condition with typical abnormalities of the distal parts of both upper and lower limbs (also designated as syndactyly Type II).

► [Repeat Expansion Diseases](#)

Definition

Systems biology comprises of a systematic approach in biology, not focused on individual genes and individual proteins, which is instead interested in analyzing whole systems of genes or proteins, by capturing information from many different elements of the studied system.

► [Microarrays in Plant Genomics](#)

► [Protein Microarrays as Tool for Protein Profiling and Identification of Novel Protein-Protein Interactions](#)

Synteny/Syntenic

Definition

Synteny/syntenic refers to the property of occurring on the same chromosome. Conservation of synteny means that a group of genes that is on a single chromosome in

Systole

Definition

Systole defines the contraction of the chambers of the heart, by which blood is driven into the aorta and pulmonary artery after diastole.

► [Cardiac Signaling: Cellular, Molecular and Clinical Aspects](#)

T Cell

- ▶ [T Lymphocyte](#)

T Lymphocyte

Definition

T lymphocyte (T cell) is a type of immune cell specialized for the production of surface receptors that recognize complexes of peptides that are associated with the products of major histocompatibility complex genes. T cells are divided into subpopulations: T helper-inducer cells; T Supressor cells and cytotoxic T lymphocytes.

- ▶ [Autoimmune Disease](#)
- ▶ [Autoimmunity](#)

Tachycardia

Definition

The term tachycardia characterizes an enhanced frequency of pulse of the heart.

- ▶ [Acute Intermittent Porphyria](#)

TAFs

Definition

The TAFs (TATA-binding protein (TBP) associated factors) are protein subunits of transscription factor complexes that include TBP. These proteins are specific for either RNA Polymerase I, II or III transcription.

- ▶ [Core Promoters](#)

T-Ag

Definition

T-Ag stands for the large tumor antigen (T-Ag) of the simian virus 40.

- ▶ [DNA-based Vaccination](#)

Tagging/Tag

Definition

Tagging is a biochemical/genetic method of adding a functional group (tag) to a protein that will be recognized by a labeling reagent.

- ▶ [Two-dimensional Crystallization of Membrane Proteins](#)
- ▶ [Electron Tomography](#)

TAK1

Definition

TAK1 denotes ▶ [TGFB \$\beta\$ -activated kinase 1](#), which is a member of the ▶ [MAPK](#) family and involved in TGF β , IL-1/18, ▶ [LPS](#), Mac-1 and other signaling pathways. ▶ [Interleukins and Their Receptors, a Role in Regulating Hematopoietic Stem Cells](#)

Talin

Definition

Talin refers to a 250 kDa protein that links many β integrin subunits to actin filaments. Talin binding to β

integrins may be a final common step in integrin activation.

- ▶ Focal Adhesions
- ▶ Integrin Signalling

TAMs

- ▶ Thymidine Analog Mutations

Tandem Mass Spectrometry

Definition

The term tandem mass spectrometry is also called ▶ **MS/MS** analysis. It utilizes two mass spectrometers in a series connected by a chamber known as a collision cell. The sample to be examined is essentially sorted and weighed in the first mass spectrometer, followed by collisional fragmentation induced by an inert gas, and fragment mass analysis in the second mass spectrometer. By using this method, compounds of a complex mixture can be identified. MS/MS can also be applied for identification of structures of organic molecules and for sequencing of proteins.

- ▶ Mass Spectrometry
- ▶ Mass Spectrometry: ESI
- ▶ Mass Spectrometry: Quantitation
- ▶ MS/MS
- ▶ SNP Detection and Mass Spectrometry

Tandem Repeats

Definition

Tandem repeats describe DNA units that are arranged in the head-to-tail manner.

- ▶ Repetitive DNA

Tangier Disease

GERD SCHMITZ, WOLFGANG E. KAMINSKI
Institute for Clinical Chemistry and
Laboratory Medicine, University
Regensburg, Regensburg, Germany
gerd.schmitz@klinik.uni-regensburg.de
wolfgang.kaminski@klinik.uni-regensburg.de

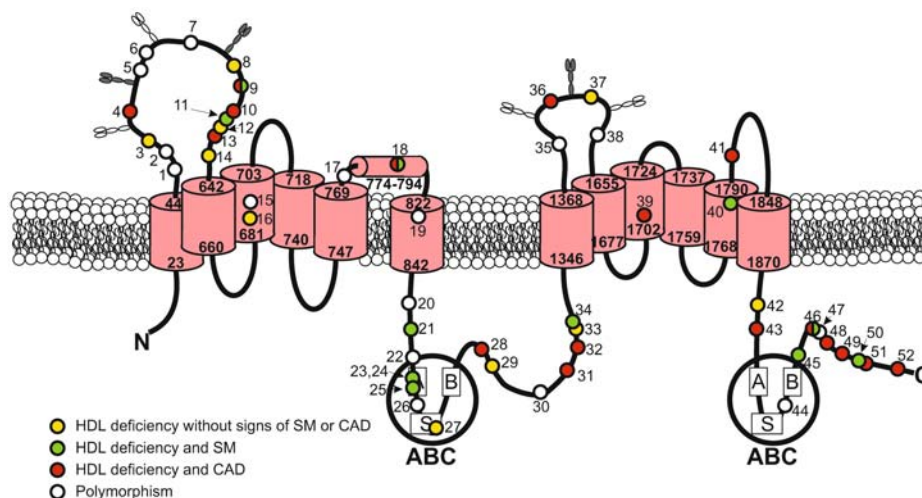
Definition

Familial HDL-deficiency syndromes such as Tangier disease (TD) are rare autosomal recessive disorders of lipid metabolism caused by mutations in the ▶ **ATP-binding cassette transporter ABCA1** (5). These syndromes are characterized by the almost complete absence of plasma ▶ **high-density lipoproteins** (HDL) and variable degrees of ▶ **hypertriglyceridemia**. TD was originally discovered in 1961 among inhabitants of Tangier Island in the Chesapeake Bay (USA), most of whom are descendants of the first settlers of 1686 (2). The clinical features of the classical TD phenotype of familial HDL-deficiency syndromes include very large, orange tonsils that have a characteristic gross and histological appearance and enlarged liver, spleen and lymph nodes, morphologic abnormalities which are the result of excessive accumulation of cholesteryl esters in these organs of the reticulo-endothelial system (RES). The principal cell types found in the spleen resemble monocyte-derived macrophages and platelets. Besides reduced HDL, ▶ **hypcholesterolemia** and abnormal ▶ **chylomicron** remnants are observed in the plasma. A second subset of patients with familial HDL-deficiency is characterized by premature atherosclerosis, which typically manifests with severe coronary artery disease (CAD). Some TD individuals present both with splenomegaly and CAD.

Characteristics

Defective ABCA1 Causes TD and Familial HDL-Deficiency Syndromes

Because of the well-established critical antiatherosclerotic properties of plasma HDL, considerable efforts were made throughout the past decades to identify the molecular defect underlying familial HDL-deficiency syndromes such as TD. The major breakthrough came in 1999, when it was discovered that mutations in the human ABCA1 gene cause these syndromes, thus identifying ABCA1 as a major regulator of HDL metabolism (5) (Fig. 1). In agreement with these observations, a recent study revealed that mice lacking functional ABCA1 exhibit plasma lipid alterations that are concordant with those found in TD (4).



Tangier Disease. Figure 1 Schematic structure of human ABCA1 and the association of known mutations with HDL-deficiency syndromes. Filled circles indicate mutations, open circles represent single nucleotide polymorphisms (numbered by their positions relative to the N-terminus). ABC, ATP-binding cassette; CAD, coronary artery disease; SM, splenomegaly.

Other findings in ABCA1 null mutant mice include vitamin K deficiency, hemorrhagic diathesis, reduced platelet counts and renal failure (4). Moreover, selective disruption of ABCA1 in circulating cells demonstrated that macrophage ABCA1 exerts substantial antiatherosclerotic activity in atherosclerosis. Importantly, this effect is independent of plasma HDL.

Originally, ABCA1 (formerly designated ABC1) was identified in the mouse in 1994. It is the defining member of the A subclass of ATP-binding cassette (ABC) transmembrane transporters which currently includes 12 members which share similar structural features. ABCA1 is expressed in a multitude of human organs with the highest expression levels in placenta, liver, lung, adrenal glands and fetal tissues. A characteristic feature of ABCA1 is that it is upregulated by cholesterol influx in macrophages and suppressed by HDL mediated cholesterol export. The human cDNA codes for a 2261 amino acid polypeptide with a molecular weight of 220 kD. Structurally, ABCA1 exhibits all the characteristics of a full-size transporter i.e. it consists of two tandemly oriented transmembrane domain-ABC subunits (Fig. 1). ABCA1 shows highest amino acid homology (54%) with the recently cloned leukocyte transporter ABCA7 and the retina-specific ABCA4 (52%). The human gene maps to chromosome 9q31.1 and is composed of 50 exons spanning a region of approximately 149 kb.

Association Between Mutation Topology and Clinical Phenotypes in Individuals with TD

More than 50 mutations within the coding region of the human ABCA1 gene have thus far been identified

(Fig. 1). When looking at the topological distribution of known mutations within the ABCA1 protein product, two findings deserve specific attention. First, mutations within the ABCA1 gene appear to occur in clusters. Major mutation clusters can be found in the C-terminal moiety of the first extracellular domain, the N-terminal ATP binding cassette and the C-terminus (Fig. 1). Secondly, mutations in the first extracellular domain and the C-terminus mutations are frequently associated with a cardiovascular phenotype. In contrast, amino acid exchanges in or in the proximity of the N-terminal ATP binding cassette appear to coincide with splenomegaly. These observations raise the possibility that amino acid substitutions within a few selected domains of the ABCA1 protein may, to a significant degree, determine the type of clinical presentation. Accordingly, one may postulate a critical function for the C-terminal moiety of the first extracellular domain and the C-terminus in the development of the cardiovascular phenotype.

ABCA1 appears to be localized in Golgi derived vesicles and on the plasma membrane (4). Surface expression of ABCA1 is up-regulated in macrophages by cholesterol loading. Recent evidence indicates that ABCA1 and *cdc42* are associated with a Lubrol-detergent resistant raft subfraction, whereas ABCA1 is not detectable in Triton-resistant rafts. In addition, the fact that ABCA1 is detectable in cytosolic vesicles and the Golgi compartment of unstimulated fibroblasts also raises the intriguing possibility that it is a mobile molecule that may shuttle between the plasma membrane and the Golgi as a constituent of a vesicular transport route. This is in line with the observation that

the trans-Golgi network in ABCA1 deficient cells is significantly dilated. Initial studies on the biological role of ABCA1 supported the view that this transporter, like MDR1 and MDR3, functions as a plasma membrane lipid pump. This was based on experiments showing an increase in cholesterol and phospholipid export under conditions of forced expression of ABCA1 in ABCA1 null mutant cells from patients with genetic HDL-deficiency which characteristically display the reverse scenario (4). However, recent work from our laboratories indicated that the ATP turnover of ABCA1 occurs at a very low rate whereas Mg^{2+} -dependent ATP binding induces conformational changes. Based on this information it is likely that ABCA1 acts rather as a regulator facilitating cholesterol/choline-phospholipid export within the cellular lipid export machinery than exerting a genuine translocator function. It will be exciting to elucidate the exact molecular mechanisms by which ABCA1 mediates the export of lipid compounds from the cell. In this context, a most critical aspect is the question as to which molecular partners interact with and thus potentially modulate ABCA1 function. Work from our laboratory has provided potential clues by demonstrating that ABCA1 interacts with a C-terminal PDZ binding domain in the [▶ \$\beta_2\$ -syntrophin/utrophin complex](#) that may couple ABCA1 to the [▶F-actin cytoskeleton](#). Moreover, evidence has recently been presented that ABCA1 function depends on binding to [▶Fas-associated death domain \(FADD\)](#) protein, an adaptor molecule involved in death receptor signal transduction (1).

Cellular and Molecular Regulation and Clinical Relevance

The TD Gene ABCA1 Controls Susceptibility to Atherosclerosis

Macrophages play a key role in the initiation and progression of atherosclerotic lesions. In the nascent lesion, they transform into foam cells through the excessive accumulation of cholesteryl esters. Dysfunctional lipid homeostasis in macrophages and foam cells ultimately results in the breakdown of membrane integrity and cell death. An interesting clue as to how ABCA1 may be implicated in the control of monocyte/macrophage targeting in TD comes from the observation that apoAI mediated lipid efflux in ABCA1 deficient cells is paralleled by the down-regulation of the protein cdc42 and filopodia formation. Cdc42, like rho and rac, is a member of the family of small GTP-binding proteins that controls formation of filopodia. [▶Rho proteins](#) induce the formation of stress fibers and focal adhesions and rac proteins regulate formation of lamellipodia and membrane ruffles. It is thus conceivable that ABCA1 modulates cellular mobility of monocytes/macrophages through this mechanism and

thus may affect recruitment of monocytes into the vessel wall. At the vascular wall level, this regulator function on filopodia formation may even extend to platelets, vascular smooth muscle cells and endothelial cells since these cells are also capable of expressing ABCA1.

The finding that ABCA1 is up-regulated in human macrophages during sustained uptake of cholesterol and that it facilitates cholesterol export from the cell render it a likely antagonistic player in foam cell development and the formation of the lesions of atherosclerosis. A number of transcriptional control elements have been identified in the ABCA1 promoter region. Among these, the [▶zinc finger protein ▶ZNF202](#) appears to function as a major repressor of ABCA1 transcription. Other factors that regulate the transcription of ABCA1 include [▶oncostatin M](#) and [▶geranylgeranyl pyrophosphate](#) (3). In addition, several [▶cis-acting elements](#) have been identified that control the expression of ABCA1. These include an E-box and consensus sequencing that are known to interact with the transcription factors Sp1/3, however, no [▶PPAR \$\alpha\$ - or \$\gamma\$ -consensus elements](#) are present in the ABCA1 promoter region.

Among patients with TD, one subgroup develops premature CAD, whereas another one presents predominantly with splenomegaly indicating differences in macrophage targeting to tissues in ABCA1 deficiency, which may be a reflection of the nature of the mutation in the ABCA1 gene. Based on this, it stands to reason that ABCA1 *per se* determines monocyte targeting into tissues. In recent experiments using chimeric [▶LDLR-/- mice](#) lacking ABCA1 in their blood cells, we tested the hypothesis that the selective disruption of ABCA1 in circulating cells, in particular macrophages, affects macrophage targeting into the vascular wall and lesion formation *in vivo*. Our results demonstrated that the absence of ABCA1 from leukocytes only is sufficient to induce aberrant monocyte recruitment into the spleen identifying ABCA1 as a critical leukocyte factor in the control of monocyte targeting. Moreover, it could be shown that LDLR-/- chimeras deficient in ABCA1 develop significantly larger and more advanced atherosclerotic lesions compared to chimeric LDLR-/- mice with functional ABCA1 (6). Of particular importance was the observation that targeted disruption of leukocyte ABCA1 function did not affect plasma HDL cholesterol levels. These results thus provided direct evidence that leukocyte ABCA1 exerts significant anti-atherosclerotic activity which is clearly independent of plasma HDL. Given the fact that macrophages are by far the predominant inflammatory cell type within lesions of atherosclerosis, these data strongly suggest that the observed increased susceptibility to atherosclerosis is largely attributable to the absence of

ABCA1 from lesional monocytes/macrophages. The view that macrophage ABCA1 functions as an important antiatherosclerotic factor has also been supported by more recent studies. For example targeted inactivation of ABCA1 in macrophages leads to increased lesion formation and foam cell accumulation in [▶apoE null mutant mice](#). Furthermore, in a reciprocal approach using ABCA1 over-expressing transgenic [▶apoE null mice](#), it has been shown that forced expression of human ABCA1 results in significantly smaller, less complex lesions in [▶apoE knockout mice](#). Interestingly, ABCA1 is highly expressed in differentiating megakaryocytes and circulating platelets indicating that monocytes/macrophages are not the only source of ABCA1 within the hematopoietic system (unpublished data). It is thus possible that altered platelet function due to the lack of ABCA1 may contribute to the observed anti-atherogenic effect.

References

1. Büchler C, Bared SM, Aslanidis C et al (2002) Molecular and functional interaction of the ATP-binding cassette transporter A1 with Fas-associated death domain protein. *J Biol Chem* 277:41307–41310
2. Fredrickson DS, Altrocchi PH, Aviola LV et al (1961) Tangier disease. *Ann Intern Med* 55:1016–1031
3. Langmann T, Porsch-Ozcurumez M, Heimerl S et al (2002) Identification of sterol-independent regulatory elements in the human ATP-binding cassette transporter A1 promoter: role of Sp1/3, E-box binding factors, and an oncostatin M-responsive element. *J Biol Chem* 277:14443–14450
4. Orsó E, Broccardo C, Kaminski WE et al (2000) Transport of lipids from Golgi to plasma membrane is defective in Tangier disease patients and Abc1 deficient mice. *Nat Genet* 24:192–196
5. Schmitz G, Kaminski WE, Orsó E (2000) ABC transporters in cellular lipid trafficking. *Curr Opin Lipidol* 11:493–501
6. van Eck M, Bos IS, Kaminski WE et al (2002) Leukocyte ABCA1 controls susceptibility to atherosclerosis and macrophage recruitment into tissues. *Proc Natl Acad Sci USA* 99:6298–6303

TAP are involved in the transport of peptides from the cytosol to the endoplasmic reticulum.

[▶DNA-based Vaccination](#)

TAP Tag

Definition

TAP tag stands for Tandem affinity purification tag. It refers to an affinity tag combination of *Staphylococcus aureus* protein A, a calmodulin binding peptide and a specific cleavage site of tobacco etch virus (TEV) protease.

[▶Affinity Chromatography and In Vitro Binding \(Beads\)](#)

TAP/NXF1

Definition

TAP/NXF1 stands for the export receptor for mRNA in metazoans, which functions as a heterodimer with a protein partner termed p15 or NXT1. TAP/NXF1–p15 heterodimers are structurally unrelated to [▶karyopherins](#) and do not require [▶Ran-GTP](#) to bind cargo.

[▶Nuclear Pore Complex](#)

[▶RNA Export](#)

Target Gene

Definition

Target gene designates a gene that codes for a protein whose activity, when inhibited (or, more rarely, increased), would affect the disease course.

[▶High-Throughput Approaches to the Analysis of Gene Function in Mammalian Cells](#)

TAP

Definition

Two genes that are located in the MHC class II region are called “transporters that are associated with antigen processing 1 and 2” (TAP–1 and –2) and encode proteins of the ATP binding cassette (ABC) family.

Target Registration DataBase

Definition

The TargetDB is a database, kept at the [▶Protein Data Bank](#) (PDB) at Rutgers University, which mirrors the status of work in 20 [▶structural genomics](#) centers around the world. The TargetDB is automatically

updated by direct access to web sites maintained by the participating centers. It can be used to monitor the success of an individual center or of the entire ►[structural genomics](#) community in terms of genes selected, cloned and expressed, proteins obtained in soluble form, purified and crystallized, the number and outcome of X-ray diffraction or NMR experiments, and the number of protein structures determined and submitted to the PDB. As of August 03, 2004, the TargetDB listed 924 completed crystal structures and 343 NMR structures, 940 of which had already been deposited with the PDB. <http://targetdb.pdb.org/>

►[Structural Genomics: Structure-to-Function Approaches](#)

genome via homologous recombination. As a consequence, essential elements of the target gene are deleted by the selection cassette.

►[Transgenic and Knockout Animals](#)

Tat

Transactivation protein of HIV.

►[Retroviruses](#)

TargetDB

►[Target Registration DataBase](#)

Targeted Gene Disruption

Definition

Experimental replacement of an endogenous gene by a mutated (non-functional) version within the otherwise unchanged genome is called targeted gene disruption. Embryonic stem cells altered in this way allow the generation of mouse strains that stably carry the single gene mutation in the germ line, and therefore propagate it in a mendelian fashion to the offspring. Complete loss of the gene function in homozygously mutated mice often mimics deficiency states in humans.

►[Muscle Development](#)

►[Peutz-Jeghers Syndrome](#)

TATA Binding Protein

Definition

TATA binding protein (TBP) is a protein that binds to a consensus TATA-box sequence element, and is found in transcription factor complexes involved in RNA Polymerase I, II and III transcription.

►[Core Promoters](#)

►[RNA Polymerase I](#)

TATA Box

Definition

The TATA box is a consensus sequence in the promoter region of many eukaryotic genes that is recognized by the TATA binding protein and specifies the position where transcription is initiated by the RNA polymerase.

►[Core Promoters](#)

Targeting Vector

Definition

A targeting vector is a recombinant DNA construct that consists of a selection cassette, tailored by DNA arms homologous to genomic DNA, which flanks critical elements of a target gene. When introduced into ►[ES cell](#) the targeting vector integrates into the ES-cell

Tauopathies

Definition

Tauopathies are neurodegenerative diseases that are characterized by linked mutations of the tau protein. Tau binds to and stabilizes microtubules. Its functional deficiency is suggested to result in certain forms of

Alzheimer disease, frontotemporal dementia, progressive supranuclear palsy and corticobasal degeneration. Degenerating neurons contain neurofibrillary tangles.

► [Cytoskeleton](#)

Tau-Protein

Definition

Tau-protein denotes a microtubule binding protein that is involved in the development of Alzheimer Disease.

► [Cardiac Signaling: Cellular, Molecular and Clinical Aspects](#)

► [Tauopathies](#)

Tax

The transactivation protein of ► [Human T-cell leukemia virus \(HTLV\)](#).

► [Retroviruses](#)

Taxanes

Definition

Taxanes are natural anticancer products acting as chemotherapeutics by pathological generation and stabilization of microtubules based on their affinity, thus hindering mitosis. They act in a cell cycle dependent manner. Commonly used taxanes include

► [Paclitaxel](#) and [docetaxel](#).

► [Multi-Drug Resistance](#)

Taxon (pl. Taxa)

Definition

Taxon is the taxonomic unit in the natural classification. It groups individuals of a species or sets of species.

► [Genetic Code](#)

TBCE Gene

Definition

The tubulin-specific chaperone E (*TBCE*) gene encodes one of several ► [chaperone](#) proteins that are required for the proper folding of α -tubulin subunits and the formation of α - β -tubulin heterodimers.

► [Hyper- and Hypoparathyroidism](#)

TbetaR-I/TbetaR-II (T β R-I/T β R-II)

Definition

T β R-I/T β R-II stands for TGF- β receptors Type I and Type II.

► [Receptor Serine/Threonine Kinases](#)

► [TGF- \$\beta\$](#)

TBP

► [TATA Binding Protein](#)

TBP-Associated Factors

Definition

The TAFs are protein subunits of transcription factor complexes that include TBP. These proteins are specific for either RNA polymerase I, II or III transcription.

► [RNA Polymerase I](#)

T-Cell Epitope

Definition

The T-cell epitope is an antigenic determinant recognized and bound by the T-cell receptor. Epitopes

recognized by the T-cell receptor are often located in the inner, unexposed side of the antigen, and become accessible to the T-cell receptors after proteolytic processing of the antigen.

► [Peptide Chips](#)

T-Cell Factors

Definition

T-Cell Factors (TCFs) comprise of a family of ► [HMG](#) (high mobility group) containing transcription factors that include TCFs and LEF-1 (► [LEF/TCF Family](#)) (lymphoid enhancer factor-1) (► [LEF/TCF Family](#)), and function as transcriptional repressors in the absence of symbol.

► [Wnt/Beta-Catenin Signaling Pathway](#)

► [β-Catenin](#)

TCF

► [LEF/TCF Family](#)

TCFs

► [T-Cell Factors](#)

Telocentric Chromosome

Definition

Telocentric chromosome is a chromosome on which the centromere is located at one end.

► [Centromeres](#)

Telomerase

KENKICHI MASUTOMI, WILLIAM C. HAHN

Department of Medical Oncology, Dana-Farber Cancer Institute, Boston, MD, USA

william_hahn@dfci.harvard.edu

Synonyms

Telomerase Reverse Transcriptase; *TERT*

Definitions

► [Telomeres](#) consist of large arrays of short, G-rich sequences and specific binding proteins. These nucleoprotein structures compose the ends of linear eukaryotic chromosome and play important roles in chromosome replication, the maintenance of chromosomal stability and cell lifespan regulation. In mammals, the telomere sequence is TTAGGG. Telomeres are synthesized by a ► [ribonucleoprotein](#) ► [reverse transcriptase](#) called telomerase. Catalytically active telomerase is composed of two core components, an RNA subunit (telomerase RNA component: TERC) and a catalytic protein subunit (telomerase reverse transcriptase: TERT).

Characteristics

Biochemical Characteristics

Telomerase synthesizes telomeric repeats using a template provided by TERC. Recent evidence suggests that like many other reverse transcriptases, telomerase exists as a dimer (1). Several proteins associate with these two core components (TERC and TERT), but none of these proteins are necessary for telomerase activity *in vitro*.

Biological Characteristics

Telomeres shorten with successive rounds of cell division. Telomere attrition to critically short lengths leads to end-to-end chromosomal fusions and induces ► [apoptosis](#). Constitutive expression of telomerase stabilizes telomere length. Experimentally, constitutive expression of TERT directly immortalizes some types of human cells (2) and cooperates with specific combinations of oncogenes to transform human cells (3).

Clinical Relevance

Cancer

Telomerase activity is strongly associated with human cancer and most human cancer cell lines exhibit stable telomere lengths during propagation in culture. These observations suggest that telomere maintenance by

telomerase is an important factor in the development of cancer. However, recent work indicates that telomere biology plays a complex role in cancer development, serving both to suppress and facilitate malignant transformation.

Limitation of Replicative Lifespan and Suppression of Tumor Formation

Normal human cells exhibit a restricted replicative lifespan when propagated in culture. In primary human cells, telomeres shorten with successive cell division and this shortening correlates with their replicative capacity. Since most asynchronously proliferating normal human cells fail to maintain stable telomere lengths, these observations indicate that alterations in telomere state trigger ►replicative senescence, an irreversible proliferative arrest, and limit replicative capacity. This telomere-induced limit suppresses malignant transformation.

Telomere Shortening Promotes Genomic Instability and Tumor Initiation

Telomere shortening eventually leads to telomeres that are too short to protect chromosomes from fusion. Cells that harbor critically short telomeres exhibit increased genomic instability and most die by apoptosis at a period called ►crisis. However, rare cells survive crisis. Such survivors typically exhibit gross changes in karyotype called ►aneuploidy and are immortal. Despite exhibiting critical telomere shortening upon entering crisis, cells that survive crisis maintain stable telomere lengths by the activation of telomerase. Thus, critical telomere shortening induces chromosomal instability that promotes, among other events, the acquisition of mechanisms to maintain stable telomere lengths. This stabilization of telomere length permits immortalization and facilitates further malignant progression. In addition, the genomic instability induced by critical telomere shortening facilitates the further accumulation of other genetic alterations that contribute to malignant progression.

Telomerase Activation and Tumor Promotion

Although the expression of *hTERT* alone facilitates the acquisition of the immortal state, telomerase alone fails to transform human cells. However, telomerase expression cooperates with certain oncogenes to induce cell transformation. The observation that inhibition of telomerase in human cancer cells limits the proliferation of human cancer cell lines and induces cell death suggests that telomerase activation is a critical event during malignant transformation. Recent work suggests that telomerase activation contributes to transformation both by stabilizing telomere lengths and by affecting cell physiology in a manner independent of its effects on telomere length.

Telomerase as a Therapeutic Target

Since telomere maintenance plays an important role in cancer development, targeting telomeres and telomerase is a potential target for cancer therapeutics. Preclinical studies targeting telomerase by pharmacological or biological means show much promise but further efforts are necessary to develop useful clinical agents.

Dyskeratosis Congenita

Dyskeratosis congenita (DKC) is a rare inherited condition, characterized by bone marrow failure, abnormal skin pigmentation, abnormal nail growth, premature aging and an increased risk of malignancy, particularly of the skin and gastrointestinal tract. Most patients succumb to bone marrow failure and/or cancer. Epidemiological studies demonstrate two patterns of inheritance, X-linked and autosomal dominant. The major X-linked form of the disease is due to mutations in a nucleolar protein, dyskerin, which participates in the processing of ribosomal RNA (4). Dyskerin is also found in the telomerase complex, and cell lines derived from patients with the X-linked form of DKC are deficient in telomerase activity. Germline mutations in the RNA component of telomerase (*hTERT*) are found in the autosomal dominant form of DKC (5). Thus DKC is a disease of telomerase deficiency.

- Molecular Aging Research
- Breast Cancer
- Cellular Senescence

References

1. Wenz C, Enenkel B, Amacker M et al (2001) Human telomerase contains two cooperating telomerase RNA molecules. *EMBO J* 20:3526–3534
2. Bodnar AG, Ouellette M, Frolkis M et al (1998) Extension of life-span by introduction of telomerase into normal human cells. *Science* 279:349–352
3. Hahn WC, Counter CM, Lundberg AS et al (1999) Creation of human tumour cells with defined genetic elements. *Nature* 400:464–468
4. Mitchell JR, Wood E, Collins K (1999) A telomerase component is defective in the human disease dyskeratosis congenita. *Nature* 402:551–555
5. Vulliamy T, Marrone A, Goldman F et al (2001) The RNA component of telomerase is mutated in autosomal dominant dyskeratosis congenita. *Nature* 413:432–435

Telomerase Reverse Transcriptase

- Telomerase

Telomere

Definition

A telomere (also called telomeric repeats) is a region of highly repetitive DNA at the end of a eukaryotic chromosome. They are formed by tandem repeats of short DNA sequences and associated proteins. Telomeres facilitate complete DNA replication of the chromosome, and prevent the chromosomal ends from joining with other chromosomes. The shortening and possible disappearance of telomeres is believed to be the basis for cessation of cell division in many cells, which eventually leads to senescence. One reason why continuously proliferating cancer cells are immortal, is thought to be that their telomeres do not shorten with each round of cell division due to the action of telomerase.

- [Breast Cancer](#)
- [Cellular Senescence](#)
- [DNA Helicases](#)
- [Repetitive DNA](#)
- [Telomerase](#)
- [Transposons](#)

Telophase

Definition

Telophase is the last stage of mitosis, when one complete set of chromosomes arrive at each spindle pole. They are enclosed within the nuclear envelope, and become a new daughter nucleus.

- [Centromeres](#)
- [Mitosis](#)

Temperature Compensation

Definition

When the period of a circadian rhythm remains relatively stable with changes in temperature this is called temperature compensation.

- [Circadian Clocks](#)

Temperature Factor

Definition

The usual model of motion used to refine crystal structures assumes that each atom vibrates harmonically with an amplitude proportional to the square root of the temperature factor. The temperature factor is a fitting parameter used in protein structure refinement.

- [3D Structure Determination by X-Ray](#)

Temperature-Sensitive Mutant

Definition

Temperature-sensitive mutant refers to an organism that has a wild-type phenotype at a permissive temperature, but a mutant phenotype at a restrictive (non-permissive) temperature.

- [Mutagenesis Approaches in Medaka](#)

Template Matching

Definition

Template Matching is cross-correlation of a template with the data volume data in order to identify components in the data set that are not visible by inspection alone.

- [Electron Tomography](#)

TER

- [Trans-Epithelial Electrical Resistance](#)

Teratogens

Definition

Teratogens are substances (environmental exposures) that can lead to congenital structural and/or functional anomalies.

► Cleft Lip Palate

Terminal Ductal-Lobular Unit

Definition

The mature human female breast contains thousands of hormone-sensitive, potentially milk-producing micro-organs, called lobules. In the nonpregnant state they vary in number from 1 to 8. Each lobule is drained by a terminal duct attached to the main duct system. The peripheral epithelial units are called the terminal ductal-lobular units (TDLU). A TDLU consists of the terminal branch of the ductal tree, and the 30 to 50 ductules branching there from and forming a lobule.

► Breast Cancer

Terminator

Definition

The terminator is a sequence element at which RNA polymerase stops transcription.

► RNA Polymerase I

TERT

► Telomerase

Tertiary Structure

Definition

Tertiary structure refers to the three-dimensional structure of a polypeptide chain, resulting from the folding and arrangement of alpha helices and beta pleated sheets. The tertiary structure is determined by the amino acid sequence, in dependence on the distinct surrounding milieu. The folded structure is stabilised by non-covalent interactions between different parts of the polypeptide chain.

► Classification of Active Centers

Testis-Determining Gene

► SRY – Sex Reversal

Testosterone

Definition

Testosterone is the strongest androgen (testicular hormone) produced in the Leydig cells of the testicles. Testosterone is responsible for the development of males and governs, together with its derivative dihydrotestosterone, the appearance of the physical traits of the male.

► SRY – Sex Reversal

Tetany

Definition

Tetany describes a clinical syndrome that is characterized by muscle twitches, cramps, carpopedal spasm of increasing severity, laryngospasm and seizures. Tetany usually results from low serum levels of ionized calcium (or magnesium) causing irritability of the central and peripheral nervous systems.

► Hyper- and Hypoparathyroidism

Tetraspanins

Definition

Tetraspanins are transmembrane proteins that span the membrane four times. They contain specific conserved residues in their transmembrane domains, one small and one large extracellular loop, and are involved in many cellular processes such as adhesion, motility, activation and proliferation. Certain tetraspanins interact with integrin receptors.

- ▶ Hemidesmosomes
- ▶ Integrins

composed of 2 modules. The CAK, Cyclin-dependent kinase Activating Kinase module, is active independently of TFIH, and it phosphorylates and activates cyclin-dependent kinases involved in cell-cycle control. The core TFIH module comprises of two ATP-dependent DNA Helicases XPB and XPD. Defects in XPB and XPD are associated with genetic inherited diseases such as ▶Xeroderma Pigmentosum, Cockayne syndrome and trichothiodystrophy. Patients affected by such diseases are deficient in ▶DNA repair, and therefore sensitive to UV light exposure. Furthermore, some of them are prone to develop cancers.

- ▶RNA Polymerase II Transcription

Tetratricopeptide Repeat Motif

Definition

The tetratricopeptide repeat (TPR) motif is a protein-protein interaction module of 34 amino acids with the consensus sequence 4 (W/L/F), 7 (L/I/M), 8 (G/A/S), 11 (Y/L/F), 20 (A/S/E), 24 (F/Y/L), 27 (A/S/L), and 32 (P/K/E), which is often arranged in tandem arrays. It is found in multiple copies in a number of functionally different proteins having specific interactions with partner proteins.

- ▶Peptidyl Prolyl Cis/Trans Isomerases

TGF

Definition

TGF stands for transforming growth factor. It is a member of the family of growth factors originally identified by their ability to induce malignant transformation of cells in culture. They affect cell proliferation but also synthesis of extracellular matrix proteins in cells. TGF α belongs to the EGF family of ligands for receptor tyrosine kinases. TGF β is the prototype and the collective name for a superfamily of secreted disulfide-linked dimeric polypeptides, which are ligands for receptor serine/threonine kinases. TGF β 's exhibit potent morphogenetic roles during embryogenesis. They modulate growth, apoptosis and differentiation of most adult cell types.

TGF- β and related family members bind to receptor heterodimers that activate the ▶Smad proteins to bind to DNA.

- ▶Growth Factors
- ▶Receptor Serine/Threonine Kinases
- ▶Transcription Factors and Regulation of Gene Expression

TFIID

Definition

TFIID is the multi-protein transcription factor that contains ▶TBP and TBP associated factors (▶TAFs) and binds to the TATA, ▶Inr and ▶DPE.

- ▶Core Promoters

TFIIH

Definition

TFIIH refers to the general transcription initiation factor that is involved both in transcription and in repair of damaged DNA by nucleotide excision. It is

TGF β Activated Kinase-1

Definition

TGF β Activated Kinase-1 (TAK1) denotes a TGF β -activated kinase that has been shown to phosphorylate ▶Nemo like kinase (NLK).

- ▶Wnt/Beta-Catenin Signaling Pathway

Th-1 Biased T Cell Response

Definition

A Th-1 biased immune response specifically primes specific T cell responses associated with strong IFN γ production.

►DNA-based Vaccination

TH1/Th2 Cells

Definition

Th1/Th2 cells are subsets of T helper lymphocytes, they are defined by their capabilities for production of different sets of cytokines.

►Rheumatoid Arthritis

Thalassaemia

Definition

Thalassaemia is an inherited disorder characterized by defective haemoglobin synthesis (Haemoglobinopathies), which is caused by a mutation in the gene coding for the α or the β globin subunit. In β -thalassaemia, homozygous patients are affected by a most severe transfusion dependent anaemia (thalassaemia major), whereas carriers with one inherited gene are healthy (thalassaemia minor). α -thalassaemias are genetically more complex, because the α -globin gene locus contains two highly linked virtually identical genes. When only one or two of the four genes are deleted, carriers are healthy. When three genes are deleted, a haemolytic anaemia of varying severity develops (Hb H disease). A deletion of all four genes usually results in stillbirth (Hb Barts hydrops fetalis).

►Polyadenylation

Thanatophoric Dysplasia

Definition

Thanatophoric dysplasia refers to a dominant lethal form of chondrodysplasia characterized by very short

limbs and flat vertebral bodies. Activating mutations within fibroblast growth factor receptor 3 (FGFR3), a receptor tyrosine kinase, are responsible for the disease.

►Bone Disease and Skeletal Disorders, Genetics

Therapeutic Immunization

Definition

Therapeutic vaccines are administered to patients post-exposure to control chronic infection.

►DNA-based Vaccination

Thermodynamic Integration

Definition

Thermodynamic integration (TI) denotes a statistical thermodynamical method to calculate the free energy difference between two states A and B by gradually transforming the potential of state A into that of state B according to a defined path variable or coupling parameter. The free energy change is obtained by integrating over ensemble averages of the derivative of the potential with respect to the coupling parameter.

►Molecular Dynamics Simulation in Drug Design

Thermodynamic Properties of DNA

G. ERIC PLUM^{1,2}, JENS VÖLKER²,
KENNETH J. BRESLAUER¹

¹Department of Chemistry and Chemical Biology,
Rutgers, State University of New Jersey, Piscataway,
NJ, USA

²IBET, Inc., Columbus, OH, USA
kjbdna@rci.rutgers.edu

Definition

DNA, like other biopolymers, can assume a large number of different structures, a feature often referred to as structural polymorphism. Since structure generally dictates or modulates biological function, it is

important to know the relative stabilities of the different structural forms. Such assessments require thermodynamic data.

Today we know that DNA molecules are far more polymorphic than originally implied by the elegant Watson-Crick double helical structure. In addition to this fully paired duplex structure, DNA also can form imperfect duplexes with misaligned templates or hairpin loops, as well as duplexes with damaged/modified bases, mismatches or gaps. DNA can also form higher ordered structures, such as triplexes, tetraplexes and organized aggregates. The relative stabilities of these polymorphs depend on base composition, base sequence, strand length, solution conditions and the stability of interactions with effector molecules, including proteins.

Recent advances, perhaps most visible in the Human Genome Initiative, have presented new and compelling means and objectives to exploit the exquisite specificity of nucleic acid [hybridization](#). Precise knowledge of the temperature dependent stability of nucleic acid (and nucleic acid analog) complexes is necessary for the rational design and optimization of hybridization probes for use in high throughput screening, nucleic acid based diagnostics and therapeutics and numerous emerging biotechnology applications. While the subject of this article is DNA duplex thermodynamics, the principles described are general for nucleic acid complexes and, with appropriate modification, can be applied to nucleic acid-protein and nucleic acid drug interactions.

The set of thermodynamic parameters describing a DNA system can be used for 2 purposes, to understand the forces that stabilize the various forms of DNA and to develop predictive models of structure and interactions. Thermodynamic analysis is based on characterization of the macroscopic properties of the system containing the DNA, which under [equilibrium](#) conditions define a [thermodynamic state](#). The thermodynamic state is described by a set of [state functions](#). Because only changes in the state functions are measurable, thermodynamic analysis describes the differences between thermodynamic states. The measured thermodynamic parameters depend on the differences between the states and not on the path taken from one state to the other. Selection of an appropriate reference state is critical for the successful application of the thermodynamic model. To compare thermodynamic descriptions of different systems, one must be careful to establish a common [reference state](#). Typically the fully disordered, single stranded state observed at high temperature is used in DNA thermodynamic studies. It is reasonable to assume that sequence effects are irrelevant under these conditions. In principle, the state functions can be dissected into contributions from various forces, although rarely is

sufficient information available to do so with confidence. However, part of the power of the thermodynamic approach is that one need not invoke a microscopic model to make use of the information.

Thermodynamic Parameters

A variety of parameters are used to describe a thermodynamic system. The most important state function is the [free energy change \(\$\Delta G\$ \)](#), which represents the relative stability of two states. The free energy change can be dissected into contributions from the [enthalpy change \(\$\Delta H\$ \)](#) and the [entropy change \(\$\Delta S\$ \)](#). The enthalpy change is dominated by short-range interactions, including van der Waals contacts and base stacking interactions, with a small net contribution from hydrogen bonding between the paired bases. The enthalpy change describes the temperature dependence of the free energy change, i.e. the shape of the melting curve, and is the most readily measurable of the thermodynamic parameters. The entropy change (ΔS) represents changes in order associated with the two states. The entropy term includes contributions from conformational changes, although it may be dominated by rearrangements of associated solvent molecules and counterions. Empirical correlations between changes in the solvent accessible surface area of macromolecules and the [heat capacity change \(\$\Delta C_p\$ \)](#) at constant pressure have been reported. The heat capacity change is related to the temperature dependence of the enthalpy and entropy changes. Although it is not a true thermodynamic parameter, the [melting temperature \$T_m\$](#) is used frequently and erroneously to describe stability. Because it describes the midpoint of the thermally induced order disorder transition the T_m has some utility. The T_m value represents the thermal stability as opposed to the thermodynamic stability described by ΔG . Because it depends on concentration (for duplexes and higher order structures) and because it does not describe the shape of the melting transition, the T_m is not useful for predicting the relative stability of nucleic acid complexes at different temperatures. A complete thermodynamic description of the system is necessary to make accurate predictions of relative stabilities.

Characteristics

Nearest Neighbor Model

While one need not consider the structure of DNA to apply the thermodynamic approach, interpretation of the thermodynamic parameters in terms of a structural model is useful so long as one recognizes that the thermodynamic parameters do not depend on the model. The stability of a DNA duplex based on its structure can be predicted at several levels of detail. Duplexes tend toward greater stability as the GC base pair content increases. However, large stability

differences are observable between duplexes of like or similar base pair composition. Therefore sequence information must be included to provide accurate predictions of duplex stability. The [▶nearest neighbor model](#) has proved to be a very useful approach to including sequence effects on DNA thermodynamics. In the nearest neighbor model the thermodynamic parameters are assumed to be divisible into pairwise contributions from each set of neighboring base pairs. Of the 16 possible nearest neighbor base pairs only 10 are unique and need to be determined experimentally (2). Several sets of experimentally determined DNA nearest neighbor thermodynamic parameters have been published, including those from the Gotoh, Breslauer, Klump, Blake, Benight, Sugimoto and SantaLucia laboratories (4). While there is some variation for specific sequences, these data sets provide roughly equivalent predictions, as shown by Benight (4). For short oligonucleotides the thermodynamic parameters can be predicted by adding values from a table of nearest neighbor contributions (1) (including small corrections for certain conditions, such as self complementary oligonucleotides). Because the nearest neighbor model does not take explicit account of helix length it must be used with caution when applied to intermediate length oligonucleotides. Additivity of the nearest neighbor values fails for polymeric DNA as polymer length exceeds the length of the cooperative melting unit; however, the same data set of pairwise thermodynamic values can be used within a polymer model (6).

Limitations of the Nearest Neighbor Model

Excepting some scattered data, consideration of the important issues of template misalignment, i.e. mismatches and loops, and damaged or modified bases is not to date included in the nearest neighbor data sets. In addition, the nearest neighbor data sets are specific for particular solution conditions, with the concentration of monovalent cations being the primary variable. Use of the nearest neighbor values for the stability of duplexes under different solution conditions should be done with caution, although for most solution conditions the effects will scale roughly independent of sequence. Thermodynamics addresses only equilibrium states. While the path between states is not reflected in the thermodynamic parameters, it is important for describing some aspects of nucleic acid behavior and for determining the appropriate model to derive model-dependent thermodynamic data (5). There are two contributions to the formation of a duplex, [▶helix initiation](#) and [▶helix propagation](#). As the two strands come together a few base pairs form an [▶initiation complex](#); this is a [▶bimolecular process](#) and depends on DNA strand concentration. Subsequent to the helix initiation step base pairing propagates to form the

complete double helix; this process is pseudomonomolecular, because the strands have already found each other, and does not depend on strand concentration. Formation of short oligonucleotide duplexes is dominated by the initiation step and is strongly dependent on concentration. Formation of polymeric duplexes is dominated by helix propagation and is effectively independent of concentration. Both short oligonucleotide and polymer formation are well described by traditional models. The formation of intermediate length helices depends in a complex way on both helix initiation and propagation. Current models do not describe such molecules satisfactorily.

Measurement of Thermodynamic Parameters

The preferred method for measuring thermodynamic parameters for DNA is [▶differential scanning calorimetry](#), wherein the apparent heat capacity of the DNA-containing solution is measured directly as a function of temperature. A heat capacity anomaly is observed over the temperature range where the strands separate. The area under the DSC curve is proportional to the energy holding the duplex together. Thus integration of the DSC curve provides a direct measurement of the ΔH value associated with the disruption of the duplex (3). From the shape of the curve one can derive values for ΔS , ΔC_p and T_m . The value of ΔG at any temperature is readily determined from the familiar relation $\Delta G = \Delta H - T \Delta S$. Note that the calorimetric approach requires no assumptions about the mechanism of DNA strand separation; one need know only the DNA concentration to normalize the DSC [▶thermogram](#). The requirement for relatively large amounts of DNA and the requirement for expensive instrumentation have inhibited the widespread use of the calorimetric methodology.

By assuming that the duplexes under study dissociate in a [▶two-state process](#), one can apply the [▶van't Hoff equation](#) ($\Delta H = -R(\partial \ln K / \partial (1/T))$, where R is the gas constant and K is the equilibrium association constant) to extract thermodynamic parameters from experimental data that quantify the relative amounts of duplex and single stranded DNA as a function of temperature. Because the data are not sufficient to extract the additional parameter, when employing the van't Hoff model, it is generally assumed that $\Delta C_p = 0$. From DSC data it is known that ΔC_p is finite, but small. The most common application of the van't Hoff methodology is to optical absorbance *versus* temperature curves. The absorbance of ultraviolet light by duplex DNA is reduced relative to the corresponding single strands, thus providing a means of quantifying their relative populations. Several methods of analysis provide van't Hoff model-dependent thermodynamic parameters from UV melting curves (3). The primary advantage of this method relative to calorimetry is the availability

of suitable instrumentation. The data derived can be reliable if the design of the experiments explicitly considers the assumptions of the model.

Clinical Relevance

Biological Consequences

Thermodynamic driving forces provide one avenue by which one can attempt to understand and ultimately perhaps manipulate nucleic acids *in vivo*. Although solution conditions in most *in vitro* studies differ significantly from *in vivo* conditions (water activity, high protein and nucleic acid concentrations, disproportionately high concentrations of numerous small effector molecules (ions, metabolites), increased viscosity etc.) some correlation between the thermodynamic properties of nucleic acids *in vitro* and *in vivo* exist. The incorporation of incorrect bases during DNA replication resulting in mispairing, be it opposite normal bases or opposite chemically damaged bases, is at least partially determined by the thermodynamic stability of such base mismatches within the polymerase active site. Recognition of DNA damage and subsequent repair by the cellular repair machinery may, in part, depend on the impact of the defect on thermodynamic stability at and near the defect site. Several debilitating genetic diseases, such as the triplet repeat expansion diseases, may be traced to misfolding of nucleic acids into competing alternative secondary structures during critical biological processes. The relative thermodynamic stability of the misfolded secondary structure in comparison to conventional ►Watson-Crick DNA appears to play a critical role in the development of such diseases. *In vitro* the specific association of nucleic acids with repair and regulatory proteins and/or small ligands/drugs critical for biological function differs from more general nonspecific association due to a delicate balance of specific and nonspecific interactions. It is likely that a similar balance of thermodynamic driving forces also controls the interactions of proteins and ligands with nucleic acids *in vivo*.

Practical Consequences

Insight into the thermodynamic properties of nucleic acids forms the basis of a number of critical applications in modern biotechnology, molecular biology and molecular medicine. In the following we highlight a few of the more important technological advances in which understanding nucleic acid thermodynamics plays a critical role. Common to all these techniques is the need to control interactions between nucleic acids with a high degree of fidelity, which is only possible through the rational application of the thermodynamic models discussed above.

Polymerase Chain Reaction

The ►polymerase chain reaction (PCR), a workhorse of modern molecular biology and molecular medicine, requires the careful design of primers specific for the sequence domains to be amplified and a careful choice of reaction conditions (hybridization temperatures, denaturation temperatures, primer concentration etc.) based on thermodynamic considerations. Failure of PCR experiments can frequently be traced to unrecognized thermodynamic features of the primer or DNA target, including unrecognized secondary primer binding sites, competing higher order secondary structures or misincorporation of bases during amplification.

Single Nucleotide Polymorphisms Detection

The importance of ►single nucleotide polymorphisms (SNPs) in determining individual susceptibility to various disease causing agents and pharmaceuticals is becoming increasingly more apparent. Numerous technologies are being developed to identify and screen for SNPs on a large scale with the hope of ultimately providing individualized medicine to patients. Almost all of these techniques require directly or indirectly (through PCR amplification of the relevant sequences) the application of thermodynamic data as part of the detection process. Of those methods that use thermodynamics as a detection tool, denaturing HPLC or denaturing PAGE methods are based on the ability to distinguish populations of DNA molecules based on their differences in thermal stability. Another class of assays relies on detection of fluorescence or fluorescence energy transfer of dyes attached to DNA probes. The assays are dependent on the specificity of the probe for the target with the added complication that the fluorescent dyes may contribute to the overall thermodynamics of strand association.

DNA Chip Technology

►DNA chip technology is based on the specific hybridization of DNA or RNA to short DNA probes immobilized at well-defined positions on the DNA chip. It is used with increasing frequency in genomics, proteomics and systems biology applications to understand systems-wide expression patterns of genes and also in the identification of disease causing agents, in forensics and potentially in sensor technologies. The optimal conditions for hybridization reactions and the potential for false positives through undetected mismatches or alternative secondary structure formation are determined by the thermodynamics of probe/DNA target interactions. Hybridization to targets immobilized on a surface is not described adequately by current thermodynamic models, which are based on data derived from and for reactions in free solution.

DNA Nanotechnology

The rational design and construction of novel and complex nucleic acid structures not necessarily found in biological systems holds promise for allowing the assembly of nanometer scale materials of increasing complexity and potentially unusual properties. The controlled, modular, stepwise construction of such structures relies for success on a precise understanding of the thermodynamics governing strand interactions.

References

1. Breslauer KJ, Frank R, Blocker H et al (1986) Predicting DNA duplex stability from the base sequence. *Proc Natl Acad Sci USA* 83:3746–3750
2. Goldstein RF, Benight AS (1992) How many numbers are required to specify sequence-dependent properties of polynucleotides? *Biopolymers* 32:1679–1693
3. Marky LA, Breslauer KJ (1987) Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers* 26:1601–1620
4. Owczarzy R, Vallone PM, Gallo FJ et al (1997) Predicting sequence-dependent melting stability of short duplex DNA oligomers. *Biopolymers* 44:217–239
5. Plum GE, Breslauer KJ, Roberts RW (1999) Thermodynamics and kinetics of nucleic acid association/dissociation and folding processes. In: Kool E (ed) *Comprehensive Natural Products Chemistry*, vol 7. Elsevier Science Ltd., Oxford, UK, pp 15–33
6. Wartell RM, Benight AS (1985) Thermal Denaturation of DNA Molecules: A Comparison of Theory with Experiment. *Phys Rep* 126:68–107

Thermodynamic State

Definition

Thermodynamic state comprises of the amounts of all components of a system as well as its physical variables, i.e. temperature-, pressure- and volume-dependences.

► [Thermodynamic Properties of DNA](#)

Thermogram

Definition

Thermogram defines the plot of apparent heat capacity versus temperature that results from a calorimetric experiment.

► [Thermodynamic Properties of DNA](#)

THF

Definition

THF stands for tetrahydrofolate. It is a carrier of one-carbon units, utilized in reactions 3 and 9 for purine biosynthesis and by thymidylate synthase.

► [Nucleotide Biosynthesis](#)

Thioester-Linkage

Definition

Thioester-linkage (or thioester bond) defines a very labile ester-bond between the COOH-group of a fatty acid and a -SH-group of a cysteine residue in the polypeptide chain (S-Acylation). An ester-type linkage can be experimentally distinguished from an amide bond by treatment of the acylated protein with hydroxylamine. This compound cleaves ester-linked fatty acids from the protein, whereas amide-linked fatty acids are not affected.

► [Fatty Acid Acylation of Proteins](#)

► [Sumoylation](#)

Thioether

Definition

Thioether is a functional group within an organic molecule consisting of a sulfur atom linked to two different carbon atoms.

Thioflavin T

Definition

Thioflavin is a fluorescent dye that can cross the blood-brain barrier and selectively labels amyloid deposits in the Alzheimer diseased brain

► [In Vivo Imaging of Transgenic Mice with Fluorescent Protein Expression](#)

Thiolate

A hydrosulfide attached to carbon.

► [Proteases and Inhibitors](#)

Thiol-Dependent Peroxidases

LEOPOLD FLOHÉ, TIMO JAEGER
MOLISA GmbH, Magdeburg, Germany
l.flohe@t-online.de

Definition

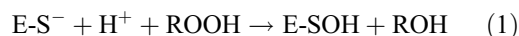
Thiol-dependent peroxidases are defined as enzymes that reduce H_2O_2 , alkyl hydroperoxides and peroxytrite at the expense of thiols with formation of disulfides.

Characteristics

General Features

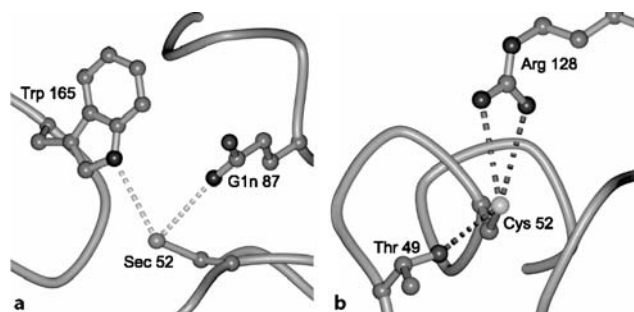
Two families of peroxidases that use thiols as donor substrates to reduce H_2O_2 and other hydroperoxides are widespread in nature: the ► [glutathione peroxidases](#) (GPx) and the peroxiredoxins (Prx). The common denominator of these enzyme families is that they do not contain any prosthetic group such as heme but use a highly activated, i.e., dissociated cysteine or selenocysteine residue for catalysis. During catalysis their active-site thiol or selenol moiety is oxidized by a hydroperoxide to a sulfenic or selenenic acid derivative, respectively; the latter reacts with a thiol to form a (seleno)disulfide, which reacts with a second thiol to

regenerate the ground state of the enzyme, as outlined below:



Apart from being unrelated in sequence, the two enzyme families differ in their strategy to force the active site SH or SeH group into dissociation. In the glutathione peroxidases, typically the sulfur (selenium) of a (seleno)cysteine residue is hydrogen-bridged to the imino nitrogen of a tryptophan and the amido nitrogen of a glutamine residue (Fig. 1a) (1); in the ► [peroxiredoxins](#), dissociation of the thiol is induced by the positive charge of a nearby arginine and by hydrogen-bonding to a threonine or, less often, to a serine residue (Fig. 1b) (2). Due to the intrinsically low pK of selenocysteine the selenium-containing peroxidases are highly efficient with k_1 values above $10^7 \text{ M}^{-1}\text{s}^{-1}$ (Eq. 1). The corresponding k_1 values in peroxiredoxins, which almost exclusively work with thiol catalysis, are one to three orders of magnitude smaller and overlap with those of the sulfur homologues of the selenium-containing GPx-type enzymes.

For the reductive part of the catalytic cycle (Eqs. 2, 3) evolution has created a diversity that makes it difficult to correlate substrate specificities with molecular clades. Typical Prx substrates are the vicinal thiols of CXXC motifs, as they are present in ► [thioredoxins](#), ► [tryparedoxins](#) or the AhpFs of bacterial alkyl hydroperoxide reductase systems, but the mono-thiol glutathione has also been identified as substrate of some Prxs (2). Inversely, the terminology glutathione peroxidase turned out to be misleading for some members of the GPx family. The first member of this family to be recovered, now called GPx-1 in mammals,



Thiol-Dependent Peroxidases. Figure 1 Catalytic triads of thiol-dependent peroxidases. (a) Catalytic triad of GPx-type peroxidases. Position numbers refer to those of bovine GPx-1. (b) Catalytic triad of peroxiredoxins. Position numbers refer to TXNPx of *Leishmania donovani*. The cysteine (Cys173) that complements the intersubunit active site in typical two-Cys-peroxiredoxins is not shown.

is indeed highly specific for glutathione, but GPx-3 equally accepts thioredoxin and glutaredoxin as substrates, GPx-4 can react with protein thiols including those of GPx-4 itself, the GPx of *Plasmodium falciparum* proved in functional terms to be a thioredoxin peroxidase and, in analogy, GPx-type proteins in *Trypanosoma* species preferred trypanredoxin over glutathione as substrate (3).

Glutathione Peroxidases (1)

Mammalian genomes disclosed the existence of 6 distinct GPx-type proteins with evidently diverse biological functions. Four of them, GPx-1, GPx-2, GPx-3 and GPx-4 are selenoproteins, GPx-5 is a cysteine homologue, GPx-6, depending on species, either contains cysteine or selenocysteine.

GPx-1 is the most abundant one. It was the first selenoprotein that was discovered in mammals. It complements catalase in the detoxification of H_2O_2 and other soluble hydroperoxides and, according to knockout studies, is a pivotal cellular device to cope with oxidative stress. Unchallenged *gpx-1* (-/-) mice do not display any phenotype, but are highly sensitive to exposure to redox cyclers or conditions triggering H_2O_2 formation such as LPS administration or viral infections. If exposed to avirulent coxsacki strains, they develop a cardiopathy that reminds one of Keshan disease, a cardiomyopathy that is endemic in certain areas of China and is attributed to selenium deficiency. GPx-1 is one of the selenoproteins that declines fast upon selenium restriction. It is therefore justified to assume an impairment of hydroperoxide detoxification as an early event in selenium deficiency.

GPx-2 is largely restricted to the gastro-intestinal epithelium. Again *gpx-2* (-/-) mice do not show any obvious phenotype. Even double knockout mice devoid of GPx-1 and GPx-2 develop normally as long as they are raised under germ-free conditions. Upon contact with normal gastro-intestinal flora however, they develop inflammatory bowel disease. This observation points to a role of GPx-2 in regulating the synthesis of inflammatory mediators, possibly of leukotrienes. Alternatively, an enhanced apoptosis at the tip of the mucosal villi, where GPx-2 is lowest, may become unmasked if the general defense against oxidative challenge is impaired due to lack of GPx-1. GPx-3 is an extracellular protein of so far unknown function. Present in an environment with vanishingly low substrate levels (GSH or thioredoxin), it can hardly be considered as an enzyme in charge of balancing oxidative stress. It rather might be a redox sensor that monitors extracellular hydroperoxide production, e.g. resulting from activation of phagocytes, and may modulate the response of receptors at the cell surface. GPx-4 is unique in reducing hydroperoxy groups of lipids in biomembranes and has for long been

implicated in the prevention of lipid peroxidation. More recently, specific roles became evident (see below). Knockout of *gpx-4* resulted in early embryonic lethality, a phenomenon that is not easily explained by any of the known GPx-4 functions. Depending on the use of distinct promoters and translation starts, *gpx-4* is translated into a cytosolic protein or isoforms equipped with a mitochondrial or nuclear leader sequence. It remains to be worked out which of the different forms of GPx-4 account for its vital importance.

GPx-5 is an epididymal protein that is built under androgen control. But its role remains obscure, as does that of the recently detected GPx-6.

Non-vertebrate GPx-type proteins are usually cysteine homologs with poor peroxidase activities and poorly defined biological roles. Exceptions are the GPxs of the fowl poxvirus and of *Schistosoma mansoni*, which contain selenocysteine. The latter appears to be involved in reproduction, since it is only found in the female gender and is restricted to the vitellin glands of the female reproductive tract.

Peroxiredoxins (2)

The first peroxiredoxin to be discovered was the “thiol-dependent antioxidant protein” of yeast in the laboratory of Earl Stadtman. It later became characterized as a peroxidase using the CGPC motif of thioredoxin as reductant. In the yeast peroxidase system three proteins are required to transfer the reduction equivalents from NADPH to the hydroperoxide, thioredoxin reductase, thioredoxin and the thioredoxin peroxidase. Similar cascades of oxidoreductases were later found to operate in mammals also. In parallel, the alkyl hydroperoxide reductase component AhpC of enterobacteria was identified as peroxiredoxin. Together with a disulfide reductase, AhpF, it constitutes a simple peroxidase system that reduces hydroperoxides at the expense of NADPH. The reduction equivalents are transferred from NADPH *via* FAD to a CXXC motif of AhpF, the latter serving as substrate for AhpC. The most complicated redox cascade fuelling a peroxiredoxin was identified in trypanosomatids. Here a flavin-containing disulfide reductase, trypanothione reductase, reduces oxidized trypanothione, that is the bis-glutathionyl derivative of spermidine; the reduced trypanothione reduces a thioredoxin-related protein called trypanredoxin; the reaction center of trypanredoxin is a CPPC motif and thus an ideal substrate to keep a peroxiredoxin-type trypanredoxin peroxidase in the reduced state ready to scavenge any hydroperoxide (3). In the meantime peroxiredoxins have been identified in all domains of life. They all have the potential to reduce hydroperoxides at the expense of thiols, but their substrate specificity, in particular for the donor substrate differs substantially. The family comprises at least six molecular clades clearly

separated by sequence and, in part, by catalytic mechanism and donor substrate specificity (4). So far the oxidizable cysteine of the active site was found to be replaced by selenocysteine in one case only (*Eubacterium acetamidophilum*).

The examples mentioned above belong to the huge subfamily of typical two-cysteine-peroxiredoxins (2-Cys-Prx). Their oxidizable cysteine is located about 50 residues from the N-terminus and is, as in most Prxs, embedded in a strictly conserved VCP motif. This subfamily contains a second strictly conserved cysteine near the C-terminus that mostly is also flanked by V and P. This second cysteine serves as the first reductant of the oxidized C-terminal one during the catalytic cycle according to Eq. 2. The disulfide bridge between the two cysteine residues can, however, only be formed if two invertedly orientated subunits of the homooligomeric proteins interact, which is only possible by substantial redox-dependent conformational changes. As a result of those, the exposed proximal cysteine that constitutes the peroxidatic catalytic triad shown in Fig. 1b becomes hidden upon oxidation, whereas the distal conserved cysteine is surface-exposed to react with another thiol according to Eq. 3. Thus, the catalytic triad (Fig. 1b), which is common to all peroxiredoxins, is only one part of the catalytic site. To complete the catalytic cycle a novel intersubunit reaction center is formed that allows specific interaction with the donor substrate. In the 2-Cys-Prx subfamily, this is typically the most surface exposed cysteine of the CXXC motif of thioredoxin, tryparedoxin or AhpF.

A variation of this theme is provided by the so-called atypical 2-Cys-Prxs, e.g. the "thiol peroxidases (TPx)" of *Escherichia coli* and *Mycobacterium tuberculosis*. Here the oxidized N-proximal cysteine has been proposed to react with a cysteine of the same subunit. Again the final reductants are thioredoxins (4). This subfamily therefore is functionally monomeric, while a dimer is the minimum catalytic unit of the typical 2-Cys-Prxs. These minimum catalytic requirements do not mimic the actual quaternary structure. The atypical 2-Cys-Prx of *M. tuberculosis* is dimeric and the typical 2-Cys-Prxs tend to form rings composed of 5 dimers (Fig. 2a, b), which had been seen in electron microscopy long before the identification of peroxiredoxins as peroxidases and were then termed "porins" (2, 5). In a minority of the peroxiredoxins only the N-proximal cysteine is conserved. In these 1-Cys-Prxs, the entire catalysis (Eqs. 1, 2, 3) must be achieved by the catalytic triad (Fig. 1b) without the aid of a co-reaction cysteine residue. In this respect, a 1-Cys-Prx acts like a GPx-type peroxidase and, interestingly, the few examples of this subfamily that have been characterized so far proved to share the substrate specificity of the classical glutathione peroxidase, GPx-1, they accepted the mono-thiol GSH and not

thioredoxin. It appears premature, however, to deduce any reliable rules as to the correlation of molecular clades and substrate specificity, since out of the more than 400 Prxs known by sequence less than two dozen have been thoroughly investigated for their metabolic context.

Molecular Interactions

Apart from GPx-1, thiol-peroxidases tend to react with protein SH groups. As outlined above, protein SH groups are the preferred substrate of most peroxiredoxins. In some cases such interactions are, however, not part of the catalytic cycles. Typical examples are the reaction of GPx-4 with a surface-exposed SH group of GPx-4 itself that results in polymerization, a phenomenon that contributes to the transformation of the peroxidase into an enzymatically inactive structural protein during sperm maturation. The insoluble material forming the mitochondrial capsule is characterized by Se-S bridges between GPx-4 molecules. Probably analogous cross-links with cysteine-rich proteins contribute to the stability of the capsule. The nuclear variant of GPx-4 has been implicated in chromatin condensation, which is also presumed to involve formation of Se-S bonds (1).

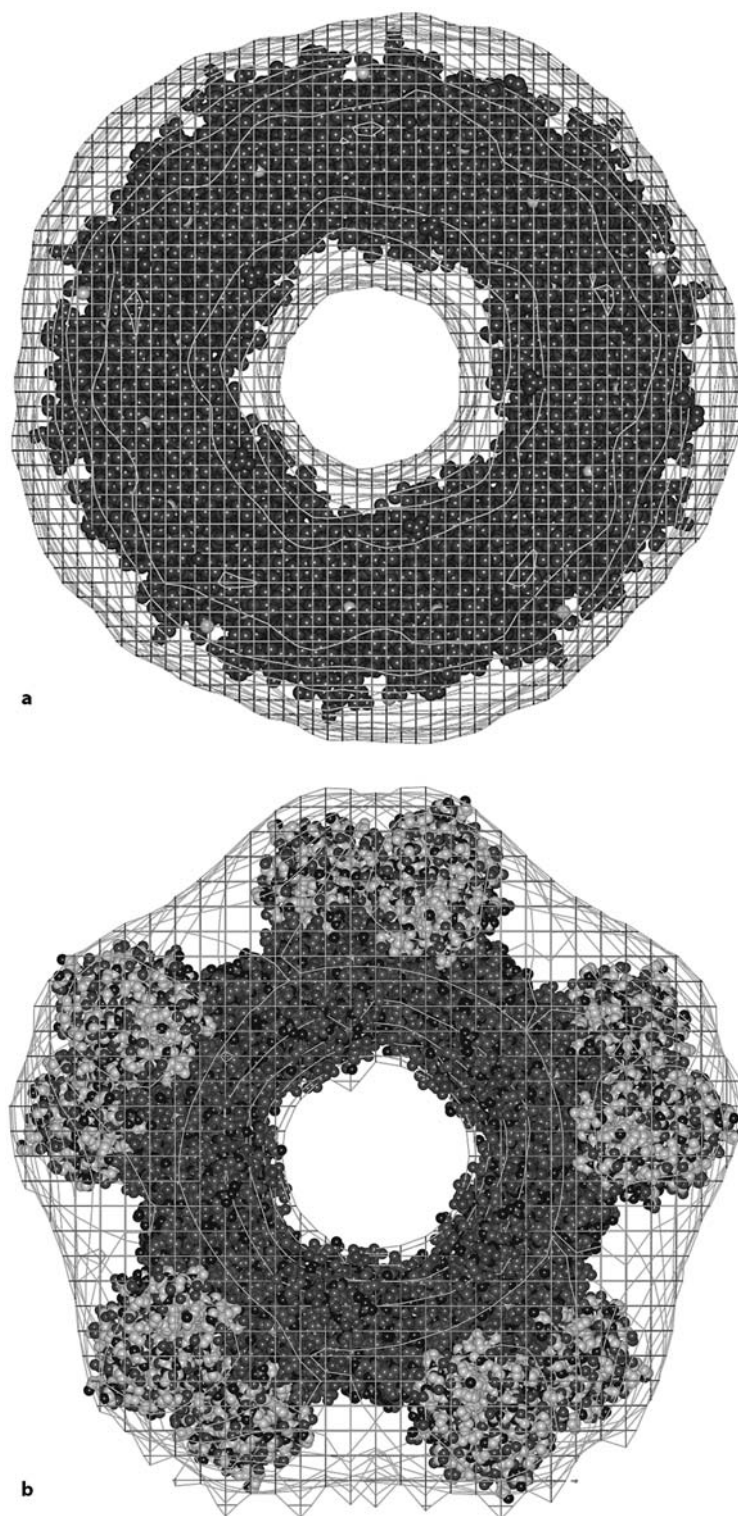
Human PrxII, III and VI were reported to interact specifically with cyclophilin A. In the case of PrxII, reduction of a cyclophilin disulfide bond was verified (2).

Regulatory Mechanisms

General Considerations

Peroxidases are commonly called antioxidant enzymes, which, however, is only true in exceptional cases. Their ability to reduce H_2O_2 , alkyl hydroperoxides or peroxynitrite does not necessarily imply that their biological role consists in balancing oxidative stress. Of the long known heme peroxidases, catalase appears to be in charge of detoxifying H_2O_2 generated in peroxisomal metabolism. In contrast, myeloperoxidase uses H_2O_2 to produce hypochlorite, which is a stronger oxidant, and the thyreoperoxidases require H_2O_2 for thyronine production. The thiol-dependent peroxidases have for long been considered to be the more reliable cellular antioxidant devices since they reduce ROOH in a clean stoichiometric reaction without the aid of transition metal cofactors that are prone to produce free radicals by univalent redox transitions.

The two-electron transitions consistently catalyzed by the thiol-dependent peroxidases have not, however, precluded these enzymes from also adopting functions that can not possibly be classified as antioxidant defense. The most striking example is the use of oxidants by GPx-4 to built up a structural protein that is essential for the fertilization capacity of sperm. Also, there is increasing evidence that oxidants like H_2O_2 and alkyl hydroperoxides are mediators in unstressed



Thiol-Dependent Peroxidases. Figure 2 Shape of a typical 2-Cys-Prx. (a) shows a model of the decameric oxidized trypanothione peroxidase (TXNPx) from *Trypanosoma brucei* as supported by averaged electron microscopy images (net). (b) the molecule has been reacted with ten mutated trypanothione molecules (taken from Budde *et al.* (2003) *Biol Chem* 384:619-633).

physiology. These oxidants, however, would react promiscuously if they were not sensed by “peroxidases” that, in a redox-dependent manner, specifically interact with target molecules, to modulate their activity. Numerous observations (1, 2, 5), which are beyond the scope of this article, support the idea that the coexistence of many peroxidases from at least three families will find its explanation in the emerging field of redox-regulation of proliferation, apoptosis, and differentiation. The phosphorylation/dephosphorylation and/or protease cascades dominating these phenomena turned often out to be modulated by the redox-state of their microenvironment. The NF κ B system may be quoted as a particularly well-documented example (1, 5). For the sake of clarity it appears mandatory to differentiate between such physiological redox-processes and “oxidative stress”, a term describing a pathological situation where the production of, or exposure to, oxidants exceeds the reductive capacity of the organism. This may, e.g., result from poisoning with redox-cyclers, exposure to hyperbaric oxygen or, most commonly, from the respiratory burst triggered in phagocytes activated by pathogenic microorganisms, anaphylatoxins or the realm of proinflammatory cytokines. To balance such an oxidative burst a peroxidase has to meet several requirements. It must be ubiquitous, abundant, highly efficient and supplied with a reasonable reduction capacity. For the mammalian body, the GSH/GPx-1 system meets these criteria, while other GPx-type enzymes are restricted to special tissues and less abundant. The Prx-type enzymes are generally less efficient and, in mammals, can hardly compete with selenium- or heme-containing peroxidases and thus are more likely regulatory enzymes. In many microorganisms, which often lack selenium-containing peroxidases or catalase, the comparatively low molar efficiency of a Prx can, however, be compensated by high enzyme concentrations. There, Prxs represent the major, if not the only enzyme to cope with peroxide stress.

Regulation of Enzyme Levels

GPx-1 in mammals is reportedly up-regulated under oxidative stress presumably *via* an oxygen-responsive element. GPx-2 is transcriptionally regulated by retinoids. At the translational level, all selenium-containing GPxs are regulated by the selenium content, although in a specific way. GPx-1 and 3 decline fast upon selenium restriction, GPx-2 and 4 are more stable. The differential response to selenium is known to be due to a differential stability of the mRNAs in selenium deficiency (1). The mechanisms leading to the pronounced tissue specificity of the selenoproteins remain to be elucidated. The abundant hormone-dependent expression of GPx-4 in testis appears to be

caused not by transcriptional activation by any of the androgenic hormones but by a testosterone-dependent augmentation of spermatids. Transcriptional or translational regulation of Prx-type enzymes has not yet been studied in detail (5).

Specific Regulatory Effects of GPx-4

Many of these peroxidases have been shown to dampen signaling *via* redox-sensitive phosphorylation cascades, probably by lowering the peroxide tone. The regulatory efficiencies, however, appear to differ between these enzymes. Moderately over-expressed GPx-4, for instance, can abrogate IL-1-induced NF κ B activation, while huge variations in GPx-1 levels have a minor effect. Similarly, GPx-4 appears to be more efficient in inhibiting 5-lipoxygenase than any of its congeners. A specific role for the nuclear variant of GPx-4 in chromatin condensation is also being discussed (1). The differential effects may result from distinct preferences for hydroperoxide or thiol substrates. A regulatory event could, however, also be achieved by a direct interaction of the peroxidase with a target protein.

References

1. Flohé L, Brigelius-Flohé R (2001) Selenoproteins of the glutathione system. In: Hatfield D L (ed) *Selenium. Its molecular biology and role in human health*. Kluwer Academic Publishers, Boston/Dordrecht/London, p 157–178
2. Hofmann B, Hecht H-J, Flohé L (2002) Peroxiredoxins. *Biol Chem* 383:347–364
3. Krauth-Siegel LR, Meiering SK, Schmidt H (2003) The parasite-specific trypanothione metabolism of Trypanosoma and Leishmania. *Biol Chem* 384:539–549
4. Jaeger T, Budde H, Flohé L et al (2004) Multiple thioredoxin-mediated routes to detoxify hydroperoxides in Mycobacterium tuberculosis. *Arch Biochem Biophys* 423:182–191
5. Wood ZA, Poole LB, Karplus PA (2003) Peroxiredoxin evolution and the regulation of hydrogen peroxide signaling. *Science* 300:650–653

Thioredoxin

Definition

A thioredoxin (Trx) is a redox-active protein typically characterized by a CGPC motif, which reversibly shuttles between the dithiol and disulfide state. Thioredoxins are involved in basic anabolic functions such as ribonucleotide reduction for DNA synthesis, they contribute to antioxidant defense as substrates of

peroxiredoxins, and are implicated in many redox-sensitive regulatory processes.

- ▶ Free Radicals
- ▶ Thiol-Dependent Peroxidases

Third-Generation MDR Modulators

Definition

Third-generation MDR (multi drug resistance) modulators are molecules which prevent the drug transport out of a multi drug resistant cell by noncompetitive binding, and thereby specifically inhibiting a single ABC transporter. Intracellular drug concentration increases and thus chemosensitivity of the cell is restored. To date these modulators are used as a reversal strategy for cells and tumors possessing the multi drug resistance phenotype.

- ▶ Multi-Drug Resistance

Threshold Liability Model of Multifactorial Inheritance

Definition

The model proposes a continuously distributed genetically determined liability to the development of disease. Individuals who develop disease will have several disease susceptibility genes. Due to the properties of the normal distribution of genes, first-degree relatives will have a much higher risk of developing disease than the general population. Second-degree relatives will have a moderate risk, and third-degree relatives will have low risk. Many of the systemic and organ-specific autoimmune diseases appear to be inherited in a pattern consistent with this model.

- ▶ Autoimmune Disease
- ▶ Autoimmunity

Thromboembolism

Definition

Thromboembolism is a process in which a blood clot forms on an organ surface, such as the heart, and travels

by means of blood circulation to another organ (e.g. lungs, brain).

- ▶ Marfan Syndrome

Thrombosis

Definition

Thrombosis is a pathological intravascular generation of blood clots. Predisposition to thrombosis is conferred by exogenic factors or by increased levels of procoagulatory factors, by decreased levels of anticoagulatory factors or by impaired fibrinolysis.

- ▶ Hereditary Hemostatic Defects and Recombinant Proteins for Treatment
- ▶ Polyadenylation

Thymidine Analog Mutations

Definition

Thymidine analog mutations (TAMs) comprise of a group of six mutations in human immunodeficiency virus (HIV)-1 ▶ reverse transcriptase (RT), i.e. M41L, D67N, K70R, L210W, T215Y/F and K219Q, which confer resistance to the antiretroviral drugs zidovudine (ZDV) and stavudine (d4T). These mutations, when present with other RT mutations, are also associated with diminished antiviral susceptibility to several other nucleoside reverse transcriptase inhibitors.

- ▶ Reverse Transcriptase

Thymidine Kinase 2

- ▶ TK2

Thymidine Phosphorylase

Definition

Thymidine phosphorylase (TP) is a cytosolic enzyme that is required to maintain nucleoside homeostasis.

- ▶ Mitochondrial Myopathies

Thyroid Cancer

► Thyroid Disorders, Genetic Basis

Thyroid Disorders, Genetic Basis

PETER KOPP

Division of Endocrinology, Metabolism & Molecular Medicine, Feinberg School of Medicine, Northwestern University, Chicago, USA
p-kopp@northwestern.edu

Synonyms

Thyroid dysgenesis, dyshormogenesis, hypothyroidism, hyperthyroidism, hyperthyroxinemia, resistance to thyroid hormone, autoimmune thyroid disease, thyroid cancer.

Definition and Characteristics

The thyroid gland evolves from two distinct embryologic structures. The thyroid anlage derives from the endodermal floor of the primitive pharynx and gives rise to thyroid follicular cells. These cells migrate caudally, then expand laterally and form the two lobes of the gland. The ultimobranchial bodies, located in the fourth pharyngeal pouch, are of neural crest origin and, after undergoing a similar descent, differentiate into the parafollicular calcitonin-producing C cells. Following migration and fusion of the two cell populations, thyroid follicular cells undergo further differentiation. This is marked by the expression of genes that are essential for thyroid hormone synthesis such as those for TSH (thyrotropin) and its receptor (TSHR), the sodium-iodide symporter (NIS), thyroperoxidase (TPO) and thyroglobulin (TG). In humans, thyroid hormone is detectable in the fetus at about gestational week 11.

The thyroid is controlled by a classic hypothalamic-pituitary axis. The hypothalamic tripeptide TRH (thyrotropin-releasing hormone) stimulates the production and secretion of the glycoprotein hormone TSH (thyroid-stimulating hormone) in the pituitary. TSH in turn stimulates growth and function of thyroid follicular cells resulting in the production of the thyroid hormones thyroxine (T4) and triiodothyronine (T3). In the bloodstream, only minute fractions of T4 and T3

are found as free hormone, most is protein-bound to thyroxine-binding globulin (TBG), transthyretin (TTR) and albumin.

After uptake into peripheral cells, a process that is at least partially mediated by amino acid channels, T4 is either 5'-deiodinated by deiodinase I or II, a process that produces the more active T3, or it is modified into the inactive rT3 by deiodination of the inner ring. At the cellular level, thyroid hormones have multiple effects on differentiation, growth and metabolism of peripheral tissues. Thyroid hormone action is primarily mediated by nuclear receptors regulating gene transcription; several non-genomic actions are a focus of growing interest. Negative regulation of the hypothalamic-pituitary-thyroid axis is exerted by T3, somatostatin and dopamine.

Cellular and Molecular Regulation and Clinical Relevance

Thyroid disorders are common. In adults, the prevalence of overt hypothyroidism is ~3–20 per 1000 in females, and ~1–7 per 1000 in males. The prevalence of hyperthyroidism is ~2–19 per 1000 in females, and ~1–2 per 1000 in males.

Genetic Defects in Thyroid Development and Hormone Synthesis

Congenital Hypothyroidism (CH)

CH affects about 1:3,000–1:4,000 infants. It is usually sporadic and only rarely familial. Screening programs now permit early recognition and treatment, thus avoiding the disastrous consequences of thyroid hormone deficiency on brain development. In about 85%, CH is sporadic and associated with developmental defects referred to as thyroid dysgenesis. They include thyroid (hemi)agenesis, ectopic tissue and thyroid hypoplasia. Dyshormogenesis, defects in normal hormone synthesis, account for about 10–15% of CH. A wide spectrum of genetic defects provides a molecular explanation for a subset of the sporadic and familial defects in pituitary and thyroid development, or thyrotropin and thyroid hormone synthesis.

Mutations in Thyroid Transcription Factors and Thyroid Dysgenesis

Mutations in transcription factors that govern thyroid development and gene expression may result in syndromic and non-syndromic forms of thyroid dysgenesis. These observations are corroborated by similar observations in murine knockout models.

PAX8

Monoallelic mutations in PAX8, a paired domain transcription factor involved in thyroid development and expression of the *TPO* and *TG* genes, have been

documented and characterized in sporadic and familial patients with thyroid hypoplasia or ectopy. It is currently unclear why mutation of a single PAX8 allele is sufficient to result in CH in humans, a finding that contrasts with the observation that mice heterozygous for a disrupted *Pax8* locus do not display a pathological phenotype.

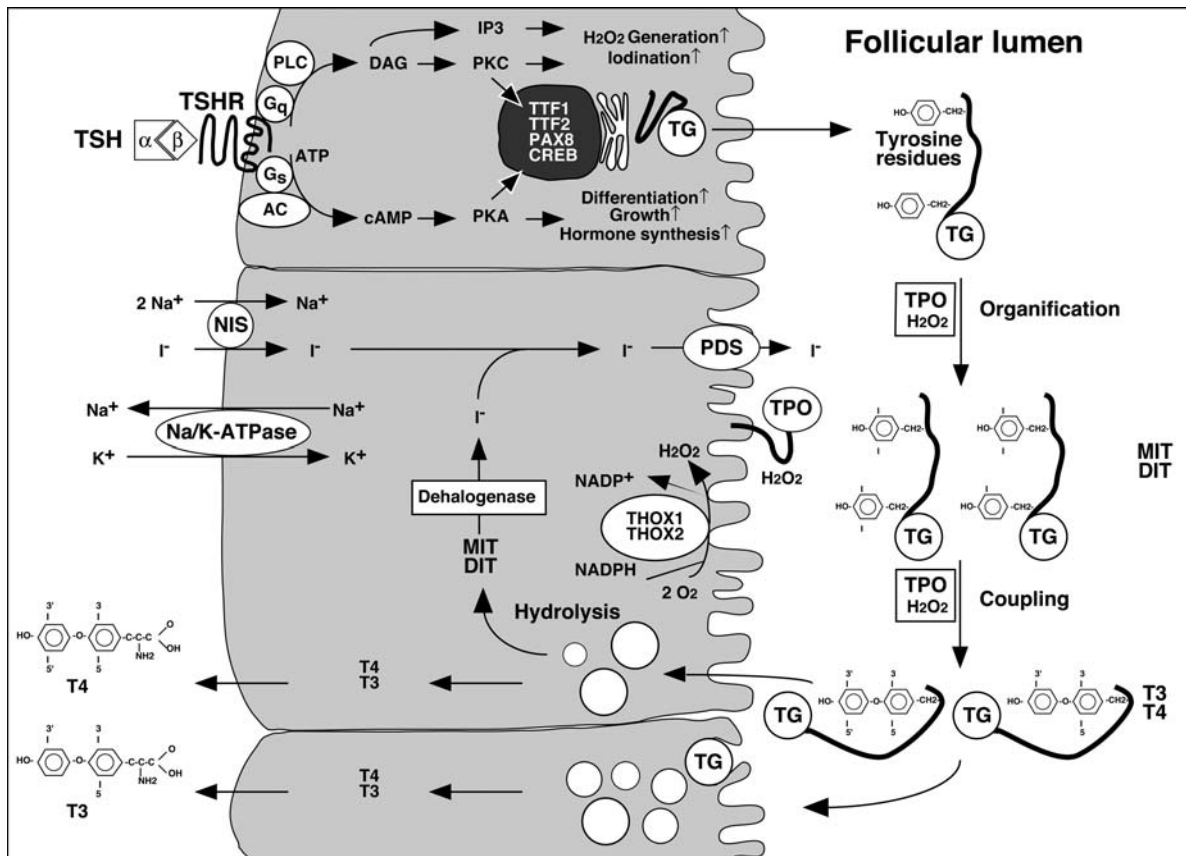
TTF1/NKX2.1

Thyroid transcription factor 1 (TTF1) is a homeobox domain transcription factor of the NKX2 family involved in the development of the gland and in transcriptional control of the *TG*, *TPO* and *TSHR* genes. It is also expressed in the lung, the forebrain and the pituitary gland. Patients heterozygous for chromosomal deletions of 1q12-13.2 or point mutations in the

TTF1 gene present with mild CH with a thyroid of normal size and location. They have neurological alterations that include ataxia or choreoathetosis, truncal apraxia and mental retardation, in combination with neonatal respiratory distress. Monoallelic *TTF1* mutations have also been identified as the molecular cause of hereditary chorea.

TTF2/FOXE1

Homozygosity for recessive mutations in the forkhead/winged-helix domain transcription factor FOXE1, commonly referred to as thyroid transcription factor 2 (TTF2), result in a syndromic form of thyroid dysgenesis with the eponym Bamforth-Lazarus syndrome. This phenotype includes thyroid agenesis, cleft palate, choanal atresia, bifid epiglottis and spiky hair.



Thyroid Disorders, Genetic Basis. Figure 1 Model of iodide transport in thyroid follicular cells. Iodide is actively transported into thyroid follicular cells by the sodium-iodide symporter (NIS) at the basolateral membrane. At the apical membrane, pendrin (PDS/SCL26A4) mediates iodide efflux into the follicular lumen. Thyroperoxidase (TPO) oxidizes iodide and subsequently iodinated tyrosyl residues of thyroglobulin (TG) in the presence of hydrogen peroxide (H₂O₂) (organification). The iodotyrosines, mono- and di-iodotyrosine (MIT, DIT) are coupled to T₄ or T₃, a reaction that is also catalyzed by TPO (coupling). Thyroglobulin is internalized into the follicular cell, hydrolyzed in lysosomes and the thyronines T₄ and T₃ are released into the blood stream.

Dyshormogenesis

Sodium/Iodide Symporter

Normal iodide uptake at the basolateral membrane by the perchlorate-sensitive sodium/iodide symporter (NIS) is a rate-limiting step in thyroid hormone synthesis. Several homozygous or compound heterozygous mutations have been identified in individuals with hypothyroidism associated with impaired iodide uptake. Many of these patients have a diffuse or nodular goiter, little or no uptake of radioiodine and a decreased saliva/serum radioiodine ratio.

PDS/SCL26A4

Efflux of iodide at the apical membrane of thyroid follicular cells is at least in part mediated by pendrin (SCL26A4), a member of the solute carrier family 26A. Mutations in the *PDS/SCL26A4* gene cause Pendred's syndrome, an autosomal recessive disorder traditionally defined by the triad of sensorineural congenital deafness, goiter and a partially positive perchlorate test. The partial discharge of radioiodine after the administration of perchlorate indicates that the gland has an impaired ability to organify iodide. Although some patients with Pendred's syndrome present with CH, the majority of individuals are clinically and biochemically euthyroid.

Thyroperoxidase

Thyroperoxidase, a glycosylated hemoprotein located at the apical membrane facing the follicular lumen, iodinates tyrosine residues in TG and facilitates the coupling of the iodinated tyrosines to generate T4 and T3. TPO defects are among the most frequent causes of inborn errors of thyroid hormone synthesis. Homozygosity or compound heterozygosity for mutations in the *TPO* gene have been reported in numerous families with a partial or total organification defect.

THOX2

The iodination and coupling reactions are dependent on H_2O_2 as an essential cofactor. Recently, two NADPH oxidases that are part of the H_2O_2 -generating system, THOX1 and THOX2, have been cloned. Heterozygous loss of function mutations in the *THOX2* gene result in mild transient CH. Biallelic *THOX2* mutations are associated with a severe phenotype and confirm that H_2O_2 is essential for iodide organification. As yet, there are no reported mutations in *THOX1*.

Thyroglobulin

Thyroglobulin (TG), a homodimeric glycoprotein, is a key element in thyroid hormone synthesis and storage. It is encoded by a very large gene spanning more than 300 kb and containing 48 exons. Recessive mutations in the TG gene have been reported in a number of

animal models and human patients. Unless treated with levothyroxine, these patients typically present with goiter in early childhood. The metabolic status is variable and, depending on the severity of the defect, the patients are hypothyroid, subclinically hypothyroid or euthyroid. The serum TG levels may be low, normal or elevated. The radioiodine uptake is elevated. In many instances, the mutated TG protein is retained in the endoplasmic reticulum resulting in a classical endoplasmic reticulum storage disease (ERSD).

Dehalogenase

After entering the follicular cell, TG is hydrolyzed and T4 and T3 are secreted into the blood at the basolateral membrane. The iodotyrosines MIT and DIT, which are much more abundant in the TG molecule, are deiodinated by an intrathyroidal dehalogenase and recycled for hormone synthesis. Several patients with leakage of MIT and DIT from the thyroid and urinary secretion of these metabolites have been identified. The disorder is recessive. Very recently, an intrathyroidal dehalogenase, which may deiodinate MIT and DIT in the thyroid, has been cloned.

Resistance to TSH

The response to bioactive TSH may be impaired at the level of the thyroid follicular cells. Total insensitivity to TSH results in a small hypoplastic thyroid gland and reduced synthesis and secretion of thyroid hormones. It is noteworthy that a similar morphologic and biochemical phenotype may occur in patients with mutations in *PAX8*, which emphasizes the limitations of morphological criteria. Resistance to TSH may be caused by various molecular mechanisms. In a subset of these patients, the molecular cause consists of recessive mutations in the *TSHR* that are partially or completely inactivating. In partial resistance, TSH is elevated, but the peripheral hormone levels are normal, a constellation referred to as *euthyroid hyperthyrotropinemia*. In these patients, the size of the thyroid is normal or enlarged. More severe homozygous or compound heterozygous inactivating mutations in the *TSHR* have been found in several patients with overt hypothyroidism and thyroid hypoplasia.

Resistance to TSH, in combination with resistance to PTH, LH, FSH and the morphological features of Albright's hereditary osteodystrophy (short stature, brachydactyly, ectopic calcifications) also occurs in pseudohypoparathyroidism Ia (PHP Ia). The molecular basis consists of inactivating mutations in the maternal copy of the *GNAS1* (*Gsa subunit*) gene, which is imprinted in a tissue-specific manner.

Unresponsiveness to TSH can also be inherited as an autosomal dominant trait, but the molecular defect remains to be defined.

Activating Mutations in the TSH receptor (TSHR)

TSH exerts its effects on thyroid follicular cells through the TSHR, a member of the G-protein coupled seven transmembrane receptors. Together with the receptors for FSH and LH it forms a distinct subfamily defined by a large amino terminal extracellular domain involved in binding the hormone. The 744 amino acid receptor is encoded by a gene containing 10 exons and localized on chromosome 14q31. The TSH receptor is primarily coupled to Gs and thus to the adenylyl cyclase cascade, which is the predominant signaling pathway for growth and function in the thyrocyte. Gain of function mutations in the TSHR lead to various characteristic pathophysiological entities.

Somatic TSHR Mutations in Toxic Adenomas and Thyroid Carcinomas

Somatic mutations in the TSHR have been found as the main molecular cause of toxic adenomas. Functionally, these mutations increase basal cAMP levels; some of the mutants also activate the phospholipase C cascade. In contrast to activating mutations in other seven transmembrane receptors, there is a striking diversity in the affected residues that are scattered over the entire transmembrane domain of the TSHR and even occur in the extracellular domain. Activating mutations in the TSHR have also been found in a small number of thyroid carcinomas with the unusual finding of increased hormone secretion. Remarkably, the mutations found in these tumors not only activate the cAMP cascade, but also the IP3 pathway. This supports the notion that concomitant activation of these two signaling cascades may promote transformation.

Activating Germline TSHR Receptor Mutations in Familial Non-autoimmune Hyperthyroidism

Activating mutations occurring in the germline give rise to familial non-autoimmune hyperthyroidism. The typical signs associated with autoimmune hyperthyroidism, i.e. Graves' disease, are absent. Several families with non-autoimmune familial autosomal dominant hyperthyroidism due to TSHR mutation have been identified.

Activating Germline TSH Receptor Mutations in Neonatal Non-autoimmune Hyperthyroidism

Congenital hyperthyroidism is usually caused by transplacental transfer of maternal antibodies in offspring of a mother with autoimmune thyroid disease. In this instance, the disease is transient and resolves within several weeks to months upon clearance of the antibodies. However, a few cases show a persistent course with severe hyperthyroidism. *De novo* germline mutations in the *TSHR* gene have been found as a cause of this unusual form of hyperthyroidism.

TSH Receptor Mutations Conferring Hypersensitivity to hCG

A form of familial gestational hyperthyroidism caused by a mutant thyrotropin receptor displaying hypersensitivity to normal levels of hCG has been identified. The mutated receptor has an increased sensitivity for hCG and thus results in activation of thyroidal hormone production during pregnancy. In addition to hyperthyroidism, the affected women presented with repeated miscarriages and hyperemesis.

Activating Mutations in the *GNAS1* (*Gsa*) Gene

Somatic gain of function mutations in the *GNAS1* gene encoding the α subunit of the stimulatory G protein (gsp mutations) lead to a constitutive activation of the cAMP pathway and to toxic adenomas. They can occur only in the thyroid or display mosaicism in the McCune-Albright syndrome (precocious puberty, polyostotic fibrous dysplasia, café au lait skin pigmentation, GH-secreting pituitary adenomas and toxic adenomas).

Binding Protein Abnormalities

Thyroid hormones circulate bound to plasma proteins, the three major ones being thyroxine-binding globulin (TBG), transthyretin (TTR) and albumin. Abnormalities in transport proteins lead to a decrease or increase in total T4 levels; free hormone levels are within the normal range and the patients are clinically euthyroid. Failure to recognize these entities results in inappropriate treatment aimed at normalizing the thyroid hormone levels. TTR variants are furthermore of clinical importance because they are associated with various forms of amyloidosis.

Thyroxin-Binding Globulin (TBG)

TBG is an acidic glycoprotein with a single binding site for T4 or T3. Like corticosteroid-binding globulin it shows a high homology with the protease inhibitors α 1-antichymotrypsin and α 1-antitrypsin and it is a member of the serine protease inhibitor (serpin) family of proteins, although it is not a protease inhibitor. Many physiological and exogenous compounds, e.g. estrogens, alter the clearance and production of TBG. Genetic TBG abnormalities are classified according to the levels of TBG into complete or partial deficiencies or TBG excess. A variety of mutations in the *TBG* gene resulting in complete or partial deficiency have been identified. In contrast, over-expression of the protein due to gene amplification leads to TBG excess.

Transthyretin

TTR is a homotetramer formed of subunits containing 127 amino acids and contains two T4 binding sites. However, negative cooperativity allows the occupation of only one site at a given time. Multiple mutations in TTR have been identified. Their impact on T4 affinity

is variable and results in unchanged, increased or reduced affinity. Many of the point mutations are associated with autosomal dominant familial amyloidotic polyneuropathy (FAP), as well as amyloid depositions in the heart.

Albumin

Albumin is a monomer of 69 kD and is associated with the transport of a wide variety of endogenous and exogenous hydrophobic compounds. Albumin has a relatively low affinity for T4 and T3, but because of its high concentration it binds up to 10% of T4 and 30% of T3. Familial dysalbuminemic hyperthyroxinemia (FDH) is the most common cause of euthyroid hyperthyroxinemia, i.e. elevated total T4, but normal free T4 levels. It is inherited in an autosomal dominant manner and studies of several families have led to the identification of two independently recurring point mutations in the albumin gene that lead to substitutions of arginine at position 218 (R218H, R218P). Crystallographic analyses reveal that the two mutations result in conformational changes that favor binding of T4 on one of the four T4-binding sites on albumin. Another mutation in the albumin gene leads to familial dysalbuminemic hypertriiodothyroninemia with an affinity for T3 that is $\sim 40\times$ higher compared to wild type albumin.

Cellular Uptake of Thyroid Hormone into Cells

Cellular uptake of thyroid hormones is at least in part mediated by channels. Point mutations and deletions in the X-chromosomal *MCT8/SLC16A2* gene have been identified in several hemizygous males with elevated T3 and TSH levels. The phenotype is remarkable for severe mental retardation, spastic quadriplegia, rotary nystagmus and impaired gaze and hearing. Heterozygous females have discrete thyroid hormone abnormalities, but no neurological alterations. The abnormal T3 elevation may be due to impaired uptake into cells such as neurons.

Peripheral Monodeiodination

In target tissues, T4 is metabolized into the more active compound T3 by intracellular 5'-monodeiodination or into the inactive metabolite reverse T3 (rT3) by 5-monodeiodination. Monodeiodination of the outer and inner ring is catalyzed by three well-characterized deiodinases (DIO1, 2, 3), enzymes that are unusual because they are selenoproteins. They contain the rare amino acid selenocysteine (Sec), which is encoded by UGA, a triplet that usually encodes a stop codon. The translation of the Sec codon requires specific stem loop sequences that are located in the 3'-untranslated region (SECIS element). Over-expression of DIO3 in infantile hemangiomas and other vascular tumors leads to

consumptive hypothyroidism through inactivation of thyroid hormone at a rate that exceeds the maximal thyroid hormone synthesis.

Thyroid Hormone Action and Resistance to Thyroid Hormone (RTH)

Thyroid hormones exert their multiple cellular effects through nuclear thyroid hormone receptors (TRs), transcription factors which act by altering patterns of gene expression both as activators and repressors. The two TR genes, TR α and TR β , are members of the steroid receptor superfamily of nuclear receptors. The TRs are encoded by separate genes that are located on chromosomes 17 and 3 and they generate several splice variants.

Resistance to thyroid hormone (RTH) is characterized by decreased responsiveness to thyroid hormone. Biochemically, the syndrome is defined by elevated free thyroid hormones and an inappropriately normal or elevated level of TSH. The clinical spectrum is highly variable and ranges from isolated biochemical abnormalities to a constellation of features that includes goiter, variable signs of hyper- and hypothyroidism, short stature and delayed bone maturation. RTH is most commonly caused by autosomal dominant mutations in TR β that exert a dominant negative effect. In one family, RTH was caused by the loss of both thyroid hormone TR β alleles. More recently, familial cases of RTH without linkage to the TR β locus have been reported, indicating non-allelic heterogeneity.

Autoimmune Thyroid Disorders (AITD)

Hashimoto thyroiditis and Graves' disease are by far the most common diseases affecting the thyroid gland. They are characterized by an immune response to thyroidal antigens, infiltration by T-cells and production of antibodies. In Graves disease, thyroid stimulating antibodies lead to activation of the TSHR and hyperthyroidism, a process involving mainly a Th2 cell response. In Hashimoto's thyroiditis, a predominantly Th1-mediated chronic inflammation leads to progressive destruction and hypothyroidism. Both disorders are found more frequently in women.

The AITD are thought to be multifactorial disorders that require a genetic susceptibility in combination with environmental triggers. Epidemiological data support a strong genetic component in the development of AITD. This is further supported by twin studies. Graves' disease has been reported to occur in about 3–9% of dizygotic and 30–60% of monozygotic twins. For Hashimoto's thyroiditis, the concordance rate for monozygotic twins is about 40%, but very low for dizygotic twins. It should be noted that monozygotic twins are not identical in terms of their immune repertoire, which may explain the absence of perfect

concordance. As in other autoimmune diseases, associations have been established between the disease and the presence of certain human leukocyte antigens (HLA) constellations, but most linkage studies have been negative. Candidate gene analyses aiming at investigating associations between variants in genes such as the *TSHR*, *TPO* and the *T-cell receptor* with AITD have been negative. In contrast, the *TG* and *CTLA-4* (cytotoxic T-lymphocyte antigen 4) genes appear to be important susceptibility genes for AITD. Whole genome screens using AITD families identified several loci or genes with evidence for linkage to AITD.

Thyrotoxic Hypokalemic Periodic Paralysis

Periodic paralysis with hypokalemia can occur in patients with thyrotoxicosis. The clinical presentation is indistinguishable from familial hereditary hypokalemic paralysis without hyperthyroidism. Analysis of several genes encoding channels that are mutated in the familial form in patients with the thyrotoxic form led to the identification of a mutation in the *KCNE3* potassium channel gene. No mutations could be found in other candidates such as the calcium channel *CACN1A5* or the sodium channel *SCN4A*.

Thyroid Cancer

Thyroid cancers are relatively infrequent neoplasms and account for ~1% of all malignancies. The incidence of thyroid cancer is 1–10 per 100,000 in most countries. Thyroid cancer is about three times more frequent in females.

The majority of thyroid malignancies are papillary thyroid carcinomas (PTC) or follicular thyroid carcinomas (FTC), both originating from follicular cells. The highly malignant anaplastic carcinoma (ATC) and several unusual variants of thyroid cancers occur less commonly. A major risk factor for the development of benign and malignant thyroid neoplasms is radiation. This has been recognized after use of external radiation for medical treatment of benign and malignant conditions in the neck and more dramatically after accidental releases of ionizing radiation from atomic explosions and accidents in nuclear facilities.

PTC, which presents with several distinct histological subtypes, may occur in the setting of some rare autosomal dominant syndromes with disseminated neoplasias, familial adenomatous polyposis (Gardner's syndrome; *APC* gene mutations), as well as Cowden's syndrome (*PTEN* gene mutations), a condition characterized by multiple hamartomas with benign and malignant breast tumors, gastrointestinal polyps, ovarian cysts and mucocutaneous papulae. The Carney complex, an autosomal dominant disorder defined by cardiac myxomas, benign endocrine neoplasms and,

more rarely, follicular cancers, is associated with activating mutations in the regulatory subunit 1A of the cAMP-dependent PKA (*PRKAR1A*). Familial aggregation of FTC seems to occur in families with dysmorphogenesis.

Several molecular mechanisms involved in the development of thyroid neoplasms have been elucidated during the last decade. Hyperfunctioning adenomas frequently harbor an activating mutation in the *TSHR* or in *Gsa*. A substantial fraction of PTC has rearrangements between the *RET* tyrosine kinase and several other genes. Other molecular alterations in PTC include intrachromosomal rearrangements of the *TRK* gene and recent studies report a high prevalence of *BRAF* gene mutations in these tumors. The molecular alterations found in follicular adenomas and FTC are distinct from PTC and include mutations in *RAS* genes, loss of heterozygosity at several loci and a rearrangement fusing the amino terminal coding region of the *PAX8* to the *PPARγ* gene. More aggressive cancers have mutations in tumor suppressor such as *Rb* and *p53*.

Acknowledgement

P.K. is supported by grant 1R01DK63024-01 from NIH/NIDDK.

References

1. Gillam MP, Kopp P (2001) Genetic regulation of thyroid development. *Curr Opin Pediatr* 13:358–363
2. Gillam MP, Kopp P (2001) Genetic defects in thyroid hormone synthesis. *Curr Opin Pediatr* 13:364–372
3. Jameson JL, Weetman AP (2005) Diseases of the thyroid. In: Braunwald E, Fauci AS, Kasper DL et al (eds) *Harrison's Principles of Medicine*, 16th edn. Chapter 320, pp 2104–2127
4. Refetoff S, Dumont J, Vassart G (2001) Thyroid disorders. In: Scriver CR, Beaudet AL, Sly WS et al (eds) *The metabolic and molecular bases of inherited disease*, vol III, 8th edn. McGraw-Hill, New York, pp 4029–4075
5. Braverman L, Utiger RD (2000) *The Thyroid A fundamental and clinical text*, 8th edn. Lippincott-Raven, Philadelphia
6. Fagin JA (2004) How thyroid tumors start and why it matters: kinase mutants as targets for solid cancer pharmacotherapy. *J Endocrinol* 183: 249–256

Thyroid Dysgenesis

► Thyroid Disorders, Genetic Basis

Thyroid Stimulating Hormone

Definition

Thyroid stimulating hormone (TSH) is an anterior pituitary hormone that stimulates the production of thyroid hormone from the thyroid gland.

► [Hypothalamic and Pituitary Diseases Genetics](#)

TI

► [Thermodynamic Integration](#)

Tight Junctions

SAIMA AIJAZ, STEVEN GOODRICK, KARL MATTER, MARIA S. BALDA

Division of Cell Biology, Institute of Ophthalmology, University College London, London, UK
m.balda@ucl.ac.uk

Definition

Tight junctions (TJs) are morphologically distinct intercellular junctional structures of epithelial and endothelial cells that restrict paracellular diffusion and, hence, allow these cells to form selective barriers between compartments of different composition. TJs not only control paracellular permeability, they also participate in the maintenance of cell surface lipid polarity and the regulation of gene expression, cell proliferation and differentiation.

Characteristics

TJs are areas of close contact between neighbouring cells and come into view as apparent hemifusions of the adjacent plasma membranes. These regions of close contacts contain intramembrane strands that form a net-like meshwork that encircles the cells. These intramembrane strands are thought to be formed by different types of transmembrane proteins that also mediate cell-cell adhesion, although not all transmembrane proteins of TJs are associated with the strands. The transmembrane components are linked to the ► [cytoskeleton](#) and different signal transduction pathways by adaptor and/or scaffold proteins that form a submembrane

junctional plaque. This complex molecular machinery is used by epithelial and endothelial cells to recognise and transduce signals that regulate paracellular permeability and guide cell proliferation and differentiation.

Molecular Interactions

Transmembrane Proteins

Four types of transmembrane proteins have been described to be associated with tight junctions: occludin, claudins, JAMs (junction adhesion molecules) and CRB-3 (one of the vertebrate homologues of the *Drosophila crumbs* protein). TJs also contain a dense plaque of highly interconnected peripherally associated cytosolic proteins. These include the ► [adaptor proteins](#) ZO-1, -2 and 3, cingulin, a dimeric protein known to interact with many different junctional components and several signalling molecules, such as protein kinases and GTP-binding proteins, whose cellular functions we are only beginning to understand (1, see Table 1).

Occludin is a 60 kD membrane protein with four transmembrane domains, as well as a short N-terminal and a long C-terminal cytoplasmic domain. Five isoforms of occludin have been described, including one lacking the fourth transmembrane domain and another with a modified N-terminal domain. Although occludin is one of the components of the intramembrane strands, it does not appear to be a key structural component but has important regulatory functions. Occludin is involved in the regulation of paracellular permeability of small and medium non-ionic molecules. Its C-terminal domain is critical for this regulatory function and interacts *in vitro* with several proteins including ZO-1, -2, -3, cingulin, OAK, PI3K and atypical PKC ζ . However, the physiological role of these interactions is not known. Phosphorylation of occludin is suggested to affect its function in different experimental systems. The extracellular domains of occludin mediate its localisation at TJs most probably *via* homotypic interactions.

Claudins are 22–26 kD integral membrane proteins that, like occludin, have four transmembrane domains with both termini in the cytosol. Claudins comprise a large family of more than 20 members in mice and humans and are expressed in a tissue-specific manner (2). Expression of individual claudins in cells lacking TJs induces networks of intramembrane particles between adjacent transfected cells similar to those observed at TJs of epithelial cells, suggesting that claudins are important structural components of the intramembrane strands. Different claudins are often co-expressed and can co-polymerise to form individual strands that can adhere to each other intercellularly. The short C-terminal domain of claudins contains an YXV sequence, a ► [PDZ domain](#)-binding motif mediating interactions with the ZO proteins and MUPP1, a

Tight Junctions. Table 1 Major Components of Tight Junctions and their Interacting Partners*

Tight junction proteins	Molecular interactions
Transmembrane proteins	
Occludin (intracellular domain)	ZO-1/2/3, cingulin, OAK, PI3K, aPKC ζ , Itch, VAP-33
Claudins (intracellular domain)	ZO-1/2/3, MUPP1, PATJ
JAMs (extracellular domain)	Integrins, Reovirus, Adenovirus
JAMs (intracellular domain)	CASK, ZO-1, cingulin, afadin/AF-6, PAR3, MUPP1
CRB-3 (intracellular domain)	Pals1, PATJ, PAR6
Cytoplasmic proteins	
Adaptor proteins	
ZO-1	occludin, claudins, JAMs, ZO-2, ZO-3, Gi proteins, ZONAB, cortactin, afadin/AF-6, actin, connexins, ZAK, α -catenin, cingulin
ZO-2	claudins, ZO-1, adenovirus E4-ORF1, DNA-binding protein scaffold attachment factor-B (SAF-B), Fos, jun, CBP, cingulin
ZO-3	claudins, ZO-1, actin, PATJ, cingulin, AF6, p120 catenin
PATJ	Pals1, ZO-3
Pals1	PATJ, crumbs, PAR6
PAR3	PAR6, aPKC ζ , JAMs, PTEN, TIAM
PAR6	PAR3, cdc42, Pals1, aPKC ζ , CRB3, TGF β -receptor
MAGI-1	adenovirus E4-ORF1, papillomavirus E6, JAMs
MAGI-2	PTEN
MAGI-3	PTEN, pro-TGF α
MUPP1	claudins, JAMs, adenovirus E4-ORF1, papillomavirus E6
cingulin	ZO-1/2/3, actin, myosin, afadin/AF-6, JAM, GEFH1
Signalling molecules	
Gi proteins	ZO-1
rab13	PKA
aPKC ζ	PAR3, PAR6, PP2A, PAR1B/EMK1
GEF-H1	RhoA, cingulin, PAK1, PAK4, 14-3-3
ZONAB	ZO-1, cdk4, Ral A
sympleskin	polyadenylation factor CstF, Heat Shock Factor 1

* Due to space limitations, not all TJ components and their interactions could be listed

protein that contains 13 PDZ domains. The extracellular loops of claudins seem to create charge-selective channels in the paracellular pathway between epithelial cells as shown by mutations of specific amino acids (3). Since claudins often differ in their ion selectivity, their expression is thought to determine a tissue's paracellular conductive properties.

JAMs are glycoproteins of approximately 40 kD and are characterised by two immunoglobulin like folds

(V_H- and C₂-type) in the extracellular domain. JAMs localise to intercellular junctions of endothelial and epithelial cells but can also be expressed on circulating leukocytes and platelets (4). The extracellular domains of JAMs bind several ligands but can also mediate homophilic adhesion. Different JAMs associate through their extracellular domain with integrins involved in the regulation of leukocyte-endothelial cell adhesion, the β 2 integrins LFA-1 and Mac-1, as well as

$\alpha 4\beta 1$. Furthermore, different viruses use JAMs as receptors. Reovirus binds to JAM-1 and adenovirus to CAR, another member of the JAM family. Reovirus binding to JAM-1 appears to be required for induction of apoptosis and activation of NF- κ B, although the precise cellular pathways involved have not yet been identified. The cytoplasmic domains of JAMs associate with several cytoplasmic partners, CASK, a calcium/calmodulin-dependent serine protein kinase, ZO-1, cingulin, AF-6 and ASIP/PAR-3. Expression of a JAM lacking the cytoplasmic domain impairs the localization of ASIP/PAR-3, suggesting a role for JAM in the stabilization of the junctional complex.

CRB-3, a 25 kD type I membrane protein, is the only one of the three vertebrate homologues of the *Drosophila* crumbs protein that has been demonstrated to associate with TJs in epithelial cells. CRB-3 is spread over the entire [apical membrane](#) and only a minor fraction of the entire cellular pool is associated with TJs. The cytoplasmic domain of CRB-3 interacts with the adaptor protein Pals1, which functions as a linker to other functional components such as PATJ and another protein complex involved in cell polarity, the PAR3-PAR6-aPKC complex.

Cytoplasmic Proteins

The cytoplasmic TJ proteins can be divided into different functional groups: adaptor proteins, which often contain PDZ domains and are known to interact with several other TJ proteins, such as ZO-1/2/3, PATJ, Pals1, PAR-3 and PAR-6; signalling molecules, such as trimeric and monomeric [G-proteins](#), atypical PKC and GEF-H1; and proteins involved in transcription and RNA processing, such as the transcription factors ZONAB and hASH1 and symplekin, a protein suggested to be involved in RNA processing (see Table 1).

Several protein-protein interactions have been characterised between these proteins, and many of these proteins also interact with membrane components as well as the actin cytoskeleton. This architectural principle is well illustrated by ZO-1, the first identified TJ component. ZO-1 has several protein-protein interaction domains, three PDZ (for PSD-95, DlgA and ZO-1) domains, one [SH3 domain](#) and one with homology to yeast guanylate kinase (GUK domain). Proteins containing these types of domains are called [MAGUKs](#) (membrane associated guanylate kinase proteins). ZO-1 interacts with claudins *via* its first PDZ domain, with occludin *via* the GUK domain and forms stable complexes with either ZO-2 or ZO-3 *via* its second PDZ domain. ZO-1 also contains a discrete actin-binding domain in its C-terminal half. The SH3 domain functions in intracellular signalling by binding to the serine protein kinase ZAK and the [Y-box](#)

transcription factor ZONAB, which in turn binds to the cell-cycle regulator CDK4 (5).

While it is not feasible to list all the known interactions here, the general principle of the junctional architecture appears to be that individual adaptor components form discrete complexes, such as the ZOs, PAR3/6 or Pals1/PATJ. These adaptor complexes function as bridges between membrane components and the actin cytoskeleton, and recruit different types of signalling components, such as aPKCs, which bind to the PAR3/6 complex. On the next organisational level, these complexes also interact with each other, resulting in a complex network of proteins. For example, the interaction between the Pals1/PATJ and PAR3/6 complexes is mediated by direct binding of the PDZ domain of PAR6 to the C-terminus of Pals1. The interaction between ZOs and the Pals1/PATJ complex occurs *via* ZO-3, which binds to one of the PDZ domains of PATJ. An exception to this architectural principle is cingulin, a dimeric actin-binding protein that can bind to different TJ components but does not appear to have any preferential binding partners. As cingulin can form short filaments *in vitro*, it might form a supporting functional scaffold (6).

Regulatory Mechanisms

The classical function of TJs is to regulate paracellular permeability and many different junctional components have been linked to this process. Paracellular permeability is a regulated process that allows selective passive diffusion of certain ions as well as small hydrophilic molecules across the junctional barrier. The main membrane components involved in paracellular permeability are the claudins and occludin (7, see Table 2).

Claudins are expressed in a tissue-specific manner and the claudin-composition of a TJ determines its ion-selectivity. For example, claudin-16/paracellin-1 is expressed in the lung and the kidney, and patients carrying a mutation in this gene suffer from hereditary hypomagnesaemia due to a deficiency in paracellular magnesium resorption in the kidney. Therefore, it has been proposed that claudin-16 forms a paracellular channel or pore that allows the diffusion of magnesium ions. Transfection experiments with epithelial cell lines further suggested that different claudins favour paracellular diffusion of specific ions. These and other observations have led to a model in which claudins form homo- and hetero-oligomers that then engage in intercellular interactions to form paracellular channels or pores. The ion-selectivity of these channels is determined by the types of claudins they contain (2, 3). Other hereditary diseases and transgenic mice models have also linked specific claudins to the normal functioning of various tissues; although, in most cases

Tight Junctions. Table 2 Major Tight Junction Components and their Functions

Tight junction proteins	Suggested functions
Transmembrane proteins	
Occludin	cell-cell adhesion, paracellular permeability, restriction of apical/basolateral lipid diffusion, leukocyte transmigration, regulation of cell proliferation
Claudins	cell-cell adhesion, paracellular ion conductivity
JAMs	cell-cell adhesion, leukocyte transmigration, regulation of junction assembly, viral receptors
CRB-3	cell polarity and apical membrane biogenesis
Cytoplasmic proteins	
ZO-1	regulation of paracellular permeability, gene expression, cell proliferation and cell density
ZO-2	regulation of cell proliferation and transcription
ZO-3	regulation of junction assembly
PATJ/Pals1	cell polarity and apical membrane biogenesis
aPKC ζ /PAR3/6	regulation of junction assembly, cell polarity
Gi proteins	regulation of junction assembly
rab13	regulation of junction assembly
GEF-H1	activation of RhoA, regulation of paracellular permeability, cell proliferation
ZONAB	regulation of gene expression, cell proliferation and cell density
sympleskin	regulation of gene expression and RNA polyadenylation

it has not been determined whether the phenotype is caused by differences in junctional ion permeability. For example, claudin-14 was identified by positional cloning as the gene mutated in a form of human hereditary deafness and claudin-14 deficient mice show outer hair cell degeneration in the organ of Corti. It has been proposed that hair cell degeneration in the absence of claudin-14 is due to a switch in junctional ion-selectivity.

In the cases of claudins-1 and -5, mouse models suggest that their absence causes more general defects in permeability. Mice lacking claudin-1 die shortly after birth due to water loss across the skin, and the absence of claudin-5 causes leakage of small tracers across the brain endothelium. It is not clear how the absence of these claudins causes permeability defects and whether they possess important structural and/or regulatory roles in these tissues. Nevertheless, mice deficient in claudin-11/oligodendrocyte-specific protein exhibit neurological and reproductive defects and the intra-membrane strands of TJs of oligodendrocytes and Sertoli cells are reduced, indicating that, at least in some tissues, specific claudins are required for morphologically normal TJs.

Occludin is also involved in the regulation of paracellular permeability, but its role appears to be more important for non-ionic rather than ion permeability. The activity of occludin is regulated by its C-terminal cytoplasmic domain since expression of occludin mutants lacking this part in cultured cells lines results in stimulation of size-selective tracer permeability. It is thought that this regulatory mechanism involves phosphorylation events as well as the actin cytoskeleton since the C-terminal domain of occludin binds to protein and lipid kinases, as well as to actin filaments and cytoskeletal linkers. Occludin does not appear to be of structural relevance since occludin-deficient mice are viable. Nevertheless, these mice exhibit a complex phenotype and many tissues and organs appear to be defective due to chronic inflammations or altered differentiation. While the mechanisms causing these defects are unknown, the chronic inflammations might reflect a role of occludin in leukocyte transmigration, which was suggested by studies in epithelial cells in culture. Since occludin has also been linked to the restriction of trans-junctional lipid diffusion, some of these phenotypes might have been caused by a loss of cell surface lipid polarity (5).

Paracellular permeability is a regulated process that involves different signal transduction pathways and components such as PKCs, PKA, and trimeric and monomeric GTP binding proteins. These signalling pathways can act specifically on TJ proteins, on the actin cytoskeleton or on both. For example, the phosphorylation of the C-terminal domain of occludin appears to be important and has been proposed to modify cytoskeletal interactions, however, the responsible protein kinase(s) are not known. Regulation of contraction of the actin cytoskeleton by RhoA has been shown to regulate TJ formation and paracellular permeability. This is thought to involve activation of the RhoA effector Rho-kinase, which in turn increases myosin phosphorylation, resulting in myosin activation and contraction of the actin cytoskeleton. TJ-specific activation of RhoA can be mediated by the ►[guanine nucleotide exchange factor](#) GEF-H1, a member of the *Dbl* oncogene family (5).

Transmigration of leukocytes across epithelial and endothelial barriers requires the coordinated opening and closing of TJs to avoid a loss in the barrier function. JAM-1 is one of the TJ components that regulate this process. Antibodies against JAM-1 inhibit leukocyte transmigration in skin inflammation and in a model of mouse meningitis (8). Although JAM-1 interacts with a number of cytoplasmic components, it is not clear which ones are involved in this regulatory process. Occludin has also been linked to leukocyte transmigration across epithelia and this involves its N-terminal cytoplasmic domain. How the N-terminal domain of occludin facilitates transmigration is not known, and it is not clear whether occludin plays a similar role in endothelia.

More recently, tight junctions have also been linked to the regulation of epithelial differentiation and establishment of cell polarity. One pathway involves ZO-1 and the interacting Y-box transcription factor ZONAB. ZO-1 and ZONAB regulate expression of the growth factor co-receptor *erbB-2* in a cell-density dependent manner and function in the control of cell proliferation and cell density. ZONAB also binds to the G1/S phase regulator CDK4. Functional studies suggest that one mechanism by which ZO-1 can inhibit cell-cycle progression is by inhibiting the nuclear accumulation of the ZONAB/CDK4 complex. Although ZONAB accumulates in the nucleus of proliferating cells, it is not yet clear whether it regulates the expression of genes that are required for proliferation. Nevertheless, these observations suggest that TJs, like ►[adherens junctions](#), function in the suppression of proliferation (5).

The evolutionarily conserved crumbs pathway is a TJ-associated signalling module that regulates the establishment of epithelial cell polarity. In *Drosophila*, crumbs binds to a cytosolic complex consisting of stardust and *Drosophila* PATJ, and was identified

because of its essential role in epithelial morphogenesis. Over-expression of Crumbs in *Drosophila* expands the apical cell surface and leads to the disruption of cell polarity of photoreceptors. Recent work has found a human homologue of *Drosophila* Crumbs, known as CRB-1, to be mutated in two eye disorders, retinitis pigmentosa and Leber congenital amaurosis. In mammalian epithelial cells, CRB-3 binds to a TJ-associated protein complex consisting of Pals1, the homologue of stardust, and mammalian PATJ. As with the role of this pathway in *Drosophila*, the CRB-3/Pals1/PATJ-pathway appears to regulate junction assembly and biogenesis of the apical membrane.

Another evolutionarily conserved signalling module at TJs that regulates cell polarity is the PAR3/PAR6/aPKC complex, which also plays a role in TJ biogenesis. PAR-3 directly associates with the cytoplasmic domain of some JAMs, resulting in the recruitment of the PAR3/PAR6/aPKC complex to cell-cell junctions. PAR6 interacts with GTP-bound Cdc42 (cell-division control protein 42), a protein essential for cell polarity that becomes activated upon the induction of cell-cell adhesion. Binding of PAR6 to Cdc42 is thought to result in the activation of aPKC, which is required for the formation of distinct tight and adherens junctions by an as yet unknown mechanism. Furthermore, the direct interaction of PAR6 with Pals1, one of the components of the CRB-3 complex, is also regulated by Cdc42. Experimental evidence suggests that interactions between individual polarity complexes are required for epithelial polarity and formation of tight junctions.

In summary, this article attempts to point out molecular interactions and regulatory mechanisms that are relevant for the functioning of TJs in health and disease. TJs are formed by a complex network of membrane proteins, cytosolic adaptors and signalling molecules, as well as cytoskeletal linkers. While we now know many of the TJ components, the molecular interactions and regulatory mechanisms are only poorly understood. One of the recent exciting discoveries in the TJ field has been the identification of multiple signalling modules that localize to this intercellular junction and regulate gene expression and cell proliferation, as well as cell polarity and differentiation. It will be a great challenge to investigate these signalling modules at a molecular level and understand their roles in healthy and diseased tissues and organs.

Acknowledgements

We thank the members of our laboratories for critical reading of the manuscript. The authors' research is supported by The Wellcome Trust, Cancer Research UK, BBSRC (Biotechnology and Biological Sciences Research Council), and Fight for Sight.

► [Epithelium](#)

► [Glial Cells and Myelination](#)

References

1. Gonzalez-Mariscal L et al (2003) Tight junction proteins. *Prog Biophys Mol Biol* 81:1–44
2. Tsukita S, Furuse M (2002) Claudin-based barrier in simple and stratified cellular sheets. *Curr Opin Cell Biol* 14:531–536
3. Anderson JM, Van Itallie CM, Fanning AS (2004) Setting up a selective barrier at the apical junction complex. *Curr Opin Cell Biol* 16:140–145
4. Bazzoni G (2003) The JAM family of junctional adhesion molecules. *Curr Opin Cell Biol* 15:525–530
5. Matter K, Balda MS (2003) Signalling to and from tight junctions. *Nat Rev Mol Cell Biol* 4:225–236
6. D'Atri F, Citi S (2002) Molecular complexity of vertebrate tight junctions. *Mol Membr Biol* 19:103–112
7. Cereijido M et al (2000) Molecular physiology and pathophysiology of tight junctions. I. Biogenesis of tight junctions and epithelial polarity. *Am J Physiol Gastrointest Liver Physiol* 279:G477–482
8. Dejana E (2004) Endothelial cell-cell junctions: happy together. *Nat Rev Mol Cell Biol* 5:261–270

TIGRFAMs

Definition

TIGRFAMs are protein families based on Hidden Markov Models (HMMs).

► [Protein Databases](#)

TIM22/23 Complexes

Definition

Proteins that are imported into the mitochondrial matrix must cross both the outer and inner mitochondrial membranes. The TIM22/23 complexes (translocases of the inner membrane), together with translocases of the outer membrane (TOM), are essential for protein import.

► [Mitochondria – Biogenesis and Structural Organization](#)

Time-of-Flight Mass Spectrometer

Definition

In a time-of flight mass spectrometer, ionized species are collectively pulsed into a flight tube with identical kinetic energy. High mass ions will drift slower through the flight tube than light mass ions, and the mass

differentiation is measured by exact determination of the drift times.

► [Mass Spectrometry: Quantitation](#)

TIR Domain

► [Toll/IL-1 Receptor \(TIR\) Domain](#)

Tissue Micro Array

Definition

Tissue micro arrays are slides containing various types of small tissues embedded in paraffin. TMAs allow the tissue-specific localization of proteins or RNA and DNA by immunohistochemical and molecular studies. They are furthermore useful to find candidates for novel molecular targets such as a cancer gene.

► [Immunological Methods, the Use of Antibodies to Discover the Function of Novel Gene Products](#)

Tissue-Specific Genes

Definition

Tissue-specific genes are genes that are expressed in only a subset of tissues or developmental stages, and express proteins with functions specific for certain cell types or organ systems.

► [Multifactorial or Common Diseases](#)

► [CpG Islands](#)

TK2

Definition

Thymidine kinase 2 (TK2) is a protein that phosphorylates nucleotides in mitochondria. Deficiency of mitochondrial thymidine kinase (TK2) is associated with mitochondrial DNA (mtDNA) depletion, and manifests by severe skeletal myopathy in infancy.

► [Mitochondrial Myopathies](#)

TLR

► Toll-Like Receptor

TM (Transmembrane)

Definition

TM (transmembrane) refers to the retrovirus protein (encoded by the env gene) that crosses the virion envelope and interacts with the target cell membrane to mediate the entry of the virion into the cell.

► Retroviruses

TM (Transmembrane) Transactivating Proteins

Definition

TM transactivating proteins are proteins in retroviruses, such as Tat and Tax, which interact with the host's transcriptional machinery to stimulate LTR-directed transcription.

► Retroviruses

TMA

► Tissue Micro Array

TNF

► Tumor Necrosis Factor

TNF Receptor/ Fas Signaling Pathways

HARALD WAJANT

Department of Molecular Internal Medicine, Medical Clinic and Polyclinic II, Würzburg, Germany
harald.wajant@mail.uni-wuerzburg.de

Definition

The tumor necrosis factor receptor (TNFR) family encompasses a structurally related group of type I cell surface receptors that are involved in a wide variety of immunoregulatory mechanisms but also show increasing relevance in developmental processes. Signaling pathways triggered by TNFRs lead, besides other effects, to ►apoptosis induction or activation of cJun N-terminal kinase (JNK), p38-►mitogen-activated protein kinase (p38-MAPK) or nuclear factor-kappaB (NF-κB).

Characteristics

TNFRs are characterized by a conserved “cysteine-rich domain” (CRD) of about 40 amino acids. Several copies of this motif are found in the extracellular domains of these receptors. Each CRD has typically six conserved cysteines that are part of three intramolecular disulfide bridges. TNFRs are normally activated by ligands of a corresponding phylogenetically related group of trimeric proteins – the members of the TNF ligand family (1). There is also evidence that some environmental stimuli including bile acids and UV can trigger at least some members of the TNFR family in a ligand-independent manner. The members of the TNF ligand family, with the exception of lymphotoxin-α, are primarily expressed as type II membrane proteins (1, 2). In most cases soluble ligands can be derived from the membrane-bound form either by proteolytic processing or by alternative splicing. Importantly, in some cases the soluble and membrane-bound form of a ligand (e.g. TNF) can show considerable differences in their bioactivities (2). Most, if not all TNFRs are self-associated in the absence of their cognate ligand by an N-terminal pre-ligand binding assembly domain (PLAD). Thus, ligand binding induces reorganization of the pre-assembled TNFR complexes in a way that allows the recruitment of intracellular adaptor proteins enabling TNFRs to stimulate a variety of intracellular signaling pathways.

Molecular Interactions

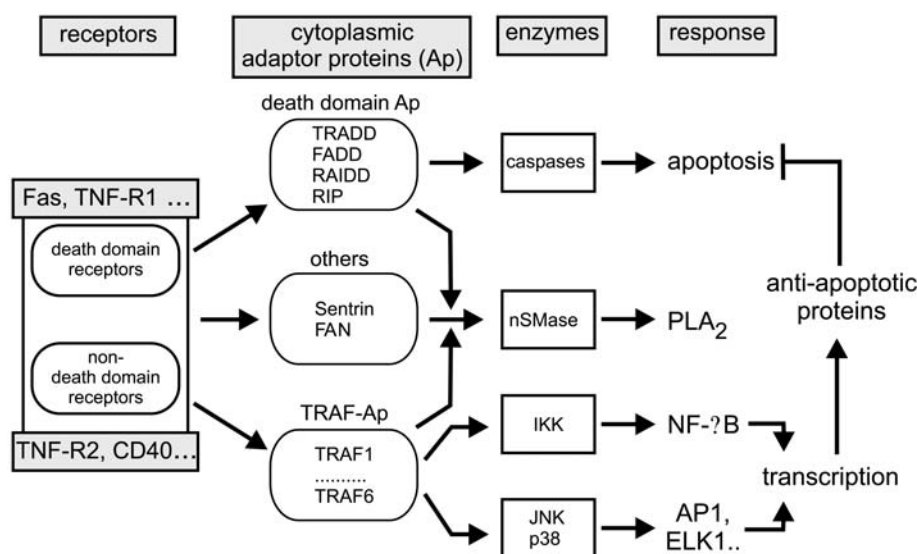
The TNFRs can be divided into two groups, the death receptors and the non-death receptors. The death receptors induce apoptosis in a variety of cells with critical involvement of a conserved part of their

cytoplasmic domain, which has been named the “death domain” (3). In contrast, the non-death receptors lack this structural hallmark and predominantly appear as molecules mediating the activation of NF- κ B and various MAPK cascades. Remarkably, the classification of the TNFR family into death receptors and non-death receptors is partially reflected by the use of two distinct categories of cytoplasmic adaptor proteins. Thus, death receptor triggering leads to the recruitment of death domain-containing adaptor proteins whereas the stimulation of non-death receptors results in the interaction with members of the **tumor necrosis factor receptor-associated factor** (TRAF) family (4).

The molecular mechanisms of death receptor-induced apoptosis have been best studied for **Fas**, TNF-related apoptosis inducing ligand (TRAIL) receptor 1 (TRAIL-R1) and TRAIL-R2, which trigger cell death by a common pathway (3). Upon stimulation of these receptors by their cognate ligands, they interact *via* their death domain with the death domain of the adaptor protein Fas-associated death domain protein (FADD). As well as its death domain, FADD contains a death effector domain, which mediates the secondary recruitment of caspase-8 and caspase-10 into the receptor-signaling complex (3). As several TNFRs are complexed upon ligand binding, the death-inducing signaling complexes (DISC) of TRAIL-R1, TRAIL-R2 and Fas contain several FADD and caspase-8/10 molecules. The dramatically enriched local concentration of caspase-8/10 within the DISC facilitates their activation by induced proximity. Active caspase-8 can

then directly trigger the execution phase of apoptosis by processing of caspase-3. However, dependent on cell type and circumstances, an additional caspase-8 stimulated amplification loop may be necessary to allow robust caspase-3 activation and apoptosis induction. This amplification loop is initiated by caspase-8 mediated cleavage of the Bcl2 family member Bid resulting in the release of a truncated Bid fragment. This fragment in turn is able to trigger Bax/Bak-dependent release of mitochondrial pro-apoptotic factors (e.g. cytochrome c, SMAC/Diablo) into the cytoplasm, leading to the formation of the **apoptosome**, a multi protein complex that activates caspase-3, and inhibition of anti-apoptotic molecules. Active caspase-3 not only executes the apoptotic program but also stimulates caspase-8 processing thereby closing an amplification loop. The cell-type specific importance of the mitochondria-dependent amplification loop for death receptor-induced apoptosis can be experimentally demonstrated by introducing Bcl2, which blocks mitochondria-dependent apoptosis (3). Cells in which death receptor-induced apoptosis is dependent on the mitochondrial amplification pathway, are protected by Bcl2, whereas in cells where caspase-8 robustly activates caspase-3 the same molecules fails to confer protection.

The most intensively studied cellular response predominantly stimulated by non-death domain-containing TNF receptors is activation of transcription factors of the NF- κ B/Rel family. In their inactive state homo- or hetero-dimers of NF- κ B/Rel proteins are sequestered



TNF Receptor/ Fas Signaling Pathways. Figure 1 Signaling pathways engaged by death receptors and non-death receptors of the TNF receptor superfamily. Please note that in some cell types activation of AP1 and/or NF- κ B leads to transcription of ligands of death receptors, thus exerting a pro-apoptotic activity in addition to its anti-apoptotic function shown in the scheme.

TNF Receptor/ Fas Signaling Pathways. Table 1 The TNF and TNF receptor families Death receptors are in shadowed boxes

Receptor	Ligand(s)	<i>In vivo</i> function	Associated diseases
TNF-R1 (TNFRSF1A, p55, CD120a, TNF-R60)	TNF (TNFSF2) LT α (TNFSF1, TNF β),	Endothelial activation Regulation of osteoclastogenesis Lymphoid organogenesis Regulation of adipocyte metabolism Parasite control Defence from intracellular pathogens	Rheumatoid arthritis Crohn's disease Septic shock Obesity-associated insulin resistance Cardiovascular disease TNF-receptor-1-associated periodic syndrome (TRAPS) Neurodegenerative disorders
TNF-R2 (TNFRSF1B, p75, CD120b, TNF-R80)	TNF (TNFSF2) LT α (TNFSF1, TNF β),	T-cell activation Activation-induced cell death of CD8(+) T-cells Thymocyte- and T-cell proliferation	Crohn's disease
LT β R (TNFRSF3)	(LT α) ₁ (LT β) ₂ (TNFSF3) LIGHT (TNFSF14, HVEM-L)	Lymph node development Micro-architecture of secondary lymphoid tissue Activation-induced cell death of CD8(+) T-cells	Mediator of cerebral malaria
OX40 (TNFRSF4, CD134)	OX40L (TNFSF4, gp34)	T-cell expansion T-cell activation	-
CD40 (TNFRSF5)	CD40L (TNFSF5, CD154, TRAP)	B-cell activation B-cell differentiation to plasma cells Isotype switching	Hyper IgM syndrome Atherosclerosis
Fas (TNFRSF6A, CD95, APO-1)	FasL (TNFSF6, CD95L)	Mediator of p53-induced apoptosis Immune privilege Activation-induced cell death of CD4(+) T-cells Homeostasis of peripheral T and B lymphocytes	Autoimmune lymphoproliferative syndrome (ALPS)
DcR3 (TNFRSF6B)	FasL (TNFSF6, CD95L) LIGHT (TNFSF14, HVEM-L) TL1A (TNFSF15, VEGI)	Decoy receptor of FasL and LIGHT	Anti-apoptotic factor in colorectal cancer
CD27 (TNFRSF7)	CD27L (TNFSF7, CD70)	T-cell activation T-cell differentiation	-
CD30 (TNFRSF8, Ki-1)	CD30L (TNFSF8)	Negative selection of thymocytes T-cell regulation	Marker of Reed-Sternberg cells of Hodgkin's lymphoma
4-1BB (TNFRSF9, CD137, ILA)	4-1BBL (TNFSF9)	T-cell costimulation	-
TRAIL-R1 (TNFRSF10A, DR4, Apo2)	TRAIL (TNFSF10, Apo-2L, TL2)	Antitumor activity of natural killer cells Mediator of p53-induced apoptosis	Mutated in some cancer types
TRAIL-R2 (TNFRSF10B, DR5, Killer, TRICK2A, TRICKB)	TRAIL (TNFSF10, Apo-2L, TL2)	Antitumor activity of natural killer cells Mediator of p53-induced apoptosis	Mutated in some cancer types
TRAIL-R3 (TNFRSF10C, DcR1, LIT, TRID)	TRAIL (TNFSF10, Apo-2L, TL2)	Decoy receptor of TRAIL	-

TNF Receptor/ Fas Signaling Pathways. Table 1 The TNF and TNF receptor families Death receptors are in shadowed boxes (Continued)

Receptor	Ligand(s)	<i>In vivo</i> function	Associated diseases
TRAIL-R4 (TNFRSF10D, DcR2, TRUND)	TRAIL (TNFSF10, Apo-2L, TL2)	Decoy receptor of TRAIL	-
RANK (TNFRSF11A, TRANCE-R)	RANKL (TNFSF11, TRANCE, OPGL, ODF)	Osteoclast differentiation and activation Bone remodeling Activation of dendritic cells	Familial expansile osteolysis Paget's disease Bone loss
OPG (TNFRSF11B, OCIF, TR1, osteoprotegerin)	RANKL (TNFSF11, TRANCE, OPGL, ODF) TRAIL (TNFSF10, Apo-2L, TL2)	Regulation of osteoclast differentiation and activation Regulation of dendritic cell activation	Familial expansile osteolysis Paget's disease Bone loss
DR3 (TNFRSF12, TRAMP, Wsl-1, LARD, TR3, APO-3)	TL1A (TNFSF15, VEGI)	Negative selection of thymocytes	-
HVEM (TNFRSF14)	LIGHT (TNFSF14, HVEM-L)	T-cell regulation	Entry receptor of herpes simplex virus
NGFR (TNFRSF16)	NGF	Regulation of neuron survival Coreceptor for TrkA,B	-
BCMA (TNFRSF17)	BAFF (TNFSF13B, Blys, THANK, TALL1) APRIL (TNFSF13)		-
AITR (TNFRSF18)	AITRL (TNFSF18)	Negative regulator of T-cell activation Inhibition of CD25(+) CD4(+) T cell-mediated suppression	-
TROY (TNFRSF19, TAJ)	-	-	-
DR6	-	Inhibitor of CD4(+) T-cell proliferation and Th2 cytokine production	-
TACI	BAFF (TNFSF13B, Blys, THANK, TALL1) APRIL (TNFSF13)	Inhibitory role in B cell activation	-
EDAR	EDA-A1	Morphogenesis of hair and teeth	Hypohidrotic/anhidrotic ectodermal dysplasia
XEDAR	EDA-A2	-	-
Fn14 (TweakR)	TWEAK (TNFSF12, APO3L)	Angiogenic regulation	-
RELT	-	T-cell coactivation	-
BAFF-R	BAFF (TNFSF13B, Blys, THANK, TALL1)	B-cell development B-cell survival	Implicated in auto immune diseases

in the cytoplasm by intermolecular interaction with ankyrin repeat-containing I- κ B proteins or by intramolecular interaction with an ankyrin repeat-containing carboxy-terminal domain present in some members of the NF- κ B/Rel family (5). Activation of NF- κ B/Rel transcription factors relies on the withdrawal of these proteins from the inhibitory influence of the ankyrin repeats. This is triggered by phosphorylation- and ubiquitination-dependent proteolytic degradation of associated I- κ B proteins or by proteolytic processing of ankyrin-repeat containing NF- κ B/Rel precursors. Both events are initiated by stimuli-dependent activation of the IKK complex. The latter is a multi protein complex of about 900 kD consisting of the regulatory subunit NEMO/IKK γ , the chaperones Hsp90 and Cdc37 and two related kinases, I- κ B kinases 1 and 2 (IKK1, IKK2) that phosphorylate I- κ B proteins and ankyrin repeat-containing NF- κ B precursor proteins (5). The gap between the non-death domain-containing TNF receptors and the IKK complex is closed by adaptor proteins of the TNF receptor-associated factor (TRAF) family (2, 4). These TRAF proteins, in particular TRAF2, TRAF5 and TRAF6, can simultaneously interact with the TNF receptors, atypical PKCs (e.g. PKC ζ) and MAPKKs (e.g. NF- κ B inducing kinase, MAPK kinase kinases 1 and 3) that have been implicated in the activation of the IKK complex. However, the exact molecular mode of IKK activation by TRAF and TRAF associated proteins is not well understood yet. Perhaps, the mechanisms differ for distinct cells and receptors and may also be dependent on the availability of the huge variety of TRAF-associated regulatory proteins (4).

The central roles of death domain containing or TRAF adaptor proteins in signaling by death and non-death receptors can explain the predominant role of these receptors as apoptosis and gene inducers. In this regard the death receptor TNF-R1 represents a remarkable exception as it fulfills *in vivo* mainly non-apoptotic, gene-regulatory functions. These, for a death receptor unusual, signaling capabilities are caused by the fact that TNF-R1, in contrast to most other death receptors, directly binds the death domain-containing adaptor protein TRADD, instead of FADD. As TRADD secondarily interacts with FADD and TRAF2, it connects TNF-R1 with the apoptotic pathway but also with the NF- κ B pathway. Due to the strong anti-apoptotic capacity of the latter, apoptosis induction by TNF-R1 is normally suppressed and is only apparent when the NF- κ B pathway is impaired (5).

Most of the known *in vivo* functions of TNF receptors can be traced back to activation of NF- κ B or apoptosis induction. However, members of the TNF receptor family also stimulate several other signaling pathways leading to the activation of JNK, p38, Erks, Src, Akt or [▶sphingomyelinases](#) (e.g. cited in ref. 2 for TNF).

Although the *in vivo* relevance of these signaling pathways for TNF receptor functions is rather poorly understood, there is evidence that in particular JNK, p38, Akt and sphingomyelinase activation serves among other things to fine-tune the balance between apoptosis-induction and NF- κ B activation.

Regulatory Mechanisms

Each step in TNF receptor signaling can be the target of regulatory mechanisms enabling cells to show flexible responses to stimulation by a given member of the TNF receptor family. Corresponding to the hierarchy of events in TNF receptor-induced signaling pathways, these regulatory mechanisms can be specific for a given receptor, common to a subgroup of TNF receptors (e.g. death receptors) or affect core components of a given pathway.

Receptor-specific regulation is often achieved by control of ligand and/or receptor expression. For example, TNF-R1 and Fas are constitutively expressed on most cell types, whereas expression of the corresponding ligands is very restricted, but can be induced by a huge variety of environmental cues including pathogen-specific molecular patterns, growth factors and cytokines, but also by stimulation of TNF receptors themselves. Remarkably, the signaling pathways involved in regulation of TNF ligands are often the same as those induced by stimulation of TNF receptors, allowing the formation of self-amplifying and auto-inhibitory feed-back loops or cytokine cascades.

Another means to gain receptor-selective regulation is the control of ligand-receptor interaction by neutralizing molecules, including decoy receptors and signaling-incompetent splice forms of the receptor or ligand. There are also examples of receptor-associated intracellular proteins that selectively modulate signaling by a single member of the TNF receptor family. The effects of receptor stimulation on NF- κ B activation and apoptosis-induction can also be determined by regulatory proteins targeting the adaptor proteins that link TNF receptors to these pathways. For example, FADD recruits the proteolytically inactive caspase-8 homologous FLICE-inhibitory protein (FLIP) into the DISC of death receptors, where this protein inhibits activation of co-recruited caspase-8. Moreover, there is evidence that distinct isoforms of FLIP are involved in the activation of non-apoptotic death receptor signaling. TRAF2, the pivotal signaling intermediate of non-death receptors, is also a major target of regulatory proteins, including cellular [▶inhibitor of apoptosis protein 1](#) (cIAP1), which triggers proteosomal degradation of TRAF2, and A20 which blocks TRADD-TRAF2 interaction. Importantly, NF- κ B activation and apoptosis-induction do not act independently, but are rather tightly linked in several ways (5). While NF- κ B

activation prevents apoptosis by up-regulating anti-apoptotic proteins including FLIP and cIAPs, ongoing apoptosis in turn inhibits the NF- κ B pathway by caspase-mediated cleavage of components utilized by this pathway. Amongst others, [▶caspases](#) cleave RIP, I- κ B, IKK2, p50 and p65. In most cases the caspase-derived fragments can act in a dominant-negative manner towards their non-cleaved counterparts (2, 5). Active caspases reduce the amount of signaling intermediates necessary to transduce a NF- κ B response, but also produce peptides that actively inhibit NF- κ B activation.

An additional layer of complexity in TNF receptor signaling is due to the activation of the JNK pathway (6). This pathway leads to the activation of the transcription factor AP1, which is involved in the regulation of members of the TNF ligand and receptor families (6). Looking at the network character of NF- κ B activation and apoptosis-induction, it is evident that there must be complex and manifold crosstalk amongst the members of the TNF receptor family.

References

1. Locksley, RM, Killeen, N, Lenardo, MJ (2001) The TNF and TNF receptor superfamilies: integrating mammalian biology. *Cell* 104:487–501
2. Wajant, H, Scheurich, P (2001) Tumor necrosis factor receptor-associated factor (TRAF) 2 and its role in TNF signalling. *Int J Biochem Cell Biol* 33:19–32
3. Fesik, SW (2000) Insights into programmed cell death through structural biology. *Cell* 103:273–82
4. Wajant, H, Henkler, F, Scheurich, P (2001) The TNF-receptor-associated factor family: scaffold molecules for cytokine receptors, kinases and their regulators. *Cell Signal* 13:389–400
5. Karin, M, Lin, A (2002) NF-kappaB at the crossroads of life and death. *Nat Immunol* 3:221–227
6. Weston, CR, Davis, RJ (2002) The JNK signal transduction pathway. *Curr Opin Genet Dev* 12:14–21

TNF-Related Apoptosis Inducing Ligand

Definition

TNF-related apoptosis inducing ligand (TRAIL, APO-2 ligand) is a member of the TNF family of cytokines, which preferentially induces apoptosis in tumor but not normal cells.

[▶Apoptosis](#)

[▶Apoptosis, Regulation and Clinical Implications](#)

TNF- α

[▶Tumor Necrosis Factor- \$\alpha\$](#)

TOB Complex

Definition

A distinct group of mitochondrial outer membrane proteins consists of membrane-embedded β -barrel proteins. Insertion of β -barrel proteins into the outer membrane of mitochondria is mediated by the so-called TOB complex. Known constituents of this complex are Tob55 and Mas37.

[▶Mitochondria – Biogenesis and Structural Organization](#)

TOF

[▶Time-of-Flight Mass Spectrometer](#)

Tolerance

Definition

The brain adapts to the continued presence of the drug of abuse. Tolerance to a drug of abuse is usually characterized by the need to take increasing amounts of the drug to elicit the same effect as this dose evoked previously.

[▶Immune Tolerance](#)

[▶Addiction, Molecular Biology](#)

Toll/IL-1 Receptor Domain

Definition

Toll/IL-1 receptor (TIR) domain designates a conserved domain in the cytoplasmic tail of Toll/IL-1 family receptors, which plays a crucial role in the

activation of the signal transduction cascade upon receptor activation.

► [Interleukins and Their Receptors, a Role in Regulating Hematopoietic Stem Cells](#)

Toll-Like Receptor

Definition

Toll-like receptors (TLRs) play an essential role in the innate immune responses. Bacterial pathogens such as lipopolysaccharides (LPS, endotoxin), bacterial lipoproteins, peptidoglycans and also CpG nucleic acids activate innate immunity, as well as stimulate the antigen-specific immune response and trigger the inflammatory response. Members of the TLR gene family convey signals stimulated by these factors, activating downstream signal transduction pathways that result in transcriptional regulation and stimulate immune function. TLRs represent a phylogenetically old system for recognition of bacterial pathogens that was initially discovered in *Drosophila*.

► [Crohn Disease](#)

TOM Complex

Definition

Proteins that are imported into the mitochondrial matrix must cross both the outer and inner mitochondrial membranes. The TOM (translocases of the outer membrane) complex, together with the TIM complex, is essential for protein import.

► [Mitochondria – Biogenesis and Structural Organization](#)

Tomography

Definition

Tomography refers to a technique for the generation of cross sectional images of three dimensional structures, especially of the human body.

► [Electron Tomography](#)
► [PET](#)

Tonofibrils

Definition

Tonofibrils are comprised of bundles of keratin intermediate filaments that form cage-like structures around the nucleus, and extend from the perinuclear region to the cell periphery IFs.

► [Cytoskeleton](#)

Topoisomerase II

Definition

Topoisomerase II (Topo II) refers to an ATP-dependent enzyme that can alter DNA topology by passing one double helix strand through another via a transient double helix break, an activity necessary for the resolution of catenated sister DNA strands. It is located both at centromeres and along mitotic chromosome arms at the base of loops.

► [Chromosome Condensation](#)

TP

► [Thymidine Phosphorylase](#)

TPR Motif

► [Tetratricopeptide Repeat \(TPR\) Motif](#)

Tracer

Definition

Tracer denotes substances, which can be discriminated from systemic substances, whose distribution and

accumulation in the system under investigation can be measured by dedicated detectors. In PET, radioactive tracers are used.

►PET

Traditional Silicon-Based Protocols

Definition

Traditional Silicon-Based Protocols are protocols such as deposition, etching and bonding applied in the development of circuits and chips for use in micro-electronics. These techniques are well developed for silicon and require special facilities and equipments.

►Proteomics in Microfluidic Systems

TRAF

Definition

Tumor necrosis factor (TNF) receptor associated factor (TRAF) refers to an adapter protein that is involved in TNF and IL-1 receptor signaling. TRAF proteins have also been implicated in functioning as regulators of apoptosis.

►Interleukins and Their Receptors, a Role in Regulating Hematopoietic Stem Cells

►Tumor Necrosis Factor Receptor-Associated Factor (TRAF)

TRAIL

►TNF-Related Apoptosis Inducing Ligand (TRAIL, APO-2 ligand)

Trait

Definition

Trait defines a particular aspect of the phenotype that can be measured or observed directly, e.g. blood pressure or body weight.

►Large-Scale ENU Mutagenesis in Mice

Trajectory/Trajectorial

Definition

In general, a trajectory is the path traced in three dimensions by a body moving as a result of an external force. In mass spectrometry, the body is an ion and the trajectory may describe its way from the ion source to the detector. A molecular dynamic trajectory is generated during a molecular dynamics simulation and consists of a set of frames of coordinates and velocities the represent the movement of the atoms over time. It can be used to generate an average structure and analyze fluctuations of geometries and time-dependent processes in molecules.

►Molecular Dynamics Simulation in Drug Design

Trans-Acting Factor

Definition

Trans-acting factor refers to a protein or RNA-protein-complex that regulates transcription by binding to the DNA sequences, in the vicinity of the structural portion of a gene, which are required for gene expression. These DNA sequences, e.g. promoter sequences or operators, are called cis-acting elements.

►Alternative Splicing

Transactivation Function

Definition

Transactivation function describes a component of a transcription factor's protein structure that is necessary to initiate transcription.

►Nuclear Hormone Receptors: Regulators of Gene Transcription and Cellular Signaling

Transacylation

Definition

Transacylation is the transfer of an acyl residue from acyl-CoA, or an acyl-donor to the hydroxy group of the acceptor mediated by acyltransferase.

►High-HDL Syndrome

Transcript

Definition

A transcript is the messenger RNA (mRNA) copy of a gene, generated by an RNA polymerase.

- ▶ [Microarrays in Pancreatic Cancer](#)
- ▶ [RNA Polymerase I](#)

Transcript Elongation

- ▶ [Transcription Elongation](#)

Transcription

Definition

Transcription describes the synthesis of an RNA molecule on a DNA template, catalysed by RNA polymerases, thereby transferring the genetic message encoding a protein from DNA to RNA.

- ▶ [Cap-Independent Translational Control](#)
- ▶ [Chromatin Remodelling](#)
- ▶ [Common Diseases Genetics](#)
- ▶ [Core Promoters](#)
- ▶ [DNA Helicases](#)
- ▶ [Duchenne Muscular Dystrophy](#)
- ▶ [Genetic Code](#)
- ▶ [Fragile X Syndrome](#)
- ▶ [Full Length cDNA Sequencing](#)
- ▶ [Leucine Zipper Transcription Factors: bZIP Proteins](#)
- ▶ [Proteomics in Human-Pathogen Interactions](#)
- ▶ [Repeat Expansion Diseases](#)
- ▶ [RNA Polymerase I](#)
- ▶ [RNA Polymerase III](#)
- ▶ [Single-Cell Gene Expression Profiling: Cell-level Biology by Multiplexed Expression Fluorescence in Situ Hybridization](#)
- ▶ [Splicing](#)
- ▶ [SRY – Sex Reversal](#)
- ▶ [Transcription Factors and Regulation of Gene Expression](#)
- ▶ [Transcription Elongation](#)
- ▶ [Two-Hybrid System](#)

Transcription Coupled Repair

Definition

Following DNA damage, it is essential for a cell to remove damage from actively transcribed regions of the DNA. In human cells repair is more rapid in the transcribed DNA strand compared with the non-transcribed DNA strand. This preferential repair of the transcribed strand is called transcription coupled repair (TCR). TCR is regarded as a subclass of nucleotide excision repair. Transcription coupled repair is defective in patients suffering from [▶Cockayne Syndrome](#).

- ▶ [DNA Repair Mechanisms](#)
- ▶ [Nucleotide Excision Repair](#)

Transcription Elongation

YUKI YAMAGUCHI, TADASHI WADA,
HIROSHI HANDA

Tokyo Institute of Technology, Graduate School of Bioscience and Biotechnology, Midori-ku, Yokohama, Japan
hhanda@bio.titech.ac.jp

Synonyms

Transcript elongation; transcriptional elongation

Definitions

Transcription is a complex process that can be divided into several mechanistically distinct steps: (1) assembly of preinitiation complex (transcription preinitiation), (2) promoter opening, (3) transcription initiation, (4) promoter clearance (promoter escape), (5) transcription elongation and (6) transcription termination. During the transcription elongation step, RNA polymerase that has left a gene promoter tracks along the DNA to synthesize up to two million bases of nascent RNA.

Characteristics

Molecular Mechanism of Transcription Elongation

Several RNA polymerases have been described, bacterial RNA polymerase, archaeal RNA polymerase and three eukaryotic RNA polymerases. All these polymerases show significant sequence similarity and

may have a common evolutionary origin. Therefore, the basic mechanisms of transcription elongation are likely to be very similar among these polymerases. Because bacterial RNA polymerase and eukaryotic **▶RNA polymerase II** have been extensively characterized, this essay will focus on these polymerases.

Bacterial RNA polymerase, in its 'core enzyme' form, comprises five proteins; two α subunits and one each of β' , β and ω subunits. The σ subunit, a constituent of the 'holo enzyme', which plays a critical role in transcription initiation, is considered to dissociate from the core enzyme at the onset of transcription elongation. However, recently this classic σ cycle model has been challenged (1). Eukaryotic RNA polymerase II comprises twelve subunits, Rpb1~12. Rpb3 and Rpb11 are homologous to bacterial α , whereas Rpb1, Rpb2, and Rpb6 are homologous to bacterial β' , β and ω respectively.

During transcription elongation, RNA polymerase forms a ternary complex together with template DNA and nascent RNA. The term 'transcription elongation complex' is often used to refer to this ternary complex or to this complex and associated regulatory proteins as a whole. Within the transcription elongation complex, 12–15 bases of DNA are unwound to form the transcription 'bubble' and 3'-terminal 8–9 bases of RNA transcript are contained by forming a hybrid with the template strand of DNA (2). The growing 3' end of the RNA transcript is usually maintained at the active site of RNA polymerase.

Transcription elongation is considered to proceed as follows. An NTP complementary to the next position on the DNA template fills the open substrate-binding site of RNA polymerase. A phosphodiester bond is formed between the 3'-hydroxyl group of RNA and the α -phosphate of the NTP. Then, RNA polymerase translocates one nucleotide forward or downstream with respect to the nucleic acids.

Regulation of Transcription Elongation

Transcription elongation is positively and negatively regulated by cis-acting elements (3). The cis-acting elements are mainly located on nascent transcripts and affect the polymerases that have read through the corresponding DNA elements. Types and roles of RNA elements have been extensively characterized in prokaryotes, while much less is known about these elements in eukaryotes. By forming structures that destabilize the transcription elongation complexes, RNA elements often cause pause, arrest or termination of RNA polymerases without the aid of protein factors. In this context, pause is a transient inactive state of RNA polymerase. Arrest is a more persistent state that eventually leads to termination unless reactivated by a positive **▶transcription elongation factor** such as

bacterial GreB or eukaryotic TFIIS. This type of RNA element is found in the bacterial *his* operon, for example. Other types of RNA elements serve as platforms for RNA-binding proteins. A prominent example in prokaryotes is the Rho utilization (*rut*) site, a negative regulatory element bound by the transcription termination factor Rho. In eukaryotes, the Tat response (TAR) element, a positive regulatory element found in the human immunodeficiency virus (HIV) gene, serves as a binding site for the viral activator Tat and cellular cofactors.

Transcription elongation is positively and negatively regulated by trans-acting factors as well (3, 4). A number of transcription elongation factors have been identified in both prokaryotes and eukaryotes (Table 1). Transcription elongation factors regulate the process of transcription elongation in at least three distinct ways. The first class of transcription elongation factors binds RNA polymerase to modify its catalytic activity, thereby inducing or releasing pause or arrest. This class includes NusA, NusG and GreB in prokaryotes and TFIIF, TFIIS, elongin, ELL, DSIF and NELF in eukaryotes. The second and third classes of transcription elongation factors are eukaryote-specific.

The second class regulates the phosphorylation status of RNA polymerase II (4). The Rpb1 subunit of RNA polymerase II has a unique C-terminal domain (**▶CTD**) that contains the phosphorylatable heptapeptide Tyr-Ser-Pro-Thr-Ser-Pro-Ser, which is repeated 52 times in mammals. Transcription elongation factors P-TEFb and Fcp1 are a kinase and a phosphatase specific to the CTD, respectively. Phosphorylation of the CTD is dynamically regulated during transcription and thought to play an important role in efficient transcription elongation and mRNA processing. CTD phosphorylation reverses the transcription repression imposed by the negative transcription elongation factors DSIF and NELF.

The third class of transcription elongation factors regulates transcription elongation through chromatin. In eukaryotes, packaging of nuclear DNA into chromatin poses a considerable problem for the progression of RNA polymerases. Transcription elongation factors such as FACT and protein methyltransferases Set1 and Set2 appear to assist RNA polymerase II in that process. It is proposed that these factors travel along the DNA with the transcription elongation complex and change chromatin structure through interaction with or covalent modification of histones.

Coupling of Transcription Elongation to Other Biochemical Reactions

In prokaryotes, transcription elongation is coupled to translation. In prokaryotes, transcription and translation are not spatially separated as in eukaryotes and,

Transcription Elongation. Table 1 A partial list of transcription elongation factors

Name	Function
Prokaryotes	
GreA	reactivates from arrest
GreB	reactivates from arrest
NusA	involved in termination and anti-termination
NusE	involved in termination and anti-termination
NusG	involved in termination and anti-termination
Eukaryotes	
TFIIS	reactivates from arrest
TFIIF	enhances the rate of transcription elongation
Elongin	enhances the rate of transcription elongation
ELL	enhances the rate of transcription elongation
Tat-SF1	enhances the rate of transcription elongation
DSIF/Spt4-Spt5	activates or represses transcription elongation
NELF	represses transcription elongation with DSIF/Spt4-Spt5
P-TEFb/Cdk9-Cyclin T	phosphorylates the CTD
Fcp1	dephosphorylates the CTD
Spt6	interacts with histones
FACT/Spt16-Pob3	interacts with histones H2A-H2B
Paf1 complex	recruits Set1/COMPASS
Set1/COMPASS	methyates histone H3
Set2	methyates histone H3

moreover, nascent transcripts for protein-coding genes are usually mature mRNAs. Thus, there is no conceptual difficulty in coupling between transcription and translation. Indeed, it is generally accepted that translation of bacterial mRNA initiates before completion of transcription. Translation can even affect transcription in the process called ►**attenuation**. In the *trp* operon encoding tryptophan biosynthetic enzymes, for example, availability of the amino acid tryptophan regulates the progression of ribosomes at the 5' portion of the mRNA, which, in turn, regulates the progression of the preceding transcription elongation complex. Overall, this allows feedback control of the intracellular tryptophan level.

In eukaryotes, transcription elongation by RNA polymerase II is tightly coupled to mRNA processing, i.e. 5' capping, splicing and 3'-end processing (5). These processing reactions, once thought to be

post-transcriptional events, are now believed to occur co-transcriptionally. Capping of mRNA occurs as early as when the first 20–30 nucleotides of nascent transcript have been synthesized. Splicing of introns at the 5' portion of a gene initiates before completion of transcription. 3'-end processing, i.e. transcript cleavage and polyadenylation, coincides with transcription termination. The capping enzyme stably associates with the transcription elongation complex at the early stage of transcription elongation. Moreover, its catalytic activity is stimulated by components of the transcription elongation complex, enabling efficient 5' capping. Similarly, components of the splicing and 3'-end processing machinery interact with the transcription elongation complex. An emerging view is that the phosphorylated CTD may serve as a scaffold for all these processing events.

Transcription elongation is also coupled to DNA repair in both prokaryotes and eukaryotes (6). DNA lesions within the coding region, caused by UV light or chemical mutagens, often block the progression of RNA polymerases. Stalled transcription elongation complexes are intrinsically inhibitory to DNA repair because they prevent access of repair enzymes to damaged sites. However, normal cells can remove DNA lesions in the coding region considerably faster than those in the genome overall. This process, called ►transcription-coupled repair (TCR), requires special protein factors in addition to the classic DNA repair machinery. TRCF in prokaryotes and CSB in eukaryotes are shown to be involved in TCR. These factors are proposed to function by removing stalled polymerases from DNA lesions and by recruiting repair enzymes to the lesions.

Genes and Pathways Regulated at the Level of Transcription Elongation

In prokaryotes, a number of genes and pathways are known to be regulated at the transcription elongation level. As mentioned above, many amino acid biosynthetic operons are regulated by attenuation mechanisms. In addition, synthesis of ribosomal RNA and ribosomal proteins is shown to be regulated by mechanisms that involve RNA elements and the NusA protein.

In eukaryotes, relatively little is known about target genes and pathways. However, many of the so-called ►immediate-early genes that are rapidly induced by environmental stresses or extracellular stimuli appear to be regulated at the transcription elongation level. The best-characterized example is *hsp70*, a major heat shock-inducible gene encoding a molecular chaperone. Transcription elongation on *hsp70* is normally down-regulated by DSIF and NELF at the promoter-proximal region. Upon heat shock, P-TEFb is rapidly recruited onto the gene and phosphorylates the CTD of the stalled RNA polymerase II, thereby activating transcription elongation.

Regulation of transcription elongation also plays an important role in the proliferation of some viruses (7). During the lysogenic cycle of bacteriophage λ , expression of the viral late genes is down-regulated by premature transcription termination. Upon expression of the viral proteins N and Q, they activate transcription elongation through multiple interactions with RNA elements and host transcription elongation factors, allowing for entry into the lytic cycle. The life cycle of HIV is regulated by a mechanism very similar to bacteriophage λ . The viral protein Tat induces the synthesis of full-length viral transcripts through interaction with the TAR RNA, P-TEFb and other

cellular proteins, leading to viral production. Hepatitis delta virus (HDV) is unique in that the RNA genome replicates through RNA-dependent RNA synthesis by RNA polymerase II, a DNA-dependent RNA polymerase. The viral protein HDAg activates otherwise inefficient transcription elongation on the unusual RNA template through interaction with RNA polymerase II. Thus, these viruses use the host transcription elongation machinery in some way or another for the benefit of viral proliferation.

Clinical Relevance

Cockaine syndrome is the result of a mutation in the human *CSB* gene. The clinical phenotype of this hereditary disease is characterized by postnatal growth failure, neurological dysfunction, cachectic dwarfism, photosensitivity, sensorineural hearing loss and retinal degradation.

In acute myeloid leukemia, the human *ELL* gene on chromosome 19 undergoes frequent translocation with the *MLL* gene on chromosome 11. The translocation t(11;19)(q23;p13.1) results in production of the MLL-ELL chimeric protein, which seems to have transforming activity. It is not known whether aberrant regulation of transcription elongation is responsible for the disease.

HIV is the causative agent of AIDS, while HDV is a causative agent of acute and chronic hepatitis, cirrhosis and hepatocellular carcinoma. Tat and HDAg, encoded by HIV and HDV respectively, are both essential for viral proliferation. These viral proteins and the host transcription elongation machinery are thus potential targets for the development of antiviral therapies.

References

1. Borukhov S, Nudler E (2003) RNA polymerase holoenzyme: structure, function and biological implications. *Curr Opin Microbiol* 6:93–100
2. Gnatt A (2002) Elongation by RNA polymerase II: structure-function relationship. *Biochim Biophys Acta* 1577:175–190
3. Uptain SM, Kane CM, Chamberlin MJ (1997) Basic mechanisms of transcript elongation and its regulation. *Annu Rev Biochem* 66:117–172
4. Conaway JW, Shilatifard A, Dvir A et al (2000) Control of elongation by RNA polymerase II. *Trends Biochem Sci* 25:375–380
5. Proudfoot NJ, Furger A, Dye MJ (2002) Integrating mRNA processing with transcription. *Cell* 108:501–512
6. Svejstrup JQ (2002) Mechanisms of transcription-coupled DNA repair. *Nat Rev Mol Cell Biol* 3:21–29
7. Yamaguchi Y, Delehouzee S, Handa H (2002) HIV and hepatitis delta virus: evolution takes different paths to relieve blocks in transcriptional elongation. *Microbes Infect* 4:1169–1175

Transcription Elongation Complex

Definition

Transcription Elongation complex is a ternary complex of RNA polymerase, template DNA, and nascent RNA, which is formed during transcription elongation. Sometimes the term refers to the ternary complex and associated regulatory proteins as a whole.

► [Transcription Elongation](#)

- [Homeodomain Transcription Factors](#)
- [Hypoxia Inducible Factors](#)
- [Leucine Zipper Transcription Factors: bZIP Proteins](#)
- [Muscle Development](#)
- [NFκB Pathway](#)
- [Protein/DNA Interaction](#)
- [RNA Polymerase III](#)
- [Two-Hybrid System](#)
- [Winged Helix Transcription Factors](#)
- [Zinc Finger Transcription Factors](#)

Transcription Elongation Factor

Definition

Transcription Elongation factor is a transcription factor that regulates the process of Transcription Elongation.

► [Transcription Elongation](#)

Transcription Factors and Regulation of Gene Expression

RAINER RENKAWITZ

Institute for Genetics, Justus-Liebig-University,
Giessen, Germany

rainer.renkawitz@gen.bio.uni-giessen.de

Transcription Enhancer

► [Enhancer](#)

Transcription Factor

Definition

Transcription factors regulate gene activity by binding to DNA regulatory elements in target genes. Basal transcription factors bind to the basal promoter of target genes, while regulatory transcription factors bind sequences further upstream from the initiation site.

The unique combination of binding sites and transcription factors used determines when and where a gene is active. The term transcription factor (TF) usually stands for DNA-binding transcription factors (activators or repressors) but strictly speaking also includes the non-DNA binding adaptor proteins such as the mediator complex.

- [Bone and Cartilage](#)
- [Drosophila Model of Cardiac Disease](#)
- [Enhancer](#)
- [Familial Hypercholesterolemia](#)
- [Growth Factors](#)
- [Hedgehog Signalling](#)
- [Helix-Loop-Helix Transcription Factors](#)

Definition

Whole genome sequencing has clearly shown that organism complexity arises from the complexity of the regulatory network controlling gene activity, rather than from the invention of new genes. Key components of gene regulation are ► [transcription factors](#), which can be defined by their location relative to the transcriptional start site of a particular gene.

First, general or basal transcription factors are required for basic ► [promoter](#) function (► [preinitiation complex](#) ► [PIC](#)) and are recruited directly to the vicinity of the transcriptional start. These complexes may have a different composition depending on the gene and the tissue (1).

Second, transcriptional activators or repressors can bind either to promoter upstream binding sites or to DNA binding sites (► [enhancers](#) or ► [silencers](#)) separated from the promoter by thousands of base pairs. Several enhancers may act in synergy with promoter upstream elements or may lead to a reduced gene activity in combination with silencers (1).

Characteristics

Basically, transcription factors are defined by two properties, they bind to DNA and they influence gene activity. The ability to regulate gene activity is often not well defined and usually is mediated by cofactors or cofactor complexes interacting with the activation or the repression domain of the transcription factors. DNA binding is mediated by a limited number of different ► [DNA binding domains](#), which are used to group transcription factors into different classes, as

explained below. This grouping is according to the generally accepted classification as explained in detail in ►<http://www.gene-regulation.com/pub/databases/transfac/cl.html>. In addition to the clearly defined classes described below, other transcription factors exist which cannot clearly be grouped into these categories.

Basic Domains

The basic region dictates DNA-sequence specificity by binding into the major groove of the DNA. This however is not sufficient for specificity, since usually the recognition helix is in contact with only 3–5 nucleotides. An adjacent dimerization domain, which allows for hetero or homodimerization, is the prerequisite for sequence specific DNA binding mediated by both basic domains. Amphipathic alpha helices that allow for dimerization are either ►[helix-loop-helix](#) motifs (b►[HLH](#)) or ►[leucine zipper](#) structures (b►[ZIP](#)) consisting of repeated leucine residues at every seventh position. Human genome sequencing has revealed about 300 genes coding for basic-helix-loop-helix and basic-leucine zipper proteins (2).

Zinc-Coordinating DNA-Binding Domains

Two zinc-coordinating DNA-binding domains differing in size, composition and function are present in each molecule of the ►[nuclear receptors](#). These receptors bind to lipophilic hormones such as steroids, thyroid hormone or others. The DNA binding motifs are two ►[zinc finger domains](#), each finger comprising 4 cysteine residues coordinating one zinc ion. The sequence between the first two cysteines of the second finger mediates dimerization upon DNA-binding.

A related, but different zinc finger motif comprises two cysteine and two histidine residues coordinating one zinc ion. This structure is exemplified by the TFIIIA and Krüppel factors and is found in many variants of DNA binding domains harboring a range of 2 to 36 individual fingers. This gene family is the largest transcription factor family in the human genome with about 900 different members (2). For those factors functionally analyzed, it has been shown that transcriptional repression is a common theme for several of these proteins, although activation has also been observed. The zinc finger protein ►[CTCF](#) contains a unique function not seen with any other human transcription factor namely the restriction of enhancer activity to specific genes (3).

►[Helix-Turn-Helix](#)

This type of DNA binding domain consists of three consecutive alpha-helix structures, with helix 3 contacting mainly the major groove of the DNA. Helices 2 and 3 resemble the helix-turn-helix structure of prokaryotic regulators. Within eukaryotes, transcription factors involved in developmental regulation

of gene expression often bind the target DNA *via* a helix-turn-helix motif, which is called a ►[homeodomain](#). This protein family comprises about 250 different members encoded in the human genome (2).

Beta-scaffold Factors with Minor Groove Contacts

These DNA binding motifs are found in important transcription factors like NFκB, p53 and STAT factors. The structure of the Rel-type DNA binding domain (as in NFκB) exhibits a bipartite subdomain structure, with each subdomain comprising a beta-barrel with several loops that form an extensive surface for contacting the major groove of the DNA.

Cellular and Molecular Regulation

Gene regulation is mediated by a complex interplay of regulatory events. For the sake of clarity, these events will be separately explained based on the type of regulatory mechanism as well on the order of events. These events are separated into those mechanisms regulating the availability of the transcription factors through ►[signal transduction](#) pathways (upstream events) from mechanisms mediating the regulatory signal from the DNA bound transcription factor to the gene (downstream events).

Upstream Events: Signal Transduction Pathways

Signal transduction in the context of gene regulation is mediating a signal from the outside of a cell to the transcription factors regulating target genes. A large group of proteins are mediating different signal transduction pathways, which are briefly summarized below and classified according to the inducing signal (4).

Growth Factors

Upon growth factors binding to the extracellular domain of ►[receptor tyrosine kinases](#) on target cells, these receptors dimerize. This induces autophosphorylation on tyrosine residues in the intracellular receptor domain, which serves as a signal to trigger a complex cascade of cytoplasmic protein phosphorylations finally ending up in the nuclear translocation or activation of transcription factors.

Peptide Hormones

►[Seven-transmembrane receptors](#) spanning the cellular membrane seven times can bind peptide hormones on the outside, which induces a conformational change of the receptor, activating G-proteins on the inside. Activated G-proteins induce the synthesis of second messengers, which again induce a phosphorylation cascade resulting in transcription factor activation.

Cytokines

After cytokine binding, ►[cytokine receptors](#) form heterodimeric complexes spanning the cellular

membrane. This activates receptor associated Janus kinases (JAKs) which phosphorylate the receptor and STAT proteins (signal transducer and activator of transcription) on tyrosine residues. The activated STAT molecules translocate as dimers into the nucleus and act as transcription factors by binding to target genes.

Transforming Growth Factor- β

Family members of **transforming growth factor- β** (TGF- β) bind to specific receptors on the outside of target cells. This induces heterodimeric receptor interaction and phosphorylation on serine/threonine residues of these receptors as well as **Smad proteins**. Activated Smads can dimerize, translocate to the nucleus and act as transcription factors.

Lipophilic Hormones

Steroid hormones (like glucocorticoid), thyroid hormone and other lipophilic hormones are transported into the cytoplasm and the nucleus by passive diffusion or active transport. Intracellular or nuclear receptors bind the lipophilic hormones specifically and are thereby activated to function as transcription factors.

Downstream Events

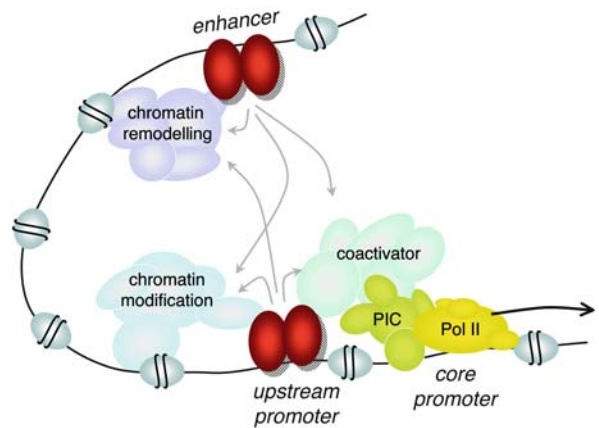
Chromatin Remodeling

Eukaryotic DNA is densely packed into a structure called chromatin. Chromatin consists of DNA wrapped around protein complexes generated from **histones** 2A, 2B, 3 and 4 (nucleosomes). Depending on the precise position of a nucleosome on a particular regulatory DNA region, sequences may not be available for specific binding of transcription factors.

Chromatin remodeling complexes carry out changes in chromatin structure by altering DNA histone contacts within a nucleosome in an ATP dependent manner. Nucleosomes may thereby be shifted on the DNA or the nucleosome contacts to specific DNA sequences may be loosened. Remodeling complexes can be divided into three classes on the basis of the similarities of their ATPase subunits to the Swi2/Snf2, Isw1, and Mi-2 proteins. Direct interactions between transcriptional activators and specific Swi/Snf subunits have been demonstrated (Fig. 1), such that selectivity between different Swi/Snf complexes is explained by specific protein/protein interactions (5).

Chromatin Modifications

In order to be recognized and read by the transcriptional machinery, chromatin has to be specifically modified. These modifications, like acetylation or methylation of the histone proteins, play a dual role. First, they alter the structure of chromatin to control transcriptional initiation and elongation. Second, they provide specific modification marks that are recognized by factors



Transcription Factors and Regulation of Gene Expression. Figure 1 Transcription factors regulate gene activity. DNA (black line) is wrapped around nucleosomes (grey oval spheres). Transcription factors (red ovals) can be bound to upstream promoter elements as well as to enhancer elements separated from the promoter region. DNA bound transcription factors can recruit multi-unit protein complexes involved in different functions. Chromatin remodeling complexes induce ATP dependent sliding of nucleosomes, thereby allowing sequence recognition by other DNA binding factors. Chromatin modification complexes enzymatically modify histones within the nucleosomes. This marks the chromatin for gene activation or repression. The coactivator complex bridges the DNA bound transcription factors with the preinitiation complex (PIC), which allows for efficient RNA polymerase II (Pol II) binding and for start site identification. The grey arrows indicate similar functions and interactions involving transcription factors bound to upstream promoter or to enhancer elements.

involved in gene activation as well as other factors mediating gene silencing.

Some modifications are relatively well understood. For example, **histone-acetyltransferases** (HAT) have been found in the context of **coactivator complexes** associated with DNA bound transcriptional activators (Fig. 1). In contrast, **histone-deacetylases** (HDACs) that mediate the function of corepressor complexes have been detected in the context of DNA bound transcriptional repressors. Histone methylation has been shown to be involved in specific functions depending on which histone and which amino acid is methylated; for example lysine-4 methylation of histone H3 occurs in the context of active genes, whereas methylation of lysine-9 within histone H3 is found on repressed genes in heterochromatic gene regions but also in some euchromatic genes. In addition, arginine methylation, serine phosphorylation and ADP-ribosylation of histones influence gene activity (6).

Interaction with the Preinitiation Complex

Genes coding for proteins are transcribed by ►**RNA polymerase II** (Pol II). Initiation of transcription is only possible after establishing multistep processes that require the assembly of a preinitiation complex (PIC). This large complex is comprised of proteins called general initiation (basal) factors, such as TFIIA, TFIIB, TFIID, TFIIIE, TFIIF and TFIIH. These factors again are multisubunit complexes, the central one of which (TFIID) contains the TATA-DNA sequence binding protein (TBP). These factors are highly conserved and guide Pol II to promoters. In some cases, sequence specific transcription factors, which are bound to promoter upstream elements or to enhancer elements, interact directly with components of the PIC increasing the efficiency of recruiting Pol II to the promoter. Other examples of interaction with the PIC are indirect and involve complexes described in yeast as Mediator and in multicellular organisms as coactivator complexes (1). Again, the purpose of such an interaction is very likely to generate a platform for high affinity binding of the PIC and therefore for increased transcriptional activity (Fig. 1).

Clinical Relevance

Precise gene regulation is required throughout development, for cell differentiation and proliferation, as well as for proper metabolism. Therefore, the temporal and spatial control of gene regulation is one of the most fundamental processes in biology as well as in medicine. Many examples of gene dysregulation are described in the context of pathological situations. Transcription factors mediating gene regulation are therefore of great clinical relevance. The current state of the knowledge of individual transcription factors can be seen at ►<http://bioinfo.weizmann.ac.il/cards/index.shtml>. Below, a few samples are presented by introducing one example for each of the transcription factor classes.

Fos/Jun

The Fos and Jun family members belong to the bZIP class of transcription factors, capable of homo- and heterodimerization. They have been identified as the two subunits that generate the transcription factor AP-1. This transcription factor regulates the expression of multiple genes involved in differentiation, apoptosis and proliferation. Therefore, constitutive activation of either Fos or Jun or both can contribute to tumor formation. The growth promoting function of Jun may be mediated by the repression of tumor suppressor genes.

Id Proteins

bHLH proteins have been found to be involved in muscle and nervous system development, as well as of other cell types. Id proteins have no basic domain and therefore cannot bind to DNA, but possess a HLH domain used for

dimerization with bHLH proteins. The resulting complex is unable to bind DNA and therefore inhibits the function of bHLH proteins. Interestingly, Id proteins are often over-expressed in several human cancers.

Glucocorticoid Receptor

The glucocorticoid receptor is a ligand dependent transcription factor of the steroid receptor superfamily, which binds to DNA and dimerizes *via* the zinc finger domain. Glucocorticoids are the end products of the hypothalamus-pituitary-adrenal axis and are required for homeostasis, cell differentiation, growth control, apoptosis and modulation of inflammatory processes. The use of glucocorticoids as anti-inflammatory and immunosuppressive drugs is widely accepted. Rheumatoid arthritis and many other rheumatoid autoimmune diseases are very sensitive to glucocorticoid treatment.

HOX Proteins

Homeodomain (helix-turn-helix motif) containing transcription factors are required for the regulation of cell proliferation, differentiation and embryonic development. Clustered homeodomain containing genes (HOX genes) determine the anterior/posterior patterning during embryogenesis. Dysregulation of HOX genes often contributes to hematological malignancies.

NFκB

NFκB and related Rel-type factors play an important role in defense against infectious diseases and cellular stress. Activation occurs through a wide variety of signals, such as bacterial products, viral infection, oxidative stress and different cytokines, to name only a few. A major mechanism for NFκB regulation is binding to inhibitory molecules (IκB), preventing the translocation of NFκB into the nucleus. Activation signals are mediated by the release of NFκB from IκB binding. Therefore it is not surprising that aberrant activation of NFκB plays a crucial role in many inflammatory diseases, especially asthma and rheumatoid arthritis. Furthermore, constitutive activation of NFκB can lead to proliferation, inhibition of apoptosis and therefore to development of cancer in various leukemias, carcinomas and adenocarcinomas.

Transcription Factor Associated Complexes

Downstream mediators of transcription factor function have not only been shown to be required for the proper function of transcriptional regulation, but furthermore, in the case of malfunction are the cause of many disorders and neoplasias. Aberrant histone deacetylation is connected to the Rett syndrome, whereas aberrant histone acetylation is associated with several neoplasias. Histone deacetylases are important targets for drugs with potential therapeutic value in the treatment of several of these diseases. Inhibitors of

HDACs are in clinical trials for cancer treatment and are of potential value for the treatment of polyglutamine neurodegenerative disorders.

References

1. Hochheimer A, Tjian R (2003) Diversified transcription initiation complexes expand promoter selectivity and tissue-specific gene expression. *Genes Dev* 17:1309–1320
2. Tupler R, Perini G, Green MR (2001) Expressing the human genome. *Nature* 409:832–833
3. Ohlsson R, Renkawitz R, Lobanenko V (2001) CTCF is a uniquely versatile transcription regulator linked to epigenetics and disease. *Trends Genet* 17:520–527
4. Pawson T, Nash P (2000) Protein-protein interactions define specificity in signal transduction. *Genes Dev* 14:1027–1047
5. Huang C, Sloan EA, Boerkoel CF (2003) Chromatin remodeling and human disease. *Curr Opin Genet Dev* 13:246–252
6. Lachner M, O'Sullivan RJ, Jenuwein T (2003) An epigenetic road map for histone lysine methylation. *J Cell Sci* 116:2117–2124

Transcription Initiation

Definition

Transcription initiation defines the site of, and/or time when, mRNA synthesis begins.

- RNA Capping
- Transcription

Transcription Unit

Definition

Transcribed DNA segment of a gene. The resulting primary transcript still includes introns which are to be removed by RNA splicing.

- Enhancer

Transcriptional Activator

Definition

Transcriptional activator designates a protein that binds specific DNA elements in the regulatory regions of an eukaryotic gene thereby activating the process of transcription.

- Protein Interaction Analysis: Variations of the Yeast Two-Hybrid System

Transcriptional Elongation

- Transcription Elongation

Transcriptional Enhancer

- Enhancer

Transcriptional Machinery

Definition

The transcriptional machinery refers to the protein complex that consists of RNA polymerase and a set of general transcription factors that are formed at the core promoter. This is also called a pre-initiation complex.

- Core Promoters
- RNA Polymerase I Transcription
- RNA Polymerase II Transcription
- RNA Polymerase III

Transcriptional Profiling

Definition

Transcriptional profiling refers to a parallel analysis of gene expression. The profiling provides information about the physiological and metabolic state of cells or organisms under various conditions.

- PNA Chips

Transcriptional Regulation

Definition

Transcriptional regulation defines the way in which RNA synthesis from a DNA template is controlled.

- Methylation of Proteins
- RNA Polymerase I Transcription

- RNA Polymerase II Transcription
- RNA Polymerase III
- Transcriptional Repression

Transcriptional Repression

GERALD THIEL, MICHAEL LIETZ, MATTHIAS HOHL
Department of Medical Biochemistry and Molecular
Biology, University of Saarland Medical Center,
Homburg, Germany
gerald.thiel@uniklinik-saarland.de

Definition

The information for the blueprint of an organism is stored in the genome. Segments of nucleic acids, the genes, encode for functional polypeptides or RNAs that are required for construction of an organism during development and maintenance of the functional organization throughout life. However, at any given time, only a fraction of these genes are actively transcribed. Moreover, the expression rate of a gene may change substantially depending on the developmental state of the organism, differentiation into a particular tissue or the presence or absence of extracellular signaling molecules such as hormones, neurotransmitters, cytokines or metabolites. Genes contain regulatory regions, ► **promoters**, ► **enhancers** and ► **silencers** that are used for the temporal and spatial control of gene transcription. The regulatory regions are binding sites for the basal transcriptional machinery, including the ► **RNA polymerase** that synthesizes RNA from the DNA template. In addition, transcription factors binding in a sequence-specific manner to the DNA of the regulatory region are required to accomplish a tight control of gene transcription. Proteins that activate transcription are termed activators and this mode of gene control is known as positive regulation. In contrast, transcription is blocked by repressor proteins and the biological function of repressors is termed negative control.

Characteristics

Transcriptional repressor proteins such as the lac or the tryptophan repressors were first discovered in prokaryotes. The DNA-tethered form of the repressor turns genes off by blocking RNA polymerase binding to the promoter or its movement along the DNA. Thus, a competition between a gene-specific repressor protein and RNA polymerase is common in *E. coli*. Although transcriptional activator proteins such as the catabolite activator protein CAP are also present in bacterial cells, the predominant mode of control is thought to be negative (4). In contrast, many eukaryotic transcriptional activator

proteins are known, establishing the view that positive regulation *via* activator proteins is the predominant mode of gene control in eukaryotes. There are two main reasons for the predominance of transcriptional activators in eukaryotes. First, the chromatin structure produces a restrictive ground state, i.e. most promoters are inaccessible to the RNA polymerases and transcriptional activators are required for eukaryotic genes to be transcribed. Transcriptional repressors would therefore exhibit a redundant biological activity. In contrast, the RNA polymerase in bacteria generally has access to the promoters, thus defining a nonrestrictive transcriptional ground state in bacteria. Second, taking a teleological view, positive regulation seems more efficient than negative regulation in a multicellular organism and seems a common sense solution. However, economical design principles do not always define cellular regulation mechanisms. Today, we know that transcriptional repression is quite common in eukaryotes, indicating that transcriptional repressors play an important role in the regulation of eukaryotic genes.

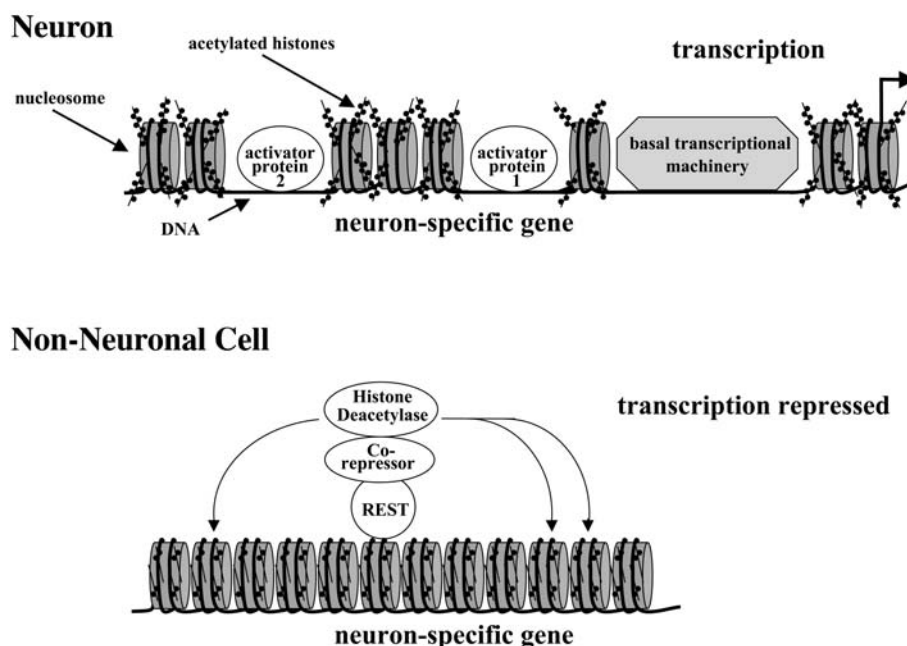
Mammalian transcriptional repressors can be classified as passive or active repressor proteins (1). Passive repressor proteins do not have intrinsic repressing activity or a portable repression domain. Rather, these proteins repress RNA synthesis by competing with transcriptional activators for DNA binding, by forming inactive heterodimers with transcriptional activators rendering them incapable of interaction with DNA or by binding to coactivators required for the transcriptional activator proteins. Thus, passive repressor proteins transmit their biological function either *via* DNA- or protein-protein interaction. The “inducible cAMP early repressor” (ICER) is an example of a passive repressor protein. ICER is encoded by the cAMP response element modifier (CREM) gene that encodes more than 10 splice variants, including activators and repressors. ICER contains the basic region leucine zipper domain necessary for dimerization and DNA binding, but lacks the two glutamine-rich activation domains necessary for transcriptional activation. Thus, ICER occupies the DNA-binding site *via* the basic region and may additionally induce the formation of non-functional heterodimers. In a broad sense, IκB can also be classified as a passive repressor because it interferes with transcriptional activation of NFκB-responsive genes by preventing nuclear translocation of NFκB.

Active mammalian transcriptional repressor proteins have an intrinsic repression activity that targets the chromatin organization of the genome. This type of transcriptional repression is activator-independent and functions over long distances. Two types of active transcriptional repression can be distinguished, transcriptional repression *via* ► **histone** deacetylation and gene silencing *via* histone methylation and ► **heterochromatin** formation.

A critical determinant in the regulation of eukaryotic genes is the structure of the chromatin. The fundamental building blocks of the chromatin are the ►**nucleosomes** with two molecules each of the histones H2A, H2B, H3 and H4 building the core histone and the 147 base pairs of DNA wrapped around this histone octamer. The single nucleosomes are linked by short stretches of DNA and by the linker histone H1. The N-terminal regions of the core histones are often modified by acetylation, methylation or phosphorylation. Histone acetylation is of major importance for the regulation of gene transcription. The acetylation of lysine residues reduces the net positive charges of the core histones, leading to a decrease in their binding affinity for DNA. The termini are subsequently displaced from the nucleosome and the nucleosome unfolds providing access for transcription factors. Deacetylation of histones by ►**histone deacetylases** removes the acetyl group from the ϵ -amino group of lysine residues of histones allowing ionic interactions between the negatively charged DNA phosphate backbone and the positively charged amino termini of the core histones. This results in a more compact chromatin structure that is not as accessible for the transcriptional machinery. While

histone acetylation and hyperacetylation has been correlated with transcriptionally active chromatin, histone deacetylation is thought to be involved in repression of transcription. Histone acetylation and deacetylation are claimed to be major regulatory mechanisms of transcription that function by modulating the accessibility of transcription factors to their binding sites on DNA. Likewise, mammalian transcriptional repressor proteins have been described that block transcription by recruiting histone deacetylases to the promoter (3). The histone deacetylases may be directly bound to the repressor protein or bind *via* adaptor proteins such as mSin3A or the nuclear corepressors N-CoR or SMRT.

The zinc finger protein RE-1 silencing transcription factor (►**REST**) that is mainly expressed in non-neuronal cell types functions as a negative regulator of neuron-specific gene transcription. REST recruits histone acetylases to neuronal genes, resulting in removal of acetyl groups from the core histones. Consequently, the neuronal genes are embedded into more tightly packed chromatin that is inaccessible for transcriptional activators (Fig. 1, bottom). In neurons, REST is expressed in extremely low concentrations.



Transcriptional Repression. Figure 1 The transcriptional repressor REST controls neuron-specific gene transcription via recruitment of histone deacetylases. In neurons, only very low levels of REST can be detected. In fact, a decrease in the REST concentration during neuronal differentiation is most probably the essential prerequisite for neuronal genes to be transcribed. The chromatin configuration of neuronal genes is open, due to an extensive acetylation of the core histones. Transcriptional activators as well as the RNA polymerase II complex can gain access to the DNA and trigger gene transcription. In nonneuronal tissue, however, REST is present and binds in a sequence-specific manner to several neuronal genes. Using the mSin3A protein as a bridge, REST recruits histone deacetylases to the transcription unit. The deacetylation of histones results in an enhanced binding between the histones and the DNA, thus impairing the binding of transcription factors to their specific DNA sequences. Moreover, further compaction of the chromatin is accomplished by inducing interactions between neighboring nucleosomes.

Therefore, the chromatin has an open configuration, allowing transcriptional activators to bind and initiate transcription of neuronal genes (Fig. 1, top). Gene transcription is thus the result of a relief of repression. DNA in vertebrates is subjected to methylation of cytidine bases mainly within the dinucleotide C^mpG. This modification of DNA is primarily associated with transcriptional repression. Methylated DNA may block access of sequence-specific transcription factors or may transform the methylated DNA sequences into a condensed state. Methylated DNA interacts with methylation-specific transcriptional repressor proteins, most prominently the MeCP2 protein that binds to single, symmetrical C^mpG pairs regardless of the sequence context. MeCP2 represses transcription *via* recruitment of histone deacetylases to methylated DNA regions, thus compacting the nucleosome array. MeCP2 thus connects DNA methylation with histone deacetylation. Inhibitors of histone deacetylases such as trichostatin A are, however, only able to partially relieve MeCP2-mediated transcriptional repression, indicating that MeCP2 may have a second way to repress transcription aside from the recruitment of histone deacetylases.

Transcriptional repression *via* histone deacetylation is gene-specific, i.e. histone deacetylases have to be recruited to the transcription unit either by sequence-specific repressor proteins or by methylation-specific transcription factors. Therefore, only a small portion of the genome is affected by histone deacetylation induced chromatin compaction at a specific time point. In contrast, gene silencing makes a larger portion of the genome inaccessible for transcriptional activation, thus acting in a regional manner. The eukaryotic genome is divided into euchromatin and heterochromatin, the latter representing a relatively inaccessible form of chromatin. How are heterochromatin domains established and maintained? Based on studies with fruit flies, yeast and mammals, the following stepwise model has been proposed. First, the amino termini of the H3 histones are deacetylated by histone deacetylases and subsequently methylated on lysine residue 9 by methyltransferases. Secondly, the H3-mLys9 modification functions as a high affinity binding site for the silencing protein, heterochromatin protein 1 (HP1), which thirdly, oligomerizes and extends the heterochromatic region into neighboring chromatin (2). The fact that terminally differentiated cells such as plasma or glial cells exhibit large-scale heterochromatinization, further indicates that gene silencing *via* formation of facultative heterochromatin is part of the genetic alterations occurring during development.

Transcriptional repression *via* recruitment of histone deacetylases and gene silencing *via* HP1 induced heterochromatin formation represent two different modes of transcriptional repression. However, the

sequences of biochemical reactions are connected, and, in fact, transcriptional repressor proteins may block transcription using both mechanisms. MeCP2 triggers histone deacetylation by recruitment of histone deacetylases, but the deacetylation of the histone tails is necessary for histone methylation to take place. Moreover, MeCP2 associates *in vivo* with [▶histone methyltransferases](#) that are able to methylate the lysine 9 residue of H3. This H3-mLys9 subsequently recruits the heterochromatin inducing protein HP1, the starting point for the establishment of facultative heterochromatin. Likewise, REST and the Krüppel associated box (KRAB) repressor proteins are reported to induce gene silencing in addition to the recruitment of histone deacetylases. The observations that the corepressor protein KRAB associated protein 1 (KAP1) and the REST corepressor protein CoREST are able to attract HP1 suggests that REST and the KRAB repressor proteins induce gene silencing that spreads beyond the transcription unit where REST or the KRAB proteins were bound in the first place.

Clinical Relevance

While transcriptional repression and gene silencing are essential for maintenance of the cellular integrity of a multicellular organism, aberrant silencing may cause diseases. In a wide variety of human cancers, for instance, a genome-wide DNA hypomethylation is observed, suggesting that this modification is the molecular reason for chromosomal instability and tumor formation. The observation that tumor suppressor genes are silenced by aberrant hypermethylation has led to the use of DNA methyltransferase inhibitors such as 5-aza-2'-deoxycytidine in the treatment of cancer. Mutations in the gene DNMT3B encoding a DNA methyltransferase are found in the human ICF syndrome (immunodeficiency, centromeric heterochromatin instability, facial anomalies) and cells of patients with this disease show an extensive loss of methylation from the pericentromeric heterochromatin. Mutations in the X-linked MeCP2 gene are the major cause of RETT syndrome, a neurological disorder characterized by the loss of speech and hand skills, microcephaly and seizures. Histone deacetylase inhibitors have been tested as a new class of anticancer drugs. Inhibition of histone deacetylases causes a hyperacetylation of histones and a diminished methylation-associated gene silencing. In addition, the histone deacetylase inhibitor valproic acid is used as an established drug in the long-term therapy of epilepsy. The wide range of diseases caused by aberrations in transcriptional repression sheds light to the importance of understanding how negative gene regulation functions in eukaryotes.

References

1. Cowell IG (1994) Repression versus activation in the control of gene transcription. *Trend Biochem Sci* 19:38–42
2. Grewal SIV, Elgin SCR (2002) Heterochromatin: new possibilities for the inheritance of structure. *Curr Opin Gen Dev* 12:178–187
3. Ng HH, Bird A (2000) Histone deacetylases: silencers to hire. *Trends Biochem Sci* 25:121–126
4. Oehler SM, Iler-Hill B (1995) Prokaryotic control of transcription: How and why it differs from eukaryotic control? In: Baeuerle PA (ed) *Inducible gene expression vol 1: Environmental stresses and nutrients*, Boston, USA, pp 1–24

Transcriptional Repressor

Definition

Transcriptional repressor denotes a protein that binds specific DNA elements in the regulatory regions of an eukaryotic gene and which represses the process of transcription.

► [Protein Interaction Analysis: Variations of the Yeast Two-Hybrid System](#)

Transcriptional Termination

Definition

Termination defines the step in which a fully mature RNA transcript is formed and released from the RNA polymerase and the DNA template.

► [Core Promoters](#)

Transcriptome

Definition

The transcriptome can be defined as the complete collection of transcribed elements of the genome present at any given moment in a cell or tissue, and its behaviour over time and cell states. In addition to mRNAs, it also represents non-coding RNAs that are used for structural and regulatory purposes. Alterations

in the structure or levels of expression of any one of these RNAs or their proteins can contribute to disease. An understanding of the transcriptome will provide a valuable tool in the research for novel drugs.

► [Automated High-Throughput Functional Characterization of Human Proteins](#)

► [Drosophila as a Model Organism for Functional Genomics](#)

► [ES Cell Differentiation as a Model System for Functional](#)

► [Medaka as a Model Organism for Functional Genomics](#)

► [Morpholino Oligonucleotides, Functional Genomics by Gene ‘Knock-down’](#)

► [Peptide Chips](#)

► [Xenopus as a Model Organism for Functional Genomics](#)

Transcytosis

Definition

Transcytosis describes a process by which endocytosed membrane proteins are directed to a different domain of the plasma membrane.

► [Cell Polarity](#)

► [Vesicular Traffic](#)

Transdifferentiation

Definition

Transdifferentiation describes the conversion of a differentiated cell into a different cell type.

► [Microarrays in Pancreatic Cancer](#)

Transdominant Genetic Effect

Definition

Transdominant genetic effect denotes the inhibition of a gene function caused by the expression of an exogenous, dominant gene.

► [Peptide Aptamers](#)

Transduction

Definition

Transduction describes the transfer of a gene from one cell to another, or from one genomic location to another, by means of bacteriophages, viruses, transposable elements and DNA injection. The genetic material is integrated into the genome of the host cell or integrated by formation of a plasmid. If integration is not achieved, the transducing DNA is unable to replicate. Transduction is applied to the transfer of bacterial and eukaryotic genes in biotechnology.

- ▶ [Clinical Gene Transfer](#)
- ▶ [Transposons](#)

Transepidermal Water Loss

Definition

Transepidermal water loss refers to constitutive loss of water vapor from the skin surface in the absence of sweat gland activity. Commonly utilized as a measure of skin permeability barrier function.

- ▶ [Heritable Skin Disorders](#)

Trans-Epithelial Electrical Resistance

Definition

Trans-Epithelial Electrical Resistance (TER) is a measure of paracellular permeability to passive ion flow and is based on a simplified circuit model of epithelia. A current pulse of known amplitude is passed across the epithelium, and the corresponding trans-epithelial voltage deflection is measured. The TER is then calculated from Ohm's law and has units of Ohm. cm^2 . Alterations in TER are considered to be a sign of altered tight junction permeability.

- ▶ [Epithelial Cells](#)

Transesterification

Definition

Transesterification is an organic reaction by which an ester is transformed into another by the hydroxyl group.

- ▶ [Splicing](#)

Transfected Cell Array

Definition

Transfected cell array refers to a micro-array based system for the functional analysis in mammalian cells of many genes in parallel.

- ▶ [High-Throughput Approaches to the Analysis of Gene Function in Mammalian Cells](#)

Transfection

Definition

Transfection describes delivery of purified nucleic acid to eukaryotic cells if naked nucleic acid or a non-viral vector is used for delivery, e.g. electroporation, lipofection or Ca-phosphate precipitation.

- ▶ [Clinical Gene Transfer](#)
- ▶ [Functional Assays](#)
- ▶ [High-Throughput Approaches to the Analysis of Gene Function in Mammalian Cells](#)

Transfer RNA

- ▶ [RNA Polymerase III](#)

Transferrin

Definition

Transferrin is a plasma protein that binds ferric iron with high affinity and delivers it to cells.

- ▶ [Hemochromatosis](#)

Transferrin Receptor

Definition

Transferrin receptor is a protein that binds the transferrin-iron complex and removes it from the

plasma. Two transferrin receptors have been identified: transferrin receptor-1 and transferrin receptor-2.

► [Hemochromatosis](#)

Transformation

Definition

Transformation defines the introduction of foreign DNA into cells by a physical method. Methods are generally quite species and strainspecific, poorly understood, and performed according to empirical recipes. The most efficient method for bacteria, and also the best defined, is electroporation.

In the context of tumorigenesis, transformation is the collection of events that leads to the loss of normal cellular function and acquisition of tumorigenic properties (malignant transformation). Transformed cells become independent of growth factors usually needed for cell growth. Normal laboratory tests for cellular transformation are used as a model to predict *in vivo* carcinogenesis. Tests are, for instance, for the ability to grow without attachment to plastic surfaces, to grow in semisterile media (e.g. agar) and the ability to form tumours when injected into nude mice.

- [Microarrays in Pancreatic Cancer](#)
- [Recombinant Protein Expression in Bacteria](#)
- [Senescence](#)
- [YAC and PAC Maps](#)

Transforming Growth Factor- β

► [TGF](#)

Transforming Protein

Definition

Transforming protein is the product of an oncogene able to induce properties of tumour cells in a sensitive cellular background. Properties of transformed cells comprise of loss of density-dependent growth control, anchorage independence and tumorigenicity.

- [Ras Signalling](#)
- [Transformation](#)

Transgene

Definition

A transgene is a cDNA or an additional extra gene introduced into the host genome. It usually contains the coding sequence and regulatory regions of a gene. The exogenous DNA fragment can be delivered to cells by naked nucleic acid transfer of viral or non-viral vectors.

- [Clinical Gene Transfer](#)
- [Transgene Silencing](#)
- [Transgenic and Knockout Animals](#)

Transgene Expression

Definition

Transgene expression means that the transgene is transcriptionally active and expresses mRNA.

- [Transgene Silencing](#)

Transgene Silencing

ERIC MILOT¹, JAMES ELLIS²

¹Centre de Recherche Guy-Bernier, Hopital Maisonneuve-Rosemont, Montreal, QC, Canada

²Developmental Biology Program, Hospital for Sick Children, Toronto, ON, Canada

ericmilot.hmr@ssss.gouv.qc.ca,

jellis@sickkids.ca

Definition

Mammalian cells can be manipulated by introduction of a ► [transgene](#) delivered as DNA or by a virus vector. Transgenic mice are created by direct DNA micro-injection into a fertilized mouse egg. In contrast, gene transfer into human cells is best accomplished using other methods of DNA transfection or by infecting the cells with a virus vector. When the transgene must persist in the target cell for many generations, retrovirus vectors have the advantage of their natural ability to integrate into the host chromosome. Retrovirus vectors are made by replacing virus genes with the gene to be transferred and then packaging the recombinant

genome into virus particles. These vectors are particularly well suited for gene transfer into stem cells, which must divide extensively to repopulate tissues with mature cells. These properties make retrovirus vectors the method of choice for developing novel stem cell gene therapies. However, gene transfer is only effective if ►**transgene expression** can be controlled appropriately, and a transgene that is not transcribed will not be functional. Such ►**transgene silencing** is very common and is dependent on several factors. Of primary importance are position effects dictated by the local chromatin structure of the DNA that flanks the integrated transgene. In addition, the specific cis-acting DNA regulatory elements that control transgene expression and the concentration of trans-acting factors that recognize them in the cell are major players that influence whether a transgene will be silent or active. Finally, ►**silencer** sequences present in retrovirus vectors are recognized very effectively in stem cells and contribute to silencing of transgenes. In order to optimize transgene expression, it is essential that the mechanisms that control transgene silencing be elucidated. To overcome transgene silencing will require the use of strong positive regulatory elements that can open ►**chromatin structure** on the transgene or that block the spread of closed chromatin into the transgene.

Characteristics

Control of Transgene Expression

The pattern of transgene expression is defined by its cis-regulatory elements such as enhancers and promoters that are bound by transacting factors. However, the interaction of transacting factors with DNA sequences can be impaired by ►**epigenetic** mechanisms that influence the organization of the chromatin structure and patterns of DNA methylation. This level of transcriptional control allows the genetic material to be in an “active or inactive” state, which can be inherited over many cellular generations.

The basic unit of chromatin structure is the ►**nucleosome**, composed of an octamer of four different histone proteins. Nucleosomal histones have a dual function in activation or repression of gene expression. In a nucleosomal configuration, genomic DNA is only partially accessible to transacting factors. However, when DNA-histone interactions are relaxed, DNA becomes more accessible to transacting factors that can bind their target DNA sequence and influence gene expression. This active chromatin has a histone code characterized by histone acetylation, methylation at specific residues as well as other covalent modifications. The nucleosome organization can be condensed through the action of the linker histone H1 and other proteins into a compact organization, commonly referred to as ►**heterochromatin**, that prevents gene expression. Such condensed chromatin has a repressive

histone code in which the histone tails are deacetylated and methylated at specific residues that differ from those on active chromatin. When the nucleosome organization is “decondensed” in ►**euchromatin**, it is permissive for transcription.

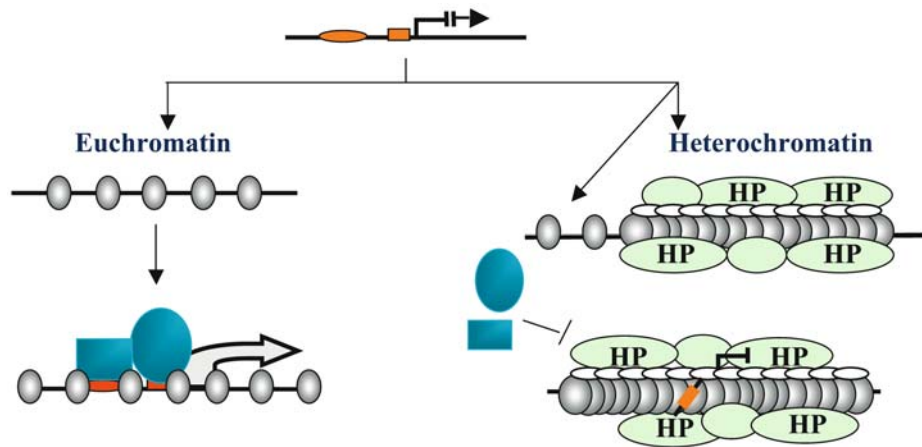
The organization of nucleosomes is regulated not only by the covalent histone modifications described above but also by remodeling complexes, replacement of core histones with histone variants and by certain transcription factors. The gene-specific recruitment of ATP-dependent chromatin remodeling and/or histone modifying complexes can be mediated by DNA interacting factors. Indeed, different DNA binding activators and repressors are capable of recruiting remodeling or modifying complexes at gene regulatory regions.

Position Effects and Transgene Silencing

Transgene integration into the genome is essentially random. They can integrate close or into a heterochromatic region thereby preventing transcriptional activation (Fig. 1). This can result in partial or complete transgene silencing, which are referred to as ►**position effects** (PE). PE frequently results from differences in the epigenetic organization of genomic regions and can be sub-categorized according to the pattern of transgene expression. If transgene silencing is complete in all cells of a particular tissue, it is referred to as stable PE. Stable PE is sometimes characterized by a cell-cycle-dependent regulation of transcription, the cell timing position effect (CTPE). CTPE is the result of a chromatin configuration that influences transcription in all cells of the tissue but only during a part of the cell-cycle. In contrast to stable PE, if transgene silencing is variable from one cell to another in the cellular population, it corresponds to an unstable PE and in particular to ►**position effect variegation** (PEV). PEV causes restricted activity of a gene in a subset of cells from a homogenous population. The silencing is randomly set and it can be inherited thereafter. The occurrence of PEV is consistent with cell-to-cell differences in the ability of heterochromatin to invade a particular transgene located in a nearby euchromatic region. This variable capacity to invade euchromatin appears to be dependent on the local concentration of heterochromatic (repressive) and euchromatic (activating) proteins. These two groups of proteins are presumed to be involved in a local competition to either repress or activate the chromatin of a transgene. A concentration change of repressive or activating proteins can result in modification of PEV (1).

Retrovirus Silencing in Stem Cells

Retrovirus vectors are subject to transgene silencing when integrated into preimplantation mouse embryos or after transfer into embryonic and adult stem cells such as hematopoietic stem cells. In more differentiated



Transgene Silencing. Figure 1 Transgene integration appears to occur randomly. If the transgene integrates close to a heterochromatic region, the heterochromatin can cover the transgene and prevent its transcriptional activation. In red are transgene regulatory elements and in blue are trans-acting factors including the transcriptional initiation complex. Grey circles are nucleosomes; in green are heterochromatin related proteins (HP) and in white are H1 histones, which are enriched in heterochromatic regions.

cell types they are highly expressed, suggesting that a gene silencing pathway active in stem cells recognizes the retrovirus DNA as foreign and either silences it or permits low level expression at best. Silencer elements have been discovered and then mutated in retrovirus vectors to increase their expression in stem cells, but even these modified vectors result in modest improvement in expression levels and can still be completely silenced. Retrovirus silencing is established within 2 days of infection into stem cells and at this stage can be reactivated by the histone deacetylase inhibitor trichostatin A, suggesting that a silencing pathway recognizes the retrovirus DNA and deacetylates histones shortly after nucleosomes form on the new provirus. Subsequently, DNA methylation is detectable on the silent provirus as well as the methyl-binding protein MeCP2 and inaccessible chromatin associated with linker histone H1 is established that completes the repressive histone code (2). These findings indicate that retrovirus silencing is a 2-step event with histone modifications occurring first prior to DNA methylation events. However, the DNA methylation step is not required for retrovirus silencing because ES cells that are null for *de novo* methylases silence retrovirus vectors with wild type efficiency.

The role of PE in retrovirus silencing is not precisely known, as retrovirus preferentially integrates into open chromatin in or near genes. However, the fact that the same retrovirus vector can be silent or low expressing at different integration sites in stem cells strongly supports a role for PE together with the retrovirus silencer elements in making the decision to be expressed or silent. The use of stronger promoters within the vector does direct increased expression

levels in stem cells, but never to the levels seen in differentiated cells. Ultimately to protect a transgene delivered by a retrovirus vector or as a DNA fragment from epigenetic silencing, the transgene has to be isolated from its host chromosomal environment. Two groups of elements are capable of preventing abnormal epigenetic regulation, ►locus control regions (LCR) and ►insulators.

Chromatin Organization and the LCR

LCR elements are involved in the expression of tissue-specific genes in mammals. The best characterized is the human β -globin LCR comprised of five Dnase I hypersensitive sites (HS). HS1 to 4 are fully formed only in erythroid cells and HS5 is found in hematopoietic cells. Erythroid specific and ubiquitous trans-acting factors bind these sites and render them hypersensitive to Dnase I by opening up the nucleosome organization locally. Indeed some factors that bind the β -globin LCR (at the different HS) can interact with histone modifying enzymes and others with chromatin remodeling complexes. The binding of these factors to each HS facilitates the formation of a LCR holocomplex that can interact with distant regulatory elements and promotes euchromatin organization within the locus. The LCR holocomplex increases transcriptional initiation at one globin promoter at a time by DNA looping. The LCR may also enhance transcriptional elongation through the gene. In transgenic mouse studies, LCR activity is revealed by expression at all transgene integration sites in a copy number dependent manner. A complete human β -globin LCR efficiently prevents PE in transgenic mice, but an LCR fragment that lacks one of the HS is still subject to PE.

Thus transgene expression regulated by an incomplete LCR can be influenced by the chromatin environment at the integration site (3).

LCRs are generally present in large DNA fragments, which is problematic for their use in retrovirus vectors to overcome transgene silencing. Therefore, different groups have “dissected” the LCR to define modules involved in chromatin activation. Since retrovirus vectors integrate at single copy, the most useful β -globin LCR element is HS3, which can activate the chromatin structure of single-copy transgenes in a chromosomal context. Small LCR cassettes that include HS2, HS3 and HS4 linked to a β -globin gene have been delivered using lentivirus vectors to correct hemoglobinopathies in mouse models. Despite this success, it appears that the lentivirus is also subject to transgene silencing as multiple integration events are required per cell to obtain therapeutic levels of β -globin gene expression. In fact, retrovirus or lentivirus DNA acts as a silencer of linked LCR β -globin transgenes in mice by imposing a repressive histone code and condensed chromatin on the transgene. Therefore, the LCR elements used in virus vectors are not capable of completely overcoming the combined effects of PE and the virus silencing pathways.

PE and Silencer Blocking by Insulator Elements

Insulators are defined as DNA sequences that protect a transgene from PE or from activation by nearby enhancers. The best-characterized vertebrate insulator is the cHS4 element located at the 5' end of the chicken β -globin locus (4). It marks a boundary between an open chromatin domain and a region of heterochromatin. This element insulates transgenes from PE and can block the ability of an enhancer to activate a promoter. Although enhancer-blocking activity by cHS4 is dependent on its CTCF binding site, this site plays no role in blocking PE. The cHS4 insulator has silencer-blocking activity when positioned between retrovirus sequences and an LCR β -globin gene in transgenic mice, showing that a combination of an insulator and an LCR can overcome retrovirus silencing. It is more challenging to incorporate functional insulators into infectious retrovirus vectors. The cHS4 element located in the LTRs of a retrovirus vector protects it from PE in mature cells, but has no effect on retrovirus silencing in ES cells. These results indicate again that PE alone cannot account for retrovirus silencing, and that effective retrovirus vectors must be designed with insulator elements flanking the transgene to protect it from both retrovirus silencers and PE. Unfortunately, it is not possible to transmit the same insulator sequence at two internal positions in a retrovirus vector. Hence, a properly insulated retrovirus vector awaits the discovery of another insulator with silencer-blocking activity to use in conjunction with cHS4 and LCR elements.

Clinical Relevance

Stem cell gene therapy using retrovirus vectors has corrected severe combined immunodeficiency in patients, but two patients have proceeded to develop leukemia caused by integration of the retrovirus vector into the *LMO-2* oncogene and activation of expression (5). This trial shows that stem cell gene therapy can cure genetic diseases but also highlights several drawbacks of retrovirus vector technology. First, any DNA regulatory element in the vector may activate neighboring genes with potentially disastrous complications. The use of enhancer blocking insulators in the LTRs and deletion of viral enhancers may serve to prevent such activation. Second, transgene silencing through PE and the action of retrovirus silencers diminishes the effectiveness of each provirus. The temptation to increase copy number to combat silencing merely increases the number of integrations and the risk of [insertional activation](#). The use of improved vectors with all silencers removed, combined with LCR elements and silencer-blocking insulators may ensure that all integration events are expressed. In this way, the number of integrations into stem cells can be minimized to prevent insertional activation. We now know that stem cell gene therapy can work in the clinic, but to make it safe with optimal expression of the transgene will require a recognition of the importance of transgene silencing in stem cells.

References

1. McMorro T, Den Wijngaard A, van Wollenschlaeger A et al (2000) Activation of the beta globin locus by transcription factors and chromatin modifiers. *EMBO J* 19:4986–4996
2. Pannell D, Osborne CS, Yao S et al (2000) Retrovirus vector silencing is de novo methylase independent and marked by a repressive histone code. *EMBO J* 19:5884–5894
3. Dillon N, Festenstein R (2002) Unravelling heterochromatin: competition between positive and negative factors regulates accessibility. *Trends Genet* 18:252–258
4. West AG, Gaszner M, Felsenfeld G (2002) Insulators: many functions, many mechanisms. *Genes Dev* 16:271–288
5. Hacein-Bey-Abina S, Von Kalle C, Schmidt M et al (2003) LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1. *Science* 302:415–419

Transgenesis

Definition

Transgenesis is the stable integration of a sequence of foreign DNA into a genome, such that transmission to

the progeny is directly coupled to the inheritance of this genome.

► *Medaka* as a Model Organism for Functional Genomics

Transgenic

Definition

A transgenic organism is an organism into which a foreign gene is introduced and successfully expressed, either in all cells, selected tissues or an additional, sometimes mutagenized, copy of an endogenous gene.

► Bone and Cartilage

► Cancer Invasion and Metastasis: Cellular, Molecular and Clinical Aspects

► Transgenic and Knockout Animals

► Wound Healing

Transgenic and Knock-out Animals

STEFAN BRITSCH

Max Delbrück Center for Molecular Medicine (MDC)
Berlin-Buch, Berlin, Germany
sbritsch@mdc-berlin.de

Synonyms

Genetically engineered animals

Definition

Animals into which additional and/or altered genetic information has been introduced are termed transgenic animals. Two principle technical approaches are used to generate transgenic animals: (i) addition of genetic information *via* microinjection of foreign DNA into the pronucleus of zygotes (ii) site-directed mutagenesis of endogenous target genes *via* ►homologous recombination in ►embryonic stem cells (►gene targeting). Transgenic organisms generated by pronuclear microinjection typically over-express the transgene, whereas animals generated by gene targeting are characterized by the deletion or modification of an endogenous target gene (►knock-out ►knock-in). So far, gene targeting is applicable only in mice. Pronuclear microinjection is used in a wide variety of different species.

Characteristics

Our understanding of the genetic mechanisms underlying normal development as well as the development of disease has been revolutionized by the availability of systematically generated transgenic animal models. This has allowed the study of individual genes by specifically altering their functions *in vivo*. The mouse has for a long time been the mammal of choice for experimental genetics and many of the pioneering transgenic techniques have been developed primarily in mice. Today, various transgenic mice are the predominant transgenic models in biomedical research (1).

Genetically engineered mice can be classified according to the type of mutation, its phenotypic consequences and the techniques by which it has been generated:

1. mice with a randomly integrated transgene (pronuclear microinjection)
 - gain-of-function mutation
 - loss-of-function mutation (dominant negative mutation; knockdown)
2. mice with site-directed mutation of an endogenous target gene (gene targeting)
 - mutation active in all tissues
 - deletion of target gene ►(knock-out)
 - modification of target gene
 - expression of novel gene under control of target gene locus (knock-in)
 - spatio-temporally restricted mutation (►conditional gene targeting)
 - conditional deletion of target gene (conditional knock-out)
 - conditional modification of target gene
 - conditional expression of novel gene under control of target gene locus (conditional knock-in)

Pronuclear Microinjection

Generation of transgenic mice by microinjection of DNA into the pronucleus of fertilized oocytes was first reported in 1981 by Gordon and co-workers (2). This method has become a simple and fast standard procedure to generate transgenic mice. In addition to the direct injection of DNA, alternative methods for gene transfer, i.e. retroviral infection, electroporation or calcium phosphate transfection have been described. After gene transfer, the zygotes are transferred into pseudopregnant female mice, which give birth to transgenic offspring after completion of embryogenesis. Some of the founder animals have stably integrated the transgene into their genome. Since integration occurs randomly, integration site(s) and number(s) of transgene copies are unpredictable. Therefore, screening of founder animals, which may

differ in transgene expression, has to be performed. The use of large DNA constructs derived from ►YACs, ►BACs or ►PACs have partially helped to circumvent these limitations. In 20–30% of cases integration of foreign DNA occurs after the first cell-cleavage, thereby creating chimeras, which express the transgene only in a subset of tissues (1).

Depending on the promoter elements used to control the transgene, pronuclear microinjection normally leads to a gain-of-function, i.e. over-expression of the transgene. By expressing a dominant-negative transgene it is possible to create the functional knockdown of an endogenous gene, although this is usually not 100% effective. Recently, transgenic expression of double-stranded RNA (RNA interference) has been shown to reduce host gene expression in mice efficiently (3). Transgene integration into the host genome sometimes interferes with endogenous gene expression and can lead to unintended additional phenotypes. Either by the specific ablation of a particular cell population *via* transgenic expression of a toxin gene or *via* expression of a reporter gene (see below), transgenic approaches have also become convenient genetic tools for cell fate analyses during embryogenesis (1).

Gene Targeting

Probably the most important limitation of the transgenic techniques described above is the fact that they do not allow the site-directed introduction of specific mutations into a host genome. Two ground-breaking discoveries made in the early 1980s have helped to overcome this limitation.

Oliver Smithies was the first to show that gene function could be directly disrupted in mammalian cells by homologous recombination, a process that had been well established for introduction of mutations into yeast cells (4). When foreign DNA homologous to host DNA is introduced into mammalian cells it can recombine (exchange) with homologous sequences of the host genome. Moreover, non-homologous sequences that disrupt critical regions of a host gene can be introduced by homologous recombination, when flanked by arms of homology.

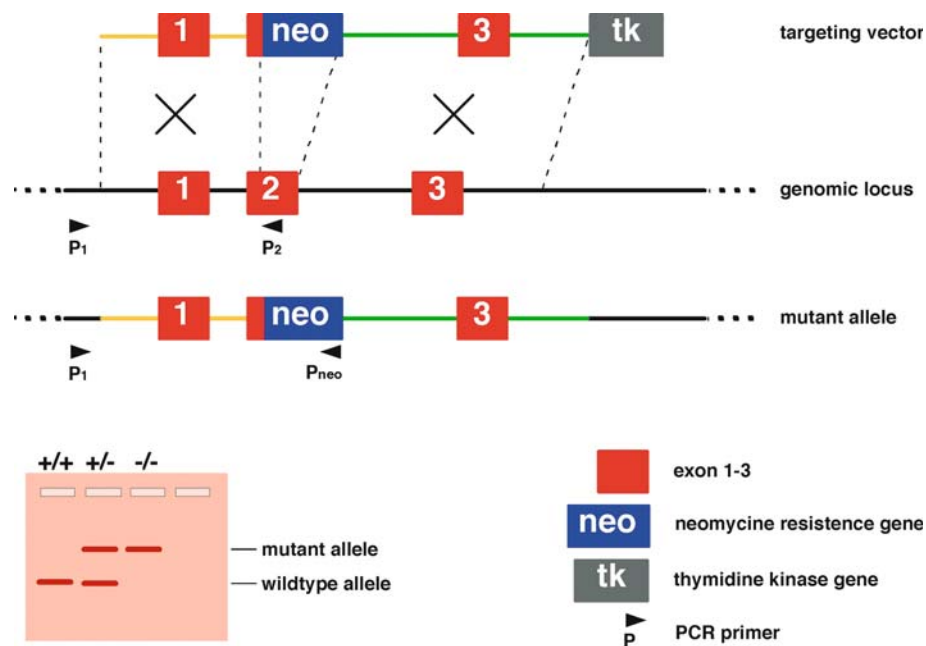
The second major discovery, made independently by two groups (5, 6) was the observation, that pluripotent, uncommitted, embryonic stem (ES) cell lines could be established from the inner cell mass of early mouse embryos. When returned into mouse embryos, ES-cells contribute efficiently to both somatic and germ cells. ES-cells can be maintained indefinitely as undifferentiated cells in culture. However, their differentiation state and germ-line competence depend critically on the culture conditions. So far, germ-line competent ES-cells have been only established in mice.

During the following years, several labs worked on combining the two techniques. In 1989 the first report of successful germ-line transmission of a mutation that had been introduced *via* homologous recombination into mouse ES-cells appeared (7). Gene targeting *via* homologous recombination in ES-cells has revolutionized mouse genetics; up to now, more than a thousand genes have been mutated in mice using gene targeting technology (for an excellent overview, see the mouse knock-out and mutation database, ►<http://research.bmn.com/mkmd>) and the number of publications related to this technology is still growing exponentially (8).

Already from the early studies it had become clear that integration of foreign DNA *via* homologous recombination in ES-cells occurs at very low frequencies and that a common competitive reaction is non-homologous integration at various sites in the host genome. Accordingly, strategies to select specifically for the rare event of homologous recombination had to be developed. This was achieved by using selection markers, for example the *neo* gene, which confers resistance to the antibiotic G418. When placed between homologous sequences the *neo* gene can disrupt a target gene and provide positive selection for cells that have taken up the DNA (Fig. 1). In order to enrich further for homologous recombination events, negative selection markers, like herpes simplex virus thymidine kinase (*tk*), are commonly used. Negative selection markers are placed outside the regions of homology so that inappropriate, non-homologous integration of the targeting construct can lead to expression of the negative selection marker (8).

Gene targeting is normally applied to generate a null-mutation that completely disrupts gene function in all tissues of an embryo throughout life. This strategy however can also be used to introduce novel genes into the disrupted gene locus, thereby creating a knock-in. Such genes could be either genetic markers, for example lacZ or ►GFP, which allow the identification of cells expressing the mutant allele and thereby the tracing of the fate of mutant cells during development, or other genes (homologues, members of gene families) that are then under the transcriptional control of the mutant locus.

Many genes have been shown to serve independent critical functions during normal embryonic development as well as in postnatal life. As a consequence, mice with null-mutations in such genes die during embryogenesis and it has been impossible to analyze additional biological functions in these knockouts. Novel conditional gene targeting strategies employ site-specific recombinase systems, for example the ►*Cre/loxP* system, to restrict mutations spatio-temporally. With this method critical regions of a target gene



Transgenic and Knock-out Animals. Figure 1 Schematic illustration of a general strategy to specifically mutate target genes via homologous recombination in ES-cells (*gene targeting*). After homologous recombination with a targeting vector, the second exon of a target gene is disrupted by a neomycin resistance gene (*neo*). The HSV thymidine kinase gene (*tk*) is lost upon homologous recombination. Dashed lines indicate regions of homology between the targeting vector and the genomic locus. Homologous recombination events can be identified by PCR screening.

are flanked in a first step by short palindromic oligonucleotide sequences, termed *loxP* sites. This modification is silent and does not interfere with normal target gene function. Mice homozygous for the floxed allele can be crossed to mice that express the recombinase *Cre* under the control of a tissue-specific or inducible promoter. Floxed DNA is specifically recognized by *Cre* and excised from the genome. Consequently, the silent mutation becomes activated only at the sites where and when *Cre* is expressed. Recently, alternative recombinases like the yeast derived Flp/FRT systems have been used in conditional gene targeting experiments (9).

Cellular and Molecular Regulation

When interpreting the phenotype of a transgenic animal, potential problems that may complicate or obscure the actual mutant phenotype have to be taken into consideration. Some of these, related to the random integration of a transgene, have been already discussed. Whenever a targeted mutation is inserted into internal exons, truncated protein may still be synthesized from the mutated gene. Residual protein may acquire novel functions and interact with unrelated proteins. Synthesis of residual protein can also be the consequence of aberrant splicing events, activation of cryptic promoters or read-through transcription followed by translation from downstream initiation codons. Production of

residual protein can be circumvented by deleting all coding exons of a target gene. However, this normally leads to the additional loss of intron sequences, which may contain regulatory elements controlling expression of unrelated, distant genes. Removal of introns may also delete so far unknown genes encoded by the opposite strand (for review, see 8). It is also well established that transcription from selection cassettes can interfere with the expression of neighboring genes. For that reason many labs remove selection markers routinely after homologous recombination.

Apart from side effects related to a particular gene locus or targeting strategy, a mutant phenotype can dramatically depend on the genetic background. This has been shown for the mutation of the EGF receptor; mutants die during early embryogenesis or can survive to birth, depending on the genetic background on which the mutation occurs. Finally, members of gene families may have redundant functions and upon deletion of individual family members related genes may compensate functionally (8).

Clinical Relevance

Transgenic animals are also of great importance for modeling fundamental disease mechanisms and for the evaluation of novel therapeutic strategies. Historically, transgenic over-expression of viral or cellular oncogenes, as well as the targeted deletion of prototype

tumor suppressor genes, like Rb or p53, in mice has been instrumental in the understanding of human cancer (10). Today, transgenic animal models are established tools for the analysis of major diseases, as for example neurodegenerative, psychiatric, infectious, cardiovascular, or even complex, polygenic metabolic diseases, like diabetes mellitus.

References

1. Nagy A et al (2003) Manipulating the Mouse Embryo. 3rd edn. Cold Spring Harbor Press, Cold Spring Harbor
2. Gordon JW, Ruddle FH (1981) Integration and stable germ line transmission of genes injected into mouse pronuclei. *Science* 214:1244–6
3. Kunath T et al (2003) Transgenic RNA interference in ES cell-derived embryos recapitulates a genetic null phenotype. *Nat Biotechnol* 21:559–61
4. Smithies O et al (1985) Insertion of DNA sequences into the human chromosomal beta-globin locus by homologous recombination. *Nature* 317:230–4
5. Evans MJ, Kaufman MH (1981) Establishment in culture of pluripotential cells from mouse embryos. *Nature* 292:154–6
6. Martin GR (1981) Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells. *Proc Natl Acad Sci USA* 78:7634–8
7. Thompson S et al (1989) Germ line transmission and expression of a corrected HPRT gene produced by gene targeting in embryonic stem cells. *Cell* 56:313–21
8. Muller U (1999) Ten years of gene targeting: targeted mouse mutants, from vector design to phenotype analysis. *Mech Dev* 82:3–21
9. Lewandoski M (2001) Conditional control of gene expression in the mouse. *Nat Rev Genet* 2:743–55
10. Jackson-Grusby L (2002) Modeling cancer in mice. *Oncogene* 21:5504–14

Internet Resources

- Mouse Genome informatics, MGI (► <http://www.informatics.jax.org/>)
- Nagy lab, Cre transgenic and floxed gene databases (► <http://www.mshri.on.ca/nagy/>)
- Mouse knock-out and mutation database (► <http://research.bmn.com/mkmd>)

Trans-Golgi Network

Definition

Proteins synthesized on ribosomes are transported through the Golgi apparatus and then reach the trans-Golgi network (TGN). In this cellular compartment, secreted and plasma membrane proteins are processed and sorted for delivery to various cellular destinations.

In most cells, the TGN has a reticular-like structure and is located in the proximity of the nucleus.

► [Diabetes Insipidus, a Water Homeostasis Disease](#)

Transition

Definition

In cases of non-silent base pair substitutions the original base pair is changed, e.g. from A to B. If A and B are both purines or pyrimidines the process is called a transition, in the case of changes from purine to pyrimidines, respectively, vice versa the process is called ► [transversion](#).

► [Sequence Annotation in Evolution](#)

► [Splicing](#)

Transition State

Definition

The transition state is the state on top of the free energy barrier that separates two states of a chemical reaction. In the context of protein folding the transition state is not one defined conformation, but an ensemble of compact conformations that are at a saddle point of free energy between the molten globule state and the native state. The transition state ensemble is characterized by many native-like contacts and an overall topology close to the native state.

► [Protein Folding](#)

Translation

Definition

Translation is the process during which the coding sequence of a gene that has been transcribed into the sequence of a messenger RNA is translated into the amino acid sequence of a protein (protein biosynthesis). This is done by the ribosome using specific adapter molecules to combine each amino acid with its corresponding codon (an RNA sequence consisting of three nucleotides), and to connect the amino acids that make up the protein.

► [Genetic Code](#)

► [Fragile X Syndrome](#)

- ▶ Full Length cDNA Sequencing
- ▶ Morpholino Oligonucleotides, Functional Genomics by Gene ‘Knock-down’
- ▶ Protein Databases
- ▶ Single-Cell Gene Expression Profiling: Cell-level Biology by Multiplexed Expression Fluorescence in Situ Hybridization
- ▶ Splicing
- ▶ Translational Control in Eukaryotes
- ▶ tRNA

Translation Elongation Factors

Definition

Translation elongation factors are protein factors that catalyze peptide bond formation and peptide translocation during the translation elongation step.

- ▶ [Translational Control in Eukaryotes](#)

Translation Initiation Factor

Definition

Translation initiation factors comprise of protein factors that catalyze the assembly of ribosomes at the translation start codon of the mRNA.

- ▶ [Translational Control in Eukaryotes](#)

Translation Start Codon

Definition

Polypeptide synthesis usually starts with the amino acid methionine, which is encoded by the AUG-triplet (start codon) nearest to the 5' end of the mRNA.

- ▶ [Translational Control in Eukaryotes](#)

Translational Bypassing

Definition

Translational bypassing describes a recoding event in which ribosomes suspend translation at a certain site,

and then resume translation downstream without decoding a block of intermediate nucleotides. It is frame independent, i.e. it may or may not cause a change of frame.

- ▶ [Genetic Code](#)
- ▶ [Translational Control in Eukaryotes](#)

Translational Control in Eukaryotes

MARTINA MUCKENTHALER¹, THOMAS PREISS²

¹Department of Hematology, Oncology and Immunology, University of Heidelberg, Heidelberg, Germany

²Victor Chang Cardiac Research Institute, Molecular Genetics Program, Darlinghurst/Sydney, Australia
martina.muckenthaler@med.uni-heidelberg.de
t.preiss@victorchang.unsw.edu.au

Definition

▶ **Translation** is a central step in the gene expression pathway and is defined as the cellular process by which the genetic information content of messenger RNA (mRNA) is read and “translated” into protein. It takes place on large ribonucleoprotein complexes called ▶ [ribosomes](#) (1). Translation is usually divided into an initiation, elongation and termination phase (2).

The central concept of translational control is that gene expression can be regulated by the efficiency of utilization of mRNA in specifying protein synthesis (3). Each phase of translation can be under the control of a number of physiological and pathological influences, however, the majority of known control mechanisms target the initiation phase. Translational control is critical for many fundamental biological processes, including cellular growth, embryonic development and the response to biological or environmental cues. Examples of translational control mechanisms with recognized clinical relevance range from cellular stress responses, through the mechanism of insulin action and on to the control of iron metabolism. Viruses can usurp the cellular translation apparatus to preferentially synthesize their own proteins. In this chapter we briefly describe the process of translation in eukaryotic cells and present a selection of known translational control mechanisms.

Characteristics

The ▶ [genetic code](#) is central to translation. Genetic information is written into the genes and transcribed into mRNA using a four letter code based on the

nucleotide bases adenine (A), guanine (G), cytosine (C) and thymine (T) or in the case of mRNA uracil (U). Sets of 3 adjacent bases on the mRNA (called codons) encode the 20 different amino acids that make up the polypeptide chains of proteins. In eukaryotes, polypeptide synthesis usually starts with the amino acid methionine that is encoded by the AUG-triplet nearest to the 5' end of the mRNA (►start codon). It stops when the translation machinery encounters one of 3 stop codons: UAG, UAA or UGA, which do not code for an amino acid. Matching the nucleotide triplets with the corresponding amino acids requires adapter molecules, the transfer RNAs (tRNAs). These are 75–80 nucleotide long RNA molecules that fold into a clover leaf-like structure. The anticodon arm of the tRNA carries a base-triplet that recognizes the codon on the mRNA *via* base pairing. The aminoacyl arm of the tRNA is covalently bound to the appropriate amino acid. Specific aminoacyl-tRNA synthetases ensure that each tRNA is only loaded with the correct amino acid. The ribosome selects aminoacylated tRNA according to the sequence of codons in the mRNA and catalyses the peptide bond formation between the growing peptide and the incoming amino acid. Eukaryotic ribosomes consist of more than 50 ribosomal proteins and 4 different ribosomal RNA (rRNA) molecules. There are 2 subunits, the 40S (Svedberg) and the 60S subunit (1, 2).

Translation is Generally Divided into 3 Phases

Translation initiation is the process that results in the recruitment of the ribosome to the mRNA (Fig. 1). Typically, it begins with the binding of the 40S ribosomal subunit and associated factors near the 5' end of the mRNA. This pre-initiation complex then moves along or 'scans' the mRNA until the AUG codon is reached. The complex then rearranges and the 60S ribosomal subunit joins to form an elongation-competent 80S ribosome. Binding of the 40S subunit is promoted by the 5' cap structure and the 3' poly(A) tail of the mRNA. The hetero-trimeric complex eIF (for eukaryotic ►initiation factor) 4F binds to the 5' cap structure. eIF4F consists of the cap-binding protein eIF4E, the ATP-dependent RNA helicase eIF4A and the multifunctional adapter protein eIF4G. eIF4G interacts not only with the other subunits of eIF4F but also binds ►the poly(A) binding protein (PABP). This mediates a pseudo-circularization of the mRNA, which is important for efficient mRNA translation (4). eIF4G also binds eIF3 and in this way attracts the 40S ribosomal subunit and an associated ternary complex, comprising the initiator-methionyl-tRNA (tRNA_{Met}), eIF2 and GTP, to the mRNA. Following scanning and AUG-codon recognition, 60S ribosomal subunit joining is catalyzed in a GTP-dependent manner by eIF2 and eIF5B (3, 5).

Translation elongation begins with an aminoacylated initiator tRNA in the peptidyl-tRNA (P) site of the ribosome. Aminoacylated tRNA with the appropriate codon-anticodon pairing binds to the aminoacyl-tRNA (A) site of the ribosome. Then, a peptide bond is formed and the resulting dipeptide is translocated from the A-site to the P-site. This leaves the A-site free to bind the next aminoacyl t-RNA and start another elongation cycle. These steps are catalyzed by the eukaryotic ►elongation factors (eEF) 1 and 2 (2).

Translation termination occurs when a ►stop codon is encountered in the A-site of the ribosome. A complex of eukaryotic ►release factors (eRF) 1 and 3 binds to the ribosome, and triggers the release of the finished polypeptide. Subsequently, the ribosome disassembles from the mRNA and dissociates into a 40S and 60S subunit (2).

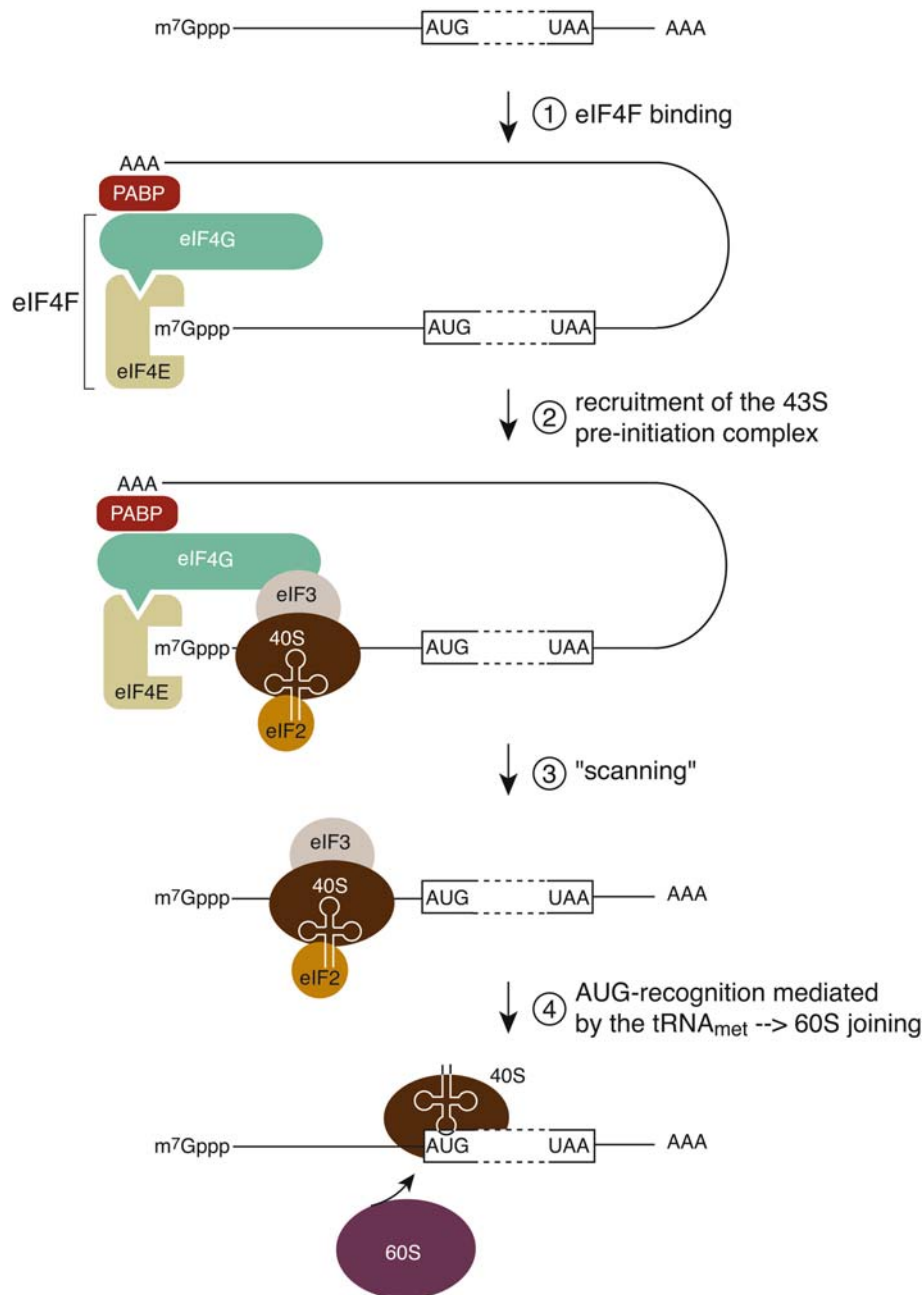
Mechanisms of Translational Control

Translational control mechanisms are often classed into two major categories, (i) global control that affects all or most cellular mRNAs in a similar manner and (ii) selective control that acts only on a subset of target mRNAs (3, 6).

Global Control of Translation

A common mechanism to stimulate or repress global translation is *via* phosphorylation of translation initiation factors (eIFs). The phosphorylation status of eIFs changes in a cell-cycle-dependent manner, responds to growth regulatory stimuli, and is affected by viral infections and cellular stress. The protein kinases and protein phosphatases that are responsible for the reversible phosphorylation are predominantly regulated *via* 2 major signal transduction cascades. The Ras/MAPK pathway regulates phosphorylation of the cap binding protein eIF4E and the PI3-K/Akt/mTOR pathway controls phosphorylation of the ribosomal protein S6, eIF4G and the 4E-binding proteins (4E-BP).

Cancer. Phosphorylation of translation initiation factors is commonly observed during malignant transformation of cells. ►Oncogenes like c-myc, ras or viral oncogenes can regulate the translation machinery. For example, expression of eIF4E and the α -subunit of eIF2 (see below) is increased in c-myc-transformed cells. eIF4E is itself considered an oncogene. The over-expression of eIF4E results in the malignant transformation of an immortalized cell line and further is able to activate the ras oncogene. In addition, eIF4E is involved in the inhibition of apoptosis, and increased levels of eIF4E are observed in a wide variety of tumors. In addition to changes in eIF4E, elevated levels of eIF4G in lung carcinomas, of eIF4A in melanomas and of eEF1A in tumors of the pancreas, colon, breast and lung are frequently



Translational Control in Eukaryotes. Figure 1 Initiation of translation Depicted is a typical eukaryotic mRNA with the post-transcriptional end modifications, a 5' cap structure and a 3' poly(A) tail. The protein-coding region is marked by start and stop codons. First, the eIF4F complex consisting of eIF4E, eIF4G, and eIF4A binds to the cap structure. The interaction between PABP and eIF4G leads to a pseudo-circularisation of the mRNA. The small ribosomal subunit (40S) is then recruited to the mRNA together with the initiation factors eIF3, eIF2 and the initiator-tRNA_{met}. This so-called 43S-preinitiation complex then moves along the mRNA in a process termed 'scanning'. The codon/anticodon interaction identifies the AUG start codon. This leads to the release of initiation factors and joining of the large (60S) ribosomal subunit. The formation of the complete 80S ribosome completes the initiation process and polypeptide synthesis as directed by the open reading frame can begin.

measured. How could an increase in translation factor activity result in the regulation of cell proliferation? A common hypothesis posits that mRNAs encoding growth factors, cellular receptors and tyrosine kinases benefit disproportionately from the increased eIF availability as they often exhibit highly structured 5' untranslated regions (UTRs). Such mRNAs are translated relatively inefficiently in the context of a limited availability of translation components.

eIF2 Phosphorylation. Viral infection, iron- or amino acid-deficiency and other cellular stress conditions inhibit global translational activity by stimulating phosphorylation of the heterotrimeric eIF2 on the α -subunit. This rapidly reduces the level of functional eIF2, by preventing the recycling of the GDP- to the GTP-bound form. The responsible kinases show great homology in their catalytic domains, but differ significantly in their regulatory domains.

During viral infections the presence of the viral genomic RNA activates the double-stranded RNA (dsRNA) dependent eIF2 α kinase (PKR). A kinase with homology to PKR (termed PERK) was recently isolated from the endoplasmic reticulum (ER). It is responsible for the shutdown of protein synthesis when unfolded proteins accumulate in the ER. Mutations have been identified in human PERK in 2 families with a rare, autosomal-recessive disease, the Wolcott-Rallison-syndrome. Mutations in PERK block beta-cell development and impair gluconeogenesis, consistent with the disease symptoms of permanent, neonatal insulin-dependent diabetes.

The heme-regulated eIF2 α kinase (HRI) is important for the iron-mediated regulation of translation in erythroid precursor cells. If iron is not sufficient for the synthesis of heme and hemoglobin HRI is activated to phosphorylate eIF α . Consequently, translation is inhibited and erythroid differentiation is blocked, resulting in microcytic, hypochromic erythrocytes. In iron-replete cells, HRI is bound to heme, which results in inhibition of its kinase activity (3).

The "Heat Shock Response". Elevated temperature, as well as a variety of other cell stresses such as exposure to heavy metals, hypoxia and lack of glucose can elicit a "[▶heat shock response](#)". A general inhibition of translation and induction of heat shock protein (hsp) expression is characteristic of the heat shock response. Heat shock proteins can prevent cell death. Many of the hsp family members are molecular chaperones, which are involved in protein folding and transport and the assembly of multiprotein complexes. During cell stress they protect proteins by being involved in protein repair and/or degradation of destroyed proteins through the ubiquitin-proteasome pathway. Global protein synthesis is shut down *via* at least 3 mechanisms, (i) phosphorylation of eIF2 α , (ii) inactivation of the eIF4F complex *via*

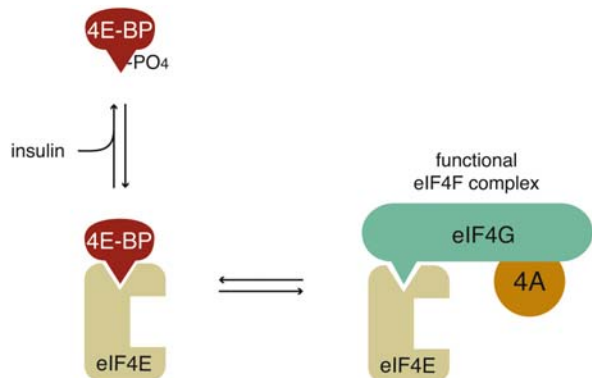
dephosphorylation of eIF4E and hypophosphorylation of 4E-BP and (iii) the unfolding of eIF4G by the hsp27/hsp70 complex, which results in the disruption of its interactions with PABP, the eIF4E-kinase Mnk1, eIF4A and eIF3. To date it is not clear how the heat shock protein mRNAs are translated under these conditions (3).

Insulin-Mediated Control. The hormone insulin increases protein synthesis in target cells. This effect is achieved by signalling through the PI3K/Akt/mTOR pathway. During active translation initiation, the cap-binding protein eIF4E binds eIF4G as part of the eIF4F complex (Fig. 1). This interaction is targeted by the 4E-BPs, which structurally mimic the eIF4E-binding site of eIF4G (Fig. 2). Hypophosphorylated 4E-BPs bind to eIF4E with high affinity and thus block eIF4F assembly and translation.

Cardiac Hypertrophy. Cardiac hypertrophy is a well-known response to increased hemodynamic load. This clinically important condition is primarily characterized by inappropriate stimulation of protein accumulation, leading to increased cell size, which underlies cardiac hypertrophy. There is mounting evidence that signal-induced phosphorylation of several translation factors mediates these pathological alterations of cellular protein synthesis.

Selective Control of Translation

mRNA-specific translational control is usually implemented by targeting ligands to regulatory elements in the 5' or 3' UTR of target mRNAs (3, 7). The best understood examples in this category of translational control mechanisms are found in the regulation of iron



Translational Control in Eukaryotes. Figure 2

Translational control by molecular mimicry Extracellular stimuli such as insulin can regulate the phosphorylation status of the 4E-BP proteins. The 4E-BPs mimic the eIF4E-binding region of eIF4G. Hypophosphorylated 4E-BP can bind to eIF4E and displace eIF4G.

metabolism, during cell growth and differentiation, and embryogenesis (3).

5'UTR Mediated Translation Control. The paradigm for 5'UTR mediated translation control is the regulation of the intracellular iron storage protein ferritin (3) (Fig. 3). Increased intracellular iron levels result in elevated ferritin expression while a decrease in iron diminishes ferritin synthesis. This regulation is mediated by a secondary structure in the 5'UTR of the ferritin mRNA, the iron responsive element (IRE). If the iron content in the cell is low, the iron regulatory proteins 1 or 2 (IRP-1 or IRP-2) bind to the IRE. This RNA-protein-interaction blocks translation. It was found that an early step of translation initiation, the binding of the 43S pre initiation complex to the cap-binding complex eIF4F is inhibited. If the cellular iron content rises, the IRP protein dissociates from the IRE and ferritin protein is synthesized. Mutations in the IRE, which block IRP-binding and therefore translational control explain the molecular defect underlying the ►[hyperferritinemia cataract syndrome](#). This disease is characterized by increased serum ferritin levels and the early formation of cataracts. In addition to iron storage, the IRE/IRP system also regulates cellular iron import and utilization. Translation of the enzyme eALAS, which catalyzes the first step of the erythroid heme biosynthesis, is regulated in a manner similar to the translation of ferritin. In contrast, the mRNA encoding transferrin receptor (Tfr) contains 5 IRE elements in its 3' UTR that regulate its stability in an iron-dependent fashion.

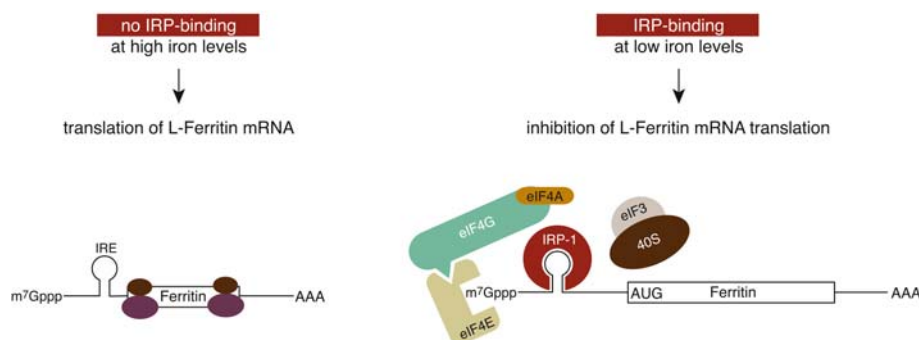
3'UTR Mediated Translation Control. A precise temporal and spatial control of gene expression is particularly important during the first hours of embryogenesis. Yet in many animals this developmental phase is characterized by a distinct absence of transcription. Many important developmental decisions that will shape the originating embryo rely solely on the precisely regulated expression of maternal mRNA

stockpiles that the oocyte has contributed to the fertilized egg. Research into these phenomena is mainly based on the use of model organisms like the worm *Caenorhabditis elegans*, the fruit fly *Drosophila melanogaster* or the mouse. Errors in the post-transcriptional regulation of maternal mRNA expression usually result in severe defects of embryonic development. Of central importance for the regulatory mechanisms involved in embryonic development are defined elements in the 3'UTRs of maternal mRNAs. These elements interact with proteins and play a role in transcript localization, stability and translational control. A hallmark of developmental control of translation is its frequent association with changes in the poly(A) tail length of the targeted mRNAs such that a short tail correlates with translational silencing and a long tail with translational activation (8).

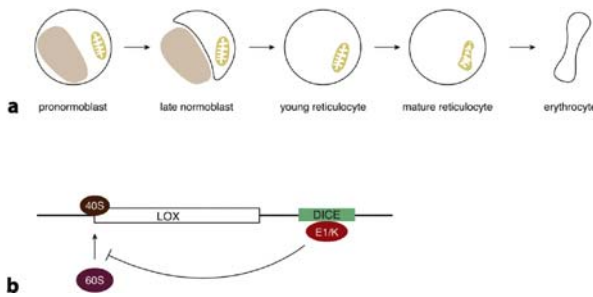
Translational regulation mediated by the 3'UTR is also observed during later differentiation events and even in somatic cells containing a nucleus (3). A mechanistically well-understood example of 3'UTR mediated translation control is the regulation of 15-lipoxygenase (LOX) translation during red blood cell differentiation (Fig. 4). LOX enzyme is involved in the destruction of internal membranes and mitochondria, a process that is important when reticulocytes mature to erythrocytes. *LOX* mRNA is transcribed during the early development of the reticulocytes – before the cell nucleus is expelled – and is then stored in the cytoplasm in an inactive form. The 3'UTR of the rabbit *LOX* mRNA contains a differentiation control element (DICE) that mediates translation repression *via* binding of the heteronuclear ribonuclear proteins (hnRNP) K and E1. This RNA/protein complex inhibits translation initiation at the 60S ribosomal subunit-joining step.

Viral Translation Control Strategies and IRES Elements

Many viral mRNAs contain ►[internal ribosomal entry sites](#) (IRES). These are RNA structures that can recruit



Translational Control in Eukaryotes. Figure 3 Translational control of ferritin synthesis. The 5' UTR of ferritin mRNA harbours a regulatory element termed IRE. When intracellular iron levels are low, the IRP proteins can bind to the IRE. This inhibits translation initiation at the level of 43S pre-initiation complex recruitment.



Translational Control in Eukaryotes.

Figure 4 Translational control of 15-lipoxygenase synthesis (LOX). (a) Schematic representation of erythropoiesis. During differentiation of red blood cells the nucleus is excluded. All mRNA molecules required for subsequent maturation steps have to be synthesised by this point. However, the mRNA encoding LOX should only be translated in mature reticulocytes when the enzyme is required for degradation of mitochondria. (b) Translation of LOX mRNA is controlled by the DICE element in the 3' UTR. hnRNP K and E1 bind to this element and inhibit the 60S ribosomal subunit joining step of translation initiation.

translation initiation complexes and ribosomal subunits to internal positions of the mRNA 5' end (4). IRES elements are functional substitutes for the 5' cap structure. The hepatitis C virus IRES can initially recruit the 40S ribosomal subunit without any requirement for translation initiation factors. However, eIF2 and eIF3 are needed for start codon recognition. Picorna viruses (hepatitis A virus, cardio- and aptho-viruses, entero- and rhinoviruses) also recruit ribosomes *via* IRES elements on their mRNAs. With the exception of hepatitis A- and the cardioviruses, they proteolytically cleave eIF4G as part of their infection strategy. This serves two purposes; it inhibits cellular translation and promotes viral translation. To recruit ribosomes to the IRES, these viruses require eIF-2, -3 and -4A as well as the C-terminal cleavage fragment of eIF4G (3, 5). In addition to viruses, IRES elements are also recognized in a growing list of cellular mRNAs. In some cases it has been documented that the translation of these mRNAs is adapted to specific cellular conditions. For example, the IRES elements of ornithine decarboxylase and *p58PITSLRE* mRNAs are only active at the transition from the G₂- to M-phases of the cell-cycle.

Another viral translation strategy is adopted by the rotaviruses, which are a main cause of diarrhea in children. Their mRNA molecules contain a cap structure but no poly(A) tail. Instead of a poly(A) tail, these mRNAs end with a short conserved sequence which is bound by the retroviral nonstructural-protein (NSP) 3. NSP3 binds to the adapter protein eIF4G and

competes for its interaction with the poly(A) binding protein (PABP). This again serves two functions, (i) it selectively inhibits the cellular, PABP-dependent translation and (ii) it stimulates viral translation, because NSP3 and eIF4G form a bridge between the 5' and 3' end of the rotavirus mRNA. A similar strategy is most likely used by the S-phase specific **▶ histone** mRNAs in Metazoa. These end with a conserved hairpin structure, which can take over many functions of the poly(A) tail. This structure is essential for histone-mRNA-translation and binds a protein, the stem loop binding protein (SLBP).

Clinical Relevance

The interactions of complex organisms with a continuously changing environment require the precise control of gene expression. The regulation of mRNA translation is an important part in nature's toolbox to achieve this control. As a result of intensive studies in basic science as well as molecular medicine, processes like the activation of target cells by insulin and links between translation and (malignant) cell growth are beginning to be understood at the molecular level. Increasing numbers of RNA elements are being identified that mediate a post-transcriptional control of gene expression and in some cases disease-causing mutations have been identified that disrupt their function. The study of the multiple viral strategies to utilize the cellular translation machinery yields an increased understanding of viral pathogenesis. Understanding the intricacies of translational control and its dysregulation in disease holds the promise of new therapeutic approaches.

References

1. Moore PB, Steitz TA (2002) The involvement of RNA in ribosome function. *Nature* 418:229–235
2. Ramakrishnan V (2002) Ribosome structure and the mechanism of translation. *Cell* 108:557–572
3. Sonenberg N, Hershey JWB, Mathews MB (eds) (2000) *Translational Control of Gene Expression*. Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York
4. Sachs AB, Sarnow P, Hentze MW (1997) Starting at the beginning, middle, and end: translation initiation in eukaryotes. *Cell* 89:831–838
5. Pestova TV et al (2001) Molecular mechanisms of translation initiation in eukaryotes. *Proc Natl Acad Sci U S A* 98:7029–7036
6. Dever TE (2002) Gene-specific regulation by general translation factors. *Cell* 108:545–556
7. Conne B, Stutz A, Vassalli JD (2000) The 3' untranslated region of messenger RNA: A molecular 'hotspot' for pathology? *Nat Med* 6:637–641
8. Mendez R, Richter JD (2001) Translational control by CPEB: a means to the end. *Nat Rev Mol Cell Biol* 2:521–529

Translational Frameshifting, Non-Standard Reading of the Genetic Code

PHILIP FARABAUGH

Department of Biological Sciences, University of Maryland Baltimore County, Baltimore, Maryland, USA
farabaug@umbc.edu

Synonyms

Programmed translational frameshifting; recoding

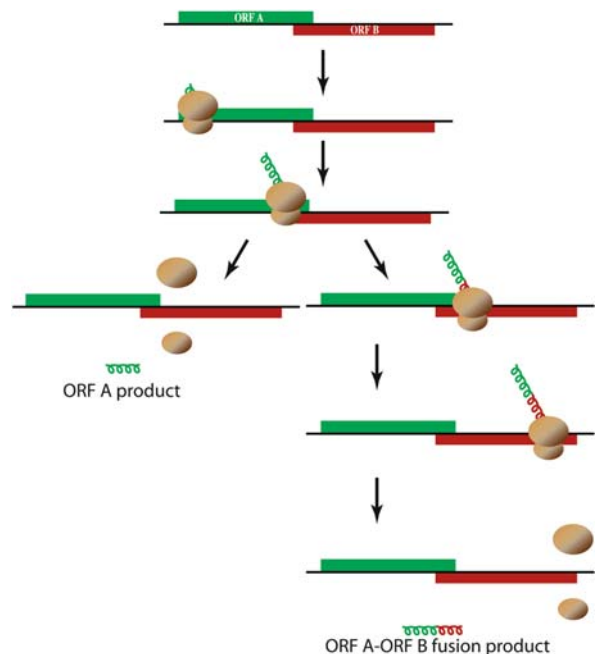
Definition

Translational frameshifting is a mechanism by which the translational machinery shifts the frame in which it decodes the mRNA while continuing to extend the polypeptide chain (Fig. 1). The result is that the mRNA does not encode the protein by a continuous run of three nucleotide codons, known as an **open reading frame** (ORF). Rather, the information encoding the protein comes from two distinct ORFs. Frameshifting is a stochastic process, meaning that each translating ribosome has a certain probability of undergoing the shift, but that only a fraction of the ribosomes does so. Ribosomes that frameshift produce a translational fusion of the two overlapping ORFs, whereas those that do not frameshift continue normal in-frame decoding and terminate at the end of the first ORF. Each of these products thus shares a common N-terminal region but has a different C-terminal domain. The frameshift occurs at special mRNA sequences that predispose the ribosome to shift reading frames. The phenomenology of frameshifting is quite complex; there are several general types of frameshifts depending on the direction and the distance of the shift and many distinct signals can induce the ribosome to shift (e.g. +1 or -1 frameshifting). Frameshifting is ubiquitous but rare, having evolved in a small number of genes in all three domains of life. Frameshifting can have a variety of purposes including allowing expression of alternative forms of proteins required for morphogenesis (especially in viruses), providing a mechanism for **autogenous translational control** or producing related proteins with distinct enzymatic characteristics.

Characteristics

The direction of the shift in frames distinguishes the two major classes of programmed frameshifts.

1. **Programmed -1 frameshifting** occurs when the ribosome shifts the reading frame by one nucleotide



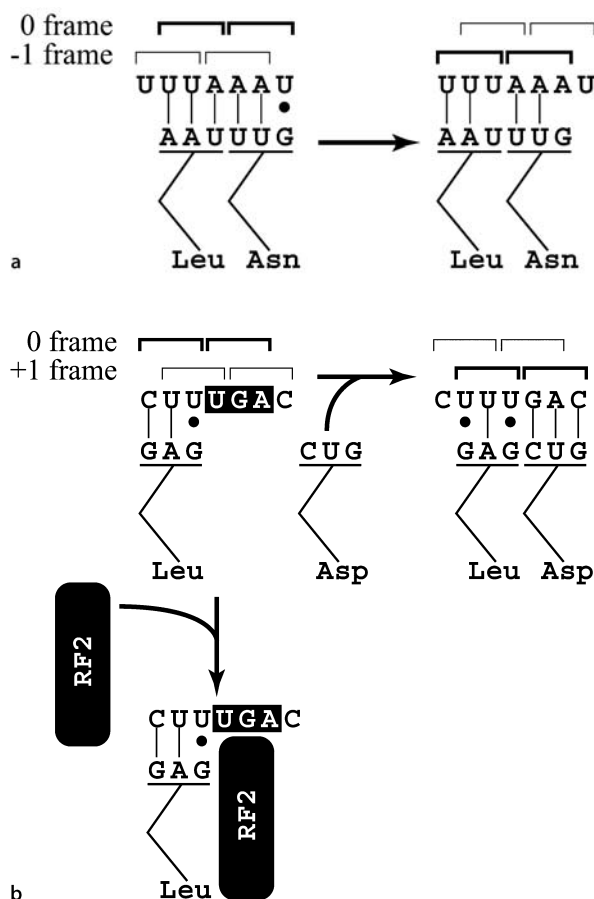
Translational Frameshifting, Non-Standard Reading of the Genetic Code. Figure 1

A gene consisting of two overlapping open reading frames (ORF A and ORF B) employs frameshifting to express two primary protein products. A ribosome is shown initiating on the upstream ORF, translating to the region of overlap between the two ORFs and undergoing two alternative events. On the left, ribosomes that do not frameshift complete expression of the protein specified by ORF A. On the right, ribosomes that frameshift express a protein that consists of an N-terminal region encoded by ORF A and a C-terminal region encoded by ORF B.

in the 5' or upstream direction. This is by far the most prevalent form of frameshifting, commonly occurring in many viruses and bacteriophages and transposons (Fig. 2A).

2. **Programmed +1 frameshifting**, which occurs when the ribosome slips one nucleotide in the 3' or downstream direction. These events are much less common, occurring in a small number of transposons and chromosomal genes (Fig. 2B).

All frameshift events share a few common characteristics. All frameshift sites include an essential sequence motif at the site of frameshifting though the general nature of that motif differs between the two major classes of frameshifts. Frameshift sites also include auxiliary non-essential sites that stimulate the frequency of frameshifting. These **frameshift stimulator sequences** also vary greatly among programmed frameshifts; for most frameshift sites this essential motif is a heptamer involving two codons and a flanking seventh base. Many frameshift signals also



Translational Frameshifting, Non-Standard Reading of the Genetic Code. Figure 2 Two types of frameshifting. (a) Programmed -1 frameshifting (here on the Rous sarcoma virus frameshift site) occurs when two ribosome-bound mRNAs (tRNA^{Leu} and tRNA^{Asn}) bound at the sequence UUA-AAU slip simultaneously one nucleotide in the 5' direction to the sequence UUU-AAA before continuing protein synthesis. (b) Programmed +1 frameshifting (here from the *prfB* gene of *E. coli*) occurs when the peptidyl-tRNA is bound to the ribosome (shown as tRNA^{Leu}). The ribosome can undergo two fates. Recognition of the in frame UGA codon by peptide release factor 2 (RF2) causes peptide release. Alternatively, slippage of tRNA^{Leu} one nucleotide in the 3' direction allows decoding by tRNA^{Asp} resulting in frameshifting.

include stimulatory secondary structures, a ▶**stem loop** or a ▶**pseudoknot** (Fig. 3). In other cases, the primary sequence stimulates frameshifting without forming any secondary structure. The stimulatory sequences affect frameshifting either by increasing the probability of the frameshift event or by decreasing the rate of the competing normal in-frame decoding event. Frameshifts only occur efficiently at these specialized sequences because at other sites normal translational

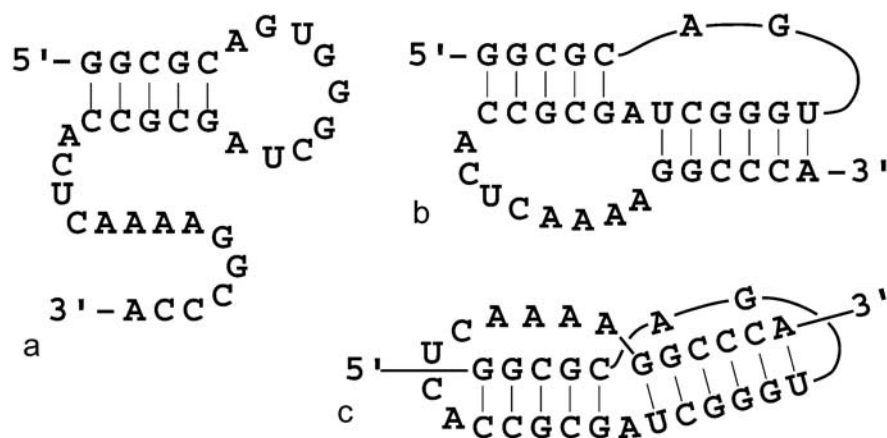
elongation occurs too rapidly to allow the ribosome to shift. Therefore, a ▶**translational pause** at the frameshift site is a universal feature of programmed events and some frameshift stimulator sequences function at least partly by increasing the duration of the pause, providing time for the kinetically slow shift in frames to occur.

Programmed -1 Frameshifting

The first described programmed translational frameshift sites were from metazoan ▶**retroviruses**. The canonical example derives from Rous sarcoma virus. That virus encodes four genes, *gag*, *pol*, *env* and the oncogene *v-src*. Transcription of the chromosomally integrated virus produces an RNA that is spliced to produce the mRNAs for the Env and Src proteins. Translation of *gag* and *pol* produces a Gag protein and much smaller amounts of a Gag-Pol translational fusion protein. Expression of the fusion requires a ribosome reading an unspliced mRNA to shift from the *gag* gene in the original reading frame (the ▶**zero frame**) into the overlapping *pol* gene in the ▶**-1 frame**, producing a fusion protein containing both *gag* and *pol* information. Fusion of the Pol protein to the Gag protein results in the packaging of the Pol protein into the nascent virus ▶**nucleocapsid**, which is important because the virus propagation requires the enzyme activities of the Pol protein to function within the particle.

The site of programmed -1 frameshifting is termed a ▶**slippery heptamer** because it is a seven-nucleotide sequence that allows slippage of ribosome-bound tRNAs from the zero frame into the -1 frame. Almost every slippery heptamer conforms to a common general structure, X-XXY-YYZ (shown as triplets in the upstream zero frame) where XXX is a triplet of any nucleotide, YYY is a UUU or AAA triplet and Z can be any nucleotide, but is species specific (Table 1). A ribosome reading the mRNA encounters the codons XXY and YYZ in the zero frame. Frameshifting occurs at the site by the ▶**simultaneous slippage** of the two tRNAs bound to these codons onto the overlapping XXX and YYY codons in the -1 frame followed by continued translation on the codon beginning at Z (Fig. 2a).

Frameshift efficiency varies greatly among the possible X-XXY-YYZ slippery heptamers. In general, AU-rich sequences tend to induce more frameshifting than GC-rich sequences. Slippage involves transient disruption of base pairing between the two tRNAs and the mRNA as they slip from the zero to the -1 reading frame. The preference for AU-richness presumably reflects the higher energy cost of disrupting GC base pairs. The correlation is not absolute; a few GC-rich heptamers strongly stimulate frameshifting and particular AU-rich heptamers stimulate very poorly, suggesting that the details of codon anticodon interaction may influence



Translational Frameshifting, Non-Standard Reading of the Genetic Code. Figure 3 Stem loops and pseudoknots are two forms of RNA secondary structures. (a) a stem loop forms by base pairing of adjacent, complementary sequences. (b) A pseudoknot forms when a downstream sequence in the RNA can base pair with a portion of the loop region. (c) A more accurate depiction of the structure of a frameshift-stimulating pseudoknot shows that the two paired regions stack on each other with a slight bend.

Translational Frameshifting, Non-Standard Reading of the Genetic Code. Table 1 Slippery heptamers that stimulate programmed -1 frameshift

Source*	Slippery heptamer
HIV-1	U-UUU-UUA
MMTV	A-AAA-AAC
17.6	A-AAU-UUU
L-A	G-GGU-UUA
dnaX	A-AAA-AAG

Symbols: HIV-1, human immunodeficiency virus-1; MMTV, mouse mammary tumor virus; 17.6, retrotransposon 17.6 of *Drosophila melanogaster*; L-A, L-A endogenous double-stranded RNA virus of *Saccharomyces cerevisiae*; dnaX, gene encoding the tau and gamma subunits of *Escherichia coli* DNA polymerase III.

the probability of slippage. This conclusion is underscored by the fact that mutations that block certain tRNA modifications can affect the efficiency of programmed -1 frameshifting. These modifications presumably function to modulate the stability of tRNA mRNA pairing and thus affecting the probability of slippage.

In almost all cases, programmed -1 frameshift sites consist of a slippery heptamer and a downstream secondary structure. That structure, usually a pseudoknot in eukaryotes and a stem loop in prokaryotes, causes a translational pause with the ribosome decoding at the slippery heptamer. However, for a pseudoknot the duration of pausing correlates poorly with the

frequency of frameshifting, suggesting that the secondary structure has an effect beyond causing pausing, though the nature of the effect remains unknown. A frameshift-inducing pseudoknot has distinctive structural features (e.g. a bend induced by a nucleotide wedged between the two stacked helices) suggesting that their 3-dimensional structure may be important to stimulate frameshifting, possibly through an interaction with some structure in the ribosome. The structural requirements of the stem loops in prokaryotes are less well understood.

Prokaryotic -1 frameshift sites usually include a third element, an upstream ▶Shine-Dalgarno site (SD-site). During initiation, base pairing between an SD-site and a complementary sequence at the 3' end of 16S rRNA in the 30S ribosomal subunit directs initiation to correct translation initiation codons. The same base pairing interaction stimulates programmed -1 frameshifting. The spacing between an SD-site and a slippery heptamer is essential and must be shorter than the spacing at initiation. The belief is that the short spacing strains the ribosome, forcing slippage of the mRNA relative to the decoding tRNAs and thereby increasing the probability of programmed -1 frameshifting.

Programmed +1 Frameshifting

Programmed +1 frameshift was first described in the gene encoding ▶peptide release factor 2 (RF2) in the bacterium *Escherichia coli* (prfB) and in ▶Ty retrotransposons in the yeast *Saccharomyces cerevisiae* (Ty elements). These frameshifts are far less common than -1 frameshifts but have been found in metazoans, fungi and bacteria. Frameshifting also requires an essential heptameric sequence but frameshifting occurs when

only one tRNA is paired with the mRNA. The site consists of two codons of the zero frame and a downstream seventh base (Fig. 2B). Frameshifting occurs after the ribosome has decoded the first codon and during a pause caused by slow recognition of the second codon. At the time of the frameshift, the first codon is in the ►ribosomal P site bound by a peptidyl-tRNA and the second codon is in the ►ribosomal A site. Slow recognition of the A site codon allows sufficient time for the mRNA to slip +1 with respect to the P site peptidyl-tRNA, which reveals a +1 shifted codon in the ribosomal A site recognized efficiently by an abundant aminoacyl-tRNA. The slippage and recognition of the first ►+1 frame codon appear to be a concerted event rather than successive.

The *prfB* frameshift is an example of autogenous translational control. The frameshift occurs after decoding of the 27th codon in the gene, with translation continuing for 340 amino acids in the shifted frame. The essential heptamer, CUU-UGA-C, includes the UGA nonsense codon recognized by the *prfB* product during peptide termination. When the RF2 concentration is high, translation terminates efficiently at the UGA, but when RF2 concentration is too low termination is slowed, allowing for stochastic slippage of the peptidyl-tRNA^{Leu} from CUU to UUU in the +1 frame. The autogenous interaction between RF2 activity and expression maintains it at the appropriate intracellular concentration. A similar feedback regulation of the ornithine decarboxylase (ODCase) antizyme gene in metazoans involves +1 frameshifting. ODCase catalyzes the first and rate-limiting step in the synthesis of the polyamines spermidine and spermine and ODCase antizyme targets ODCase for degradation. Expression of ODCase antizyme requires +1 frameshifting and the efficiency of that frameshift increases when the concentration of polyamines increases. As with *prfB*, the effect is to moderate increases or decreases in polyamine concentration. Programmed +1 frameshifting in the Ty retrotransposons accomplishes the same end as programmed –1 frameshifting in metazoan retroviruses. Expression of a Gag-Pol fusion protein requires +1 frameshifting and the fused protein assembles into a nucleocapsid structure whose function depends on the proper level of frameshifting.

Frameshifting in Chromosomal Genes and Its Effect on Genomics

Though the majority of characterized frameshifting systems involve viruses or transposable elements, a small but significant number of chromosomal genes also use the mechanism. Partly, the prevalence of ►extrachromosomal elements results from the frequent use of frameshifting in their compact genomes. It may be that chromosomal genes are less constrained to

adopt these unusual mechanisms or it may be that researchers are less adept at identifying adjacent ORFs as parts of a frameshifting system. Despite this difficulty, researchers have identified several chromosomal genes that, like *prfB* and ODCase antizyme, employ frameshifting. For example, in the yeast *Saccharomyces cerevisiae*, the *EST3* and *ABP140* genes employ +1 frameshifting meaning that in a genome encoding about 6000 genes, only about 0.03% use frameshifting. This low frequency of use of frameshifting appears to be normal, but some species may employ frameshifting much more frequently. For example, in ciliated protozoa of the genus *Euplotes* it appears that up to 10% of genes may employ programmed +1 frameshifting. Clearly, a full understanding of the structure of genomes would require identifying frameshift genes.

Several researchers have used ►bioinformatics to identify frameshift genes in completed genomes. Most have focused their work on the relatively well-defined –1 simultaneous slippage frameshift sites. This work is still in its infancy and the number of sites found by the various methods has varied quite widely.

Clinical Relevance

Programmed frameshifting occurs in a variety of medically and agriculturally relevant viruses. A good example is the ►human immunodeficiency virus-1 (HIV-1), which uses programmed –1 frameshifting in expression of its Gag-Pol protein. We know from studies with simpler, genetically tractable systems that even relatively subtle changes in the efficiency of frameshifting can severely compromise virus propagation. Furthermore, the efficiency of frameshifting is strongly dependent on features of the cellular translation system. Model system studies have shown that perturbations to the translational machinery that have no effect on bulk translation (i.e. tRNA modification) can have strong effects on frameshifting. These facts argue that the frameshifting system could be an excellent target for antiviral drug development.

References

1. Baranov PV, Gesteland RF, Atkins JF (2002) Recoding: translational bifurcations in gene expression. *Gene* 286:187–201
2. Biswas P, Jiang X, Pacchia AL et al (2004) The human immunodeficiency virus type 1 ribosomal frameshifting site is an invariant sequence determinant and an important target for antiviral therapy. *J Virol* 78:2082–2087
3. Farabaugh PJ (2000) Translational frameshifting: implications for the mechanism of translational frame maintenance. *Prog Nucleic Acid Res Mol Biol* 64:131–170
4. Kontos H, Napthine S, Brierley I (2001) Ribosomal pausing at a frameshifter RNA pseudoknot is sensitive to reading phase but shows little correlation with frameshift efficiency. *Mol Cell Biol* 21:8657–8670

Translational Inhibition

Definition

Translational inhibition denotes the interference with the synthesis of a peptide chain from its mRNA template.

- ▶ [Morpholino Oligonucleotides, Functional Genomics by Gene 'Knock-down'](#)
- ▶ [Translational Control in Eukaryotes](#)

Translational Pause

Definition

Translational pause is a transient arrest of the 5' to 3' movement of the ribosome on the mRNA.

- ▶ [Translational Frameshifting, Non-standard Reading of the Genetic Code](#)

Translationally Silent Mutation

Definition

Translationally silent mutation designates a nucleotide change that alters a codon but not the amino acid which it specifies; also termed anonymous mutation.

- ▶ [Splicing](#)

Translesion DNA Synthesis

Definition

Translesion DNA synthesis describes a DNA synthesis on a DNA molecule that contains altered or missing bases. The damage in the template strand blocks the normal replication apparatus.

- ▶ [DNA Polymerases](#)
- ▶ [Replication Fork](#)

Translocation

Definition

Translocation designates the transfer of a chromosome segment to another chromosome, or from one part of a chromosome to another part of the same chromosome. In a balanced translocation, no overall loss or gain of chromosomal material occurs.

- ▶ [Chromosome 21 Disorders](#)
- ▶ [Schizophrenia Genetics](#)
- ▶ [SRY – Sex Reversal](#)
- ▶ [X-Chromosome Inactivation](#)

Transmembrane Proteins

Definition

Transmembrane proteins contain at least one hydrophobic stretch of amino acids long enough to span a lipid bilayer.

- ▶ [Biological Membranes](#)
- ▶ [Fatty Acid Acylation of Proteins](#)

Transmissible Spongiform Encephalopathies

- ▶ [Prion diseases](#)

Transmitters

Definition

Transmitters are small molecules such as acetylcholine, glutamate or γ -aminobutyric acid acting as messengers within the brain or between nerve cells and muscles.

- ▶ [Ion Channels/Excitable Membranes](#)
- ▶ [Neurons](#)

Transposable Element

Definition

Transposable element refers to DNA sequences that can move to different positions in the genome, either by excision and re-insertion or by retrotransposition.

- ▶ DNA Recombination
- ▶ Medaka as a Model Organism for Functional Genomics
- ▶ Repetitive DNA
- ▶ Transposons

Transposition

Definition

Transposition refers to the movement of a genetic element within the host genome. In molecular terms, the best understood transposition reactions are those of phages Mu and Tn10.

- ▶ DNA Recombination
- ▶ Transposons

Transposons

ZOLTÁN IVICS, OLIVER WALISKO,
ZSUZSANNA IZSVÁK
Max Delbrück Center for Molecular Medicine,
Berlin, Germany
zivics@mdc-berlin.de

Synonyms

Transposon is a short and most commonly used term for transposable element. Transposons are also referred to as mobile genetic elements, which is a wide category including the classical transposable elements, integrating viruses, mobile introns and some other products of DNA recombination. Transposons are also called “jumping genes”, for their ability to move.

Definition

Transposons are discrete segments of DNA that have the distinctive ability to move and replicate within genomes. Transposons were discovered in the 1940's

by Barbara McClintock (who later was awarded with the Nobel Prize for this discovery) in the maize genome and have since been found ubiquitously in essentially all living organisms. The process of element movement is generally called transposition and can contribute to insertional mutagenesis, altered gene expression and recombination. Transposons make up significant fractions of genomes; for example, about 45% of the human genome is composed of sequences of a variety of different elements (1). Transposons are best viewed as molecular parasites that propagate themselves using resources of the host cell. However, unlike viruses, transposons are not infectious and their activities are therefore confined to intracellular space. Due to this restriction, transposable elements have to coexist with host cells in order to survive and therefore their activities are highly controlled. Despite their parasitic nature, there is increasing evidence that transposable elements are a powerful force in gene evolution. Indeed, about 50 human genes are derived from transposable elements (1), among them genes that are responsible for immunoglobulin gene recombination in all vertebrates.

Characteristics

Classes of Transposable Elements and Mechanisms of Transposition

Transposable elements fall into two major classes. The first is retroelements that transpose through an RNA intermediate and include long interspersed elements (LINEs), short interspersed elements (SINEs), and LTR retrotransposons. Retroelements transpose through a replicative (copy-and-paste) mode of transposition, in which the transposon does not get excised from its donor locus but instead a copy of it is produced which inserts elsewhere in the genome (Fig. 1). Thus, replicative transposition leads to an increase in the copy number of the transposon within a genome. Members of the second class are called DNA transposons because they move directly as DNA, by a conservative (cut-and-paste) mechanism of transposition, in which the element gets excised from the donor locus and is subsequently reinserted elsewhere (Fig. 1). The copy numbers of DNA transposons also increase over time, but amplification is not inherent to the transposition process itself.

The major non-LTR retrotransposon in mammals is LINE-1 (or L1). L1 has about 5×10^5 copies in the human genome, thereby making up about 17% of human genomic DNA (1). LINEs contain a polymerase II promoter and produce a polyadenylated, ▶ **bicistronic** RNA containing two open reading frames (ORFs) that encode proteins required for transposition (Fig. 2a). The element-encoded ▶ **endonuclease** generates a single-stranded nick in the target DNA, and the



Transposons. Figure 1 Schematic representation of the two major mechanisms of transposition. During conservative transposition, the element is excised from the donor DNA (red line), and integrates into a new target DNA (green line). The broken donor DNA has to be repaired by host factors and this process can result in a small “footprint” (black dot) that marks the former presence of the element in that site. Replicative transposition requires amplification of the element either by replication or by copying of the element through transcription followed by reverse transcription. The amplified element gets inserted elsewhere in the genome.

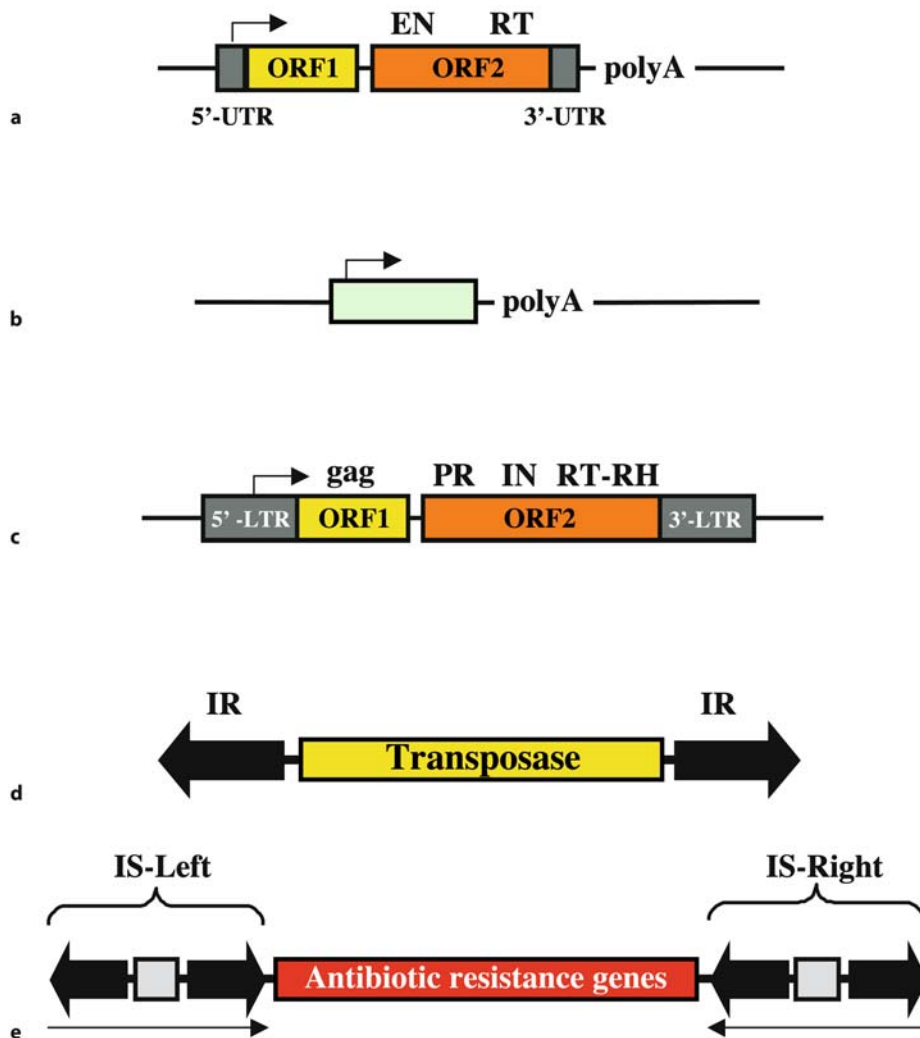
►reverse transcriptase uses the nicked DNA to prime reverse transcription from the 3' end of the L1 RNA (2). Transposition-competent L1s in humans are about 6 kb long. However, because reverse transcription is frequently incomplete, the majority of L1s are truncated and thus non-functional. The LINE retrotransposition machinery is believed to be responsible for most reverse transcription activity in the genome, including retrotransposition of the non-autonomous SINEs and generation of ►processed pseudogenes.

SINEs are short (about 100–400 bp) retrotransposable elements that contain a polymerase III promoter and encode no proteins (2) (Fig. 2b). These non-autonomous transposons are thought to use the LINE machinery for transposition. The vast majority of known SINEs are derived from tRNA sequences, with the exception of the human *Alu* element, which is derived from the 7SL component of the ►signal recognition particle. *Alu* elements were originally identified as repetitive DNA elements in human DNA renaturation curves and contain a recognition site for the restriction enzyme *AluI*. *Alu* elements are represented in the human genome with $>1 \times 10^6$ copies, which make up about 11% of the total genome. *Alu* is the only active SINE in humans.

LTR-retrotransposons are similar to retroviruses in that they are flanked by long terminal direct repeats that contain transcriptional regulatory elements that drive the expression of at least two ORFs similar to the *gag* and *pol* genes (Fig. 2c). LTR-retrotransposons comprise two major families, Ty1/*copia* and Ty3/*gypsy*, based on the founder elements in the yeast *Saccharomyces cerevisiae* (Ty1 and Ty3) and in *Drosophila* (*copia* and *gypsy*). LTR-retrotransposons are widely distributed in eukaryotes, and make up about 8% of the human genome. Transposition occurs through reverse transcription of the retrotransposon RNA, and integration of the resultant cDNA into a new location by the

integrase protein. Endogenous retroviruses (ERVs) appear to have been recently active in the mammalian genome.

The simplest DNA transposons are composed of short terminal inverted repeats flanking a transposase gene (Fig. 2d). The transposase binds to the inverted repeats, and catalyzes ‘cut-and-paste’ transposition (Fig. 1). Bacterial insertion sequence (IS) elements usually encode no function other than the transposase, which distinguishes them from bacterial transposons. Transposons encode, in addition to the transposase, another function that often confers antibiotic resistance (Fig. 2e). Transposons are flanked by inverted or direct repeats, which are in many cases structurally similar to IS elements (Fig. 2e). Although DNA transposons are widespread in prokaryotes, they are less abundant in eukaryotes. For example, only 2–3% of the human genome is composed of DNA transposons and there is no indication for their activity in the past 50 million years (1). Active DNA transposons are transitory components of genomes. This is because these elements accumulate inactivating mutations over evolutionary time. Inactive copies accumulate in the genome, because transposase can mobilize inactive (nonautonomous) elements. The expansion of inactive transposon copies results in regulatory mechanisms that limit transposition rates (3). Eventually, the last remaining transposase-producing (autonomous) copy gets mutated, which seals the fate of the given transposon in the genome. Thus, DNA transposons are doomed to die out and, in order to survive, they must be able to invade and colonize virgin genomes. This process is mediated by ►horizontal transmission from one species into the germline of another species. For example, the P transposable element in *Drosophila melanogaster*, an everyday research tool of fly geneticists, is believed to have been horizontally transferred from *Drosophila willistoni* some time in



Transposons. Figure 2 Structures and organization of the main types of transposable elements. (a) Non-LTR retrotransposon. The element consists of a 5' untranslated region that has promoter activity (arrow pointing towards the downstream genes), which is required to drive transcription of the element-encoded genes. ORF1 encodes a nucleic acid binding protein. ORF2 encodes an endonuclease (EN) and a reverse transcriptase (RT). The element has a polyA tail. (b) A typical SINE. The element is a small, RNA-derived pseudogene, which is transcribed from an RNA polymerase III promoter within the element (arrow). The element has a polyA tail. (c) LTR-retrotransposon. The element consists of long terminal repeats (LTRs) similar to those of retroviruses. The LTRs flank two open reading frames. ORF1 encodes the group specific antigen (gag), ORF2 encodes a protease (PR), an integrase (IN), and a reverse transcriptase-RNaseH (RT-RH) function. (d) DNA transposon. The central transposase gene (yellow box) is flanked by terminal inverted repeats (IRs, shown as black arrows). The IRs contain the binding sites for the transposase and sequences that are required for transposase-mediated cleavage. (e) Composite bacterial transposon. The element consists of antibiotic resistance genes (red box) flanked by two copies of an insertion sequence (IS) element that contains the transposase gene (yellow boxes). The arrows underneath indicate the inverted orientation of the IS elements.

the last century. Currently, there are only three active DNA transposons known from vertebrates, the Tol2 element in the medaka fish, and the *Sleeping Beauty* and *Frog prince* elements that have been reconstructed from inactive elements in fish and frog genomes, respectively (4, 5).

Impact of Transposable Elements on the Host Genome

A general feature of transposable elements is that they can replicate independently of the cellular replication cycle, and new copies can emerge at new locations in the genome. Thus, mobile elements can cause insertional mutagenesis if they land within a gene (6), but

they can also lead to altered gene expression and genetic recombination. Insertion of a transposable element into the protein-coding region of a gene (exon) can disrupt gene function. For example, bacterial IS elements were identified as DNA insertions in the *E. coli gal* operon which cause highly polar mutations. The *Drosophila* P element was found to cause mutations at the *white* locus (and have since been used as a mutagenic agent in the laboratory) and the DNA transposon Tc1 is the main cause of mutation in the nematode *Caenorhabditis elegans*. In contrast to direct insertion of a transposable element into exon sequences of a gene, the element can be inserted into upstream regulatory sequences of a gene and thereby alter the regulation of its expression. For example, insertion between the core promoter and adjacent enhancer regions would increase the distance between these regions and thus negatively affect promoter activity. Another damaging aspect of transposable elements is that repeated, dispersed copies of homologous sequences can promote secondary rearrangements (6), which can result in deletions, duplications and inversions. This potential of dispersed transposon copies to promote homologous recombination can be even more damaging to the genome than a *de novo* insertion.

Transposable elements do not only do harm, but are also a creative force. In *Drosophila*, [telomere](#) maintenance is not brought about by telomerase, but by repeated transposition of two non-LTR retrotransposons, HeT-A and TART, into chromosome ends. The acquisition of new transposon insertions can donate regulatory elements to genes or even lead to the evolution of new genes. L1 elements can carry non-transposon sequences into new places, a process that can contribute to “exon shuffling” and thus to gene evolution (2). This is because L1 transcription can read through the native transcription termination site of the element into flanking genomic sequences. It is estimated that about 0.5–1% of the human genome may have been generated by L1-mediated [transduction](#) of 3′ flanking sequences (1). The L1 retrotransposition machinery can also mediate reverse transcription and genomic insertion of host gene mRNAs, resulting in processed pseudogenes. Some of these insertions can give rise to functional processed genes. Approximately 50 human genes have evolved from transposable elements, mostly from DNA transposons (1). These include the RAG1 and RAG2 immunoglobulin gene recombinases and the [centromere](#)-binding protein CENPB. Thus, although transposable elements have not been selected for conferring selective advantage to the host, they can contribute useful functions to genomes.

Regulation of Transposition

Transposition is under strict control, and the accumulation of hundreds or thousands of transposon copies can

take millions of years. This is because transposition can potentially endanger the survival of the host organism and consequently that of the transposable element. Therefore, transposons and their hosts have coevolved and developed strategies that reduce the negative effects on the host but ensure proliferation of the element. The mechanisms of transpositional regulation are diverse (8). The expression of transposon-encoded factors can be kept in check by transcriptional silencing, through DNA methylation. Post-transcriptional silencing can also contribute to down-regulation of factors that are required for transposition. For example, [RNA interference](#) (RNAi) suppresses gene expression in a wide variety of organisms, and has been shown to constitute an antiviral defence mechanism in plants. RNAi has been proposed to be a major mechanism for transposon silencing in nematodes and perhaps contributes to transposon regulation in other animals as well. Another form of regulation is site-selective insertion of transposons into “safe” places in the genome. For example, the Ty retrotransposable elements in yeast show considerable site-specificity of insertion, thereby reducing the negative impact of transposition. Transposition of the bacterial transposon Tn5 and that of P elements and *mariner* elements in *Drosophila* can be regulated by repressor proteins, which are truncated or point mutant versions of the transposase polypeptide (3). For example, defective transposases can compete with wild type transposase for binding sites located in the transposon ends. Furthermore, P element transposition is restricted to the fly germline, by [alternative splicing](#) of the transposase mRNA. A particularly interesting feature of the bacterial Tn7 element is that it does not insert into DNA that already contains a copy of Tn7, a phenomenon called target immunity. Target immunity helps to avoid multiple copies of the element in the same DNA molecule, which might result in deleterious recombination between the two elements.

Taken together, there are a great variety of mechanisms that put a limit on transpositional activity. The outcome of this regulation is that transposable elements move at very low frequencies.

Clinical Relevance

About 1 in 600 mutations in humans is estimated to arise from retrotransposon-mediated insertion. The major causative agent of endogenous genomic insertions is L1 (6). An average human being has 80–100 retrotransposition-competent L1s, which belong to a particular subfamily of these elements in the human genome. Results also suggest that a relatively small number of very active L1s comprise the bulk of L1 activity (6). A current estimate for transpositional frequencies in humans is that about 1 in 8 individuals harbour a new L1 insertion (2). New, disease-causing

insertions of L1 in humans were in fact the first retrotransposition events detected in mammals. These insertions occurred in the blood clotting protein Factor VIII, dystrophin, APC and β -globin genes.

Alu elements continue to amplify at a rate of about one insertion every 200 new births. New insertion events can lead to genetic disorders including haemophilia, neurofibromatosis, cholinesterase deficiency, breast cancer and leukaemia (7). *Alu* element insertion is estimated to contribute to about 0.1% of human genetic diseases. The large number of *Alu* elements within the human genome also provides ample opportunity for homologous recombination events between dispersed *Alu* repeats. These events can result in deletion or duplication of exons in a gene and other chromosomal abnormalities. This mode of mutagenesis is estimated to account for 0.3% of human genetic diseases, including Fabry disease, Duchenne's muscular dystrophy, ADA deficiency and a variety of cancers.

There is considerable interest in technologies that allow the delivery and expression of genes in certain tissues or organs *in vivo* for the correction of genetic diseases. Paradoxical sounding perhaps, but transposable elements can be harnessed as useful vectors for gene therapy. In particular, the *Sleeping Beauty* transposon (4) has been shown to be a promising gene delivery vehicle for therapeutic applications. *Sleeping Beauty* mediates efficient and stable insertions of transgenes into human (and other vertebrate) chromosomes (4) and ensures persistent, life-long transgene expression at therapeutic levels. In light of the intrinsic ability of transposons to cause undesired mutations in endogenous genes, one particularly relevant concern in human applications is safety. With the use of technologies that allow designed targeting of transposable elements into safe locations in the human genome, *Sleeping Beauty* and other transposons might evolve into simple, efficient and safe vectors for stable gene transfer for the correction of human genetic diseases.

References

1. International Human Genome Sequencing Consortium (2001) Initial sequencing and analysis of the human genome. *Nature* 409:860-921
2. Moran JV, Gilbert N (2002) Mammalian LINE-1 retrotransposons and related elements. In: Craig NL, Craigie R, Gellert M et al (eds) *Mobile DNA II*. ASM Press, Washington pp 836-869
3. Hartl DL, Lozovskaya ER, Nurminsky DI et al (1997) What restricts the activity of mariner-like transposable elements? *Trends Genet* 13:197-201
4. Ivics Z, Hackett PB, Plasterk RH et al (1997) Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells. *Cell* 91:501-510
5. Miskey CS, Izsvak ZS, Plasterk RH, Ivics Z (2003) The Frog Prince: a reconstructed transposon from *Rana pipiens* with high activity in vertebrates. *Nucleic Acids Res* 31:6873-6881
6. Kazazian Jr (1999) An estimated frequency of endogenous insertional mutations in humans. *Nat Genet* 22:130
7. Deininger PL, Batzer MA (1999) *Alu* repeats and human disease. *Mol Genet Metab* 67:183-193
8. Fedoroff N (2002) Control of mobile DNA. In: Craig NL, Craigie R, Gellert M (eds) et al *Mobile DNA II*. ASM Press, Washington, pp 836-869

Trans-Splicing

Definition

Trans-splicing describes a form of splicing that is mechanistically quite similar to the removal of introns, in which a short leader is spliced onto the 5' ends of mRNAs.

► *C. Elegans* Genome: Comparative Sequencing

Transvection

Definition

Effect of regulatory DNA elements, notably enhancers and silencers, from one DNA molecule to another one, i.e., *in trans* instead of *in cis*. Transvection was first described in *Drosophila* as an effect from one chromosome to its paired homolog.

► Enhancer

Transversion

Definition

In cases of non-silent base pair substitutions the original base pair is changed, e.g. from A to B. If A is a purine and B a pyrimidine, or vice versa the process is called a transversion. For comparison see

► transistion.

► Sequence Annotation in Evolution

► Transition

TRE

Definition

TRE stand for TPA-responsive element. It refers to an AP-1 binding site first defined in the human collagenase gene to mediate induction by the phorbol ester TPA.

► [Jun/Fos](#)

Treadmilling

Definition

Treadmilling describes a pattern of growth of microtubules or F-actin. It is characterized by net addition of tubulin heterodimers or G-actin at the plus end, and the balanced net loss from the minus end.

► [Cytoskeleton](#)

Trichothiodystrophy

Definition

Trichothiodystrophy (TTD) is a rare, autosomal recessively transmitted, multisystem disorder associated with defects in nucleotide excision repair. Diagnostic hallmark is short, brittle hair, low in sulfur and cystine, because of impaired synthesis of high-sulfur matrix protein. Patients suffer from increased UV sensitivity due to the defects in the nucleotide excision repair system. The disease is not cancer prone.

► [DNA-Repair Mechanisms](#)

Trinucleotide Repeat

Definition

Trinucleotide repeat is a stretch of DNA containing a tandemly repeated sequence motif of three nucleotide bases, e.g. the CGG repeat in the ► [FMR1](#) gene of fragile X mental retardation syndrome.

► [Fragile X Syndrome](#)

► [Repeat Expansion Diseases](#)

Trinucleotide Repeat Expansion Disease

► [Polyglutamine Disease, the Emerging Role of Transcription Interference](#)

Triple Resonance NMR Experiment

Definition

Triple resonance NMR experiment describes a NMR experiment using the proton, carbon and nitrogen nuclei in a biomolecule to create spectra for the assignment of resonances and the extraction of structurally relevant parameters. These are standard experiments for the structure determination of biomolecules in the context of structural genomics.

► [Multidimensional NMR Spectroscopy](#)

Trisomy

Definition

Trisomy is a chromosomal aberration that is characterized by the existence of a third chromosome besides the normal diploid set of chromosomes. The most prominent disease resulting from a trisomy of chromosome 21 is ► [Down Syndrome](#) due to non-disjunction of chromosomes at meiosis.

► [Chromosome 21, Disorders](#)

► [Gene Duplications](#)

tRNA

ESTHER LIZANO, MARIO MÖRL
Institute for Biochemistry, University of Leipzig,
Leipzig, Germany
elizano@uni-leipzig.de
moerl@uni-leipzig.de

Definition

In the central dogma of biology, the information of a gene is copied into a ► [messenger RNA](#) (► [trans-](#)

cription) and subsequently translated into protein. The sequence of such an mRNA is read in nucleotide triplets (codons) that correspond to individual amino acids in the resultant protein. Until the 1950s, however, it was not clear how such a base sequence is translated into a sequence of amino acids. With the advent of biochemical studies on protein synthesis, the participation of small “soluble” RNA molecules became evident. These function as adaptors that recognize specific codons in the mRNA and bring the correct amino acids to the corresponding positions in the nascent protein. Due to this role in transfer of genetic information into protein sequence, these small RNA molecules were named transfer RNA (tRNA) (1). Since proteins contain different amino acids (20 standard amino acids plus 2 rare ones), every organism needs a corresponding set of different tRNA molecules. These tRNAs have certain common characteristics, since they all have to interact with the protein synthesis machinery (the ribosome and accessory translation factors), but they also must have very specific and individual features, in order to interact exclusively with one type of amino acid and one sort of codon on the mRNA. These different requirements are mirrored in the typical secondary and tertiary structure of these adaptor molecules.

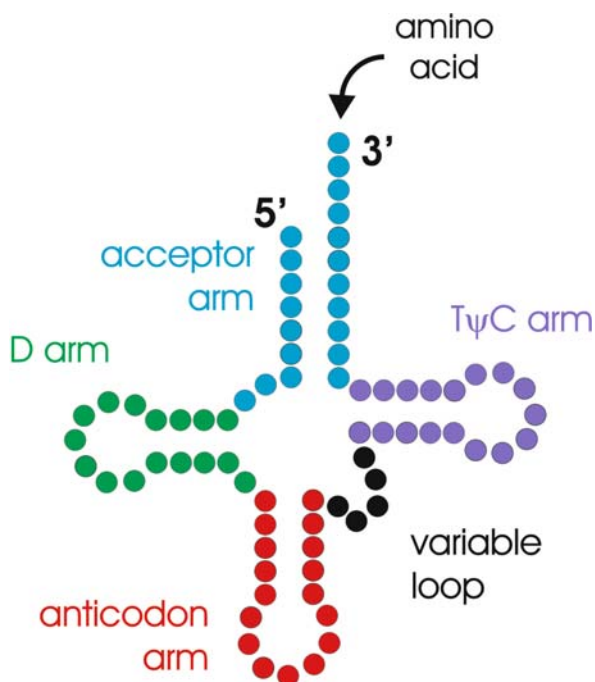
Besides their participation in protein synthesis, tRNAs can have additional functions in some organisms, like the tRNA for glutamate, which is used as a substrate for the synthesis of tetrapyrroles (like chlorophyll and heme) in plants, archaea and many bacteria (2).

Characteristics

Whether members of the prokaryotic, eukaryotic or archaeal kingdoms, all organisms have tRNA molecules of a very similar size and structure. They are rather small transcripts, consisting of a single-stranded polynucleotide chain with an average length of 76 nucleotides (in a range from 60 to 95 nucleotides). In addition to the four standard bases, tRNAs contain many modified bases important for the structure and function of the molecules.

Secondary Structure

Complementary bases allow tRNAs to form internal hydrogen bonds (secondary H-bonds) between different parts of the polynucleotide sequence. These interactions force the linear molecule to adopt a secondary structure in the form of a cloverleaf (Fig. 1). The base pairs involved usually follow the classical Watson and Crick rules (A-U and G-C), but also include non-conventional interactions like G-U, G-ψ and A-ψ base pairs (ψ represents pseudouridine, a modified base). The cloverleaf structure consists of four different arms composed of double-stranded stem



tRNA. Figure 1 Characteristic secondary structure of a tRNA molecule. The hydrogen bonds formed between complementary bases in the tRNA lead to the formation of the typical cloverleaf structure, with the acceptor stem (blue) and the 3 “leaves” D-arm (green), anticodon-arm (red) and TψC-arm (purple). The additional variable loop is labelled in black. The aminoacylation of the tRNA 3’end is indicated.

regions and single-stranded loops. Each arm is named after its function or characteristic feature:

1. The acceptor arm is the “branch” of the cloverleaf and is formed by the pairing of bases adjacent to the 5’- and 3’-termini of the RNA. In addition, the 3’-end carries a stretch of unpaired nucleotides ending with the invariant sequence C-C-A. The terminal A residue of this sequence has an essential function as the part of the tRNA that carries the cognate amino acid.
2. The D-arm is one of the “leaves”. It is named after the modified base dihydrouridine, found frequently in this tRNA region.
3. The TψC-arm, another “leaf”, is identified by the presence of the conserved triplet sequence T-ψ-C in many tRNAs.
4. The anticodon arm is located opposite to the acceptor stem. In the center of its loop, it carries the anticodon triplet, which is complementary to a specific codon in the messenger RNA. This codon – anticodon interaction brings the designated tRNA to the proper position of the mRNA and, as a consequence, the corresponding amino acid to the correct site of the growing polypeptide.

Since these structural elements are composed of a rather conserved number of nucleotides, a useful numbering system for tRNA base positions has been established based on the tRNA consensus structure. However, the secondary structure described contains an additional variable domain, the variable loop, located between the T ψ C and the anticodon arm. Hence, tRNAs are divided into two classes: Class 1 represents molecules with a variable loop of 4–5 bases, whereas Class 2 tRNAs have 10–24 bases in this domain. Interestingly, although tRNAs vary in their primary sequence, almost all of these molecules can adopt the cloverleaf structure. Exceptions include some mitochondrial tRNAs that lack the D- or T ψ C-arm and hence carry just two of the three “leaves”.

Positions in the nucleotide sequence can be classified by degree of conservation between different tRNAs. Invariant positions contain particular bases in almost every tRNA (>90%), while other positions are occupied by either a ►purine or a ►pyrimidine and are therefore called semi-invariant. The functions of these invariant and semi-invariant positions cannot be deduced from the secondary structure, but become evident when the tertiary structure is analyzed.

Tertiary Structure

The first three-dimensional structure of a tRNA resulted from crystallographic studies on the phenylalanine tRNA of yeast, and this molecule has been a great source of information on RNA structure (3). The cloverleaf secondary structure folds into a compact L-shaped tertiary structure (Fig. 2) stabilized by an intricate network of hydrogen bonds (tertiary H-bonds) involving many of the invariant and semi-invariant bases. Many of the bases are stacked and about half of them are involved in base pairings. The L-shape is the result of two coaxial helices, one formed by stacking interactions of the acceptor- and T-stems and the other by the stacking of the anticodon- and D-stems. The juxtaposition of the D- and T ψ C-loop forms the elbow region where several Mg²⁺ binding sites are found; both T ψ C and Mg²⁺ contribute to the stabilization of this region by interacting with the D-loop. In this L-shape, one end carries the anticodon (to interact with codons on the mRNA), while the other end is charged with the amino acid (to interact with the nascent protein). Therefore, both functional sites are at a maximum distance from each other. Surprisingly, the above-mentioned tRNAs lacking one of the cloverleaf domains can also adopt an L-like three-dimensional shape and so can function as regular tRNA molecules in protein synthesis.

The correct tertiary structure of the tRNA is a prerequisite for the interaction with specialized enzymes that are either involved in the maturation of the transcript into a functional tRNA (processing enzymes,

see below) or that charge it with the cognate amino acid (aminoacyl tRNA synthetases) (4). The L-shape is also important for the interaction with translational cofactors that transport the charged tRNA to the ribosome.

Modified Nucleotides

In addition to the four standard bases, more than 80 different modified nucleotides have been identified in tRNAs, with changes ranging from simple methylation to whole restructuring of the purine ring. These modifications are found in all parts of the molecule and are synthesized by specific tRNA-modifying enzymes (5). While modifications like dihydrouridine or pseudouridine are constant features in all tRNAs, others are specific for a particular tRNA. The modifications are introduced post-transcriptionally and are considered to be important for specificity, fidelity and/or efficiency of the various inter- and intramolecular interactions in which tRNAs are involved. A direct effect can be observed in the anticodon, where modifications influence the base-pairing interactions with the corresponding mRNA codons. An example is the tRNA for isoleucine in *Escherichia coli*, where the first position of the anticodon, a cytidine, is modified to lysidine (2-lysylcytidine). With this single modification, the codon specificity is changed from methionine to isoleucine. Consequently, in order to avoid amino acid misincorporation during translation, the tRNA carrying this modification has to be charged with isoleucine instead of methionine.

tRNA Maturation and Aminoacylation

In all organisms analyzed thus far, tRNAs are not transcribed as functional molecules, but are synthesized as large precursor transcripts with additional 5'- and 3'-sequences. During maturation, these extra sequences are removed by exo- and/or endo-nucleolytic activities. An interesting processing enzyme involved in the removal of the 5'-leader is RNase P, an enzyme consisting of RNA and protein subunits, which in some cases has catalytic activity in the RNA part. Other steps in maturation include modification of nucleotides (see above) and addition of the triplet CCA at the 3'-terminus by tRNA nucleotidyl transferase. Since this CCA terminus undergoes frequent turnover, the tRNA nucleotidyl transferase is also responsible for the maintenance of the 3'-end in order to keep the tRNA molecule functional for aminoacylation.

The charging of tRNAs with their cognate amino acids is catalyzed by aminoacyl-tRNA synthetases. For each sort of amino acid, there exists one of these highly specific enzymes in the cell. The synthetases charge the tRNAs in a very specific manner that underlies a ►proof reading mechanism in order to keep mischarging at a very low level. This is necessary since mischarged tRNAs would lead to the incorporation of

amino acids at wrong positions in the protein, resulting in deleterious non-functional products.

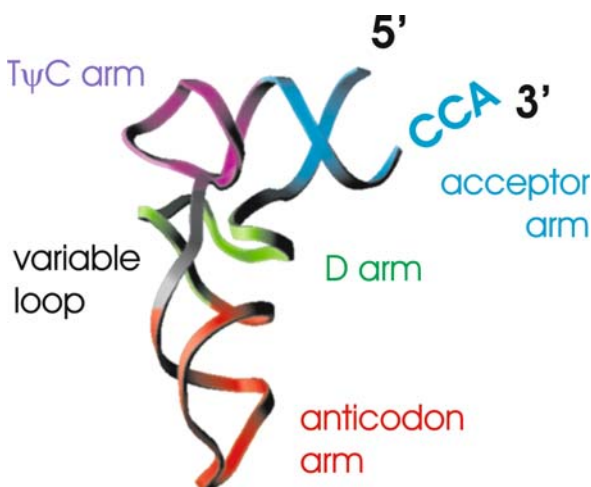
Suppressor tRNAs

Protein synthesis terminates when the translational machinery, the ribosome, reaches a codon in the mRNA that is not recognized by a tRNA, but rather by release factors that stop further elongation of the growing polypeptide. These so-called stop codons have the sequence UAG, UAA or UGA. Interestingly, certain mutational alterations of the tRNA's anticodon allow the tRNA to read such a stop codon and to incorporate its amino acid, leading thereby to a “read through” of the stop codon and suppressing the termination of protein synthesis. This means that these mutant tRNAs can suppress “nonsense” mutations in protein coding genes as follows. If a mutation changes a codon of the mRNA into a stop codon, it is no longer possible to synthesize the full-length protein encoded in this gene, since the new stop codon leads to premature termination of translation. However, if such a mutated tRNA can read this stop codon, it suppresses the premature termination, allowing the ribosome to complete the protein. Therefore, mutated tRNAs of this sort are called “nonsense suppressor tRNAs” (6).

A second category of suppressor tRNAs are the so-called missense suppressors. A mutation in a codon does not necessarily lead to a stop codon, but often changes the identity of the codon to that of another amino acid. As a result, a different amino acid is incorporated into the growing protein at this position. In this situation, a complementary mutation in the anticodon of the tRNA can restore the base pairing interaction with the mutated codon in the mRNA. This then allows the tRNA to deliver the correct amino acid to this position in the protein, again suppressing the effect of the original mutation in the protein-encoding gene. Along with their interesting role in biology, these suppressor tRNAs have turned out to be very useful tools for addressing many questions in molecular genetics.

Clinical Relevance

tRNA molecules are not only interesting research subjects for molecular biology, but are also of significant clinical relevance, since many mutations in these genes are associated with certain diseases. In ►mitochondria (cellular organelles with their own genome and translational machinery), many ►point mutations in tRNA genes have been implicated in a variety of multisystemic neuromuscular diseases and other severe disorders. Around 100 disease-correlated mutations in tRNA genes are currently described. An example is the mitochondrial gene that encodes for leucine tRNA, a particular hot spot with 16 pathogenic base changes that lead to diseases like ►MELAS (7). Mutations with the most severe effects are often found



tRNA. Figure 2 tRNA tertiary structure. The 3-dimensional L-shaped structure is formed and stabilized by tertiary hydrogen bonds between individual regions of the tRNA. The 4 helical domains of the cloverleaf secondary structure are arranged in two continuous helices: acceptor stem and T ψ C-arm build up one arm of the L, while D- and anticodon-arm form the other. This structure is stabilized by several stacking interactions. Note that the part reading the mRNA codon (anticodon loop) and the region carrying the amino acid (acceptor stem) are at maximum distance, which is required for protein synthesis at the ribosome.

in conserved elements and either affect invariant nucleotides or disrupt base pairs in double helical regions. Most of these changes are transitions (replacing a purine by a purine or a pyrimidine by a pyrimidine), which represent “mild” mutations in many protein-encoding genes. Here, however, these mutations can influence tRNA structure and recognition elements for interacting enzymes and therefore lead to severe negative effects on tRNA function. Due to the central role of mitochondria in energy metabolism, these mutations affect predominantly tissues with a high energy demand such as neuronal and muscular tissues, thus resulting in diseases like encephalopathies, myopathies or cardiopathies.

Another involvement of tRNAs in diseases is *via* their recruitment as primers for the replication of ►retroviruses like HIV. This can occur because the viral genome has a region that is complementary to the 3'-end of tRNA for lysine. By base pairing of this tRNA part with the viral RNA, the tRNA is used as a starting point for the ►reverse transcription of the viral RNA genome into the DNA form, which is a prerequisite for the replication and amplification of the virus. The host tRNA molecule therefore plays a central role in viral propagation in the cell and is also packed into the virion particle to be immediately available to the virus upon infection of another cell.

Hence, tRNA molecules are of clinical interest not only as a result of their association with certain hereditary diseases, but also due to their involvement in the replication of pathogenic retroviruses that can cause pandemic diseases like AIDS and possibly cancer.

- ▶ [Cap-Independent Translational Control](#)
- ▶ [Mitochondrial Myopathies](#)
- ▶ [RNA Polymerase III](#)
- ▶ [Transfer RNA](#)

References

1. Khorana HG (1995) Transfer RNA: Discovery, Early Work, and Total Synthesis of a tRNA gene. In: Söll D, Rajbhandary UL (eds) tRNA: Structure, Biosynthesis, and Function. American Society Microbiology Press, Washington, pp 5–16
2. Jahn D, Verkamp E, Söll D (1992) Glutamyl-transfer RNA: A Precursor of Heme and Chlorophyll Biosynthesis. Trends Biochem. Sci 17:215–218
3. Robertus JD, Ladner JE, Rhodes JTFD et al (1974) Structure of Yeast Phenylalanine tRNA at 3 Å Resolution. Nature 250:546–551
4. Giegé R, Sissler M, Florentz C (1998) Universal Rules and Idiosyncratic Features in tRNA Identity. Nucleic Acids Res 26:5017–5035
5. Björk GR, Ericson JU, Gustafsson CED et al (1987) Transfer RNA Modification. Ann Rev Biochem 56:263–287
6. Beier H, Grimm M (2001) Misreading of Termination Codons in Eukaryotes by Natural Nonsense Suppressor tRNAs. Nucleic Acids Res 29:4767–4782
7. Florentz C, Sissler M (2001) Disease-Related versus Polymorphic Mutations in Human Mitochondrial tRNAs: Where is the Difference? EMBO Rep 2:481–486

TRP Channel

Definition

TRP (transient receptor potential) channels comprise of a superfamily of cation channels conserved in mammals, flies, worms and yeast. The various TRP-proteins bear sequence and structural similarities to the founding member of this superfamily, the transient receptor potential (TRP), a light activated cation channel in *Drosophila* photoreceptor. Visual transduction in *Drosophila* is mediated by three distinct channel proteins, TRP (missing in the *trp* mutant), TRP-like (TRPL) and TRP γ . So far, more than 25 TRP genes have been identified in mammals. Almost all TRPs are supposed to form ion channels that are widely expressed in the nervous system. Members of the TRP channel family are probable mediators of the store operated calcium entry (SOC) pathway in mammalian cells. TRP channel primary structures predict six

transmembrane spanning domains, with a pore domain between the fifth (S5) and sixth (S6) segments, and both C and N termini located intracellularly. These monomers are thought to coassemble to form homo- or heterotetrameric channels.

- ▶ [Photoreceptors](#)
- ▶ [Polycystic Kidney Disease Autosomal Dominant](#)

True Hermaphroditism

- ▶ [SRY – Sex Reversal](#)

Tryparedoxin

Definition

Tryparedoxins are redox-active proteins of the thioredoxin superfamily. In contrast to typical thioredoxins, their active site motif is CPPC. In the parasite family of trypanosomatids, tryparedoxins functionally substitute for thioredoxins in hydroperoxide detoxification and ribonucleotide reduction.

- ▶ [Thiol-Dependent Peroxidases](#)

TSA

Definition

Tyramide signal amplification (TSA) sometimes called CARD (Catalyzed Reporter Deposition) is an enzyme-mediated detection method that utilizes the catalytic activity of horseradish peroxidase (HRP) to generate high-density labeling of a target protein or nucleic acid sequence *in situ*. The TSA method has been reported to increase the sensitivity by up to 100- fold compared to conventional avidin-biotinylated enzyme complex procedures.

- ▶ [Protein Microarrays as Tool for Protein Profiling and Functional Proteomics](#)

TSE

- ▶ [Prion Diseases](#)
- ▶ [Transmissible Spongiform Encephalopathies](#)

TSG

- Tumor Suppressor Genes

TSH

- Thyroid Stimulating Hormone

t-SNARE

Definition

t-SNARE denotes SNAREs present on the target membrane (► SNARE Proteins).

- Cell Polarity

TTD

- Trichothiodystrophy

T-Tubulus

Definition

T-tubulus refers to invaginations of the surface membrane of a muscle cell. T-tubules are about 0.04 µm in diameter and are transversely arranged in the muscle fiber.

- Muscle Contraction

Tuberous Sclerosis

DAVID J. KWIATKOWSKI
Brigham & Woman's Hospital, Boston, MD, USA
dk@rics.bwh.harvard.edu

Definition

Tuberous sclerosis (TSC) is an autosomal dominant genetic disorder characterized by development of

► hamartomas and hamartias in several organ systems (1). The term hamartoma is used to indicate that there is overgrowth of relatively mature appearing cells in TSC lesions, in distinct contrast to common malignancies. TSC hamartomas involve the skin, brain and kidneys in most patients. However, the clinical manifestations of TSC are quite variable, so that there are formal diagnostic criteria, as listed below (2).

Major Features

1. Facial angiofibromas or forehead plaque
2. Nontraumatic ungual or periungual fibroma
3. Hypomelanotic macules (three or more)
4. Shagreen patch (connective tissue nevus)
5. Multiple retinal nodular hamartomas
6. Cortical tuber*
7. Subependymal nodule
8. Subependymal giant cell astrocytoma
9. Cardiac rhabdomyoma, single or multiple
10. Lymphangiomyomatosis+
11. Renal angiomyolipoma+

Minor Features

1. Multiple, randomly distributed pits in dental enamel
2. Hamartomatous rectal polyps
3. Bone cysts
4. Cerebral white matter radial migration lines *
5. Gingival fibromas
6. Nonrenal hamartoma
7. Retinal achromic patch
8. 'Confetti' skin lesions
9. Multiple renal cysts

Definite Tuberous Sclerosis Complex: Either two major features or one major feature plus two minor features.

Probable Tuberous Sclerosis Complex: One major plus one minor feature.

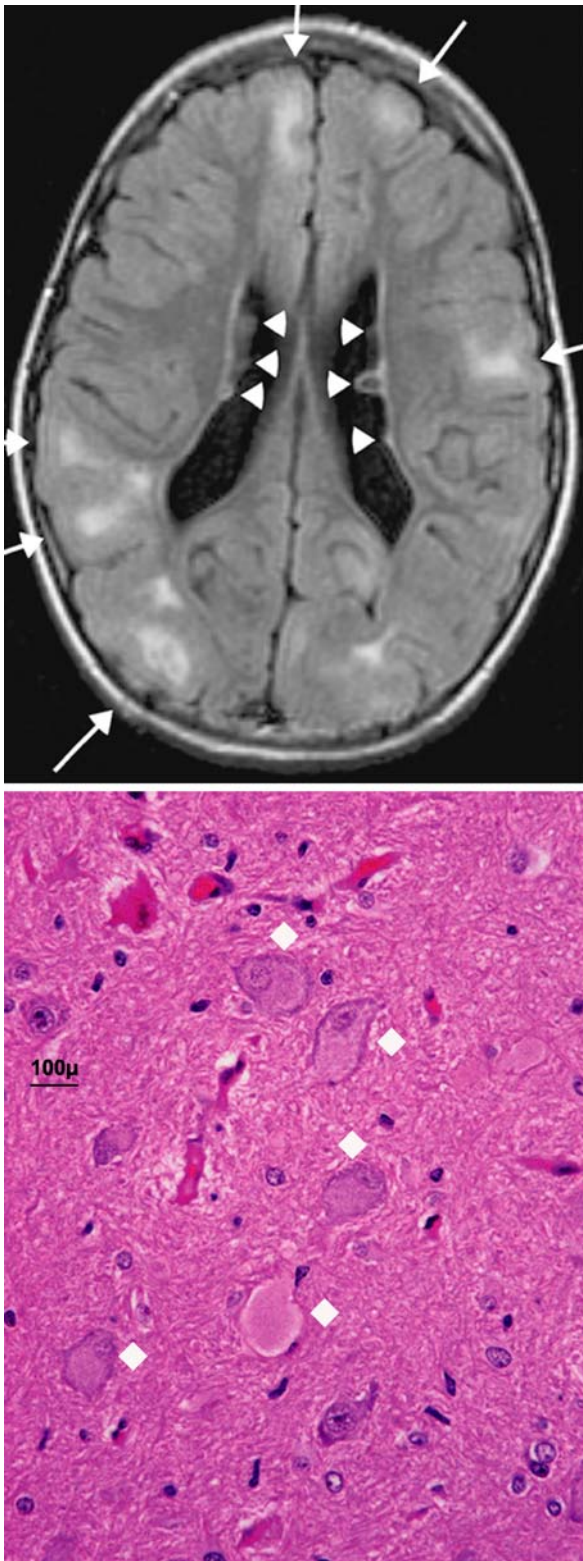
Possible Tuberous Sclerosis Complex: Either one major feature or two or more minor features

Characteristics

Although several organ systems may be affected by hamartomas, the vast majority of TSC hamartomas display limited growth potential and do not require intervention. A small fraction display persistent growth, necessitating surgical control. However, progression to malignancy is very rare in TSC, and has been seen only in TSC renal hamartomas, termed

*When cerebral cortical dysplasia and cerebral white matter migration tracts occur together, they should be counted as one rather than two features of tuberous sclerosis.

+When both lymphangiomyomatosis and renal angiomyolipomas are present, other features of tuberous sclerosis should be present before a definite diagnosis is assigned.



Tuberous Sclerosis. Figure 1 Brain lesions in tuberous sclerosis. (top) Brain MRI of an 8-year-old boy using the FLAIR technique demonstrates several cortical tubers (white, cortical lesions, indicated by arrows) and

angiomyolipoma (AML), at a frequency of about 2% of all TSC patients. Another remarkable feature of TSC hamartomas is that they first appear at different ages during the patient's life, and occasionally spontaneously resolve. For example, cardiac **rhabdomyomas** are often present at birth, but then typically disappear during childhood.

Brain Involvement

The cardinal feature of TSC is the **cortical tuber**, which is a distinctive form of brain hamartia (Fig. 1, top) (1, 3). Grossly, tubers are firm, smooth, somewhat raised, pale lesions of the cerebral cortex and can number up to forty in one brain, ranging in size from a few millimeters to several centimeters. The second common brain lesion seen in TSC is the **subependymal nodule** (SEN), typically found in the lateral ventricle. They occur as smooth nodules projecting into the ventricle and typically calcify in patients past the age of 10.

Histologically, cortical tubers are characterized by architectural disarray with disruption of cortical lamination (Fig. 1, bottom) (1, 3). Tubers contain a mixture of cells, including normal and dysplastic neurons, astrocytes and the hallmark giant cell. Giant cells extend short aberrant processes and express both glial and neuronal proteins. Dysplastic neurons in tubers show disrupted radial orientation and abnormal dendritic arbors and express both typical neuronal and embryonic neuroepithelial proteins. SENs consist of large cells in a vascular stroma, with occasional cells that attain giant cell proportions with aberrant and/or multiple nuclei. **Subependymal giant cell astrocytomas** (SEGA) develop from a SEN in 5–10% of TSC patients. Although there is growth to a size >1 cm, there is no histological progression. These lesions must be treated by surgical resection to prevent hydrocephalus and other adverse effects.

The major presenting symptoms and signs of cerebral TSC are a variety of epileptic seizures (1, 3). These seizures have their origin in the cortical tubers or in nearby transitional cortex. About two-thirds of TSC patients first present with **infantile spasms** usually between the ages of 3 and 9 months. There is a correlation between the number of cortical tubers in a TSC patient and both a younger age at onset of seizures and severity of seizures.

Mental retardation and developmental problems are serious consequences of brain involvement by cortical tubers. Mental retardation is variable but can be

subependymal nodules lining the ventricles (arrowheads). Courtesy of J. Egelhoff, Cincinnati, OH. (bottom) High power view of a cortical tuber shows giant cells (white diamonds). Cortical surface is to the right. Courtesy of J. Chan, Boston, MA.

profound, and rarely occurs in the absence of a generalized seizure disorder. Developmental problems are very common in TSC, and include autistic behavior, hyperactivity, sleep disorders and aggressive behavior.

Kidney and Skin Involvement

► **Angiomyolipomas** (AMLs) are the most common lesion in the TSC kidney, being seen in about 75% of children by adolescence using ultrasound imaging. These lesions consist of a variable mixture of aberrant vessels, smooth muscle, and fat. AMLs are typically bilateral and of variable size. Lesions >5 cm diameter are not rare, and large lesions are associated with a significant risk of bleeding as well as loss of nephrons due to compression.

Cutaneous findings are present in the great majority of TSC patients and are the most easily identified sign of the disease. Most lesions are of minor clinical significance, but facial angiofibromas can be a significant cosmetic issue.

Hypomelanotic macules or white spots typically have a lance-ovate shape (ash-leaf) and are most common over the trunk and buttocks. Three or more of these lesions are very unusual in the general population and thus are considered a major diagnostic criterion (see list under Definition). Histopathological examination shows reduced number, pigment content and size of melanosomes within the melanocytes in a hypomelanotic macule. The lesions are present at birth and do not change, though they are more easily observed following sun tanning.

Facial ► **angiofibromas** (formerly known as adenoma sebaceum) are red to pink papules or nodules with a smooth surface that are found in a malar distribution and extending down to the chin. They typically first appear between the ages of 2 and 6 and progress to a variable extent during puberty. Histological findings are dermal fibrosis and angiogenesis with occasional large glial appearing cells. Forehead fibrous plaque is a larger lesion related to facial angiofibroma, and shagreen patch is also a related lesion found typically on the upper buttock.

Ungual fibromas are another common cutaneous manifestation of TSC, being seen in most adults over the age of 30 years, but rarely prior to adolescence. These lesions are red- or flesh-colored papules or nodules occurring in the finger or toe nail regions. Their histopathology is similar to that of facial angiofibromas.

Other Sites of Involvement

► **Lymphangiomyomatosis** (LAM) is a disorder seen almost exclusively in females, which is characterized by smooth muscle cell proliferation and cystic changes in the lung parenchyma. Most LAM patients also have AMLs in the kidney or elsewhere in the abdomen.

LAM presents clinically as progressive dyspnea, or acute dyspnea due to pneumothorax. The natural history of LAM is poorly defined. There are women in whom it is progressive and fatal. Recent studies indicate that about 40% of adult TSC women have CT scan evidence of LAM. It is likely that most of those in whom it is asymptomatic and detected by CT will not have significant symptoms from this condition.

The retina is involved by hamartomas in about 50% of TSC patients. Cardiac rhabdomyomas are also found in most TSC infants when echocardiographic screening is done. These lesions can be up to several centimeters in diameter and consist of glycogen-filled myocytes. In most cases, there are no symptoms and the lesions spontaneously decrease in size or disappear over time. Less often, rhabdomyomas cause heart failure due to obstruction, affect myocardial function by replacement or cause rhythm disturbances and surgical resection is considered.

Cellular and Molecular Regulation

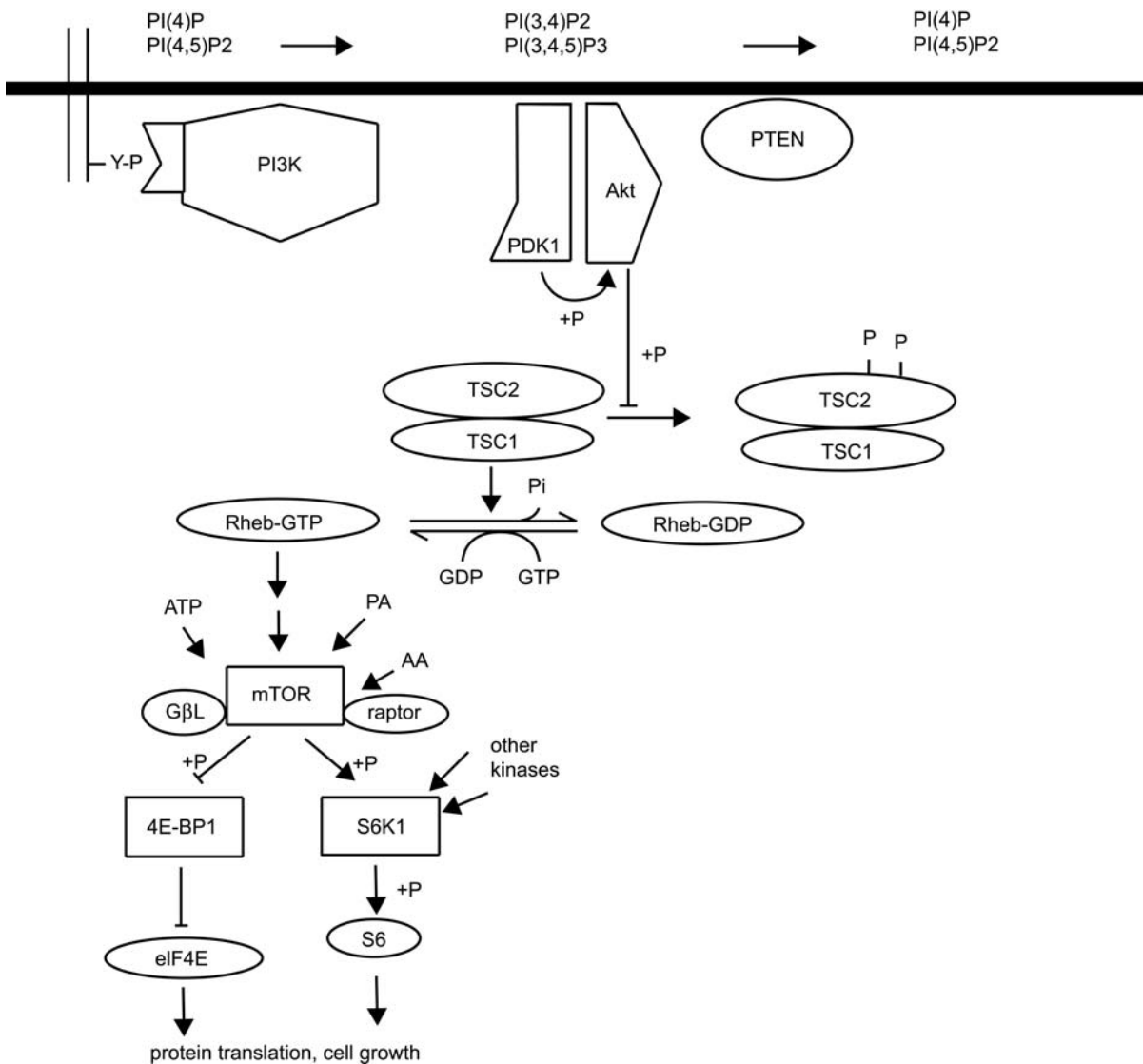
Clinical Genetics of TSC

TSC occurs as an autosomal dominant disorder. Large families with the condition are rare due to reduced reproduction among those affected. About two-thirds of TSC cases are sporadic, representing new mutations. TSC occurs in up to 1 in 6,000 live births without apparent ethnic clustering.

TSC Genes, Mutations, and the Two Hit Mechanism

TSC occurs due to inactivating mutations in either of *TSC1* or *TSC2*. These genes are both relatively large, and encode the proteins hamartin (130 kD) and tuberlin (200 kD), respectively. Over 600 *TSC1* and *TSC2* mutations (over 300 unique) have been identified, and comprehensive mutational analyses indicate that about 85% of TSC patients will have a mutation in either *TSC1* or *TSC2* identified (4, 5). *TSC2* mutations are about 4.2× as common as *TSC1* mutations, reflecting a higher germline mutation rate. Clinical surveys also indicate that patients with *TSC1* mutations have symptoms and clinical features that are milder on average than patients with *TSC2* mutations, although there is considerable overlap in the two sets of patients (5). Linkage studies provide evidence against a third TSC gene. Patients in whom mutations cannot be found are probably partially explained by ► **mosaicism**, which compromises efforts at mutation detection.

TSC hamartomas, particularly renal AMLs, often display ► **loss of heterozygosity** (LOH) for the wild type allele of either *TSC1* or *TSC2*, consistent with a two hit mechanism for complete inactivation of either *TSC1* or *TSC2* (4). LAM lesions require microdissection of the SMCs to demonstrate LOH events and such methods have been used to show that patients with LAM but without other features of TSC also have two



Tuberous Sclerosis. Figure 2 Signaling pathway model for the function of TSC1 and TSC2 in mammalian cells. A phosphorylated growth factor receptor is shown at upper left, to which a PI3K molecule is binding. This leads to conversion of the indicated phosphoinositides to 3'-phosphoinositides, which leads to recruitment of Akt to the membrane in a position where it can be phosphorylated and activated by PDK1 and a second kinase. PTEN functions to terminate this signaling pathway by acting as a 3' phosphatase on these phosphoinositides. Activated pAkt phosphorylates TSC2, which inactivates its GAP activity. When active, TSC1/TSC2 complex serves as a GAP for Rheb, reducing levels of Rheb-GTP. Rheb-GTP activates mTOR by an uncertain mechanism (thus 2 arrows). ATP, phosphatidic acid (PA), and amino acids (AA) all influence mTOR activity, although the sensing mechanisms are unknown and probably indirect. Active mTOR phosphorylates 4E-BP1 and S6K1. p4E-BP1 releases from eIF4E, permitting formation of the eIF4F translation complex. pS6K1 phosphorylates S6 and together they activate the translational machinery. For simplicity, only the main pathway involving TSC1, TSC2, and mTOR is shown. Arrows indicate positive actions and bars represent negative actions.

hit inactivation of *TSC2*. Cortical tubers do not show evidence of LOH, even with microdissection and alternative mechanisms of gene inactivation or cell admixture probably explain this finding. LOH is also seen in the tumors (kidney cystadenomas, liver

hemangiomas and extremity angiosarcomas) that develop in *Tsc1*^{+/-} and *Tsc2*^{+/-} mice.

Consistent with a relative lack of malignancy in TSC patients, limited surveys of human cancer specimens have failed to show evidence of either *TSC1* or *TSC2*

mutation, with the possible exception of bladder carcinoma.

TSC1/TSC2, PI3K, and mTOR Signaling

Major progress has occurred in recent years in our understanding of the biochemical function of the *TSC1* and *TSC2* gene products, initiated by seminal studies in *Drosophila* (6). Work in mammalian cells has led to the model of PI3K-Akt-TSC1/TSC2-Rheb-mTOR signaling that is shown in Fig. 2. In normal quiescent cells, PI3K and Akt are inactive, the TSC1/TSC2 complex is active as a GTPase activating protein (GAP) for Rheb, there are low levels of Rheb-GTP and mTOR is inactive. In response to growth factor stimulation, PI3K and Akt become activated, TSC2 is phosphorylated by Akt and the TSC1/TSC2 complex becomes inactive as a GAP, so that Rheb-GTP levels rise, stimulating mTOR. In cells lacking TSC1 or TSC2, there is no GAP for Rheb and Rheb-GTP levels are high, leading to constitutive activation of mTOR and phosphorylation of S6K1 and 4E-BP1. This model captures most of the current understanding of the roles of TSC1, TSC2, Rheb and mTOR in this pathway, but simplifies many aspects, omitting the multiple targets of Akt and the complex hierarchical phosphorylation of S6K1 that regulates its activity, as just two examples. One clinical correlate of this model is that hamartomas that occur in TSC both in patients and mouse models typically express pS6 and pS6K1, signposts of activation of mTOR.

The molecular basis of epileptogenesis in cortical tubers is uncertain, beyond the general lack of organization of the cortex in these lesions. There is increased transcription of genes encoding glutamatergic receptors in dysplastic neurons and giant cells, with reduced expression of gamma-aminobutyric acid (GABA)-ergic receptors. These expression changes may contribute to epileptogenesis in TSC (3). Recent work indicates that the tuber giant cells and SEN cells express pS6 and pS6K1 like other TSC lesions, suggesting that there is complete inactivation of *TSC1* or *TSC2* with activation of mTOR as in other TSC hamartomas.

Other Functions of TSC1/TSC2

The TSC1/TSC2 complex is 330 kD in size and the GAP domain of TSC2 comprises about 10 kD of this complex. This alone suggests that there are other functions for the complex. TSC1 has been reported to bind to ezrin and other ERM family proteins, and appears to be involved in adhesion events and rho signaling to the actin cytoskeleton. TSC2 has been reported to have a role in the membrane localization of polycystin-1 in renal epithelial cells. A role for the TSC1/TSC2 complex in beta-catenin signaling has also been noted. Whether any of these observations are

independent of or relate to the role of TSC1/TSC2 in the PI3K signaling pathway is unknown.

Clinical Relevance

TSC is a disorder that affects approximately 40,000 people in the USA. It is remarkable in several respects; the frequency and severity of associated epileptogenesis, the unique predisposition to autism and the unusual spectrum of associated tumors including LAM. The placement of the TSC genes in the PI3K-Akt-TSC1/TSC2-Rheb-mTOR signaling pathway is a major advance and highlights the potential importance of the *TSC1* and *TSC2* genes in other disorders affecting this pathway, including diabetes mellitus. This recent pathway discovery also provides the possibility of several pharmacological approaches that may be effective for TSC. These include drugs already in use in the clinic, including rapamycin and analogues, farnesyl transferase inhibitors, angiogenesis inhibitors and interferon- γ . Current ongoing and planned trials of these agents in both preclinical models and TSC patients are of considerable clinical and basic interest.

Acknowledgements

Supported by the NIH NINDS (NS31535, NS24279), the LAM Foundation, and the Rothberg Courage Fund.

References

1. Gomez M, Sampson J, Whittemore V (1999) The tuberous sclerosis complex. Oxford University Press
2. Roach ES, Gomez MR, Northrup H (1998) Tuberous sclerosis complex consensus conference: revised clinical diagnostic criteria. *J Child Neurol* 13:624–8
3. Crino PB, Henske EP (1999) New developments in the neurobiology of the tuberous sclerosis complex. *Neurology* 53:1384–90
4. Cheadle JP, Reeve MP, Sampson JR et al (2000) Molecular genetic advances in tuberous sclerosis. *Hum Genet* 107:97–114
5. Dabora SL, Jozwiak S, Franz DN et al (2001) Mutational Analysis in a Cohort of 224 Tuberous Sclerosis Patients Indicates Increased Severity of TSC2, Compared with TSC1, Disease in Multiple Organs. *Am J Hum Genet* 68:64–80
6. Kwiatkowski DJ (2003) Rhebbing up mTOR: new insights on TSC1 and TSC2, and the pathogenesis of tuberous sclerosis. *Cancer Biol Ther* 2:471–6

Tubulin

► Alpha-Tubulin/Beta-Tubulin

Tumor Angiogenesis

Definition

Most tumors need blood vessels in order to receive oxygen and nutrients necessary for their unrestricted growth. Tumors are able to build a functional vascular system by recruiting, stimulating and remodeling the host's vascular bed.

► [Angiogenesis](#)

inflammation and the control of cell proliferation, differentiation and apoptosis. It mainly exerts pro-inflammatory effects, but is also involved in anti-inflammatory mechanisms. Binding of TNF α to its two receptors, TNFR1 and TNFR2, results in signaling to at least three distinct effector pathways, and to activation of caspases and two transcription factors, AP-1 and NF κ B.

► [Apoptosis](#)
 ► [AP-1](#)
 ► [Catalytic RNA](#)
 ► [Crohn Disease](#)
 ► [Cytokines](#)
 ► [Growth Factors](#)

Tumor Necrosis Factor

Definition

As the name implies, tumor necrosis factor (TNF) can induce cell death. However, this cytokine has multiple functions and is involved in pro-inflammatory as well as antiinflammatory mechanisms.

► [Apoptosis](#)
 ► [Apoptosis, Regulation and Clinical Implications](#)
 ► [Cytokines](#)
 ► [Growth Factors](#)

Tumor Promoter Genes

Definition

Tumor promoter genes are normal or altered genes, the activation (stimulation of transcription) of which contributes to the initiation and development of tumors in general, or to separate aspects of tumor progression such as ► [Cancer Invasion and Metastasis: Cellular, Molecular and Clinical Aspects](#).

Tumor Necrosis Factor Receptor-Associated Factor

Definition

The TNF-receptor-associated factor (TRAF) family is a phylogenetically conserved group of scaffold proteins, with a pivotal role in the signal transduction of the TNF receptor superfamily and the interleukin-1 receptor/Toll-like receptor superfamily. TRAF proteins mediate activation of NF κ B and mitogen-activated protein kinases and serve as a docking platform for a variety of regulators of these signalling pathways.

► [TNF Receptor/Fas Signaling Pathways](#)

Tumor Suppressor Gene

Definition

Tumor suppressors comprise of a group of genes whose gene products reduce the probability of malignant transformations by repressing cell proliferation. As the name implies, they inhibit tumor formation. Loss of their function by mutations or loss promotes malignant transformation. According to the two-hit model by Knudsen, a mutation in only one allele of a tumor suppressor ("first hit") has no effects on the phenotype unless followed by a further mutation of the second allele ("second hit"). Frequently, mutation of one allele consists of a specific alteration of the coding sequence of a tumor suppressor resulting in a nonfunctional protein, whereas mutation of the other allele involves larger chromosomal alterations such as deletions, gene conversion and others. Alternative mechanisms of tumour suppressor inactivation are epigenetic silencing by methylation, expression blocks induced by upstream regulatory factors or defects in the regulatory region of the gene, e.g. preventing transcription factor binding.

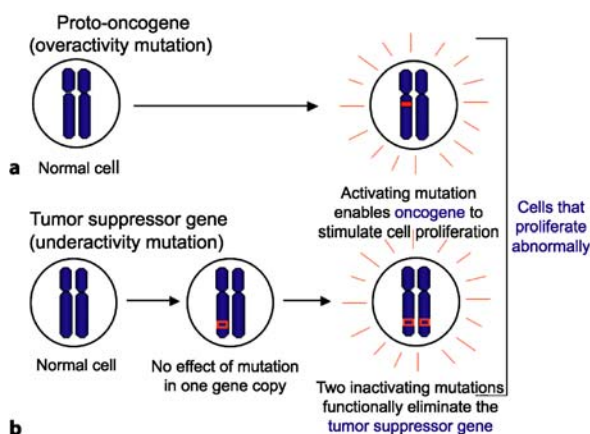
Tumor Necrosis Factor- α

Definition

Tumor necrosis factor- α (TNF- α) is a pleiotropic cytokine that plays an important role in immunity,

Sometimes, mutations of tumor suppressors occur in an inherited, familial form resulting in a predisposition of patients for development of multifocal tumors at a low age. Prominent tumor suppressors are APC and p53.

- Breast Cancer
- Cancer Invasion and Metastasis: Cellular, Molecular and Clinical Aspects
- Colorectal Cancer
- Double-Strand Break Repair
- Genomic Information and Cancer
- Hedgehog Signalling
- Mouse Genomics
- Neurofibromatosis Type 1 (NF1), Genetics
- Protein/DNA Interaction
- Ras Signalling
- RNA Polymerase III
- Tumor Suppressor Genes



Tumor Suppressor Genes. Figure 1 Uncontrolled cell growth as a consequence of mutations in proto-oncogenes and/or tumor suppressor genes. Adapted from Alberts et al. (2002) *Molecular Biology of the Cell*, Chapter 23, p.1334.

Tumor Suppressor Genes

SUSAN WIJNHOFEN, ANNEMIEKE DE VRIES
National Institute of Public Health and the Environment, Laboratory of Toxicology, Pathology and Genetics, Bilthoven, The Netherlands
susan.wijnhoven@rivm.nl, annemieke.de.vries@rivm.nl

Definition

►Cancer arises as the result of an accumulation of inherited and somatic mutations in cancer-critical genes. These genes can be divided in two broad classes, depending on whether the cancer risk arises from too much activity of a gene or too little. Mutation of a single copy of a ►proto-oncogene can have a dominant, growth-promoting effect on a cell. In the case of a ►tumor suppressor gene, the cancer-causing mutations are generally recessive; both copies of the gene must be removed or inactivated in the normal cell before an adverse effect is observed (Fig. 1).

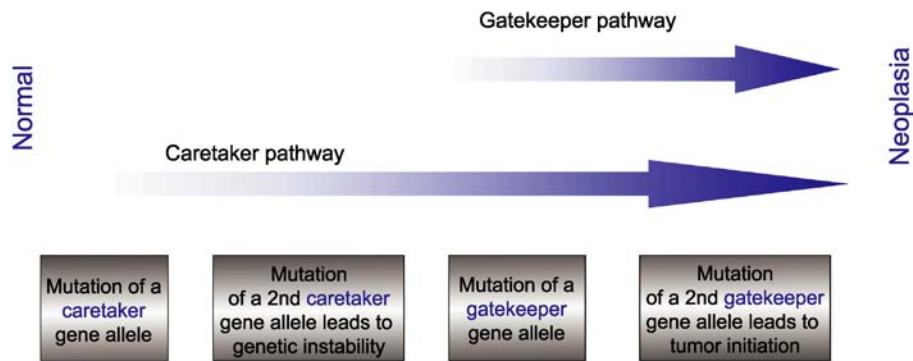
Tumor suppressor genes (TSG) encode proteins functioning in growth regulatory and/or differentiation pathways. The protein products of TSG are needed for proper cell functioning, and as a consequence, loss of their function can cause uncontrolled cell growth and ultimately tumors. Therefore, TSG are mainly detected by deletions or other inactivating mutations in tumors. The most compelling evidence for their important tumor suppressive role however is provided by certain ►hereditary cancers in which patients with the disease develop tumors that have lost both alleles, thereby lacking an active tumor suppressor gene. Also the progression of a wide range of sporadic and hereditary

cancers may be associated with inactivating changes in TSG (1). Mutations in TSG can either directly or indirectly increase the risk of cancer development (Fig. 2) (2). Mutations in so-called ►caretaker genes affect DNA damage repair, leading to generalized genomic instability, increased mutation rates and ultimately cancer. Examples of tumor suppressor genes that probably have an impact on DNA damage repair include BRCA1 and BRCA2, both associated with breast cancer development. Mutations in ►gatekeeper genes, such as those that regulate the ►cell cycle or those that promote ►apoptosis, can lead directly to a cellular malignant change, as these functions are the final checkpoints in the maintenance of DNA integrity. Key members of this class of tumor suppressors are ►p53 and ►RB.

Characteristics

Identification of Tumor Suppressor Genes

In contrast to the relatively straightforward approaches to the identification of oncogenic alleles in cancer, the identification of tumor suppressor genes has proved to be quite difficult. Oncogenes, with a ►gain-of-function mutation driving the cell towards cancer (Fig. 1), can be detected by their growth promoting effect and/or transforming capacity when added to the genome of a suitable type of tester cell. This can be performed by, for example, DNA transfection or through infection with a viral vector. In the case of a tumor suppressor gene, a different approach is needed. By studying rare cancers that run in families (hereditary cancers), a few key tumor suppressor genes (RB and p53) with universal relevance have been discovered (see also



Tumor Suppressor Genes. Figure 2 Pathways to neoplasia. In the caretaker pathway, multiple genes need to be inactivated in the cell to become neoplastic. First, two alleles of a caretaker gene are mutated. The inactivation of the caretaker gene subsequently accelerates the accumulation of the mutations in gatekeeper genes. In contrast, in the gatekeeper pathway, only one gene inactivation is needed to initiate neoplasia. Adapted from Kinzler and Vogelstein. (1997) *Nature*, 386, p. 763.

below). However, when there is no such clue from a hereditary syndrome, TSG have to be identified by virtue of their absence from tumor cells. Current methods depend on scanning the genomes of cancer cells for signs of gene loss. These methods are often employed in concert and involve cytogenetic studies of constitutional chromosomal alterations in cancer patients, linkage analysis to localize genes that predispose to cancer and loss of heterozygosity (or allelic loss) studies undertaken on matched pairs of normal and cancer tissues. Loss of heterozygosity (LOH) appears the most common molecular genetic alteration in human cancer. In sporadic tumors, up to 50% of all chromosomes may have undergone LOH events.

An increasing number of tumor suppressor genes with critical roles in cancer predisposition, and associated with a hereditary cancer syndrome, have been identified and cloned (Table 1) (10).

RB and Retinoblastoma

The key insight that led to the discovery of *RB*, the first identified tumor suppressor gene, came from studies of a rare type of human cancer called ►retinoblastoma. Retinoblastoma occurs in childhood with tumors developing from neural precursor cells in the immature retina. In the hereditary form of this disease, multiple tumors arise in both eyes, whereas in the non-hereditary form a single tumor arises in each eye. All retinoblastoma patients (both hereditary and non-hereditary) have an abnormal ►karyotype, with a deletion of a specific region of chromosome 13. Using the known location of the chromosomal deletion associated with the disease, it was possible to clone and sequence the *RB* gene, whose loss appeared to be critical for the development of the cancer. In patients who suffer from the hereditary form of the disease, a deletion or loss of

function mutation is present in one copy of the *RB* gene in all cells of the body, which makes these cells *RB* heterozygous and as such predisposed to become cancerous (3). Analysis of a variety of somatic tumors revealed mutations in *RB* in several common types of tumors, including carcinomas of the lung, breast, and bladder. In all of these tumors, complete loss of *RB* is frequently a major step in the progress towards malignancy. The RB protein turns out to be a universal regulator of the cell-cycle that is normally expressed in almost all cells of the body (4).

p53 and the Li-Fraumeni Syndrome

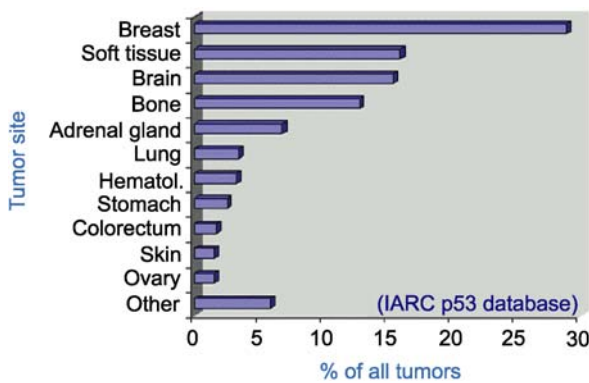
The ►Li-Fraumeni syndrome (LFS) is a rare ►autosomal, dominantly inherited disorder characterized by the diagnosis of bone or soft-tissue ►sarcoma at a relatively young age. Already in 1969, Li and Fraumeni identified the first 4 families in which sibs or cousins developed a childhood sarcoma, together with striking histories of breast cancer and other neoplasms. This suggested a new familial cancer syndrome of diverse tumors, including soft tissue sarcomas, breast cancer, brain tumors, osteosarcomas, leukemia and adrenocortical carcinoma (Fig. 3) (5). Since then, a lot of other families have been identified with the classic phenotype of LFS (bone- or soft-tissue sarcoma at an early age with one first-degree relative with early-onset cancer or sarcoma diagnosed at any age). In LFS-like (LFS-L) families the phenotype of LFS is not expressed completely. These families show a combination of multi-cancer phenotypes, but fail to meet the stringent criteria for LFS. However, common to all these families is the occurrence of a variety of cancers at a relatively early age (5). Germ-line alterations in the *p53* tumor suppressor gene, located on chromosome 17, have been observed in the majority (~80%) of LFS families and

Tumor Suppressor Genes. Table 1 Human (candidate) tumor suppressor genes associated with familial syndromes (see for extensive list (10))

Gene	Cancer syndrome	Principal tumors
RB1	Retinoblastoma	Retinoblastoma, osteosarcoma
<i>p53</i>	Li-Fraumeni	Sarcomas, breast and brain tumors
APC	Familial adenomatous polyposis	Adenomatous polyps, colon cancer
WT1	Wilms' tumor	Nephroblastoma
NF1	Neurofibromatosis type I	Neurofibromas, sarcomas, gliomas
NF2	Neurofibromatosis type II	Schwannomas, meningiomas
VHL	von-Hippel Lindau	Renal cell (clear cell) pheochromocytomas, hemangiomas
<i>CDKN2A</i> <i>P16^{Ink4a}</i>	Familial malignant melanoma	Melanoma, pancreatic cancer, other?
<i>CDKN2A</i> <i>P14^{ARF}</i>	Familial malignant melanoma	Melanoma, pancreatic cancer, other?
BRCA1	Familial breast cancer	Breast and ovarian cancer
BRCA2	Familial breast cancer	Breast and other
TSC1	Tuberous sclerosis I	Renal and brain tumors
TSC2	Tuberous sclerosis II	Renal and brain tumors
PTEN	Cowden syndrome, Bannayan-Riley-Ruvalcaba syndrome	Hamartoma, glioma, prostatic, endometrial cancer
CDH1(E-cadherin)	Familial gastric carcinoma	Gastric and lobular breast
CYLD	Familial cylindromatosis	Cylindromatosis
EXT1	Multiple exostoses I	Exostoses, osteosarcoma
EXT2	Multiple exostoses II	Exostoses, osteosarcoma
FH	Hereditary leiomyomatosis and renal-cell cancer	Leiomyomatosis, renal
MADH4	Juvenile polyposis	Gastro-intestinal polyps
MEN1	Multiple endocrine neoplasia type I	Parathyroid adenoma, pituitary adenoma, pancreatic islet cell, carcinoid
NBS1	Nijmegen Breakage syndrome	NHL, glioma, medulloblastoma, rhabdomyosarcoma
PTCH	Nevoid basal-cell carcinoma syndrome	Skin basal cell, medulloblastoma
SDHB	Familial paraganglioma	Paraganglioma, pheochromocytoma
SDHC	Familial paraganglioma	Paraganglioma, pheochromocytoma
SDHD	Familial paraganglioma	Paraganglioma, pheochromocytoma

in a proportion (~20%) of LFS-L families. These mutations in *p53* are inherited as autosomal dominant, so affected individuals are heterozygotes with a missense *p53* mutation in one allele. The mutations

are thought to behave as ► **dominant-negatives**, inhibiting the function of the wild type *p53* protein (6). However, more than half of all tumors carrying *p53* mutations have lost the remaining wild type allele



Tumor Suppressor Genes. Figure 3 Tumor spectrum in Li-Fraumeni patients.

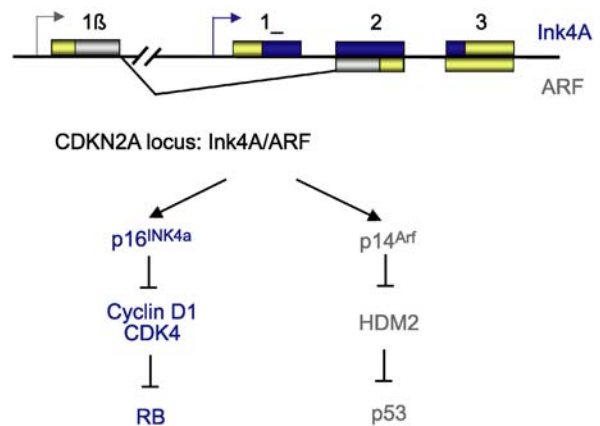
(►loss of heterozygosity, LOH), suggesting that complete loss of p53 is essential for tumor development and/or maintenance.

Apart from Li-Fraumeni patients, p53 mutations are also frequently found in ►sporadic tumors. The first p53 mutation in an individual with sporadic cancer was reported as early as 1989. Astonishingly, it now seems that more than half of all human cancers lose p53 function by mutation. Due to the increased number of sequences of human cancers that have been added to a database, more than 10,000 somatic tumorigenic p53 mutations are known at this moment. Interestingly, certain ►hotspot mutations are frequently observed, with the specific mutation dependent on the tumor type (7).

In normal cells, the induction of a functional p53 protein in response to stress leads to either reversible inhibition of cell growth through the induction of a cell-cycle arrest enabling DNA repair, or the induction of apoptosis if DNA damage is too severe. Based on these functions, p53 is also frequently referred to as the “guardian of the genome” (8).

Molecular Interactions/Regulatory Mechanisms

Although it is clear that all tumor suppressor proteins exhibit a variety of functions, the cellular functions and regulators of a number of TSG remain largely undefined. However, the regulation of both *RB* and *p53* is well elucidated. These two transcription factors are regulated by the *CDKN2A* locus, a locus that encodes two different proteins, p16^{INK4a} and p14^{ARF}. Both proteins are necessary to regulate the cellular lifespan via two different pathways, the *RB* and *p53* pathways respectively (Fig. 4). In the *RB* pathway, expression of p16^{INK4a} functionally inhibits the cyclin D-CDK4/6 complexes that regulate *RB*. p14^{ARF} inhibits the ubiquitin ligase HDM2, leading to elevated levels of p53. Both pathways are known to be



Tumor Suppressor Genes. Figure 4 The human Ink4A/Arf locus, encoding two different proteins. First exons 1α and 1β are transcribed from different promoters resulting in separate 5'ends of the Ink4A and Arf RNA transcripts. Both are spliced to the acceptor splice site in exon 2. As such, 2 different proteins are produced regulating the *RB* or the *p53* pathway. Adapted from Sherr (2001) *Nature Reviews Cancer*, 2, p. 732.

important in the regulation of the process of immortalization of cells, an essential prerequisite for the formation of a tumor cell (4).

RB

The *RB* protein and its relatives p107 and p130 constitute the “pocket protein” family of cell-cycle regulators. Of this protein family, only *RB* has been demonstrated to have real tumor suppressor properties. It inhibits cell proliferation by preventing entry into the DNA synthesis (S) phase. The main targets of the pocket proteins are E2F transcription factors that are blocked from activating DNA replication, nucleotide metabolism genes such as cyclin A, E and D1 and thymidine kinase. However, it has recently been discovered that *RB*, p107 and p130 are also involved in the regulation of terminal differentiation of various tissues and cell types. *RB* appears to play an important but complex role in multiple stages of the differentiation process including irreversible exit from the cell-cycle, protection from apoptosis, induction of cell type-specific gene expression and maintenance of the post-mitotic state by interaction with many diverse transcriptional regulators (9).

p53

In unstressed cells, p53 is present in a latent form and is maintained at low levels through targeted degradation. However in response to damaged DNA, nucleotide depletion, hypoxia and several other genotoxic stresses,

p53 accumulates in the nucleus and is activated to exert its function as a transcription factor to prevent tumor development. The way p53 performs its tumor suppressor role involves diverse cellular processes. The repertoire of p53 activities includes regulation of the cell-cycle, induction of apoptosis, ►senescence, facilitating DNA repair and antagonizing angiogenesis. Many of these functions are mediated by transcriptional activation of target genes by p53. On the basis of these functions that all regulate cell growth and maintain genome integrity, the tumor suppressor p53 is also known as the “guardian of the genome” (8).

Conclusions and Outlook for the Future

An important goal of cancer research has been the identification of so-called cancer genes, i.e. genes that are frequently mutated in and, as a consequence, causally related to human cancers. Knowledge about these genes has evolved enormously, and thus far, a total of 291 genes has been identified representing more than 1% of all human genes. With the development of new and highly innovative genomics technologies such as micro-array expression analysis and comparative genome hybridization, it is to be expected that many more cancer genes, including TSG, will be identified in the near future. In addition, revealing the human genome sequence, as well as that from other organisms such as the mouse, will undoubtedly simplify the identification of these genes. Finally, generating tissue-specific knockout mouse models using Cre-loxP based technologies will enable a more detailed and *in vivo* analysis of the functions of these genes.

References

1. Fearon E (1998) Tumor Suppressor Genes. In: Vogelstein B, Kinzler KW (eds) *The Genetic Basis of Human Cancer*. Mc-Graw-Hill Companies Inc. USA, pp 229–236
2. Kinzler KW, Vogelstein B (1997) Gatekeepers and Caretakers. *Nature* 386:761–763
3. Cavenee WK, Dryja TP, Phillips RA et al (1983) Expression of recessive alleles by chromosomal mechanisms in retinoblastoma. *Nature* 305:779–784
4. Sherr C (2001) The Ink4a/ARF network in tumour suppression. *Nat Rev Mol Cell Biol* 2:731–737
5. Malkin D (1998) The Li-Fraumeni Syndrome. In: Vogelstein B, Kinzler KW (eds) *The Genetic Basis of Human Cancer*. Mc-Graw-Hill Companies Inc. USA, pp 393–407
6. Varley JM, Evans DGR, Birch JM (1997) Li-Fraumeni syndrome – a molecular and clinical review. *Brit J Cancer* 76:1–14
7. IARC TP53 mutation database: ►<http://www.iarc.fr/p53>
8. Vousden KH, Lu X (2002) Live or let die: The cell's response to p53. *Nat Rev Cancer* 2:594–604

9. Lipinsky MM, Jacks T (1999) The retinoblastoma gene family in differentiation and development. *Oncogene* 18:7873–7882
10. Futreal PA, Coin L, Marshall M et al (2004) A census of human cancer genes. *Nat Rev Cancer* 4:177–183

Tumor Viruses

Definition

Tumor viruses comprise of a heterogeneous group of viruses defined by the causative role of its members in naturally occurring malignant diseases, or their capacity to induce tumours in experimentally infected laboratory animals. Most of them are hepadnaviruses, papillomaviruses, polyomaviruses, adenoviruses, gamma herpesviruses or retroviruses.

►Viral Oncogenesis

Tumorigenesis

Development of a tumor out of a normal tissue/cell type.

►Microarrays in Pancreatic Cancer

Tunicamycin

Definition

Tunicamycin is a nucleoside antibiotic from *Streptomyces Lysosuperificus*, which acts in eukaryotic cells to inhibit N glycosylation by inhibiting the first step in the synthesis of the dolichol linked oligosaccharide, i.e. by preventing the formation of Dol-P-P-GlcNAc from Dol-P and UDP-GlcNAc.

►Glycosylation of Proteins

Two Channel Microarrays

Definition

Two channel microarrays defines a particular usage of microarrays, where two differently labelled nucleic

acid extracts are hybridised onto the array simultaneously, for measuring their relative abundance.

► [Microarray Data Analysis](#)

Two-Dimensional Crystal

Definition

Two-dimensional crystal characterizes a regular assembly of three-dimensional molecules in a planar array.

► [Two-dimensional Crystallization of Membrane Proteins](#)

Two-dimensional Crystallization of Membrane Proteins

THOMAS BRAUN, THOMAS C. KAUFMANN,
HERVÉ-W. RÉMIGY, ANDREAS ENGEL
M.E. Müller
Institute for Structural Biology, Biocenter,
University of Basel, Basel, Switzerland
andreas.engel@unibas.ch

Definition

To determine the structure of a membrane protein, two-dimensional (2D) or three-dimensional (3D) crystals are required. 2D crystallization assembles membrane proteins into highly ordered 2D arrays in the presence of lipids. In this case, membrane proteins are embedded in their native environment, the lipid bilayer, and are thus in a functional state. Their structure is then assessed by ► [electron crystallography](#).

Description

Membrane proteins comprise more than 30% of the proteome of higher organisms. With characteristic dimensions of 5–10 nm they are membrane-embedded nanomachines that fulfill key functions such as energy conversion, solute transport, secretion and signal transduction. Their central role in a wide range of diseases and in cell-cell communication may explain the fact that 70% of all drug targets are membrane proteins. While the structures of more than 10,000 soluble proteins are solved, the number of membrane

protein structures is smaller than 100. This lack of structural information is related to the instability of membrane proteins in a ► [detergent](#)-solubilized state, making the growth of 3D crystals difficult. To understand the function of membrane proteins knowledge of their structure is urgently required. 2D crystals of purified membrane proteins reconstituted in the presence of lipids provide a close to native environment. Electron crystallography allows the atomic structure of a membrane protein packed in a perfect 2D crystal to be elucidated at atomic resolution. In contrast, ► [atomic force microscopy](#) (AFM) provides the surface structure of membrane proteins from tightly packed, but disordered membranes at sub-nanometer resolution. Combining these methods the structure and function of membrane proteins can be directly assessed.

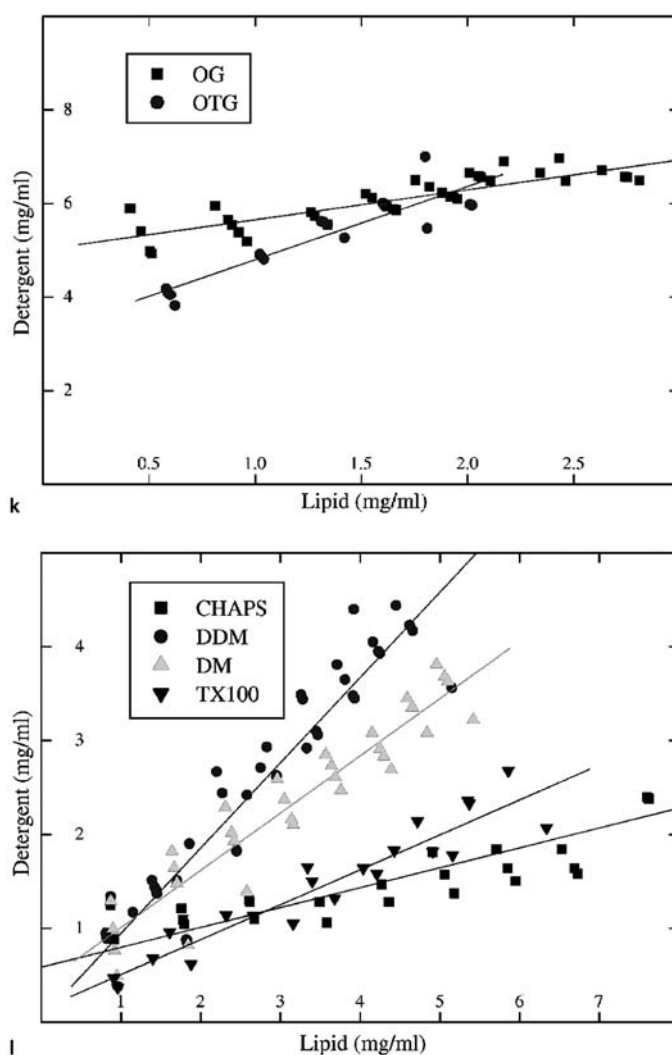
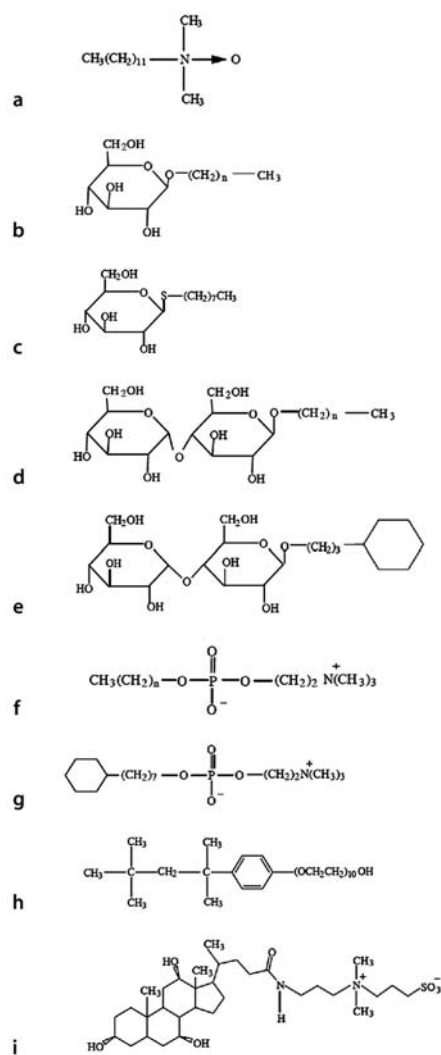
To obtain 2D crystals the protein of interest has to be expressed in mg quantities, solubilized and purified in a detergent and finally reconstituted into an artificial lipid bilayer. The interactions of lipids, proteins and detergents are complex and difficult to predict. Thus, 2D crystallization is still an experimental art, in spite of accumulated knowledge summarized in recent reviews (1, 2, 3) each new membrane protein requires specific optimizations to be crystallized. In this chapter the general trends and current techniques of 2D crystallization are described, emphasizing the importance of membrane ► [solubilization](#).

Membrane Protein Purification

Detergents

Detergents are amphiphilic molecules with limited water solubility as monomers. However, above their ► [critical micellar concentration](#) (CMC), the structure of such molecules promotes clustering into micelles. Thereby their hydrophobic part is buried in the micelle core and is shielded from water by the hydrophilic micelle surface. The CMC is determined by the balance between the solvation energy provided by the hydrophilic head group and the entropic cost of transferring the hydrophobic tail into water. Above the CMC, hydrophobic substances can be solubilized by embedding them in the hydrophobic core of the micelles. Depending on the structure of the detergent, a micelle can assume a spherical or rather a disk-like shape, accommodating about hundred or only a few molecules. Accordingly, the solubilization capacities of different detergent types for membrane proteins and lipids vary significantly, and are dictated by their stereochemistry (Fig. 1; see ► www.anatrace.com and LeMaire et al. 2000).

Let's consider the properties of some frequently used detergents. The first group concerns detergents having an alkyl moiety, which all form spherical micelles accommodating about 100 detergent molecules. The



Two-dimensional Crystallization of Membrane Proteins. Figure 1 The structure of commonly used detergents and their lipid solubilization capacity. Three groups are distinguished and discussed in the text. Alkyl based detergents are the most widely used detergents for membrane protein crystallization (a, b, c, d, f). Cyclohexyl based detergents (e, g) have recently been developed and are mild as result of the cyclohexyl group. They are to some extent similar to Triton X-100 (h). Sterol based detergents (i) exhibit a bulky hydrophobic group and accommodate only a few molecules per micelle. Turbidimetry has been used to determine the dissolution point of DMPC in the presence of different detergents (k, l) (7).

zwitterionic dimethyl-amine-N-oxide type detergents (Fig. 1a) have the smallest head group of all detergents shown in Fig. 1. The often-used dodecyl variant (DDAO), also known as lauryl-dimethyl-amine-N-oxide (LDAO; Fig. 1a) exhibits a CMC of 0.14 mM and has an almost cylindrical outer surface and a cross-section of about 0.34 nm^2 (4). β -D-glucopyranoside type detergents have a larger head group and thus a conical shape (Fig. 1b). Octyl- β -D-glucopyranoside (OG) is a frequently used detergent, because it is rather mild in spite of its high CMC (25 mM). Nevertheless, decyl- β -D-glucopyranoside (CMC=2.2 mM) and

dodecyl- β -D-glucopyranoside (CMC=0.13 mM) are often used as well. Octyl- β -D-thio-glucopyranoside (OTG; CMC=9 mM; Fig. 1c) is interesting because of its specific interaction with lipids, which promotes formation of large 2D crystals (1). β -D-maltoside type detergents (Fig. 1d) have a cross-section of about 0.47 nm^2 (4). Octyl- β -D-maltoside (CMC=23 mM), decyl- β -D-maltoside (DM; CMC=1.6 mM) and dodecyl- β -D-maltoside (DDM; CMC=0.15 mM) are commonly used. DDM is a particularly mild detergent, which allows solubilization and purification of fragile membrane proteins. This is the result of appropriate

head group properties and the lipid-like hydrophobic moiety. However, combining a phosphocholine head group with an alkyl tail (Fig. 1f) appears to provide an even milder detergent suitable for membrane protein solubilization. FOS-CHOLINETM detergents are available as octyl up to (CH₂)₁₅ variants, providing a large range of CMCs. Alkyl-polyoxyethylene (C_nE_m) type detergents exhibit an even larger head group. C₁₂E₆ for example, has a cross-section of about 0.66 nm² (4) and has been used for the crystallization of several membrane proteins.

The second group of interest concerns detergents with a bulky hydrophobic tail as results of a cyclohexyl group terminating the (CH₂)_n chain (Figs. 1e, g). This bulky tail is expected to prevent the destabilization of the ternary structure of typical α -helical membrane proteins, because such a tail may not be able to wedge in between the α -helices. The development of the corresponding CYMALTM and CYGLUTM detergent series has been fostered by the success of Triton detergents in functional reconstitution of membrane proteins (Fig. 1h).

The third group considered has an altogether different, steroid-based hydrophobic moiety also found in the bile salts. The structure of one representative of this group, CHAPS (Fig. 1i) shows the trihydroxy-oxocholan structure, a flat disk with 3 hydroxyls on the top side and a hydrophobic bottom side, to which the zwitterion dimethylammonio-propane sulfonate is attached. For such detergents only a few molecules are required to form a micelle, ideally two CHAPS molecules could stack together to shield the hydrophobic oxocholan surface from the aqueous environment. These detergents have a large capacity to solubilize lipids and are relatively mild in spite of high CMC values.

Membrane Protein Solubilization and Purification

The solubilization of lipids has been studied in depth using different physical methods (for recent reviews see BBA Volume 1508, 1-251), leading to a model defining a three-stage solubilization process. In stage I, non-micellar detergent partitions into the phospholipid bilayer, whereas in stage III the phospholipid is fully solubilized by uptake into detergent micelles. In the intermediary stage II, phospholipid membranes saturated with incorporated detergent coexist at thermodynamic equilibrium with mixed phospholipid-detergent micelles saturated with phospholipid. This process is well characterized by light scattering and microcalorimetry for different lipid/detergent systems and described by thermodynamic models (5, 6). The **dissolution points** of dimyristoyl-phosphatidylcholine (DMPC) for different detergents as determined by turbidimetry (7) (Figs. 1k, l) illustrate the lipid solubilization capacity of frequently used detergents.

Beyond the **saturation point** detergent molecules cooperatively interact with the bilayer to produce filamentous lipid assemblies that probably carry detergent caps at their ends. This has been observed by cryo-electron microscopy as well as light and neutron scattering experiments.

In the presence of both lipids and proteins, the situation is more complex (4). Above the critical solubilization concentration of the detergent, lipid and protein containing units exist as mixed lipid/detergent micelles and detergent solubilized proteins, covered by detergent and tightly associated lipids. Mild detergents surround the hydrophobic sector of integral membrane proteins by a micelle-like structure without impairing the protein's function. Depending on the detergent - protein and lipid-protein affinities, the hydrophobic sector of the protein can bind detergent even at concentrations below the CMC, which displaces lipids from the protein surface. More favorable is the situation where tightly associated lipids remain bound to the protein, preserving the native state even after solubilization. Since the stereochemistry of all constituents dictates these interactions, it is difficult to predict whether a certain detergent preserves the structure and function of a membrane protein. Although low CMC detergents are milder than high CMC detergents, the properties of both head group and hydrophobic tail define the stability of the solubilized membrane protein.

The choice of the detergent is most critical and has to be tested for every protein individually. It should be stressed that the availability of a stable, solubilized and pure membrane protein is the absolute prerequisite for successful 2D crystallization. To explore this, the membrane preparation containing the protein is divided into aliquots and solubilized with different detergents. Samples are centrifuged at high speed and both the pellet and supernatant are analyzed on SDS-PAGE. Western blotting, e.g. with antibodies against the affinity **tag**, then reveals the fate of the protein: If most of the protein is found in the pellet, the detergent may have destabilized the protein or only incompletely solubilized the membrane. In the latter case, pre-washing the membrane to strip off peripheral membrane proteins may help. This is achieved by washing the membranes in buffers of different types (high salt, low salt, EDTA, cholate, alkaline pH).

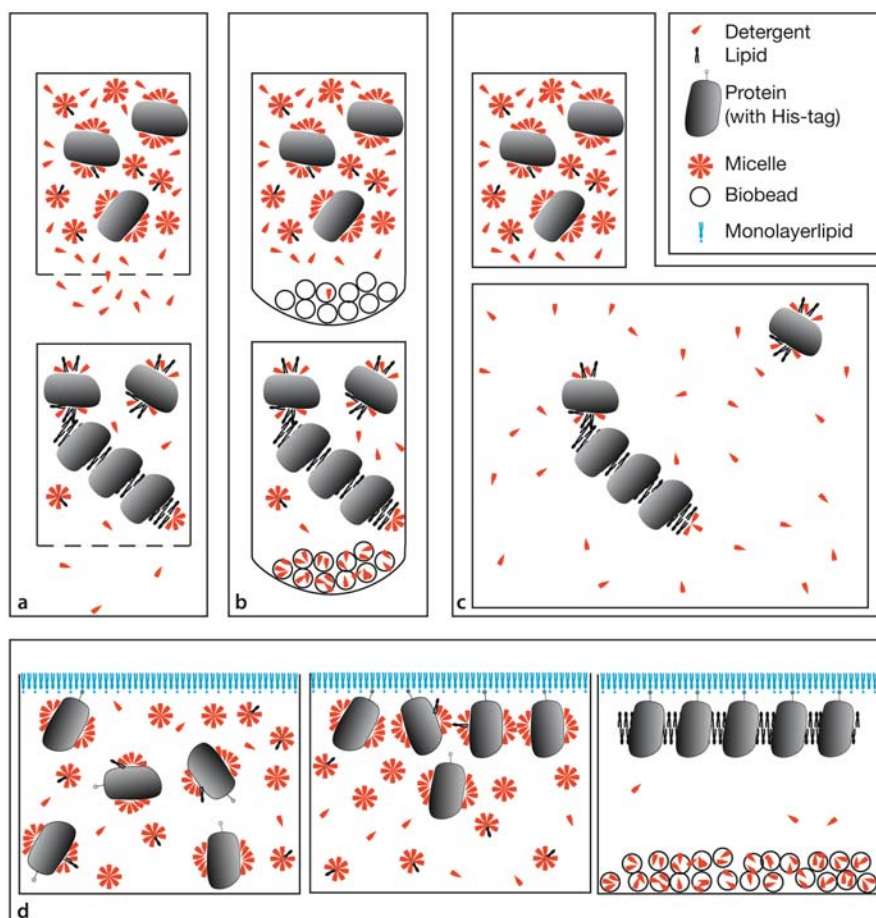
The selected detergents must be tested for their suitability to purify the protein. For purification, classical chromatographic methods can be used but affinity purification methods are generally milder, thus preserving the native state of the protein. Small aliquots of purified protein are analyzed for their oligomeric state by gel filtration or ultracentrifugation and for protein stability. The best method is to prepare negatively stained samples of the protein solution and

observe them by transmission electron microscopy. By this method, small aggregates are detected, which form after a few days and are not visible by eye. As an alternative, parts of the sample are centrifuged at high speed and the protein concentrations of the pellet and the supernatant are subsequently measured.

Two-Dimensional (2D) Crystallization

To assemble 2D crystals in the presence of lipids, the membrane protein is reconstituted in the bilayer. During reconstitution, the protein concentration in the lipid phase is drastically increased and conditions are selected that promote the precise regular packing of the

membrane protein, a most critical parameter being the lipid-to-protein ratio. This process is the reversal of the membrane solubilization process described above. Therefore, the respective affinities between the membrane protein, lipids and detergent determine the outcome of the reconstitution experiment. Fine-tuning of these interactions is achieved by careful selection of pH, ionic strength, counterions and osmolytes. The speed of detergent removal is another critical factor, which is difficult to control in most reconstitution schemes. Many protocols designed to reconstitute membrane proteins into 2D crystals have been described (Fig. 2).



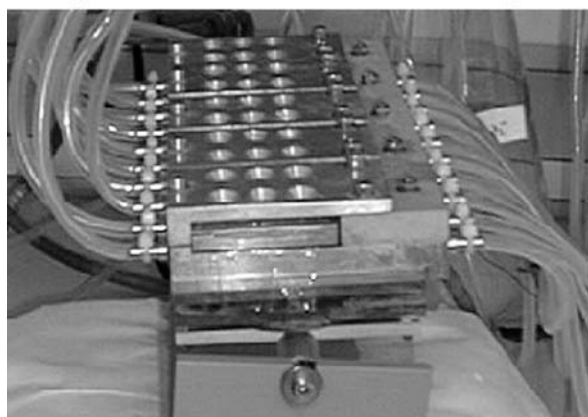
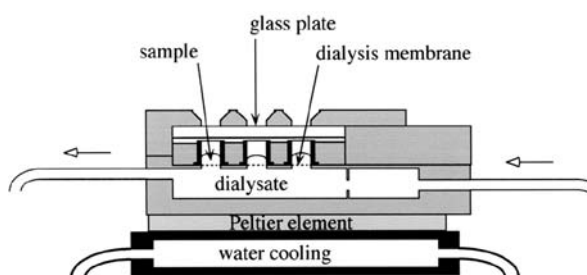
Two-dimensional Crystallization of Membrane Proteins. Figure 2 2D crystallization modes. All modes are based on the principle of bringing the detergent concentration in the aqueous phase below the CMC, forcing the detergent in the mixed micelles to partition in the aqueous phase. As result, mixed micelles merge to form larger structures and ultimately 2D crystals. (a) Dialysis can be used to remove the detergent provided its CMC is >1 mM. (b) Bio-Beads adsorb detergent molecules and can be used for all detergents. Bio-Bead driven 2D crystallization is particularly successful with low CMC detergents. (c) Dilution is a well-known method for functional reconstitution of membrane proteins. In spite of dilution it is also suitable for 2D crystallization, because the protein is highly concentrated after integration in the bilayer. (d) The monolayer technique combines the Bio-Bead method with crystallization at the air-water interface. This method works only with low CMC detergents because of the necessity of preserving the lipid monolayer. The latter incorporates special lipids having a high affinity for the solubilized protein, e.g. by recognition of a specific tag.

Dialysis Driven 2D Crystallization

Dialysis is the most widely used technique in 2D crystallization trials (1, 2). Usually small sample compartments are dialyzed against large buffer volumes. To improve the reproducibility of crystallization conditions, a temperature-controlled continuous flow dialysis apparatus can be employed (Fig. 3). Temperatures of up to 40°C sometimes improve the crystal quality, but this requires the proteins to be integrated in the lipid bilayer. Such heat treatment promotes reordering of the molecules into more regular arrays. The major draw-back of the dialysis method is the long dialysis time needed to remove low CMC detergents, making it practical only for medium to high CMC detergents (typically $\text{CMC} > 1 \text{ mM}$). In spite of these limitations, dialysis has led to many highly ordered crystals (1, 2).

► Bio-Beads

Hydrophobic adsorption of detergents by polystyrene beads (Bio-Beads SM2, Bio-Rad) works for removal of low CMC detergents and has been used to generate highly ordered 2D crystals of membrane proteins (1).



Two-dimensional Crystallization of Membrane Proteins. Figure 3 A temperature controlled open dialysis system for 2D crystallization. The upper panel shows the cross-section of the device, while the lower panel displays the machine allowing simultaneous reconstitution in 30 wells. Each well accommodates about 100 μl and the overall size of the system is about 10 cm by 20 cm.

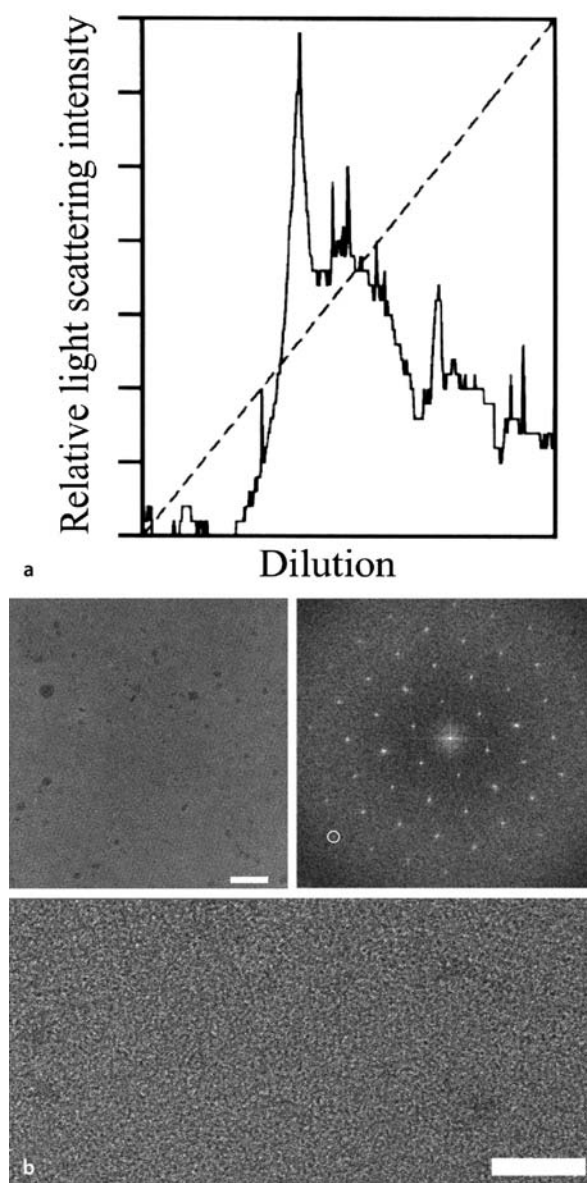
The non-specific adsorption of lipids is about 100–200 \times lower than the specific adsorption of detergents (2–4 mg phospholipids/g Bio-Beads SM2). The rate of detergent removal is directly linked to the weight of Bio-Beads used, but also to the working temperature, the rate doubles with every 15°C. To maintain a reproducible adsorption property, precise weighing of the Bio-Beads has to be performed on freshly blotted beads that must not dry out. Examples of crystals produced by the Bio-Bead method are given in a recent review (1).

Dilution Driven 2D Crystallization

Diluting a solution of protein, lipid, and detergent decreases the concentrations of all components by equal factors, until the detergent concentration in the aqueous phase drops below the CMC. Fig. 4 shows an example where the concentration of octyl-polyoxyethylene (8-POE) was decreased by dilution and the formation of structures of different sizes was monitored using turbidimetry (7). Porin OmpF assembled in the presence of DMPC only if the dilution rate was slow. During dilution, assembly of mixed micelles started shortly before the CMC of 8-POE was reached. Upon further dilution the co-existence of crystalline sheets of variable sizes up to several micrometers and of small, spherical vesicles was observed. Invariably, the large structures all exhibited crystallinity while small vesicles had a similar size to those produced in the absence of proteins. Crystallization by the dilution method requires a significant dilution of the protein and thus rather high initial protein concentrations. But the dilution method allows the process to be arrested when the micelle-bilayer transition is reached, extending the time in which an ordered assembly of the components can take place. In addition, the dilution method is suitable for low CMC detergents whose concentration is brought close to the CMC by the use of Bio-Beads prior to the dilution experiment.

Growth of 2D Crystals at Interfaces

The possibility of engineering tags to either terminus of a recombinant membrane protein greatly facilitates its purification. In addition, such a tag provides a handle to adsorb the membrane protein to a lipid monolayer formed at the water air interface and composed of spacer lipids and lipids designed to bind the hapten grafted to the protein. The method developed for soluble proteins has recently been used successfully to crystallize membrane proteins. Advantages concern unidirectional incorporation in the bilayer and the minute amounts of protein required. The lipid film consists of a mixture of a spacer lipid and – for a protein bearing a His-tag – a Ni-chelating lipid. Lipids are solubilized in an organic solvent, and spread onto a small buffer volume in a Teflon well



Two-dimensional Crystallization of Membrane Proteins. Figure 4 Dilution driven reconstitution of 2D crystals. (a) The turbidity profile of a dilution experiment with mixed micelles of DMPC and porin OmpF solubilized in octyl-polyoxyethylene (POE) reveals a sharp increase in the scattered light upon reaching the CMC of octyl-POE. Upon further dilution, the light scattering decays. (b) Negatively stained OmpF crystals obtained after a ten fold dilution over 2 h and phospholipase A2 treatment (top left, overview; top right, diffraction pattern with the marked spot at $(1.8 \text{ nm})^{-1}$; bottom, high magnification). The scale bars represent 120 nm (top left) or 90 nm (bottom).

of approximately 4 mm diameter and 1 mm depth. The solubilized membrane protein is injected into the buffer with excess detergent and incubated

for several hours. Excess detergent is removed with Bio-Beads (3).

Functional & Structural Assessment of 2D Crystals

Functional Analyses

Since membrane proteins are reconstituted in a lipid bilayer during 2D crystallization, all vesicle based functional assays may directly be applied. Thus, the function of the protein may be assessed in its crystallized form, an essential advantage of 2D crystallization. Another advantage is the precise knowledge of active units per surface area, allowing quantitative functional analyses to be achieved. This approach has been used to assess the water transport capacity of aquaporin-1 (AQP1; Fig. 5a).

Determination of Surface Structure and Dynamics by Atomic Force Microscopy

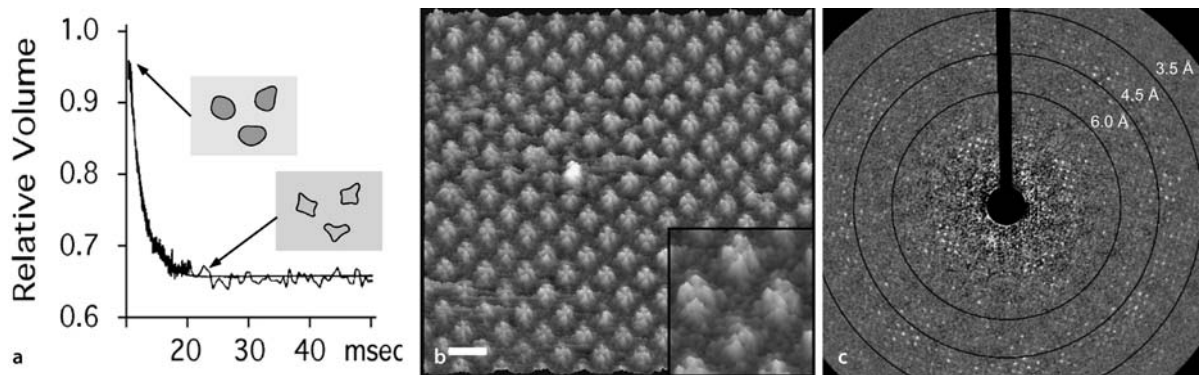
The atomic force microscope (AFM) allows biomolecules to be observed and manipulated under native conditions. It produces images with an outstanding signal-to-noise ratio and addresses single molecules while the sample is in a buffer solution. Topographs that reveal subnanometer details as well as the surface dynamics of membrane proteins reconstituted in a bilayer have been recorded. The surface structure of a tetragonal AQP1 crystal (Fig. 5b) reveals the topography of both the cytoplasmic and the extracellular surface, because AQP1 tetramers are packed in up and down orientations.

Determination of the Atomic Structure by Electron Crystallography

Electron diffraction is the ultimate test of crystal order. A diffraction pattern of an AQP1 2D crystal reveals diffraction spots to a resolution better than 3.5 Å (Fig. 5c). Electron crystallography of such crystals has led to the first atomic structure of a human membrane channel.

Perspectives

2D crystallization is an interesting approach for assessing the structure and function of membrane proteins. Although the preparation of a stable, solubilized and pure membrane protein must initially be achieved as in the growth of 3D crystals, speedy incorporation of the protein into the bilayer during 2D crystallization enhances the chances of restoring and preserving the native protein structure. A vision to be further developed is the *in vitro* expression of a membrane protein into an essentially empty lipid vesicle. *In situ* crystallization starting from such vesicles enriched in the protein of interest seems feasible. The decisive advantage of this approach is the circumvention of the solubilization step required in all other methods.



Two-dimensional Crystallization of Membrane Proteins. **Figure 5** Functional and structural assessment of aquaporin-1 (AQP1) 2D crystals. (a) Rapid mixing of fluorescein-loaded vesicular 2D crystals with a 20% sucrose solution at time 0 results in an osmotically driven rapid outflow of water and vesicle compression. Concentration-induced quenching of the fluorescence indicates the relative volume change. From the average size of the vesicles, the transported water volume as well as the number of AQP1 channels can be calculated. (b) AQP1 tetramers are packed in a tetragonal crystal with $p422_1$ symmetry as revealed by the topography recorded by AFM. The structures of the extracellular, highly corrugated surface and the cytosolic surface are distinctly different. The scale bar represents 15 nm. An electron diffraction pattern of such tetragonal crystals documents their excellent crystalline order, revealing diffraction spots beyond 3.5 Å resolution.

References

1. Rigaud J, Chami M, Lambert O et al (2000) Use of detergents in two-dimensional crystallization of membrane proteins. *Biochim Biophys Acta* 1508:112–128
2. Ringler P, Heymann B, Engel A (2000) Two-dimensional crystallization of membrane proteins. In: Baldwin SA (ed), Oxford University Press
3. Levy D, Chami M, Rigaud JL (2001) Two-dimensional crystallization of membrane proteins: the lipid layer strategy. *FEBS Lett* 504:187–193
4. Le Maire M, Champeil P, Moller JV (2000) Interaction of membrane proteins and lipids with solubilizing detergents. *Biochim Biophys Acta* 1508:86–111
5. Heerklotz H, Seelig J (2000) Correlation of membrane/water partition coefficients of detergents with the critical micelle concentration. *Biophys J* 78:2435–2440
6. Lichtenberg D, Opatowski E, Kozlov MM (2000) Phase boundaries in mixtures of membrane-forming amphiphiles and micelle-forming amphiphiles. *Biochim Biophys Acta* 1508:1–19
7. Remigy HW, Caujolle-Bert D, Suda K et al (2003) Membrane protein reconstitution and crystallization by controlled dilution. *FEBS Lett* 555:160–169

Definition

In **▶two-dimensional gel electrophoresis** (2-DE) proteins are separated in polyacrylamide gels by two independent parameters, charge (isoelectric point) and size (molecular mass). The best resolution is obtained if isoelectric focusing (**▶IEF Isoelectric Focusing**) under denaturing conditions is combined with sodium dodecylsulfate polyacrylamide gel electrophoresis (**▶SDS-PAGE**). There are two common alternative versions of isoelectric focusing, non-equilibrium pH gradient electrophoresis (NEPHGE) using carrier ampholytes to build up the pH gradient in the electric field and **▶immobilized pH gradient IEF** using immobilized ampholytes (immobilines) (3).

Description

Although already developed in 1975 (O'Farrell) as a procedure that allowed a holistic view of biology, two-dimensional electrophoresis (2-DE) technology is only poorly automated in comparison to robotized genome technology. Manual **▶proteomics** procedures have reached a high performance and using 2-DE gels it is possible to resolve up to 10,000 **▶protein species** with a gel size of 30 cm×40 cm with a sensitivity of low numbers of molecules per cell. Protein species are separated in polyacrylamide gels using two independent separation criteria, isoelectric point in the first dimension and molecular mass in the second dimension. Only after complete denaturation by urea, dithiothreitol and detergents, are proteins in a well-defined structure and does their separation result in

Two-Dimensional Gel Electrophoresis

PETER R. JUNGBLUT
Max Planck Institute for Infectious Diseases,
Protein Analysis, Berlin, Germany
jungblut@mpiib-berlin.mpg.de

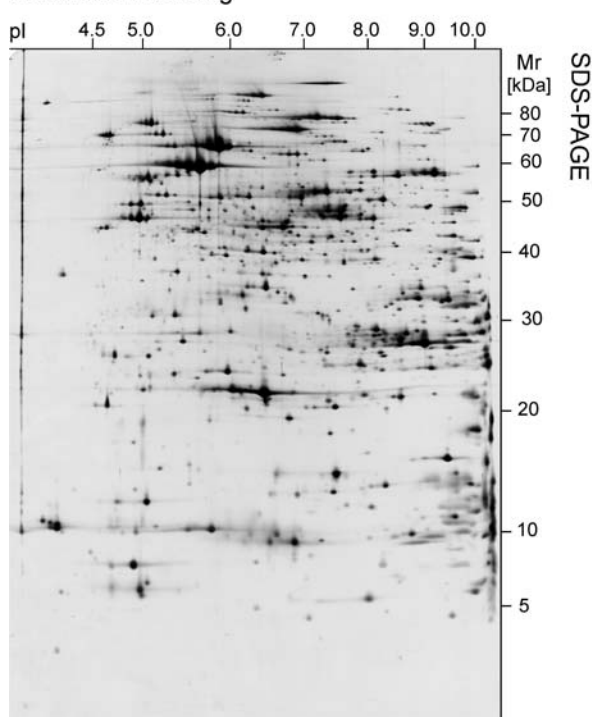
distinctly shaped spots in 2-DE gels. Conformational effects on charge and molecular mass (Mr) have to be minimized to obtain optimal resolution. For separation according to the isoelectric point, isoelectric focusing is the method of choice. Two technologies are in use, non-equilibrium pH-gradient gel electrophoresis (NEPHGE), which uses carrier ampholytes and gives optimal resolution, and focusing with immobilines, immobilized ampholytes, which is easier to perform. Focusing is combined with sodium dodecylsulfate polyacrylamide gel electrophoresis (SDS-PAGE). The proteins are detected within the gels using silver stain, Coomassie Blue, negative staining, fluorescent stains or after radioactive labeling by autoradiography. Fig. 1 shows a result of a 2-DE separation after silver staining of the gel. In this experiment, cellular proteins of *Helicobacter pylori* have been separated. Each spot of this protein pattern represents at least one protein species. Quantification is possible by comparing the optical densities of the spots, but is clearly improved by isotopic or ►fluorescence labeling.

Proteins are modified during (co-translational) or after synthesis (post-translational). Each of these modifications leads to a new chemical structure for the protein. Each chemical structure of a protein has been defined as a 'protein species' (2). So one gene may result in the production of several hundred protein species. If two protein species have a different charge or Mr, 2-DE allows the separation at the protein species level. For separation the Mr-difference should be higher than 100 D. The exchange of one amino acid with a charged side chain for a neutral one or a post-translational modification such as a phosphorylation is sufficient for separation on a 2-DE gel. Current 2-DE technology was reviewed by Görg et al. (2004).

The separation range for Mr is between 5 and 150 kD, depending on the pore size of the applied gels and for pI between 4 and 10.5 depending on the ampholytes used. The detection sensitivity increases from Coomassie Brilliant Blue to silver staining and radioactive detection from 100 ng to 1 ng and down to several molecules per cell, respectively.

Proteins on 2-DE gels may be identified by mass spectrometry or after ►blotting by immunostaining or N-terminal protein sequencing. As a rule of thumb all proteins stainable with Coomassie Blue can be identified, if the protein sequence is within a database. 2-DE separates the proteins at the protein species level and allows ►subtractive analyses between 2-DE patterns of different biological situations to detect variations in amounts of protein species. Relative quantitative comparisons are performed by comparing the optical densities of the resulting spots or by fluorescent (DIGE, fluorescence difference gel electrophoresis) or ►isotopic labeling. The results of a subtractive 2-DE analysis are stored and retrieved in

Isoelectric Focusing



Two-dimensional Crystallization of Membrane Proteins. Figure 1 Protein pattern of *Helicobacter pylori* 26695 cellular proteins separated by 2-DE. Each spot represents at least one protein species. The identity of the spots in this pattern may be found by clicking on the spots within the 2D-PAGE database (►<http://www.mpiib-berlin.mpg.de/2D-PAGE/>).

2-DE databases such as the WORLD-2DPAGE database (►<http://au.expasy.org/ch2d/>) or the centralized ►2D-PAGE database (►<http://www.mpiib-berlin.mpg.de/2D-PAGE/>).

Clinical Applications

With its microscope-like character 2-DE was applied for marker search in many diseases such as cancer, infectious and heart diseases. In principle the proteins of two biological situations are separated by 2-DE and differences in the presence or in the intensity of the spots are determined by visual evaluation or software-supported image analysis. An example is the comparison of 2-DE patterns from control tissue with diseased tissue. All variations in the 2-DE pattern from the diseased tissue are defined as disease-associated variations.

For example: N-methyl-N-nitroso urea induced rat hepatoma cells produce an embryonal form of aldose reductase, which is not present in 2-DE patterns of control cells (5). The *in vivo* situation was investigated in complete organs, biopsies, tissues or single cells.

The complexity of protein composition decreases from organs to cells and therefore the chance of detecting cancer-associated proteins increases. Calgranulin B was present only in 2-DE patterns of colorectal carcinoma as compared with mucosa of the colon. Cell groups may be prepared by laser capture microdissection to obtain unique tissue. Because single cells do not contain sufficient amounts of protein, several hundred microdissections have to be accumulated. From microdissected normal and malign prostate epithelial cells, 40 potentially dysregulated proteins were found by 2-DE, e.g. 14.3.3 protein, a protein of signal transduction, the structure proteins tropomyosin and cofilin and the enzyme aldehyde dehydrogenase.

At first glance the application of 2-DE seems to be more complicated for infectious diseases because in addition to the diseased organism we also have to consider the proteome of the infecting microorganism. Fortunately the microorganisms can be analyzed in culture and show a complexity of 500–5000 genes which is adequate for 2-DE analysis. Between 10 and 50% of the predicted proteins were already identified on single 2-DE gels for bacteria. With this coverage there is a good chance to detect virulence-associated proteins by a comparison of 2-DE patterns of virulent strains with those of attenuated strains. Other strategies are to look for the most abundant proteins or surface proteins, which have a high probability of coming in contact with the host's immune system, or direct detection of the antigens by blotting of the proteins from a 2-DE gel onto a membrane, where the antigens are detected by sera of patients infected by the microorganism (4). This procedure is named immunoblotting. Antibodies from the sera react with the antigens and a second antibody against the constant region of immunoglobulins reacts with the antigen antibody complex. This second antibody is coupled with an enzyme, which catalyses the reaction of an uncolored substrate to a colored product. Because the antigen pattern represents the antigens of the pathogen and the antibodies of the host, it also represents the immunoproteome of the host pathogen interaction.

The heart with its relatively unique muscle tissue is a clearly defined organ and therefore predestined as a starting material for questions on the molecular functioning of a complete organ using proteomics. To this end, an attempt has been made to elucidate cardiomyopathy associated protein variations using proteomics. In spite of the unexpectedly high genetic variation, reproducible changes in intensity were recognized for several proteins. During the search for disease-associated proteins for dilated cardiomyopathy (DCM) it became obvious that ►proteomic signatures have to be determined. Heat-shock protein 27 (Hsp27) was separated into 59 protein species and changes in the amount of some of these were recognized to be

cardiomyopathy-associated using immunoblot techniques. From the many animal models investigated, results from a bovine model are remarkable. For bovine DCM, the most notable change was a sevenfold increase in the enzyme ubiquitin carboxyl-terminal hydrolase. This could result in increased protein ubiquitination in the diseased state, leading to proteolysis via the 26S proteasome pathway. An increased amount of this enzyme was also detected in human DCM heart tissue.

Therapeutic Consequences

2-DE is a powerful technique for detecting disease-associated and immuno-relevant protein species, which are potential candidates for diagnosis, vaccination and therapy. After 2-DE these candidates have to be evaluated and further reduced to the most relevant markers. For example in the case of *Helicobacter pylori*, the causative agent of gastric carcinoma, the immuno-relevant proteins were reduced to about 500 by ►immunoproteomics starting with 1800 protein species on a 2-DE gel. These 500 proteins were further reduced by the following criteria, most abundant proteins, surface localization, distribution in 15 *H. pylori* isolates, *Helicobacter* specific proteins and T-cell epitope containing proteins. Only two proteins fulfilled these criteria optimally. These two proteins showed a promising vaccination effect in a vaccination experiment in the mouse. Clinical trials are now underway. This example shows that 2-DE is an important first step in the search for medically relevant proteins. The number of the potential candidates may be reduced by 2-DE and the resulting candidates need further evaluation and reduction.

References

1. O'Farrell PH (1975) High resolution two-dimensional electrophoresis of proteins. *J Biol Chem* 250:4007–4021
2. Jungblut P, Thiede B, Zimny-Arndt U et al (1996) Resolution power of two-dimensional electrophoresis and identification of proteins from gels. *Electrophoresis* 17:839–847
3. Görg A, Postel W, Günther S (1988) The current state of two-dimensional electrophoresis with immobilized pH gradients. *Electrophoresis* 9:531–546
4. Krah A, Miehle S, Pleissner K-P et al (2004) Identification of candidate antigens for serologic detection of *Helicobacter pylori* infected patients with gastric carcinoma. *Int J Cancer* 108:456–463
5. Zeindl-Eberhart E, Jungblut P, Otto A et al (1994) Characterization of tumor-associated protein variants during hepatocarcinogenesis in the rat. *J Biol Chem* 269:14589–14594
6. Görg A, Weiss W, Dunn MJ (2004) Current two-dimensional electrophoresis technology for proteomics. *Proteomics* 4:3665–3685

Two-Dimensional Polyacrylamide Gel Electrophoresis

Definition

Two-dimensional gel electrophoresis (sometimes abbreviated 2D-PAGE) separates proteins (or other compounds) first by their isoelectric point (i.e. charge) and then by size, usually in a polyacrylamide gel.

- ▶Protein-Protein Interaction
- ▶Protein Interaction Analysis: Chemical Cross-Linking
- ▶Two Hybrid System
- ▶Two-Dimensional Gel Electrophoresis

Two-Hit Model

The two-hit model is a model for oncogenesis developed by Alfred G. Knudson. This model suggests that cancer develops through loss of function of both the maternal and paternal alleles (copies) of a cancer susceptibility gene. In inherited forms of cancer (e.g. retinoblastoma), there is a loss of either the maternal or paternal allele within the egg or sperm from which an individual is conceived, thus every daughter cell thereafter has ‘one hit’ at this genetic locus. The loss of the remaining allele, or ‘second hit’, may occur in time as the result of an environmental trigger, or by chance during DNA replication. Sporadic cancers begin with two normal copies of a cancer susceptibility gene. Two successive acquired mutational events (‘two hits’), occurring in a single cell at each allele of a cancer susceptibility gene, are necessary for the development of cancer.

- ▶Inherited Mental Retardation Syndromes

Two-Hybrid System

CAROLINA S. VOLLERT, PETER UETZ
Institute for Genetics, Forschungszentrum Karlsruhe,
Karlsruhe, Germany
peter.uetz@itg.fzk.de

Synonyms

Interaction trap

Definition

Experimental procedure to detect protein-protein interactions that uses two ▶fusion proteins (“hybrids”) that reconstitute an active ▶transcription factor (or some other active protein) when the two fusion proteins interact (Fig. 1). The reconstituted transcription factor binds to the ▶promoter of a ▶reporter gene in the nucleus and switches on its ▶transcription. Consequently, reporter gene activity can be detected or measured.

It has been estimated that more than 50% of all protein interactions described in the literature have been detected using the yeast-two hybrid system.

Characteristics

The yeast two-hybrid system was first described 1989 by S. Fields and O-K. Song (1). The basic concept emerged from previous experiments on transcription factors, which usually contain separable DNA-binding ▶domains (DBD) and transcriptional activation domains (AD) (Fig. 1). This property can be exploited to detect protein-protein interactions.

Basic Principle of the Classical Yeast Two-Hybrid System

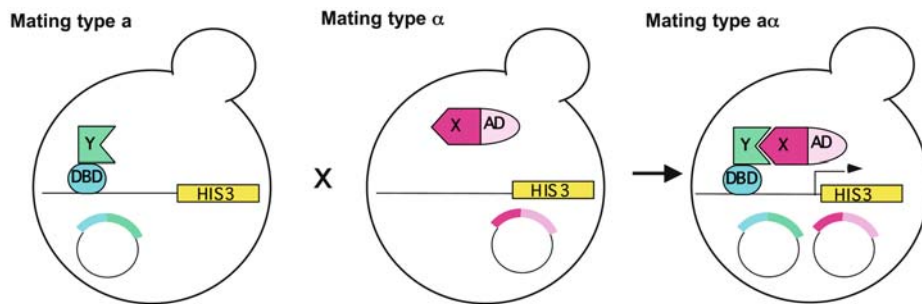
In order to test if two proteins X and Y interact, they are expressed as fusion proteins (hybrids) with a transcriptional activating domain (AD, ▶prey) and a DNA-binding domain (DBD, ▶bait), respectively. The term “two-hybrid” derives from these two ▶chimeric proteins.

The bait and prey fusions both contain a ▶nuclear localization signal (NLS) which allows the import of these proteins to the nucleus (targeting non-nuclear proteins to the nucleus). A physical interaction between X and Y leads to the reconstitution of a functional transcription factor which binds to the promoter of a reporter gene (Fig. 1). Consequently, transcription of the reporter gene is switched on. A protein-protein interaction is therefore translated into reporter gene activity that can be detected or measured.

The two-hybrid system can be used for the detection of essentially any protein-protein interaction, independent of the function of the corresponding proteins (but see limitations below).

Advantages and Limitations of the Yeast Two-Hybrid System

The yeast two-hybrid system became one of the most popular technologies for the detection of protein-protein interactions because it is fairly simple, rapid and inexpensive (it avoids the costly protein purification and antibody development needed in the traditional biochemical methods). No previous knowledge about the interacting proteins is necessary for a screen



Two-Hybrid System. Figure 1 Classical Yeast Two-Hybrid System. A protein of interest, Y, is expressed in yeast as a fusion to a DNA-binding domain (DBD, “bait”; circles denote expression *plasmids*). Another protein of interest, X, is fused to a transcriptional activation domain (AD, “prey”) and expressed in a different cell of opposite mating type (α). The two yeast strains are mated to combine the two fusion proteins in one cell. If proteins Y and X interact in the resulting diploid cells, they reconstitute a transcription factor that activates a reporter gene (here *HIS3*) and therefore allows the cell to grow on selective media (here media lacking histidine).

to be performed. Finally, the system can be scaled up to high-throughput usage.

Limitations

Some classes of proteins are not suitable for analysis by the yeast two-hybrid system. For example, transcriptional activators may activate transcription without any interaction. Another class of troublesome proteins are those containing hydrophobic transmembrane domains, which may prevent the proteins from reaching the nucleus. To overcome this limitation one of the alternative membrane-associated two-hybrid systems may be used, such as the split-ubiquitin assay or the SOS recruitment assay (Figs. 4, 5).

Other proteins may require modification by cytoplasmic or membrane associated enzymes in order to interact with binding partners. Alternative methods could also help in this case such as the three-hybrid system (Fig. 3).

► False Positives and ► False Negatives

The two-hybrid system can produce false positives, that is reporter gene activity where no protein-protein interaction is involved. Frequently, such false positives are caused by bait proteins that act as transcriptional activators. These types of activators are easily detectable when analysed against “empty”-prey fusions, that is preys containing only the transcriptional activating domain. If the reporter gene has been switched on in this case, then different procedures as mentioned below can be tested to reduce this unspecific background activity. Other false positives may be caused by proteins that lead to non-specific interactions for largely unknown reasons. Some bait or prey proteins may affect general colony viability and hence allow a cell to grow under selective conditions and activate reporter gene activation. Mutations or other random

events of unknown nature may be invoked as potential explanations as well. Overall, for extremely few cases of false positives the underlying mechanism is known. A number of procedures have been developed in order to identify or avoid false positives altogether, including the utilization of multiple reporters, independent methods of specificity testing or simply repeating assays to make sure a result is reproducible. False negatives involve physiological protein-protein interactions that are not detected by two-hybrid assays. They may arise by steric hindrance of the two fusion proteins so that transcriptional activation is prevented. Other explanations include instability of proteins or failure of nuclear localization.

Variations of the Two-Hybrid System

Based on the two-hybrid paradigm numerous variations have been developed to overcome the limitations of the classical yeast two-hybrid system, such as the analysis of membrane proteins. A more comprehensive overview of variations is shown in Table 1.

Only some important variations are described as follows.

Reverse Two-Hybrid and Split Hybrid System

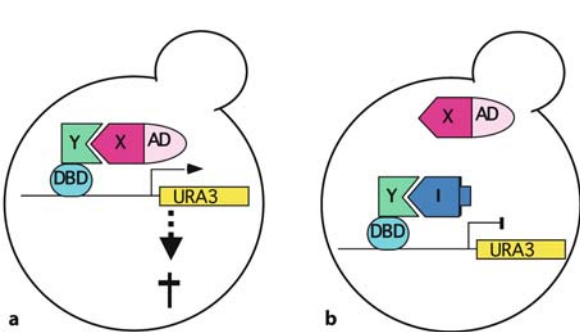
The “reverse” two-hybrid system has been invented to select for disrupted two-hybrid interactions e.g. by ► [mutations](#), drugs or competing proteins (Fig. 2). In this system the interaction of X and Y proteins induces the transcription of a reporter gene that confers toxicity to the yeast.

Three-Hybrid System

In this yeast two-hybrid variation, a third protein (Z) is expressed along with the DBD and AD fusions (Fig. 3). Expression of the reporter gene is used to select for interactions that occur only in the presence of this protein.

Two-Hybrid System. Table 1 Comparison of different methods for identifying protein-protein interactions

Method	Advantages	Disadvantages
Two-Hybrid	Simple and inexpensive	Limitation analysing transcription factors and membrane proteins
	Coverage of low abundant proteins	
	Suitable for high throughput	
Affinity purification + Mass Spectrometry	Identification of protein complexes	Expensive, complexes do not provide information on topology
<i>In vitro</i> binding (Coimmunoprecipitation, GST-pulldowns etc.)	Defined conditions	Potentially non-physiological conditions
Protein Chips	Defined conditions Suitable for high throughput	Requires purification of protein

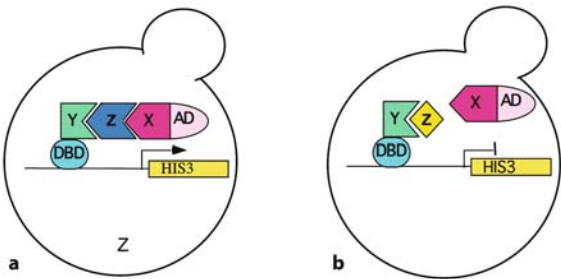


Two-Hybrid System. Figure 2 Reverse Two-Hybrid System. The interaction of bait and prey in the yeast cell is lethal (a), selecting for yeasts where the interaction is disrupted (b). Selection can be induced by the addition of FOA (5-fluoro-orotic acid), which is converted to the toxic compound 5-fluorouracil by the URA3 gene product.

Protein Z may be directly involved in the interaction as shown in Fig. 3 but may also allow the interaction between X and Y *via* a post-translational modification of these proteins (such as phosphorylation). SenGupta et al. (8) developed a three-hybrid system to detect and analyze RNA-protein interactions in which the binding of a bifunctional RNA molecule links the DBD and AD hybrid proteins and activates transcription of the reporter gene. This system is known as RNA three-hybrid system.

SOS Recruitment System (SRS)

This membrane-associated two-hybrid system makes use of the Ras pathway in yeast (9). When localized at the plasma membrane, the yeast Ras guanyl nucleotide exchange factor (RGEF) cdc25 stimulates GDP/GTP exchange on Ras and promotes downstream signalling events that ultimately lead to cell growth. A mutant

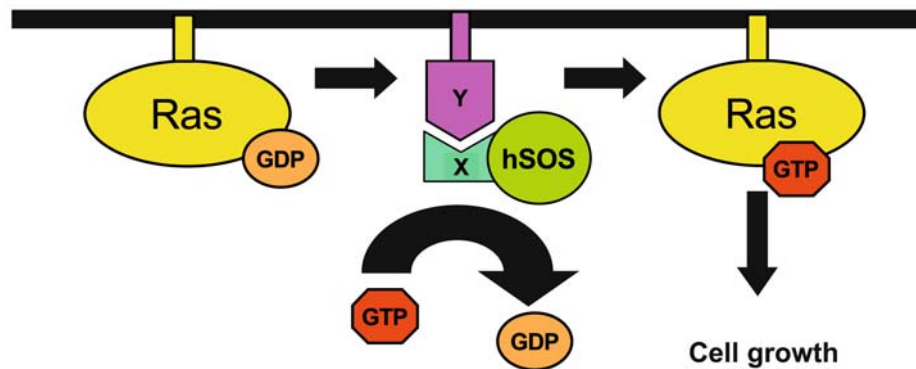


Two-Hybrid System. Figure 3 Three-Hybrid System. A third protein (Z) is expressed along with the DBD and AD fusions. Expression of the reporter gene is used to select for interactions that occur only in the presence of this protein (a). This third protein alternatively can prevent the formation of a two-hybrid complex (b). Alternatively, Z may be a hybrid RNA molecule with part of the sequence binding to Y and the other to X (“RNA three-hybrid”).

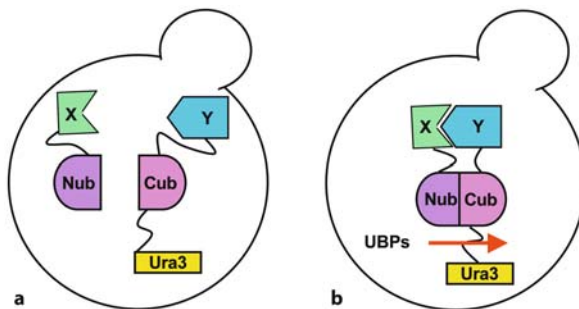
yeast strain harbouring the temperature sensitive cdc25-2 allele is still able to grow at 25°C but fails to grow at 36°C. However, the human RGEF (hSOS) when targeted to the plasma membrane efficiently complements the mutation, leading to cell growth at 36°C. In the SRS the translocation of hSOS is dependent on a protein-protein interaction; the bait X is fused to C-terminally truncated hSOS, which is active but unable to target to the plasma membrane. The bait is co-expressed with a prey Y, which can be either an integral membrane protein or a soluble protein that is anchored to the membrane by means of a ▶myristoylation signal.

Split-Ubiquitin System

Johnsson and Varshavsky (10) have developed a cytoplasmic two-hybrid assay based on ▶ubiquitin. Ubiquitin is a small protein of 76 amino acids that acts



Two-Hybrid System. Figure 4 SOS Recruitment System. Protein X is fused to a human Ras guanyl exchange factor, hSOS. Putative interaction membrane protein Y (or protein localized to the membrane by a myristoylation tag) is co-expressed. Interaction between X and Y recruits hSOS to the membrane, where it stimulates guanyl nucleotide exchange on Ras. GTP-bound Ras stimulates cell growth.



Two-Hybrid System. Figure 5 Split Ubiquitin System. Proteins fused to ubiquitin are rapidly cleaved *in vivo* by ubiquitin specific proteases (UBPs). A mutant Nub is fused to protein X and the Cub-reporter gene hybrid is fused to protein Y (a). Interaction between both proteins reconstitutes ubiquitin and leads to cleavage and release of the reporter gene (URA3) (b).

as a “tag” for [protein degradation](#). Proteins fused to ubiquitin are rapidly cleaved *in vivo* by ubiquitin-specific proteases (UBPs).

If the carboxy terminal of ubiquitin (Cub) is fused to a reporter protein and co-expressed with the amino-terminal fragment (Nub), the two halves will reconstitute the native ubiquitin, resulting in the cleavage of the reporter protein. For its adaptation to detect protein-protein interactions, a mutant Nub, unable to interact with Cub on its own, was fused to one protein and a Cub reporter hybrid was fused to its prospective interaction partner (Fig. 5). Interaction between the two proteins allowed ubiquitin to be reconstituted, leading to cleavage and release of the reporter gene. As the formation of split-ubiquitin and subsequent cleavage by UBPs do not depend on any special

localization of the proteins, this system is suitable for the investigation of membrane proteins. Up to date two membrane based split-ubiquitin systems have been described, the Ura3 based split-ubiquitin system and the transactivator based split-ubiquitin system.

A further advantage of this system is that the signal for an interaction can be changed by changing the nature of the reporter protein (e.g. transcription factor or enzyme activated by ubiquitin cleavage). Furthermore, the small size of the ubiquitin fragments in the hybrid proteins may also be advantageous because it minimizes the possibility of steric hindrance.

Up to date it is the most widely used of the alternative yeast-based two-hybrid systems.

Clinical Relevance

Two-hybrid systems are useful for the investigation of protein-protein interactions, protein-nucleic acid and protein-small molecule interactions. However, the technology is mainly used in basic research in order to understand the function of proteins. Many protein-protein interactions can cause disease when they are disrupted (e.g. haemoglobin is a tetramer that binds oxygen efficiently only when the subunits interact). Protein interactions are now being identified by biotech and pharmaceutical companies as potential drug targets, especially in combination with screens for small molecule ligands that can disrupt or modulate these protein interactions.

► [High-Throughput Approaches to the Analysis of Gene Function in Mammalian Cells](#)

► [Protein Interaction Analysis, Variations of the Yeast Two-Hybrid System](#)

► [Protein-Protein Interaction](#)

References

1. Fields S, Song O-K (1989) A novel genetic system to detect protein-protein interactions. *Nature* 340:245–246
2. Bartel PL, Fields S (1997) *The Yeast Two-Hybrid System*. Oxford University Press, Oxford
3. Mac Donald PN (2001) *Two-Hybrid Systems, Methods in Molecular Biology*. Humana Press
4. Golemis E (2002) *Protein-Protein Interactions, A Molecular Cloning Manual*. Cold Spring Harbour Laboratory Press, Cold Spring Harbour
5. Auerbach D et al (2002) The post-genomic era of interactive proteomics: Facts and perspectives. *Proteomics* 2:611–623
6. Uetz P, Pankratz MJ (2004) Protein interaction maps on the fly. *Nat Biotechnol.* 22:43–44
7. Giot L, Bader JS, Brouwer C et al (2003) Protein interaction map of *Drosophila melanogaster*. *Science* 302:1727–1736
8. Sen Gupta DJ, Zhang B, Kraemer B et al (1996) A three-hybrid system to detect RNA-protein interactions in vivo. *Proc Natl Acad Sci USA* 93:8496–8501
9. Aronheim A, Zandi E, Hennemann H et al (1997) Isolation of an AP-1 repressor by a novel method for detecting protein-protein interactions. *Mol Cell Biol* 17:3094–3102
10. Johnsson N, Varshavsky A (1994) Split ubiquitin as a sensor of protein interactions in vivo. *Proc Natl Acad Sci USA* 91:10340–10344
11. ► <http://www.fccc.edu/research/labs/golemis/interaction-trapinwork.html>
12. ► <http://depts.washington.edu/sfields/projects/YPLM/>
13. ► <http://www.two-hybrid.com>

Two-Photon Excitation

Definition

Two-photon excitation refers to a non-linear fluorescence excitation phenomenon, where two photons of half the energy required for an electronic transition are simultaneously absorbed by a fluorophore. In this case, their energy can be added and the excitation facilitated, although the wavelength is approximately twice as large as the absorption maximum (usually, in the infrared).

► FCS

Two-Site Immunoradiometric Assay

Definition

A two-site immunoradiometric assay is an assay using two different antibodies (one of which is radiolabelled) targeting the different ends of a polypeptide.

► Hyper- and Hypoparathyroidism

Two-State Process

Definition

Two-state process designates a process that involves no thermodynamically significant intermediate states. The two-state hypothesis assumes that only 2 species occur in equilibrium whose relative populations are shifted by temperature or any other environmental parameter such as pH or denaturant agents. This assumption applies to conformational changes of many small proteins and simplifies data analysis significantly.

► Differential Scanning Calorimetry

► Thermodynamic Properties of DNA

Ty Retrotransposon

Definition

Ty retrotransposon describes any of several transposable genetic elements from the yeast *Saccharomyces Cerevisiae*, which move to new chromosomal locations by a retrovirus-like mechanism involving reverse transcription of its mRNA.

► Translational Frameshifting, Non-standard Reading of the Genetic Code

► Transposons

Type B Reactions

► Idiosyncratic Drug Reactions

Type 1 Diabetes Mellitus

Definition

Type 1 diabetes mellitus (also known as insulin-dependent diabetes mellitus (IDDM) or juvenile-onset diabetes) is the most frequent form of diabetes that is characterized by early age of onset, autoimmune mechanism and absolute requirement for insulin.

► Common (Multifactorial) Diseases

► Diabetes Mellitus, Genetics

Type 1 Pseudohypoaldosteronism

Definition

Type 1 pseudohypoaldosteronism (PHA 1) is a hereditary disease characterized by salt wasting, resulting from target organs unresponsiveness to mineralocorticoids. There are 2 forms of PHA1: the autosomal recessively inherited form is caused by mutations in the gene encoding the amiloride-sensitive luminal sodium channel, and is clinically more severe involving pulmonary symptoms. Sporadic or autosomal dominantly transmitted forms are caused by mutations in the mineralocorticoid receptor gene showing milder symptoms and generally remit with age. Autosomal recessive forms are caused by mutations in the gene encoding the amiloride-sensitive sodium channel and are clinically more severe involving also pulmonary symptoms.

► [Mendelian Forms of Human Hypertension and Mechanisms of Disease](#)

Type 1A Topoisomerases

Definition

Type 1A topoisomerases are present in all organisms. These enzymes transiently break a single strand of double stranded DNA, through a 5' phosphotyrosine intermediate, and pass another DNA strand through the break, catalyzing relaxation of negative supercoiling in DNA. Type 1A topoisomerases carry out changes in supercoiling by an enzyme bridging mechanism. The first identified topoisomerase was *E. coli* topoisomerase I, a type 1A enzyme.

► [DNA Topoisomerases](#)

Type 1B Topoisomerases

Definition

Type 1B topoisomerases are present in eukaryotes and some bacteria. These enzymes transiently break one strand of a double stranded DNA molecule, forming a 3' phosphotyrosine intermediate. Relaxation occurs through a strand rotation mechanism. Unlike type 1A enzymes, type 1B topoisomerases efficiently catalyze relaxation of both negative and positive supercoils.

► [DNA Topoisomerases](#)

Type 2 Diabetes Mellitus

Definition

Type 2 diabetes mellitus (also known as non-insulin dependent diabetes mellitus (NIDDM) or adult-onset diabetes) results from a combination of insulin resistance and a defect in insulin production. Treatment may involve diet modification and exercise alone, oral medications and/or insulin injections.

► [Diabetes Mellitus, Genetics](#)

Type 2 Pseudohypoaldosteronism

Definition

Type 2 Pseudohypoaldosteronism (PHA2) also called ► [Gordon's syndrome](#) or familial hyperkalemic hypertension) is an autosomal dominantly transmitted disorder characterized by hypertension with hyperkalemia/hyperaldosteronism, slight hyperchloremic metabolic acidosis and otherwise normal renal function. Thiazide diuretics are highly effective in this syndrome, commensurate with salt sensitivity. This form of hypertension is caused by mutations in the genes of the ► [WNK family](#) of serine-threonine kinases (WNK 1 for PHA 2C and WNK 4 for PHA 2B). The mutations in WNK1 are large intronic deletions that increase WNK1 protein expression. The mutations in WNK4 are missense and cause a loss of function. The mutations cluster in a short and highly conserved segment of the derived protein.

► [Mendelian Forms of Human Hypertension and Mechanisms of Disease](#)

Type I Transmembrane Glycoprotein

Definition

Type I transmembrane glycoprotein denotes a single-pass transmembrane protein that is glycosylated on the extracellular amino-terminal domain.

► [Desmosomes](#)

Type II Myosin

Definition

Type II myosin is the molecular motor of all muscle types, which consists of 2 heavy chains (200 kDa each) and 4 light chains (16–28 kDa each).

► [Muscle Contraction](#)

Type IIA Topoisomerases

Definition

Topoisomerases are present in all eubacteria and eukaryotes, and are dimeric or tetrameric enzymes. These enzymes generate double strand breaks during their reaction cycle. Type II topoisomerases can catalyze catenation or decatenation of double stranded DNA, and are required for separating replicated DNA molecules at the completion of DNA replication.

► [DNA Topoisomerases](#)

Type III Hyperlipidemia

Definition

Type III hyperlipidemia, also known as familial dysbetalipoproteinemia or remnant removal disease, is a rare familial disorder characterized by a combined elevation of serum cholesterol and triglycerides (TG), and a marked accumulation of remnant lipoproteins in plasma. Type III hyperlipoproteinemia is based on apoE2 homozygotes genotype.

► [High-HDL Syndrome](#)

Tyrosinase

Definition

Tyrosinase is an enzyme that is required for the production of pigments in the hair, retina and inner ear. Tyrosinase expression is visible in the albino background.

► [Mouse Genomics](#)

Tyrosine Kinases

UTE SCHAEFER, KATJA S. GROSSMANN

Max Delbrück Center for Molecular Medicine,
Berlin, Germany

uschaep@mdc-berlin.de

Definition

Tyrosine kinases are enzymes that catalyze the transfer of the γ -phosphate of adenosine triphosphate (ATP) to the hydroxyl group of tyrosine residues in a protein substrate (1). They are found in all multi-cellular organisms and participate in signal transduction pathways that regulate multiple cellular processes like cell migration, motility, differentiation, cell growth or apoptosis. Tyrosine kinases are frequently activated in human tumors and constitute novel targets for anti-tumor therapies (2).

Characteristics

The catalytic module of tyrosine kinases consists of about 260–300 amino acid residues. The crystal structure of several tyrosine kinase domains has been solved and reveals a common architecture (3, 4). The catalytic domain of a tyrosine kinase forms a classical two lobed structure. The N-terminal lobe is composed of anti-parallel β -strand sheets and an α helix, while the C-terminal lobe is mainly α helical (Fig. 1). The ATP binding pocket is located in the hinge region between the two lobes, while the substrate binding region, catalytic loop and activation loop all reside in the C-terminal lobe. Most tyrosine kinases harbor one to three tyrosine residues within the activation loop. Phosphorylation of these A loop tyrosine residues enhances tyrosine kinase activity.

According to their different modular organization, tyrosine kinases are divided into two subfamilies, the ► [receptor tyrosine kinases](#) (RTKs) and the non-receptor tyrosine kinases (NRTKs). The receptor tyrosine kinases (RTKs), like the epidermal growth factor (EGF) receptor or the insulin receptor are transmembrane glycoproteins (Fig. 2). They possess an extracellular region, which is glycosylated, the transmembrane domain and a cytoplasmic part. The extracellular domain is diverse among the different RTK subfamilies and composed of several subdomains such as the immunoglobulin (Ig), fibronectin type III or epidermal growth factor (EGF)-like domains (Fig. 2). The cytoplasmic part contains the kinase domain, regulatory sites and substrate binding sites. The activity of RTKs is regulated by binding of extracellular peptide growth factors, such as EGF or insulin, which bind and activate the EGF receptor or insulin receptor,



Tyrosine Kinases. Figure 1 Ribbon diagram of insulin receptor tyrosine kinase. Crystal structure of activated insulin receptor tyrosine kinase in complex with peptide substrate and the ATP analog AMPPNP (4). The kinase domain forms a two lobed structure. The N-terminal lobe is shown on top and the C-terminal lobe at the bottom. The α -helices are depicted in red, β -sheets in turquoise, the activation loop in blue, the peptide substrate in yellow, the ATP analog AMPPNP in pink and Mg^{2+} ions in orange. ATPNP binds to the hinge region between the N-terminal and the C-terminal lobes. The substrate binding site and the activation loop reside in the C-terminal lobe. The picture was generated by Klaus Hellmuth with Accelrys ViewerLite 5.0 according to the atom coordinates PDB accession number 1IR3.

respectively. NRTKs on the other hand are cytoplasmic or membrane associated proteins. They harbor protein-protein interaction domains involved in subcellular targeting and/or regulation of enzymatic activity, like c-Src or the Janus kinase JAK (Fig. 3). NRTKs function downstream of RTKs, G-protein coupled receptors or cell surface receptors that lack intrinsic enzymatic activity, such as B cell receptors, T cell receptors or integrins (see below).

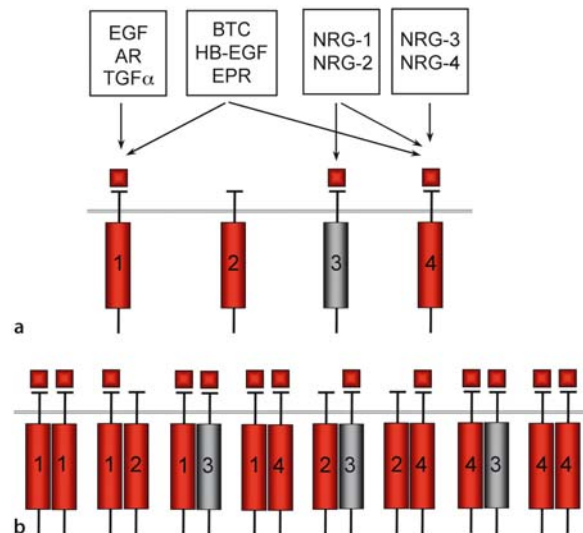
Molecular Interactions Receptor Tyrosine Kinases

Binding of extracellular ligand to RTKs leads to activation of an intracellular signaling cascade that spreads to the nucleus and leads to changes in cell

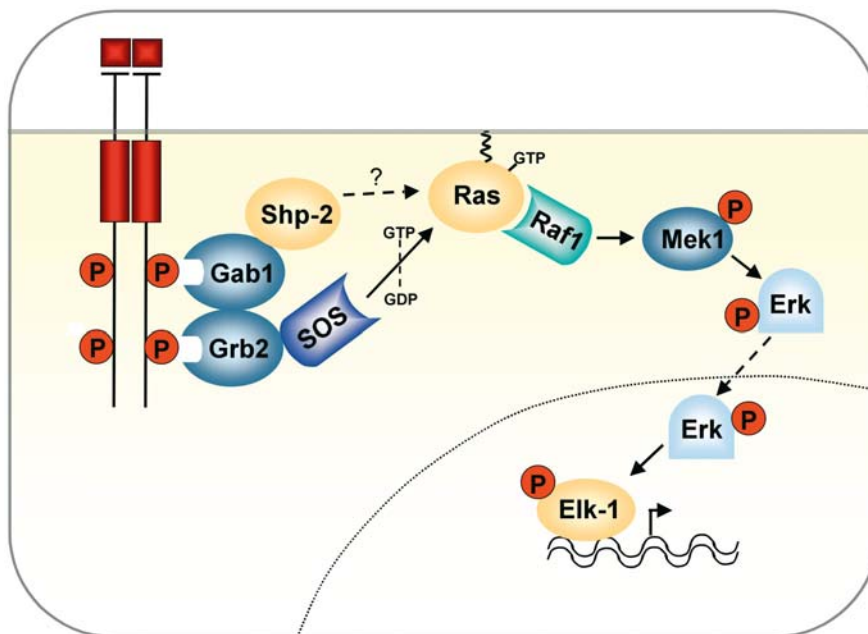
morphology and gene expression. RTKs are activated by peptide growth factors. Some receptors have only one ligand, such as the c-Met RTK, which is activated by scatter factor/hepatocyte growth factor (SF/HGF). Others such as the EGF- or FGF-receptor families recognize multiple ligands (Fig. 4a). These can be either soluble or membrane bound. Ligand binding of most RTKs induces multi-merization of the receptors. In addition to homodimerization, the EGF-receptor can also heterodimerize with other EGF-R family members, such as the RTKs ErbB2, ErbB3 or ErbB4 (Fig. 4b). Homo- and hetero-dimers of the different EGF-receptor families have various ligand binding and signaling properties (5). The EGF receptor recognizes several ligands such as epidermal growth factor (EGF), transforming growth factor α (TGF α), betacellulin (BTC) and heparin binding EGF (HB-EGF). ErbB3 and ErbB4 recognize members of the neuregulin family of peptide growth factors. Thus far, no ligand for ErbB2 has been discovered. It is assumed that ErbB2 functions primarily as a cofactor for the other ErbB family members. The ErbB3 RTK binds neuregulin growth factors, but lacks intrinsic tyrosine kinase activity. Thus it is active only in concert with other ErbB family members (Fig. 4b).

Once activated, the RTKs become autophosphorylated on tyrosine residues. This creates binding sites for proteins with phosphotyrosine interaction domains, such as src homology 2 (SH2) or phosphotyrosine binding (PTB) domains (1). SH2 or PTB domains are frequently found on proteins involved in tyrosine kinase signaling. SH2 domains are compact structures of about 100 amino acids. They bind distinct peptide sequences that are defined by a phosphotyrosine and 3–6 amino acids C-terminally. PTB domains also form compact structures, but are unrelated to SH2 domains. Their binding specificity is defined by one phosphotyrosine and 3–5 residues N-terminally. Thus the amino acid sequences surrounding the phosphotyrosine docking sites of RTKs or kinase substrates determine the combination of SH2 and PTB containing signaling proteins that are recruited and activated by different extracellular signals.

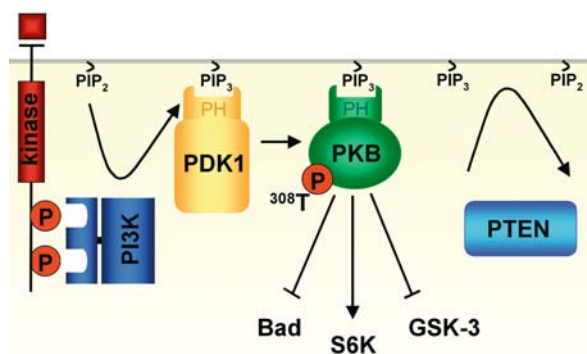
Signaling proteins recruited by RTKs can be adapter molecules, docking proteins, enzymes or even transcription factors. A common adaptor protein recruited by RTKs is growth factor receptor binding protein 2 (Grb2). Grb2 is modular in nature, consisting of a central SH2 domain flanked by a C-terminal and a N-terminal Src homology 3 (SH3) domain. Via its SH2 domain Grb2 binds specific phosphotyrosine residues. The SH3 domains of Grb2 bind other signaling proteins such as SOS, a GDP/GTP exchange factor for the GTPase Ras or the docking protein Gab1 (Fig. 5) (6). Recruitment of Grb2 by RTKs brings the Grb2 binding partner SOS in close proximity to the membrane



Tyrosine Kinases. Figure 4 ErbB receptor tyrosine kinases and their ligands. (a) EGF-R (ErbB1), ErbB2, ErbB3 and ErbB4 have different ligand binding and signaling properties. No ligand has been identified for ErbB2. ErbB3 lacks tyrosine kinase activity. EGF, epidermal growth factor; TGF α , transforming growth factor alpha; AR, amphiregulin; EPR, epiregulin; BTC, betacellulin; HB-EGF, heparin binding EGF; NRG, neuregulin. (b) Heterodimerization by ErbB family members enhances complexity of signal transduction.

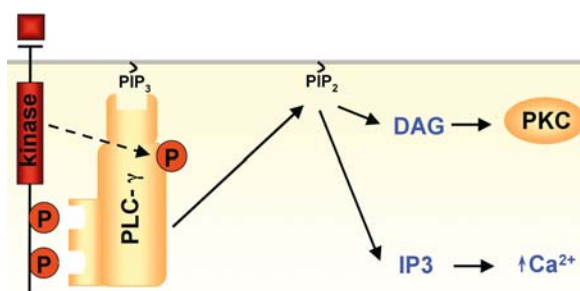


Tyrosine Kinases. Figure 5 The Ras/MAPK pathway. The adaptor protein Grb2 associates constitutively with the GDP/GTP exchange factor SOS. Grb2/SOS complex is targeted to the membrane by binding to phosphotyrosine residues of activated cell surface receptors. This leads to formation of GTP bound Ras protein, and activation of the MAP kinase pathway. Ras activates the serine threonine kinase Raf1, which then activates the MAPKK (Mek1). Mek1 phosphorylates the MAPK Erk. Phosphorylated Erk translocates to the nucleus, where it phosphorylates transcription factors like Elk-1 leading to changes in gene expression. Grb2 can also recruit the docking protein Gab1 to the membrane. Phosphorylation of Gab1 by tyrosine kinases induces binding and activation of the tyrosine phosphatase Shp2. Shp2 activation stimulates Ras by unknown mechanisms leading to sustained MAPK activity. P in red circles indicates phosphate groups.



Tyrosine Kinases. Figure 6 The PI(3)K signaling pathway. PI(3)K is activated by binding of its regulatory subunit to phosphotyrosine docking sites and subsequent tyrosine phosphorylation of its catalytic subunit. It phosphorylates the 3' position of the inositol ring to generate phosphoinositides like PtdIns (3,4,5)P₃ (PIP₃) from PtdIns (4,5)P₂ (PIP₂) and PtdIns (3,4)P₂ from PtdIns (4)P. PIP₃ binds the PH domains of a variety of proteins. PDK1 is activated by PIP₃ dependent recruitment to the membrane, where it phosphorylates and activates another serine threonine kinase PKB/Akt at threonine 308. PKB/Akt phosphorylates several target proteins that regulate cell survival, metabolism, proliferation and transcription. For instance PKB/Akt prevents apoptosis by inactivating Bad, a pro-apoptotic Bcl-2 family member. It blocks GSK-3 (glycogen-synthase kinase-3) thereby promoting cell-cycle entry and changes in glucose metabolism. Ribosomal p70 S6K is also activated by PKB/Akt and promotes cell growth through control of protein translational machinery. PTEN antagonizes PI(3)K signaling by dephosphorylation of the second messengers PtdIns (3,4,5)P₃ (PIP₃) and PtdIns (3,4)P₂ (PIP₂) at the 3' position of phosphoinositides.

(PtdIns (4)P) and phosphatidylinositol 4,5-bisphosphate (PIP₂) to generate the second messengers phosphatidylinositol 3,4-bisphosphate (PtdIns (3,4)P₂) and phosphatidylinositol 3,4,5-trisphosphate (PIP₃), respectively (1, 2). PI(3)K consists of a catalytic subunit, p110 and a regulatory subunit, p85. The regulatory subunit possesses two SH2 domains and forms a constitutive complex with the p110 PI(3)K catalytic subunit. Binding of its SH2 domains to phospho-tyrosine residues of activated receptor induces a conformational shift releasing the auto inhibitory constraint on the catalytic subunit (p110). In addition, localization of the enzyme to the plasma membrane brings it in closer proximity to its lipid substrates PtdIns (4)P and PIP₂. The reaction products PtdIns (3,4)P₂ and PIP₃ activate other signaling proteins, such as the serine/threonine kinases PDK1 and protein kinase B (PKB)/Akt (Fig. 6). Targets of PI(3)K signaling, often

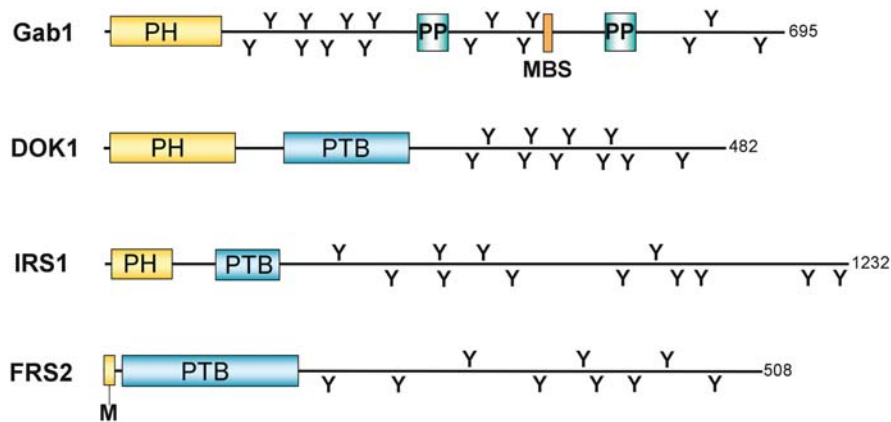


Tyrosine Kinases. Figure 7 The PLC- γ signaling pathway. Phospholipase C-gamma (PLC- γ) is recruited to the membrane by binding of its PH domain to PtdIns (3,4,5)P₃ (PIP₃) and by its SH2 domains that bind phosphotyrosine peptide motifs. Full activation of PLC- γ also requires tyrosine phosphorylation. PLC- γ cleaves the PI(3)K reaction product PtdIns (4,5)₂ (PIP₂) to generate the second messengers diacylglycerol (DAG) and inositol (1,4,5) P (IP₃). IP₃ induces the release of calcium from intracellular storage. Ca²⁺ and DAG then activate protein kinase C (PKC).

contain a Pleckstrin homology (PH) domain which binds PIP₃. Activation of PI(3)K then leads to translocation of the proteins to the plasma membrane and activation (1,2).

Phospholipase C- γ (PLC- γ) contains a SH2 domain and is directly recruited by RTKs (1, 2). It is a substrate of tyrosine kinases and activated by tyrosine phosphorylation. PLC- γ hydrolyzes PIP₂ to diacylglycerol (DAG) and inositol 1,4,5-trisphosphate (IP₃). IP₃ stimulates the release of Ca²⁺ from intracellular storage compartments, activating Ca/calmodulin-dependent kinases (1). Ca kinases together with diacylglycerol activate protein kinase C (Fig. 7).

The tyrosine phosphatase Shp2 is also an important signaling molecule downstream of RTKs (6). Shp2 contains two SH2 domains at its N-terminus and the phosphatase (PTP) catalytic domain. In the basic state the phosphatase activity is blocked sterically by the N-terminal SH2 domain. Interaction of the SH2 domains with phosphotyrosine interaction sites leads to activation of Shp2 phosphatase activity. Shp2 is a positive regulator of the mutagen activated protein kinase (MAPK) pathway. Only a few physiological substrates of Shp2 are known, but genetically Shp2 has been placed upstream of the Ras/MAPK pathway (Fig. 5). Docking proteins like insulin receptor substrate 1 (IRIS-1) or Grb2 associated binder 1 (Gab1) comprise another group of signaling proteins activated by RTK (Fig. 8). They are substrates of tyrosine kinases, but lack intrinsic enzymatic activity (1). They are characterized by an N-terminal membrane targeting domain, like a Pleckstrin homology domain (PH), a



Tyrosine Kinases. Figure 8 Docking proteins. Docking proteins are substrates of tyrosine kinases. They harbor multiple tyrosine residues. When phosphorylated, these can function as binding sites for signaling molecules. Commonly, docking proteins possess a Pleckstrin homology domain (PH) or a myristoylation site (M) that is important for their membrane targeting. In addition they contain a phosphotyrosine-binding domain (PTB) that is required for their targeting to specific signaling complexes. Gab1 lacks a classical PTB domain, but instead has a c-Met binding site (MBS) important for its targeting to the RTK, c-Met. PP indicates proline-rich regions that mediate protein-protein interactions with SH3 domains. The proline-rich regions of Gab1 bind SH3 domains of Grb2 (7).

central phosphotyrosine binding domain (PTB), which mediates specific binding to growth factor receptors and numerous tyrosine residues that serve as additional docking sites for signaling molecules. Docking proteins are recruited by binding to phosphotyrosine binding sites or by membrane targeting *via* a PH domain that binds the PI(3)K reaction product PIP₃ (Fig. 6). Phosphorylation of docking proteins may enhance the signaling properties of RTKs by providing additional phosphotyrosine binding sites to build a larger platform of signaling complexes.

STATs are transcription factors with a central DNA binding domain, a putative SH3 domain and a C-terminal SH2 domain (7). They are also substrates of tyrosine kinases and recruited *via* association of their SH2 domain with phosphotyrosine docking sites. Once phosphorylated, they dimerize by reciprocal interaction of phosphotyrosine and the SH2 domain, translocate to the nucleus and activate transcription (Fig. 9).

Non-Receptor Tyrosine Kinases

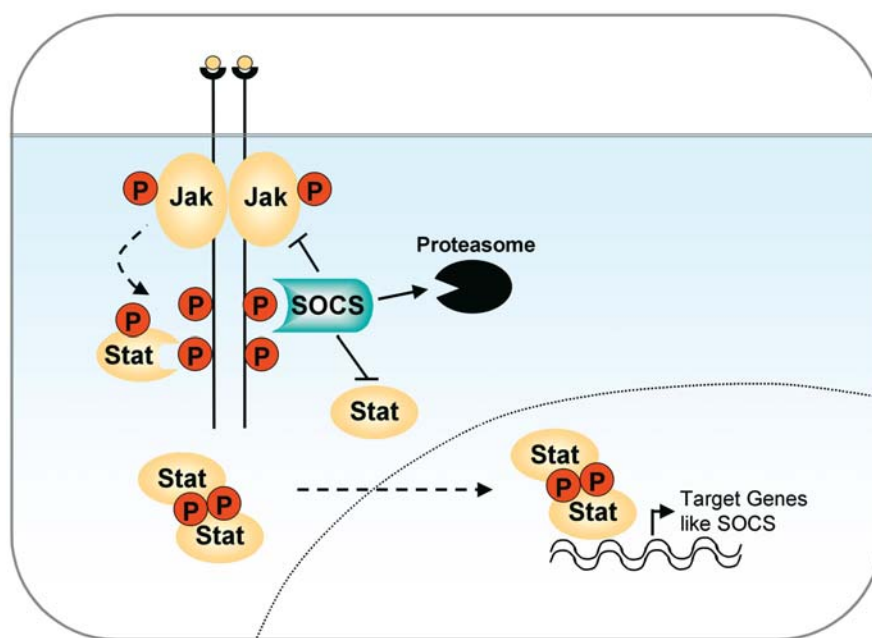
NRTKs can be activated by RTKs, but function also downstream of membrane receptors that lack intrinsic enzymatic activity (2, 3). For instance, the Janus family of NRTKs, JAKs, is commonly activated by cytokine receptors, like interleukin (IL) 2, IL3, interferon (IFN) or erythropoietin receptor (7). Janus kinases (JAKs) contain a cytokine receptor-binding domain (FERM domain) at their N-terminus and associate constitutively with specific cytokine receptors (Fig. 9). Ligand induced activation of Janus kinases leads to phosphorylation of the cytokine receptor on tyrosine residues and recruitment and activation of additional signaling

molecules, in particular STAT transcription factors and Src family kinases.

The src family kinases act downstream of the T cell and B cell receptors and the receptors for the γ portion of immunoglobulin (IgG). These receptors belong to the family of multi-chain immune recognition receptors (MIRR). They are composed of several transmembrane glycoproteins involved in ligand binding and/or signaling. Characteristically, the cytoplasmic part of the signaling components contains one or more of the tyrosine based activation motifs (ITAM), a peptide sequence with the sequence motif YX₂LX₇YX₂L (7). Ligand binding induces receptor clustering and thereby activates closely associated src kinase family members. This leads to rapid phosphorylation of the ITAM motifs. Upon phosphorylation, ITAM tyrosine residues function as docking sites and recruit additional effector molecules similar to the activity of RTKs or cytokine receptors.

Regulatory Mechanisms

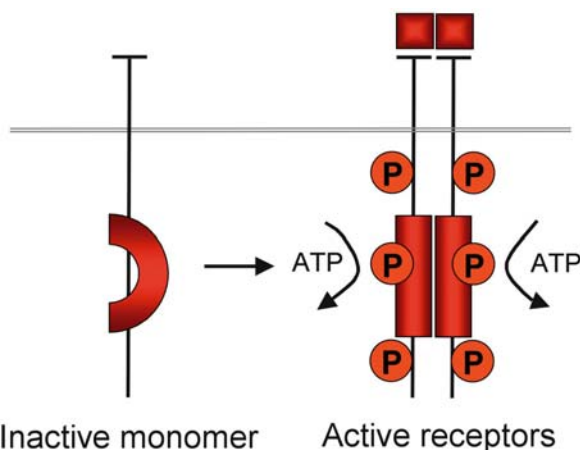
The activity of tyrosine kinases is tightly regulated (1, 2, 3). In the absence of ligand, most RTKs exist as monomers in the cell membrane. They display residual kinase activity, but are kept in a low activity state due to the activity of tyrosine phosphatases. Crystal structures of tyrosine kinases can show that in the non-phosphorylated state the A loop of the catalytic center impairs accessibility to substrate or ATP. Binding of extracellular ligand induces receptor dimerization increasing local receptor concentration. This facilitates transphosphorylation of the receptor pair. Phosphorylation of the A loop tyrosine residues



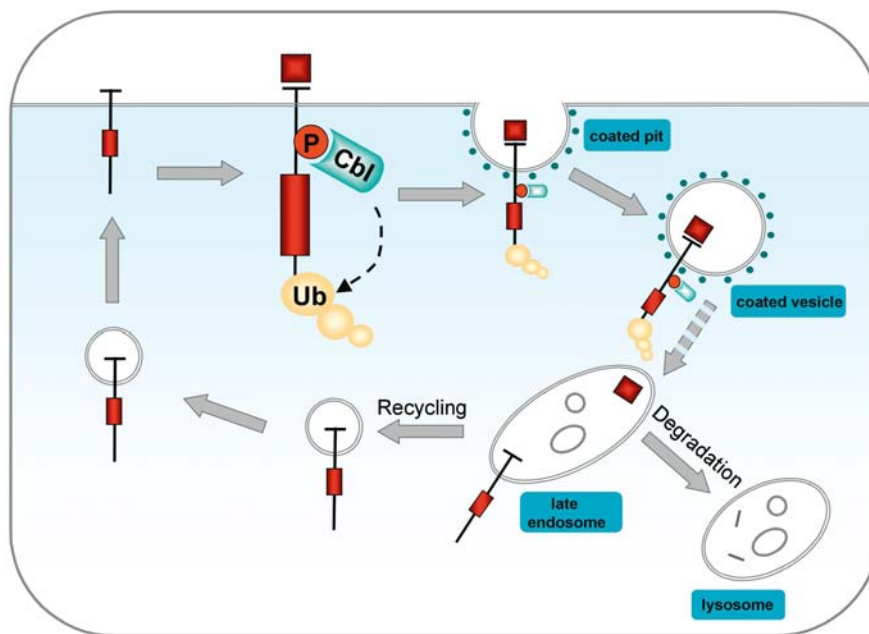
Tyrosine Kinases. Figure 9 Janus kinases. Janus kinases (JAKs) are constitutively associated with cytokine receptors, which lack intrinsic enzymatic activity. Cytokines induce dimerisation of the receptors and activation of the associated Janus kinases. Janus kinases phosphorylate the cytoplasmic tail of the cytokine receptors creating binding sites for signal transducers and activators of transcription (STATs). After phosphorylation by Janus kinases, STATs homodimerize, translocate to the nucleus and activate transcription. Suppressors of cytokine signaling (SOCS) are also target genes for STATs. SOCS proteins are involved in down-regulation of cytokine response. SOCS proteins promote degradation of JAKs by targeting them to the proteasome. They also interfere with STAT signaling by occupying STAT-binding sites on cytokine receptors.

then shifts the state of the kinase into an open, more active conformation and allows autophosphorylation to proceed (Fig. 10).

Tyrosine kinase signaling is shut down by several mechanisms. Activation of RTKs reduces the half-life of the receptor (1). Ligand bound receptors cluster in coated pits on the cell surface, undergo endocytosis and are transported into multivesicular bodies (Fig. 11). There the ligand and receptor dissociate; the receptor is recycled to the cell surface or undergoes degradation in the lysosomes. For the EGF-R and the RTK c-Met it has been shown that the c-Cbl proto-oncogene product promotes degradation of activated receptor *via* the proteasome. c-Cbl is a multi-domain adaptor protein and E3-ligase (5). It contains an SH2-like domain responsible for binding to the EGF-R and c-Met receptor at specific sites. It also has a ring finger domain which functions as an ubiquitin ligase, thereby targeting the protein complex to the proteasome and degradation. Consequently, EGF-R or c-Met mutants, which lack c-Cbl binding sites, show increased half-life and oncogenic signaling capacity.



Tyrosine Kinases. Figure 10 Activation of receptor tyrosine kinases. In the absence of ligand, RTKs exist as monomers. Ligand binding induces dimerization and a conformational shift that allows transphosphorylation of activation loop residues. This enhances tyrosine kinase activity resulting in autophosphorylation of additional tyrosine residues.



Tyrosine Kinases. Figure 11 Down-regulation of receptor tyrosine kinase signaling. The E3 ubiquitin ligase c-Cbl binds to specific phosphotyrosine binding sites of the activated EGF-R and c-Met receptor (only one receptor molecule is depicted) and leads to ubiquitination (Ub) of the activated receptor. The ubiquitinated receptors are sorted into invaginating clathrin coated pits and engulfed into coated vesicles. These vesicles shed the clathrin coat and give rise to early and late endosomes. During this maturation process, ligand and receptors dissociate due to lowered pH in the late endosome and are either recycled back to the cell surface or destined to degradation in the lysosome. Degradation is promoted by c-Cbl ubiquitination, since c-Met or EGF-R mutants deficient in c-Cbl binding are characterized by enhanced half-life.

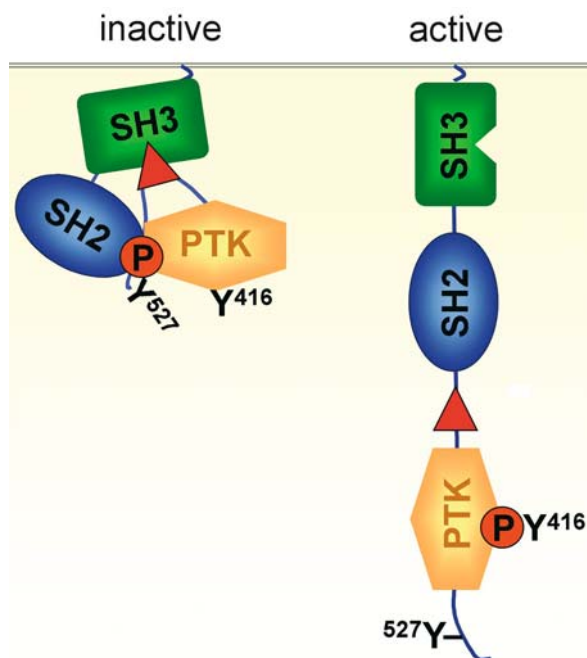
Regulation of NRTKs

The activity of NRTKs is also tightly regulated (2, 3). c-Src for instance contains an N-terminal SH3 domain, followed by an SH2 domain and the kinase domain (Fig. 12). In addition, it has two regulatory tyrosine residues Y416 and Y527(3). Y416 is located in the A loop, phosphorylation of this residue is required for full activation of src. Y527 is phosphorylated by the C-terminal src kinase (CSK) a negative regulator of src. When Y527 is phosphorylated, it forms an intramolecular interaction with its SH2 domain stabilizing association of the SH3 domain with proline-rich sequences. As a result of these intramolecular interactions, the kinase catalytic core is maintained in a closed inactive conformation. Src can be activated by binding of its SH2 domain to phosphotyrosine peptide sequences which compete with binding of pY 527 at its C-terminus, followed by transphosphorylation of Y416 in the A loop. In addition, src is activated through the activity of tyrosine phosphatases, which dephosphorylate Y527 or by association of proline-rich sequences with the SH3 domain of src.

JAK tyrosine kinase family members are non-covalently bound to cytokine receptors (Fig. 9). Receptor dimerization induced by ligand binding leads to JAK dimerization facilitating transphosphorylation of A loop tyrosine residues. These tyrosine residues function to maintain an active state until dephosphorylation by phosphatases, like SHP1 or CD45 (7). Signaling by JAKs is further down regulated by SOCS proteins (suppressors of cytokine signaling), also called CIS (cytokine induced SH2 containing cytokines) proteins. SOCS proteins contain a central SH2 domain and a C-terminal SOCS box, which functions in a multi-protein complex as an E3 ubiquitin ligase. Cytokine signaling induces SOCS protein expression and association of SOCS with phosphorylated JAKs or cytokine receptor *via* its SH2 domain (Fig. 9). The SOCS box recruits ubiquitin 3 ligases to the protein complex leading to ubiquitination of JAK and cytokine receptor and finally degradation by the proteasome (7).

► Growth Factors

► Neurotrophic Factors



Tyrosine Kinases. Figure 12 Regulation of c-src kinase activity. Src kinase is held in an inactive state by two intramolecular interactions: (i) tyrosine 527 is phosphorylated by the C-terminal src kinase (CSK), and bound to the SH2 domain, (ii) the SH3 domain interacts with a proline-rich region (red triangle) in the SH2 domain linker. Src kinase activity is induced by dephosphorylation of Y527, by binding of other proline-rich sequences to the SH3 domain of Src or by binding of its SH2 domain to other phosphotyrosine binding sites. This promotes a more open conformation of the kinase and allows autophosphorylation of Y416 in the activation loop of src resulting in full activation of src kinase activity.

References

- Schlessinger J (2000) Cell Signalling by Receptor Tyrosine Kinases. *Cell* 103:211–225
- Blume-Jensen P, Hunter T (2001) Oncogenic kinase signalling. *Nature* 411:355–365
- Hubbard SR, Hill JH (2000) Protein Tyrosine Kinase Structure and Function. *Ann Rev Biochem* 69:373–398
- Hubbard J (1997) Crystal structure of activated insulin receptor tyrosine kinase in complex with peptide substrate and ATP analog *EMBO J* 16:5573–5581
- Citri A, Skaria KB, Yarden Y (2003) The deaf and the dumb: the biology of ErbB-2 and ErbB-3. *Exp Cell Res* 284:54–65
- Schaeper U, Gehring NH, Fuchs KP et al (2000) Coupling of Gab1 to c-Met, Grb2, and Shp2 Mediates Biological Responses *JCB* 149:1419–1432
- Rane GR, Reddy EP (2000) Janus kinases: components of multiple signalling pathways. *Oncogene* 19:5662–5679

Tyrosine Sulfation of Proteins

KEVIN L. MOORE

Oklahoma Medical Research Foundation,
Oklahoma City, OK, USA
kevin-moore@omrf.ouhsc.edu

Synonyms

3'-phosphoadenylyl-sulfate; protein-tyrosine *O*-sulfotransferase; protein-tyrosine sulfotransferase; tyrosyl-protein sulfotransferase; 3'-phosphoadenylyl-sulfate; 3'-phosphoadenosine 5'-phosphosulfate; adenosine 3', 5'-bisphosphate; 3'-phosphoadenosine 5'-phosphate

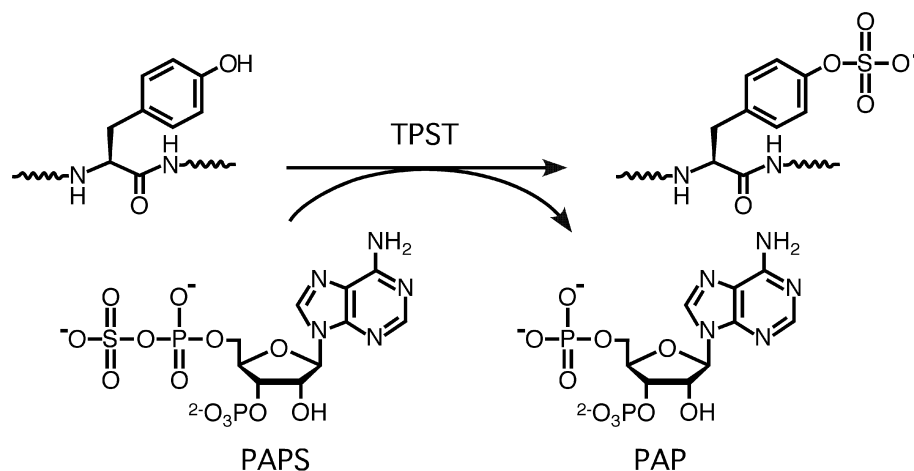
Definition

Tyrosine sulfation is one of many post-translational modifications of proteins that have been described in nature. Sulfation of other amino acids with side-chain hydroxyl groups has not been described. Sulfated protein-tyrosine residues are formed by the action of an enzyme activity called 3'-phosphoadenylyl-sulfate: protein-tyrosine *O*-sulfotransferase, commonly called tyrosylprotein sulfotransferase (TPST, EC 2.8.2.20). This enzyme activity mediates the transfer of the 5'-sulfonyl group from the universal high-energy sulfate donor, 3'-phosphoadenosine 5'-phosphosulfate (PAPS), to the side chain hydroxyl group of tyrosine residues to form a tyrosine *O*⁴-sulfate ester and 3'-phosphoadenosine 5'-phosphate (PAP). In mammalian cells, it has been estimated that tyrosine sulfation accounts for only ~5% of the sulfate incorporated into proteins. The remaining ~95% is accounted for by sulfation of various monosaccharide units in Asn-linked and Ser/Thr-linked glycans in glycoproteins and proteoglycans. A family of Golgi **▶carbohydrate sulfotransferases** mediates these sulfate conjugation reactions. For reviews on tyrosine sulfation and TPSTs see references (1, 2).

Characteristics

Species and Tissue Distribution

TPST activity and/or tyrosine-sulfated proteins have been observed in many species in both the plant and animal kingdoms, but not in prokaryotes or yeast. Complementary DNAs have been identified in many eukaryotic organisms from *Caenorhabditis elegans* to man. In most species there are two TPST isoenzymes, called TPST-1 and TPST-2. The one known exception to this rule is *Drosophila melanogaster* that has only a single TPST gene.



Tyrosine Sulfation of Proteins. Figure 1 The Tyrosylprotein Sulfotransferase Reaction. TPSTs catalyze the transfer of the sulfonyl group from the sulfate donor, 3'-phosphoadenosine 5'-phosphosulfate (PAPS), to the hydroxyl group of a lumenally-oriented protein-tyrosine residue to form a tyrosine O^4 -sulfate ester and 3'-phosphoadenosine 5'-phosphate (PAP).

The cellular and tissue distribution of TPST expression is best characterized in mouse and man. Based on Northern blot analysis of tissues and various cell lines, enzymatic assays of tissue extracts and the representation of TPST-1 and -2 expressed sequence tag clones in a wide variety of cDNA libraries, it appears that both isoenzymes are broadly co-expressed in mammalian cells. The sub-cellular localization of TPST has also been examined using a variety of techniques. These studies indicate that the enzymes are localized to the trans-Golgi network. Therefore, only protein in the secretory pathway can be tyrosine-sulfated.

Structure and Function of TPSTs

Complementary DNAs encoding TPST-1 and/or TPST-2 orthologs have been identified in many vertebrates (human, mouse, rat, dog, cow, pig, chicken, zebrafish, fugu, channel catfish and African clawed frog) and invertebrate species (*Caenorhabditis elegans*, *Drosophila melanogaster*, *Anopheles gambiae*, *Ciona intestinalis*, *Halocynthia roretzi* and *Schistosoma japonicum*). These cDNAs encode type II transmembrane proteins ranging from 340 to 380 residues in length. Each has a short < 10 residue cytoplasmic domain, a single ~17 residue membrane-spanning domain, followed by a luminal catalytic domain. Multiple sequence alignments of TPSTs from various species show that the membrane proximal 30–40 residues of the luminal domain are very poorly conserved. This domain probably represents a stem region similar to that described in many [glycosyltransferases](#). However, whether this region is dispensable

for catalytic activity awaits experimental confirmation. All TPSTs have 1 or 2 putative [Asn-linked glycosylation](#) sites. In human TPST-1 and -2 both Asn-linked sites are utilized. Site-directed mutagenesis of both Asn-linked sites (Asn to Gln) and expression of catalytically active recombinant TPST-1 and -2 in bacteria demonstrate that Asn-linked glycans are dispensable for catalysis. The number of luminal cysteine residues in TPSTs ranges from 4 in *C. intestinalis* and *H. roretzi* to 7 in *D. melanogaster*. Using the human sequences as an alignment template, the second, third, fourth and fifth cysteine residues are conserved in all species and are likely to be involved in intra-molecular disulfide bond formation. However, the disulfide-bonding pattern has not been determined.

Crystal structures of several members of the sulfotransferase family reveal two motifs, designated the 5' PSB and 3'PB motifs. These motifs are involved in binding of the 5'- and 3'-phosphate groups of the reaction product PAP, respectively (3). These motifs are conserved in TPSTs and correspond to residues 78–85 and residues 180–195 in human TPST-1.

Substrate Repertoire and Specificity of TPSTs

Since the first description of protein tyrosine O -sulfation in 1954, 64 different secreted and transmembrane proteins have been shown to be tyrosine-sulfated. The list of known tyrosine-sulfated proteins includes several adhesion molecules, coagulation factors, G-protein coupled receptors, extra-cellular matrix proteins, serpins, hormones and a variety of other proteins (see reference (2) for a comprehensive list).

In contrast to some other post-translation modifications, there is no sequon for tyrosine *O*-sulfation. Nevertheless, the dominant primary structural characteristic of known sulfation sites is the presence of several acidic amino acids within 5 residues of the sulfotyrosine. A software tool called Sulfinator ([▶www.expasy.org/tools/sulfinator/](http://www.expasy.org/tools/sulfinator/)) has been developed to predict tyrosine *O*-sulfation sites in proteins. However, the positive predictive value of this algorithm is uncertain. Nevertheless, application of this prediction tool to the human proteome would suggest that many other proteins (perhaps hundreds) may be tyrosine-sulfated (2).

Information about the differential substrate specificity of TPST-1 and -2 is limited. Nevertheless, *in vitro* enzyme assays using synthetic peptide acceptors indicate that TPST-1 and -2 differ in substrate preference. Peptides modeled on the sulfation sites of human C4a chain and heparin cofactor II are more efficiently sulfated by TPST-1 than by TPST-2. In addition, sulfation of a peptide modeled on the N-terminal domain of human CCR5, which has 4 potential tyrosine *O*-sulfation sites, has been examined by determining the time course of appearance of reaction products. Both TPST isoenzymes sulfated Tyr¹⁴ and Tyr¹⁵ first, followed by Tyr¹⁰ then Tyr³. However, TPST-1 clearly preferred Tyr¹⁴ over Tyr¹⁵ as the initial sulfation site, whereas TPST-2 preferred Tyr¹⁵ over Tyr¹⁴.

Role of Tyrosine *O*-Sulfation in Protein Function

The general consensus in the field is that tyrosine sulfate residues play a direct role in protein-protein interactions. However, evidence supporting this conclusion is not extensive. This assertion is most strongly and directly supported by three examples in which tyrosine-sulfated proteins [hirugen, P-selectin glycoprotein ligand -1 (PSGL-1, CD162), glycoprotein Ibα (CD42b)] have been co-crystallized with ligands. In the crystal structures of the hirugen/α-thrombin complex, the complex between a tyrosine-sulfated glycopeptide modeled on the N-terminal domain of PSGL-1 and the lectin-EGF domain of P-selectin and, mostly recently, the structure of the glycoprotein Ibα/α-thrombin complex (4), sulfato-oxygens of tyrosine sulfate residues in hirugen, PSGL-1 and glycoprotein Ibα are involved in direct protein-protein contacts. Among other tyrosine-sulfated proteins, there is an additional subset in which sulfation is clearly required for optimal protein function, although the precise structural basis for this requirement is not known. Thus, tyrosine sulfation is required for optimal proteolytic processing of gastrin, binding of cholecystokinin to the CCK-A receptor, proteolytic activation of factor V and VIII by α-thrombin, binding of glycoprotein Ibα to von Willebrand factor and optimal interactions between several

▶G-protein coupled receptors and their cognate ligands. For the remainder of the known tyrosine-sulfated proteins, there is no information about location, stoichiometry or the role of the sulfotyrosine residue(s) in protein function.

Clinical Relevance

The human *TPST1* and *TPST2* genes are on chromosome 7q11.21 and 22q12.1, respectively. To date, no human genetic disease has been ascribed to defects at either locus. It was hypothesized that ▶Schwachman-Diamond syndrome (SDS) might be the result of defects in the *TPST1* gene because the putative SDS gene maps to a 1.9 cM interval at 7q11 that contains the *TPST1* gene. However, detailed genetic analysis of SDS patients has excluded *TPST1* as the causative gene and SDS-associated mutations have subsequently been identified in an uncharacterized gene, *SBDS*, in this same genetic interval.

Although *TPST1* and *TPST2* genes are not known to be disease genes *per se*, mutations at or near tyrosine sulfation site(s) in any TPST substrate could theoretically produce human disease. The only example of this comes from cases of mild-moderate ▶hemophilia A that are due to missense mutations resulting in substitution of a Phe residue for Tyr¹⁶⁸⁰ in the factor VIII protein. Sulfation of Tyr¹⁶⁸⁰ at the junction of the B and A3 domain of factor VIII is required for optimal binding to von Willebrand factor. This mutation may directly explain the mild to moderate hemophilia in these patients because von Willebrand factor is the major carrier protein for factor VIII in plasma.

Other similar examples will probably come to light as more tyrosine-sulfated proteins are identified and as more examples of the functional importance of this post-translational modification in protein function are documented. A great deal of interest has recently focused on the role of tyrosine *O*-sulfation in ▶chemokine receptor function after CCR5 was shown to be tyrosine-sulfated (5). CCR5 is a seven transmembrane domain G-protein-coupled receptor that functions as a receptor for MIP-1α/CCL3, MIP-1β/CCL4 and RANTES/CCL5 and is a major HIV co-receptor. Sulfation of one or more tyrosine residues in the N-terminal extra-cellular domain of CCR5 is required for optimal binding of MIP-1α, MIP-1β and RANTES and for optimal HIV co-receptor function. In addition, sulfotyrosine-containing peptides modeled on this domain bind to the HIV gp120/CD4 complex and inhibit HIV entry into cells. Similarly, other studies indicate that sulfation of tyrosine residue(s) in the N-terminal domains of other G protein-coupled receptors, including CXCR4, CCR2B, CX3CR1, C5a receptor, C3a receptor and the thyroid-stimulating hormone

receptor, is an important requirement for optimal binding of SDF-1 α /CXCL12, MCP-1/CCL2, fractalkine/CX3CL1, C5a, C3a and TSH respectively. The N-terminal domains of all known chemokine receptors are highly acidic and contain one or more tyrosine residues. This is the dominant feature of known tyrosine sulfation sites suggesting that many chemokine receptors may require tyrosine sulfation for optimal chemokine binding. This possibility has broad pathophysiological implications because chemokine receptors and their cognate ligands play crucial roles in hematopoiesis, innate and adaptive immunity, angiogenesis, tumor growth and metastasis.

New insights into the potential clinical relevance of tyrosine sulfation of proteins are emerging from the studies of TPST-deficient mice. Targeted disruption of the *Tpst1* and *Tpst2* genes results in distinct, unexpected and pleiotropic phenotypic effects. Disruption of one *Tpst* gene does not effect transcription of the other *Tpst* gene as assessed by Northern blot analysis. *Tpst1*^{+/−} and *Tpst2*^{+/−} mice appear normal and, when interbred, yield litters of normal size with a Mendelian genetic distribution of the targeted mutation. However, *Tpst1*^{−/−} mice have ~ 5% lower average body weight than wild type littermates. *Tpst1*^{−/−} females have smaller litters than wild type females due to fetal death at 8.5–15.5 days *post coitum*, although fertility of *Tpst1*^{−/−} mice is normal (6). In contrast to *Tpst1*^{−/−} mice, the growth of *Tpst2*^{−/−} mice is severely delayed. At 4–5 weeks the body weight of *Tpst2* nulls are ~ 20% below wild-type littermates. However, they attain normal body weight by 10 weeks and otherwise appear healthy. Also in contrast to *Tpst1* nulls, *Tpst2*^{−/−} males have severely defective reproductive performance. *Tpst2*^{−/−} males mate with females, although with a longer latency than do wild type males, but they only rarely sire litters. *Tpst2*^{−/−} males have normal reproductive hormone levels (FSH, LH and testosterone); testicular histology is normal, as are sperm counts, morphology and motility. Thus, defective fertility of the *Tpst2* null male may result from defective sperm transport, capacitation, acrosomal reaction and/or fertilization *per se*.

These data demonstrate that the two TPST isoenzymes have distinct biological roles and underscore the

importance of tyrosine *O*-sulfation in several physiological processes that had not been previously appreciated, including growth and reproductive physiology.

Acknowledgement

This work was supported by Grants HL63152 and HL074015 from the United States Public Health Service.

References

1. Huttner WB, Baeuerle PA (1988) Protein sulfation on tyrosine. *Mod. Cell Biol* 6:97–140
2. Moore KL (2003) The biology and enzymology of protein tyrosine O-sulfation. *J Biol Chem* 278:24243–24246
3. Kakuta Y, Pedersen LG, Pedersen LC et al (1989) Conserved structural motifs of the sulfotransferase family. *Trends Biochem Sci* 23:129–130
4. Celikel R, McClintock RA, Roberts JR et al (2003) Modulation of α -thrombin by distinct interactions with platelet glycoprotein Iba. *Science* 301:177–179
5. Farzan M, Mirzabekov T, Kolchinsky P et al (1999) Tyrosine sulfation of the amino terminus of CCR5 facilitates HIV-1 entry. *Cell* 96:667–676
6. Ouyang YB, Crawley JTB, Aston CE et al (2002) Reduced body weight and increased postimplantation fetal death in tyrosylprotein sulfotransferase-1 deficient mice. *J Biol Chem* 277:23781–23787

Tyrosine-Kinase-Associated Receptors

Definition

Tyrosine-kinase-associated receptors comprise of cell surface receptors for a large number of cytokines.

- Growth Factors
- Tyrosine Kinases

Tyrosylprotein Sulfotransferase

- Tyrosine Sulfation of Proteins

UAP56

Definition

UAP56 (56-kDa U2AF-associated protein) and its yeast homolog Sub2 are essential splicing factors that are required for ►[spliceosome](#) assembly, and for the recruitment of Aly/REF to the exon junction complex. UAP56 also plays important roles in the export of mRNA from the nucleus to the cytoplasm.

►[RNA Export](#)

UAS-Gal4 System

Definition

UAS-Gal4 system refers to an *in vivo* system that has been adapted from yeast, and consists of two transgenic lines of *Drosophila* – a driver line (yeast transcription factor Gal4 is expressed under the control of a specific promoter) and a responder line UAS. It is promoter specific for Gal4, linked to the cDNA of choice. After the crossing of both lines, Gal4 binds to the UAS-sequence and induces the expression of the gene or RNAi construct of interest. This system allows *in vivo* expression in a spatial and temporal restricted manner.

►[Drosophila Model of Cardiac Disease](#)

UBF

Definition

UBF or Upstream Binding Factor, is an activator of RNA Polymerase I transcription in vertebrate cells. It is sometimes referred to as UBF1, a 97kDa protein, and UBF2, which is a 94 kDa splice variant.

►[RNA Polymerase I](#)

Ubiquitin

Definition

Ubiquitin is a 76-amino acid polypeptide that is conjugated posttranslationally to internal lysine residues of other proteins. It is ubiquitously present in all eukaryotes, hence its name, and is one of the most highly conserved proteins, with only 3 residues change between yeast and humans. Ubiquitination is a mechanism to induce proteasomal degradation of proteins. E3 ligases are enzymes that, together with E1 and E2 ligases, catalyse the transfer of an ubiquitin polypeptide onto proteins destined for degradation. E1 ligase activates ubiquitin, E2 ligase performs the actual transfer of ubiquitin to the target protein, and E3 ligases are involved in substrate recognition and determining the specificity. Poly-ubiquitinated proteins are degraded in proteasomes.

►[Adherens Junctions](#)

►[Colorectal Cancer](#)

►[Hemochromatosis](#)

►[Limb Girdle Muscular Dystrophies](#)

►[Nucleotide Excision Repair](#)

►[Two-Hybrid System](#)

►[Ubiquitination](#)

Ubiquitin Chains

Definition

Ubiquitin molecules can be linked to internal lysine residues of a previous ubiquitin, thus forming a chain. The linkage between the ubiquitin molecules determines the fate of the modified target protein. The most common ubiquitin chain is linked via the ϵ -amino group of Lysine 48 and the carboxy terminus in the adjacent ubiquitin, and targets the tagged protein to the proteasome for degradation or other processes such as DNA damage repair (e.g. lys63).

- Sumoylation
- Ubiquitination

Ubiquitin Proteasome Pathway

- Ubiquitination
- NF- κ B Pathway

Ubiquitination

THOMAS SOMMER, GUNNAR DITTMAR
Max Delbrück Center for Molecular Medicine, Berlin,
Germany
sommer@mdc-berlin.de

Definition

► **Ubiquitin** is a small protein acting as a molecular tag that is post-translationally attached to other proteins (targets). An enzymatic cascade, consisting of different classes of enzymes, catalyzes this covalent conjugation of ubiquitin in the presence of ATP. In this reaction, the C-terminus of ubiquitin is covalently connected to the ϵ -amino group of a lysine residue in the target molecule, resulting in an ► **isopeptide-bond**. If this modification occurs in multiple rounds, it leads to polyubiquitinated substrates, which are targeted for destruction by the 26S ► **proteasome**. Besides this so-called “classical” function, ubiquitin conjugation also serves non-classical functions as a posttranslational modifier. As a requirement for these non-classical functions ubiquitination is a reversible process through the action of deubiquitinating enzymes (DUB).

Characteristics Ubiquitin

Ubiquitin is highly conserved among all eukaryotes. Almost 96% of the amino acids are identical between yeast and man. All eukaryotes have several genes that code for ubiquitin. Ubiquitin is always expressed as a fusion protein, either with ribosomal subunits or as a multimer of several ubiquitin molecules. Under normal growth conditions, the ribosomal fusion proteins are only expressed under stress conditions and upon entry

into the stationary growth phase the multimeric form of ubiquitin is expressed.

Mature ubiquitin is released from the precursor by a proteolytic cleavage catalyzed by specialized enzymes, the ubiquitin C-terminal hydrolases (UCH). In this reaction the mature C-terminus of ubiquitin with its characteristic GG-motif is created. The conjugation of ubiquitin to a target protein occurs *via* the last glycine of the mature protein. This need for C-terminal processing prevents the modification with a precursor form of ubiquitin.

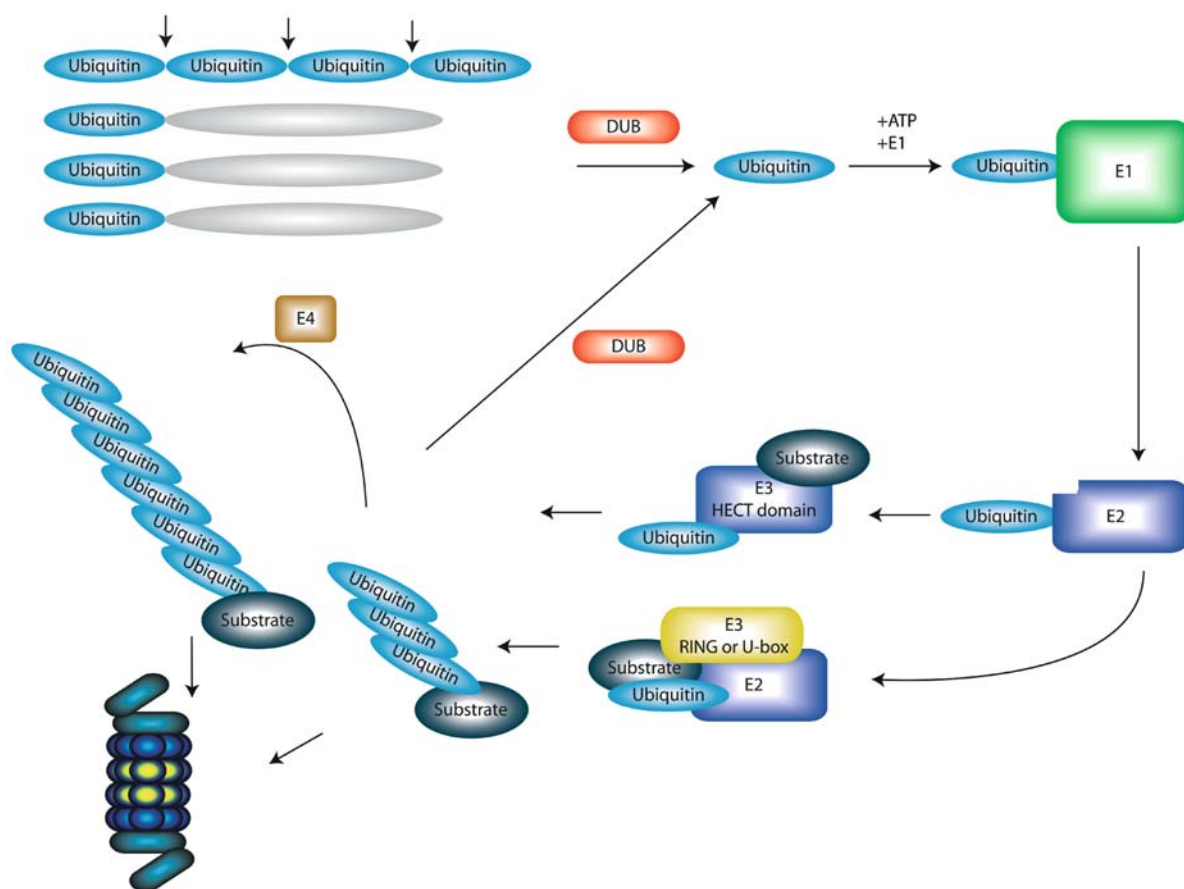
Mono-Ubiquitination and Ubiquitin-Chains

To fulfill both classical and non-classical functions the modification with ubiquitin must transfer more than a single piece of information. This information is encoded in the number of ubiquitin molecules that are added to the target protein and in the way they are connected. The most common modification with ubiquitin occurs in the form of a chain. Since ubiquitin itself contains several lysine residues, ubiquitin can be linked to a previously attached ubiquitin molecule. In this way a chain is synthesized, which can be observed in a SDS-PAGE gel as the typical “ladder”-pattern. In principle each of ubiquitin’s lysine residues can be used to build a polyubiquitin chain, but the most commonly used residue is lysine 48 (K48). K48-linked chains of at least four molecules in length are recognized by the proteasome and the target molecule is then degraded. Current knowledge suggests that only K48 and K29 linked chains are recognized by the proteasome and subsequently lead to the degradation of the target protein. In contrast a K63 linkage seems to act as a modification that alters the activity of the target protein. So far there is little knowledge about the other linkages.

In addition to polyubiquitination the attachment of a single ubiquitin has also been proven to be a vital function. This so called mono-ubiquitination has no effect on the half-life of the target protein, but, at least in some cases, acts as a sorting signal within the cell. This has been suggested for receptor internalization and for sorting into the multivesicular body pathway (mvb).

The Ubiquitination Machinery

The specificity of ubiquitin conjugation is mediated by an enzymatic cascade, which attaches ubiquitin to target proteins. This cascade consists of ubiquitin-activating enzyme (E1), ubiquitin-conjugating enzymes (E2), ubiquitin-ligases (E3) and sometimes ubiquitin-chain-elongation (E4) enzyme. As for all reversible modifying systems, enzymes exist that cleave ubiquitin from substrates. These factors comprise a family of enzymes known as deconjugating enzymes (DUB) (Fig. 1).



Ubiquitination. Figure 1 Creation of a poly-ubiquitin chain is a multi-step enzymatic reaction. Ubiquitin is cleaved of the precursor molecule by specialized enzymes releasing the mature ubiquitin (DUB). The mature ubiquitin is then bound by an ubiquitin-activating enzyme (E1) and activated by hydrolyzing an ATP molecule to AMP. In a following reaction ubiquitin is passed to a conjugating enzyme (E2). With the help of an ubiquitin-ligase (E3), the E2 can connect the C-terminus of ubiquitin with the substrate protein. Alternatively, the activated ubiquitin can be transferred to a HECT-E3 ligase, which in turn transfers the ubiquitin on its substrate. The ubiquitination reaction is repeated several times until an ubiquitin-chain is formed. In some cases the ubiquitin-chain is elongated by an elongation factor (E4). When the ubiquitin-chain is formed there are two possibilities to proceed, either the degradation of the substrate at the proteasome or the deubiquitination reaction catalyzed by deconjugating enzymes (DUB).

Ubiquitin-Activation

A single enzyme – the E1 – catalyzes the activation of mature ubiquitin. In this process ATP is hydrolyzed to AMP and pyrophosphate, which fuels the formation of a thio-ester bond between the active center of the E1 and the C-terminus of ubiquitin.

Ubiquitin-Conjugating Enzymes

In contrast to the activation by a single E1, the second step in this cascade is catalyzed by a variety of conjugating enzymes. In a trans-esterization reaction the activated ubiquitin is transferred to the E2 enzyme. Here it is covalently bound, again *via* a thio-ester bond between the active cysteine in the center of the enzyme and the C-terminus of ubiquitin. The family of

ubiquitin-conjugating enzymes is defined by the E2 catalytic domain (Smart: UbcC; Pfam: UQ_con) and has 11 members in yeast and about 60 in humans. These enzymes catalyze special chain formation (like ubc13, which catalyses K63 linkages) or have a unique localization (Ubc6, which is localized to the ER). This is part of the mechanism that provides the system with the versatility to recognize very divergent substrates without losing specificity.

Ubiquitin-Ligases

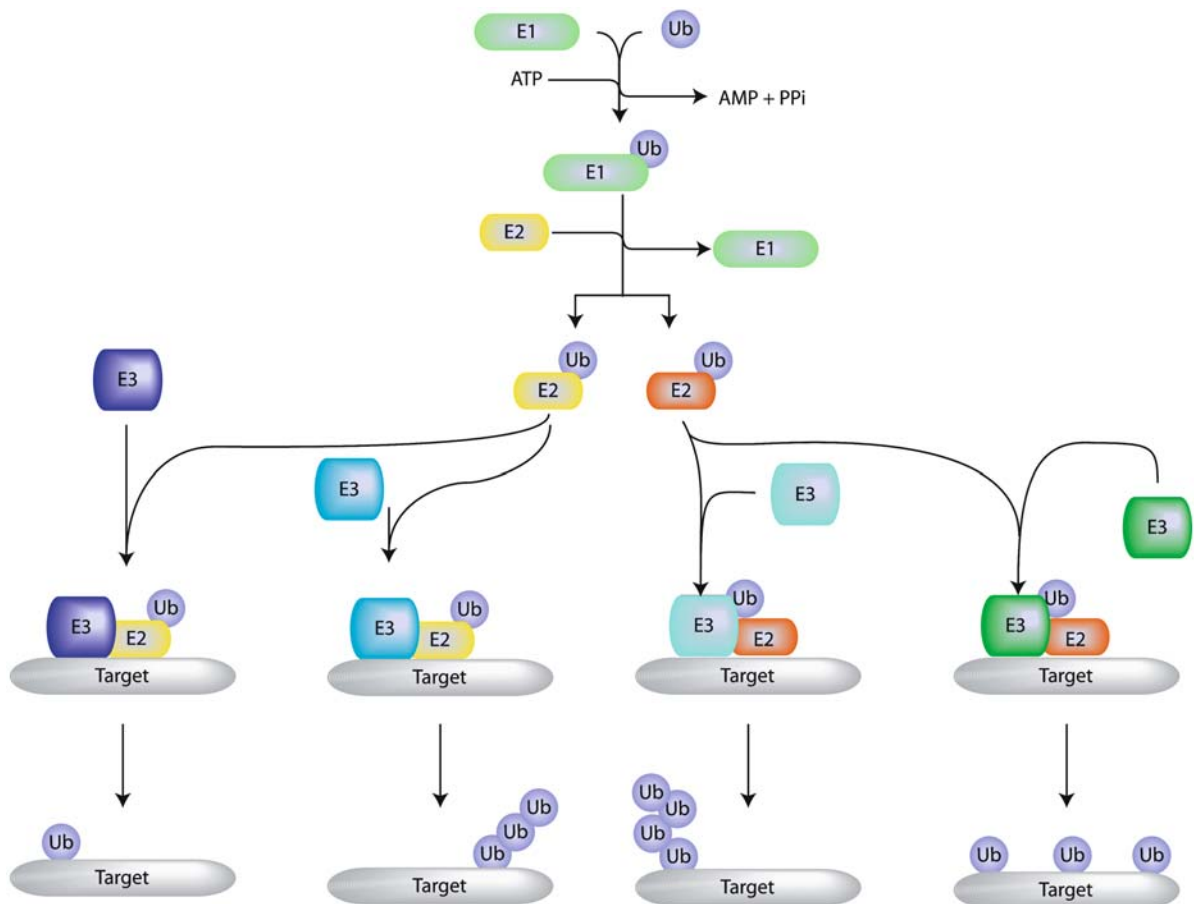
Ubiquitin-ligases catalyze the next step in the cascade. The family of ubiquitin-ligases is probably by far the largest group of enzymes and co-factors within the group of proteins that are involved in ubiquitination.

Two classes of E3 are described so far, E3 that forms a covalent intermediate with ubiquitin (► **HECT-domain**, Pfam: HECT) or E3 that acts as a mediator. The latter contains special amino acid motifs like the RING-finger (Pfam: zf-C3HC4), U-box (Pfam: U-box) or PHD (Pfam: PHD). The relatively small group of HECT-domain E3s is able to bind ubiquitin covalently. This reaction occurs within the 340 amino acid HECT domain. The transfer to the substrate will ultimately be catalyzed without an E2 enzyme present. In contrast to the HECT-E3s, ligases that contain a ring finger or a ring-finger-like motif are not able to bind ubiquitin covalently. They require the recruitment of an E2 for the transfer of ubiquitin to the substrate. The E3 acts as an adapter that recognizes the substrate and mediates the interaction with the E2, which catalyzes the transfer of ubiquitin to the substrate. An E2 enzyme can usually interact with

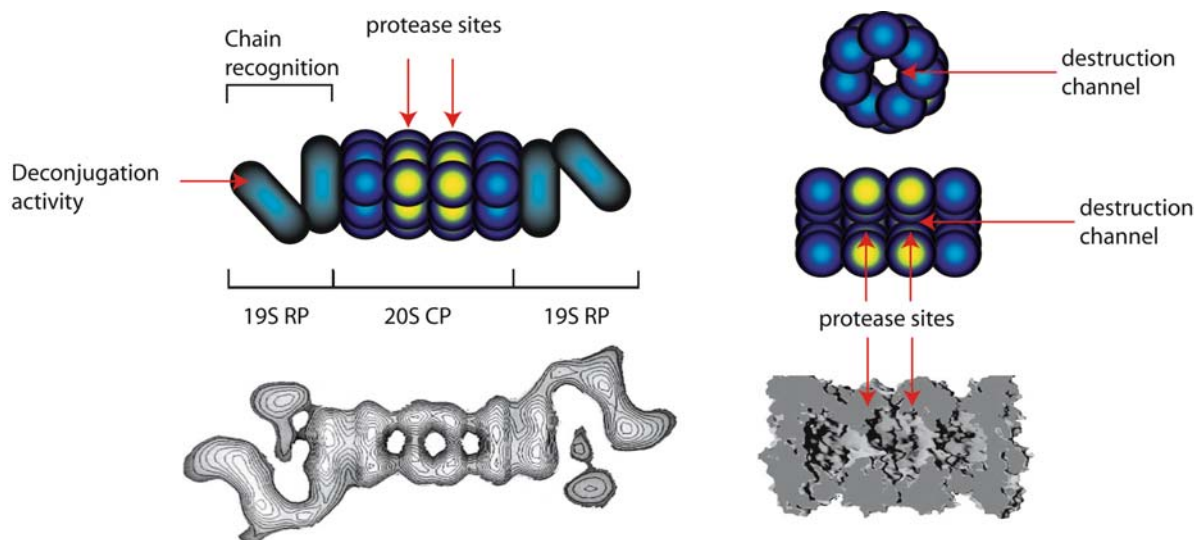
different E3s (Fig. 2). This modular system provides a wide variability, which makes it possible for the ubiquitination system to recognize very divergent substrates. An example of multimeric ring-finger-E3s is the F-box (Pfam: F-box) cullin system as seen in the ► **SCF-E3** (Skp-Cullin-F-box). The substrate recognition is transferred to a special factor the F-box protein, which interacts with the skp component of the ligase. These modules provide additional variability to the system. The family of F-box proteins has 5 members in *S. cerevisiae* and 62 in humans.

The Proteasome

The proteasome is a big barrel-shaped cytoplasmic protease (Fig. 3). The proteasome falls into two distinct subcomplexes the 20S core particle and the 19S regulatory particle. Each core particle has 6 proteolytic sites, which harbor 3 different activities (caspase-like,



Ubiquitination. Figure 2 The ubiquitin system uses a number of specialized enzymes to create different signals. In each step of the ubiquitination cascade an increasing number of enzymes are used. After ubiquitin has been activated by a single E1, in the next step a greater number of E2s can receive the ubiquitin, providing a greater range for substrate recognition. In the following step, the association with an E3, one E2 can associate with different E3 ligases, increasing the number of substrate again. Depending on the combination of E2 and E3, different signals (mono-ubiquitination, differently linked chains) can be produced on the substrate molecule.



Ubiquitination. Figure 3 The proteasome. The proteasome is a multi-subunit protease, which resides in the cytoplasm. The proteolytic sites are covered on the inside of the barrel-shaped central complex. This design prevents unwanted proteolytic activity. Every substrate has to be unfolded and threaded into the destruction channel. The unfolding activity is located in the 19S regulatory particle (RP), which sits on one or both sides of the core. Parts of the figure taken with permission from Finley (2002) *Nat Cell Bio* 4:E121-3.

tryptic and chymotryptic activities). These proteolytic sites are hidden inside a chamber in the barrel of the 20S proteasome (self-compartmentalizing protease). The sides of this chamber are closed by a gating mechanism that regulates the entry into this destruction channel. On one or both sides of its barrel-shaped structure the 20S proteasome is decorated with a regulatory particle. The regulatory particle harbors an unfolding, a deubiquitination activity and the ability of the proteasome to bind K48-linked poly-ubiquitin chains. For destruction of a substrate molecule, the poly-ubiquitinated (K48-linked) protein is recognized by the proteasome. The unfolding activity unfolds the substrate and threads the peptide-chain into the destruction channel. During this process the deubiquitinating activity removes the ubiquitin-molecule for recycling.

Deconjugating Enzymes

Tagging with ubiquitin is reversible. The family of deconjugating enzymes or DUBs catalyzes the deconjugation reaction, which always takes place in competition with the conjugation reaction. These proteases, which are specific to ubiquitin, fall into three categories, the UCH, UBP and Jamm-type proteases. The UCH proteases process the ubiquitin precursors and release the ubiquitin moiety from the fusion protein. In contrast the UBPs and Jamm-type proteases are specific for branched peptides. The UCHs

and the UBPs both belong to the group of cysteine proteases, meaning that the proteolytically active residue is a cysteine. The Jamm-proteases belong to the metallo-proteases and carry a Zn atom in their active center. A UBP and a Jamm-protease have been shown to be associated with the proteasome. Here they release the ubiquitin from the target making it available for recycling.

Molecular Interactions

UBA, UIM, UEV and CUE Domain-Proteins

The way in which ubiquitin is linked to the target molecule carries information about the nature of the substrate molecule. This information has to be decoded by the cell in one way or another. A signal for destruction of the substrate is recognized by the proteasome, which takes care of the substrate removal and the recycling of ubiquitin. Other modifications are interpreted by proteins that recognize ubiquitin by one of the domains, UBA, UIM, UEV, CUE or NZF. For all of these domains a binding to mono-ubiquitin has been shown, but the dissociation constants indicate that the UBA and the UEV domain have a preference for the binding of poly-ubiquitin. In contrast to this, the UIM, CUE and NZF domains have a higher affinity for mono-ubiquitin. But the exact binding preferences of mono-ubiquitin or poly-ubiquitin and

their preferences in terms of linkage remain to be characterized.

Regulatory Mechanisms

The ubiquitin-proteasome system is involved in a huge number of cellular processes. Here are a few examples of regulatory activity by this system.

Quality Control/Stress Response

The removal of misfolded proteins from the cell is an important and essential process inside a cell. Proteins that get misfolded during their life cycle are recognized by the ubiquitin-proteasome system and get degraded. Keeping this in mind, it is not surprising that parts of the ubiquitin system belong to the heat-shock response system. But degradation occurs not only after protein synthesis is completed, but also during synthesis and maturation of a protein. The quality control mechanisms in the endoplasmic reticulum recognize proteins that fail to fold and export them to the cytoplasm where they are degraded by the proteasome (ER associated protein degradation, ERAD).

Regulation of the Cell-Cycle

The ubiquitin proteasome system also plays an essential role in the regulation of the cell-cycle. Two important ubiquitin-ligases SCF and APC guard progression through the G1/S and the G2/M transitions. Here they recognize e.g. the phosphorylated cyclin and degrade it, moving the cell-cycle to the next step. Besides the cyclins, other important cell-cycle regulators have been described as substrates for the ubiquitin-proteasome system (e.g. securin in sister-chromatid separation).

Regulation of the Ubiquitin-Proteasome System by Other Modifications

In order to be recognized by the different parts of the ubiquitin-conjugation machinery substrates have to carry certain modifications. Probably the best-studied modifications are phosphorylations, which lead to degradation of the substrate as shown for cyclins. Another example is histone 2A, which has to be methylated in order to be ubiquitinated. For the DNA binding protein PCNA and others, a new modification has been described, the modification with SUMO-1, which has here an antagonizing effect on ubiquitination. This molecule is one of several small modifiers, which share a significant structural similarity to ubiquitin. In yeast this group has six members Smt3 (or SUMO, see chapter on [▶sumolyation](#)), Rub1, Apg8, Apg12, Urm1 and Hub1. In mammalian cells, additional ones (Fat10, MNSF, ISG15, Ufm1) have been described. All these modifiers do not act as [▶degradation signals](#), but fulfill other functions in the cell.

References

1. Hicke L, Dunn R (2003) Regulation of membrane protein transport by ubiquitin and ubiquitin-binding proteins. *Annu Rev Cell Dev Biol* 19:141–172
2. Pickart CM (2001) Mechanisms underlying ubiquitination. *Annu Rev Biochem* 70:503–533

Unc

Definition

Unc denotes a *C. elegans* mutant that displays uncoordinated movement due to muscular or neural defects.

▶[C. Elegans as a Model Organism for Functional Genomics](#)

Uncaging

Definition

Uncaging designates a technique by which an active biomolecule from a biologically inactive precursor is released using light illumination.

▶[Uncaging and Photoconversion/Activation](#)

Uncaging and Photoconversion/Activation

VOLKER HAGEN

Research Institute for Molecular Pharmacology, Berlin, Germany
hagen@fmp-berlin.de

Definition

The [▶photorelease](#) of biomolecules from photolabile biologically inactive precursors, in biology termed so-called [▶caged compounds](#), is referred to as [▶uncaging](#). When caged, the biomolecule is rendered biologically inactive by covalent attachment of a photochemically removable protecting group ("[▶caging group](#)") to the key pharmacophoric functionality. Flash photolysis using light of a specific color cleaves the modifying

group, i.e. triggers the uncaging, and rapidly activates the molecule. Light-activated caged compounds are powerful tools for controlling the release of biomolecules within organized biological systems with spatial and temporal precision by dictating when and where light exposure occurs. Biological response may then be measured with second up to nanosecond time resolution depending on the photochemistry and the system under investigation. The fact that uncaging does not require any reagent, just light, opens the possibility of dealing with extremely sensitive molecules. In addition, concentration jumps mediated by light-directed activation of caged compounds can be conducted without physically disturbing the preparation.

Description

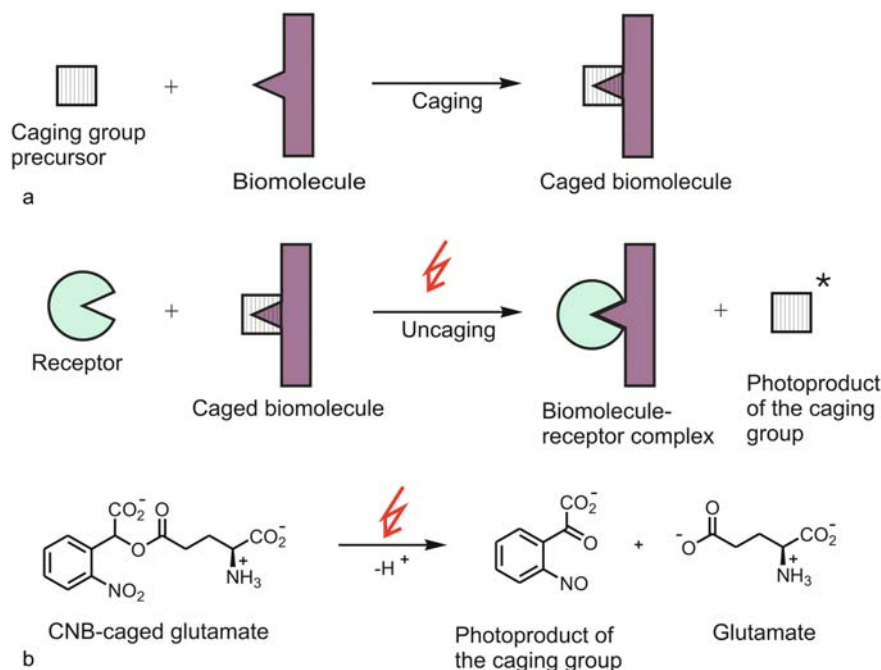
Photoremovable Protecting Groups

Deactivation or protection of functional groups and conversely their specific activation are common practice in chemistry, especially in the design and construction of polyfunctional molecules such as peptides, oligosaccharides, glycolipids and nucleotides (1). Protecting groups protect specific functional groups during synthetic steps and are removed once the protection is no longer required. Among the numerous classes of protecting groups, ►**photoremovable protecting groups** provide a general route to

regenerate the unprotected compound by photochemical means under mild, neutral conditions. A photoremovable protecting group is covalently bound to the functional group and contains a light-sensitive chromophore that can serve as a latent activator of the functional group. On irradiation with light of suitable wavelength, the functional group is converted to an active form and the light-sensitive chromophore is removed. Ideally, the wavelength of the light to be used for deprotection should not affect other parts of the molecule. A short lifetime of the excited state that is responsible for the deprotection reaction and a relatively high stability of the protecting group and the protected bond to a variety of reagents and chemical treatments are desirable. Furthermore, the yield of the deprotection step should be high and the photodecomposition process should be clean. Finally, the photoremovable protecting group should be readily introduced by accessible reagents and the protecting group photoproduct should be separable from the deprotected compound.

Caged Compounds

Caged compounds or photoactivatable bioagents are biomolecules bearing photoremovable protecting groups that inactivate or “cage” the biomolecule and then release or “uncage” it on a flash of light (Fig. 1). They are very useful tools in biophysical and



Uncaging and Photoconversion/Activation. Figure 1 Caging and uncaging of a biomolecule. (a) general principle. (b) uncaging of CNB-caged glutamate.

biochemical studies and can be applied extracellularly or incorporated into cells (2, 3). The major advantage of their use is the high speed of the photochemistry and the fact that they can be applied to organized systems, such as muscle fibers, ion channels and membranes under physiological conditions. Additionally, they are potentially powerful tools in gene expression targeting, in protein folding studies, time-resolved X-ray studies or ►[photolithographic syntheses](#) of ►[biochips](#) (4, 5, 6). A wide variety of caged species has been described, including neurotransmitters, drugs, nucleotides, DNA, RNA, amino acids, peptides, proteins, enzymes, NO and Ca^{2+} . It is important that a caged compound can be introduced to its site of action over a period of minutes (or hours) without evoking biological responses and that the biologically active molecule can then be released from the “cage” by flash photolysis extremely fast (within milliseconds or less). Because photolysis of caged compounds generates effectors *in situ*, much faster and more spatially uniform concentration jumps can be effected than is possible with traditional rapid mixing techniques. There are no diffusion problems and no spatial inhomogeneity problems with addition of substrates. Desensitization of the biological response is minimized. Furthermore, the biologically inert caged compound can be loaded into intracellular compartments by a variety of techniques and concentration jumps can be effected in an environment that is otherwise inaccessible.

Among the different photochemical protecting groups, 1-(2-nitrophenyl)ethyl (NPE), 1-(4,5-dimethoxy-2-nitrophenyl)ethyl (DMNPE), 4,5-dimethoxy-2-nitrobenzyl (DMNB), α -carboxy-2-nitrobenzyl (CNB), 2-(2-nitrophenyl)propoxycarbonyl (NPPOC), para-hydroxyphenacyl, benzoin and coumarinylmethyl groups are the most widely used caging groups. They have been successfully applied to protect carboxyl, hydroxyl, thiol, amino and carbonyl groups as well as phosphate functions in biomolecules. A special strategy for the photorelease of poor leaving groups such as alcohols or amines is to release carbonic or carbamic acid monoesters, which decarboxylate when released.

For biological applications, a caging group must undergo uncaging rapidly, in high yield and on long wavelength UV/Vis or IR activation. High absorption coefficients at wavelength above 320 nm are favorable. Furthermore, the ►[phototriggers](#) should be sufficiently soluble in aqueous solutions, stable toward solvolysis, nontoxic and biologically inert. Sometimes cell membrane permeability is required.

In practice most caged compounds do not fulfill all the requirements for perfect caged compounds and therefore it is desirable to have a range of compounds with different characteristics from which to choose.

Photoactivation

The caged precursor must be photocleavable by light at a wavelength compatible with the biological system being studied. Generally, light at wavelengths below 300 nm can cause photodamage and has limited penetrability through biological material. In the usually used one-photon excitation, a photorelease by near ultraviolet (UV) light in the 350–400 nm range is desirable. Commercially available photoactivation light sources are lasers and flash lamps. The advantages of lasers are their high power coupled with the very short pulse duration and their monochromaticity. A new technique is the two-photon excitation method, which provides better penetration into living tissues and three-dimensional control over the localization of substrate release. In two-photon photolysis, UV excitation is replaced by the simultaneous absorption of two IR photons of equivalent total energy. Two-photon uncaging used a femtosecond-pulsed, mode-locked titanium-sapphire laser. However, most caging groups do not allow two-photon uncaging. A useful caging group for two-photon excitation is the (6-bromo-7-hydroxycoumarin-4-yl)methyl (BHCM) moiety.

Photoactivation illuminators can be integrated into microscope systems and are able to focus UV light sufficient for photorelease within a small region of the specimen, even within a single cell. This capability allows the properties of subcellular microenvironments to be examined and also makes it possible to study spatial and temporal aspects of dynamic processes.

A caged compound should show high absorptivity at the wavelength used for photolysis. The quantum yield of the photoreaction, defined as the ratio of the number of bioactive molecules released to the number of protons absorbed by the caged precursor, should be between 0.1 and 1. Quantum yields may depend strongly on solvent and pH.

For time-resolved studies the release rate of the biomolecule must exceed that of the response investigated. Therefore, knowledge of the actual release rates under physiological conditions is important. Release rates are determined by the caging group, the leaving group and the mechanism of the photorelease. Generally, the generation of the biomolecules occurs in the submillisecond time range.

Selected Practical Applications

Caging and uncaging of biomolecules combined with rapid monitoring of the ensuing reaction using photomultiplier or image-based techniques are widely used for studies of the mechanisms and the kinetics of cellular processes. Caged ATP was applied to investigate the energetics of ►[muscle contraction](#) or to trigger Ca^{2+} signals that are communicated between different cell types. The caged second messengers cAMP and

cGMP were used to analyze cyclic nucleotide-dependent signal transduction pathways and their possible functional consequences. Photolabile Ca^{2+} release chelators were applied for the generation of Ca^{2+} concentration jumps in a myriad of vital physiological processes including muscle contraction, secretion, mitosis, channel gating and chemotaxis. Caged neurotransmitters as well as a variety of caged neurotransmitter agonists allowed the investigation of the detailed kinetics of receptor-mediated events and were useful for mapping receptor distributions in individual [▶neurons](#). A variety of drugs were exploited to probe drug-receptor interactions. Photoactivatable peptides bearing caging groups at essential amino acids including lysine, tyrosine, cysteine, serine, threonine, aspartic acid, glutamic acid and arginine were found to be important reagents for investigating cell signaling and controlling protein-protein interactions as well as enzyme activity. Application of caging chemistry to DNA and mRNA offered the possibility of targeting expression of genetic material with light. Caging of transcriptional activators was also a tool for spatiotemporal control of gene expression.

Methods to cage protein activity were also developed. Caged proteins were used to investigate protein structure-function relationships and protein function in living cells.

A general approach to trigger a single activation cycle of an enzyme involves chemical modification of the protein by a photosensitive protecting group that yields a deactivated enzyme. Photoremoval of the protecting group regenerates the native protein and reactivates the enzyme. This approach is exemplified in studies of fast conformational changes in enzymes and other proteins. Caged ligands can be also used to photochemically trigger an enzymatic reaction within an enzyme crystal and are thus valuable tools for studying protein dynamics at the atomic level by time-resolved X-ray. Furthermore, photochemical triggering methods are suitable for kinetic studies of protein folding.

Finally, caged compounds have also been used in the synthesis of microarrays of peptides and oligonucleotides on surfaces for use in genomics and proteomics research, utilizing photolithographic processes.

References

1. Pillai VNR (1987) Photolytic deprotection and activation of functional groups. In: Pawda A (ed) *Organic Photochemistry* vol 9. Marcel Dekker, New York, pp 225–323
2. Kao JPY, Adams SR (1993) Photosensitive caged compounds. In: Herman B, Lemasters JJ (eds) *Optical Microscopy. Emerging Methods and Applications*. Academic Press, San Diego, pp 27–85
3. Pelliccioli AP, Wirz J (2002) Photoremovable protection groups: reaction mechanisms and applications. *Photochem Photobiol Sci* 1:441–458
4. Ando H, Furuta T, Tsien RY et al (2001) Photo-mediated gene activation using caged RNA/DNA in zebrafish embryos. *Nat Genet* 28:317–324
5. Shigeri Y, Tatsu Y, Yumoto N (2001) Synthesis and application of caged peptides and proteins. *Pharmacol Ther* 91:85–92
6. Emili AQ, Cagney G (2000) Large-scale functional analysis using peptide or protein arrays. *Nat Biotechnol* 18:393–396

Unconventional Myosins

Definition

Actin-based motor proteins of the myosin superfamily that convert the chemical energy released upon hydrolysis of ATP into mechanical force. Conversely to conventional myosin II, unconventional myosins do not form filaments.

[▶Microvilli](#)

Unigene Library

Definition

Unigene library is a cDNA library in which each transcript of a given genome is represented only once, irrespective of its abundance.

[▶Medaka as a Model Organism for Functional Genomics](#)

Unit Cell

Definition

The basic element of a crystal, repeated infinitely (for an ideal crystal) in three dimensions parallel to the cell axes. The contents of the unit cell comprise the asymmetric unit and all its symmetry images.

[▶X-Ray Crystallography—Basic Principles](#)

Uniparental Disomy

Definition

Uniparental disomy refers to the situation in which both members of a chromosome pair or segments of a

- [chromosome](#) pair are inherited from one parent and neither is inherited from the other parent. There are two types of UPD; isodisomy is the presence of two identical homologues, and heterodisomy signifies two different homologues from one parent. UPD can result in an abnormal

- [phenotype](#) in some cases.
- [Prader Willi and Angelman Syndromes](#)
- [Microdeletion Syndromes](#)

Univalent

Definition

Univalent defines a chromosome that either fails to synapse with its partner or does not undergo crossing-over (chiasma formation). As a result, the chromosome segregates independently of its partner, thereby increasing the risk of nondisjunction.

- [Meiosis and Meiotic Recombination](#)

Unsupervised Analysis

Definition

Unsupervised analysis is an exploratory data analysis, such as clustering or principal components analysis, which typically has the goal of reducing the dimensionality in the data.

- [Microarray Data Analysis](#)

Unsupervised Learning

Definition

Unsupervised learning describes the finding of a hidden structure in multivariate data by statistical and machine learning methods.

- [Computational Diagnostics](#)

Untranslated

Definition

Untranslated designates a part of the protein coding region of a gene (exon) that is not used as a template for protein synthesis.

- [Familial Hypercholesterolemia](#)

5' Untranslated Region

Definition

5' Untranslated Regions refer to elements of gene or mRNA sequences that precede the AUG translational start site and that do not encode peptide information.

- [Morpholinos](#)

UPD

- [Uniparental Disomy](#)

Upper Motor Neurons

Definition

Upper motor neurons is a clinical term referring to motor neurons whose cell bodies are in the motor cerebral cortex, and that make axonal projections to lower motor neurons in the spinal cord. The corticospinal tracts are made up of the axons of upper motor neurons.

- [Hereditary Spastic Paraplegias](#)

Upper to Lower Segment Ratio

Definition

Upper to lower segment ratio is the value obtained by dividing the upper body segment (total height minus the

lower segment) by the lower segment. Lower body segment is the measured distance from pubic bone to the floor. The norms for this value vary based on the patient's ethnicity.

► [Marfan Syndrome](#)

Upstream

Definition

Upstream on the 5' side of any given site in DNA or RNA and on the N-terminal side within a polypeptide (► [downstream](#)). With respect to its transcriptional activity, upstream regions of a gene are on the 3' side of any given region in the coding (sense) strand.

► [Two-Hybrid System](#)

Uronic Acid

Definition

Uronic acid belongs to the class of acidic compounds of the general formula $\text{HOOC}(\text{CHOH})_n\text{CHO}$, typically

hexose, which contain both carboxylic and aldehyde groups, such as glucuronic acid and iduronic acid.

► [Glycosylation of Proteins](#)

UTR

Definition

UTR stands for UnTranslated Region of an mRNA. An mRNA can be divided into the ► [5'untranslated region](#), the Open reading frame (ORF) and the 3' UTR. The UTRs do not encode protein information, but have regulatory functions. Sequence motifs define e.g. the localization of the mRNA in the cell, the stability of the mRNA, and the amount of protein that is supposed to be generated from the mRNA at a given time.

► [Cap-Independent Translational Control](#)

► [Full Length cDNA Sequencing](#)

UV Signature Mutations

Definition

UV irradiation, a known mutagen and carcinogen, induces DNA alterations mainly at neighboring pyrimidines by producing dimers. This induces CC to TT or C to T base substitutions at dipyrimidine sites.

► [Skin and Hair](#)

V(D)J Recombination

Definition

V(D)J recombination is the abbreviation of V: variable, D: diversity, and J: joining. In germ-line DNA immunoglobulin, constant and variable regions of heavy and light chains are encoded by different gene segments. During B- and T cell development, the huge repertoire of immunoglobulins and T cell receptors is accomplished, in part, by the combinatorial association of three gene segments (V, D, and J) for one chain, and two gene segments (V and J) for the second chain of these heterodimeric antigen receptors.

- [Double-Strand Break Repair](#)
- [DNA Recombination](#)

V2R

Definition

V2 vasopressin receptor (V2R) mediates the antidiuretic effect of vasopressin by stimulating adenylyl cyclase activity in the collecting tubules of the kidney.

- [Diabetes Insipidus, a Water Homeostasis Disease](#)

Vaccine

Definition

Vaccination induces pathogen-specific protective immune memory. The injected vaccines can be e.g. recombinant antigens of the pathogen or antigen-expressing vectors.

- [DNA-based Vaccination](#)

Validation of Therapeutic Targets

Definition

Validation of therapeutic targets is a matter of industrial strategy. The definition of the concept of target validation is rather loose and varies significantly from one company to another. To give a consensual operational definition, validating a therapeutic target is collecting all the experimental data needed to launch a drug discovery project. To give a precise definition, more stringent than other definitions commonly found, validation of a therapeutic target is the demonstration that the use of a specific ligand against the target induces a desired phenotype. In relevant cellular and animal models of pathological situations.

- [Peptide Aptamers](#)

Valve

Definition

Valve is a component that helps in manipulating flow direction (and velocity). For example, check-valve allows fluid flow only in one direction.

- [Proteomics in Microfluidic Systems](#)

Van der Waals Interaction

Definition

Van der Waals interaction stands for weak attractive interaction of fairly short range that occurs between all atoms, because they behave like oscillating dipoles. It decays with $1/r^6$, where r is the distance between the interacting atoms. The van der Waals term in force fields normally includes a strong repulsive potential at very short interatomic distances.

- Molecular Dynamics Simulation in Drug Design
- Protein/DNA Interaction

van't Hoff Equation

Definition

van't Hoff equation designates an equation relating the temperature dependence of the equilibrium constant to the enthalpy change of a process.

- Thermodynamic Properties of DNA

Variegate Porphyria

Definition

Variegate porphyria describes an acute hepatic porphyria with protoporphyrinogen oxidase deficiency that is inherited in an autosomal recessive trait.

- Acute Intermittent Porphyria

Vascular Morphogenesis

Definition

Vascular morphogenesis describes the formation and maturation of functional blood vessels by the interaction of vascular endothelial cells with pericytes and smooth muscle cells.

- Angiogenesis

Vasculitis

Definition

Vasculitis is an inflammation of the lining of the blood vessels.

- Morbus Wegener

Vasculogenesis

Definition

Vasculogenesis is the *in situ* differentiation of endothelial precursor cells during development.

- Angiogenesis

Vasoactive Peptides

Definition

Vasoactive peptides are peptides that cause constriction or dilation of blood vessels.

- Cardiac Signaling: Cellular, Molecular and Clinical Aspects

Vasopressin

Definition

Vasopressin is a hormone that is produced in the hypothalamus and stored in the posterior pituitary gland. It increases water permeability in the distal nephrons of the kidney.

- Hypothalamic and Pituitary Diseases Genetics

vCJD

Definition

vCJD stands for the variant of Creutzfeldt-Jakob disease. This variant is most likely caused by the bovine spongiform encephalitis (BSE) agent, and probably transmitted by the consumption of contaminated food.

- Prion Diseases

Vector (Particle)

Definition

Vector (particle) refers to a circular self-replicating DNA molecule that originates from a plasmid, phage or virus used for cloning of selected foreign DNA. Vectors

are employed for the introduction of foreign DNA into host cells. In the context of clinical gene transfer, it is assumed that the vector particle contains all components necessary for delivery of the specific nucleic acid associated with the particle to a pre-conceived target, cell tissue or organ.

► [Clinical Gene Transfer](#)

► [Recombinant Protein Production in Mammalian Cell Culture](#)

► [YAC and PAC Maps](#)

VEGF

Definition

VEGF stands for vascular endothelial growth factor. It defines a growth factor that stimulates proliferation of endothelial cells and participates in the formation of new blood vessels.

► [Growth Factors](#)

► [Angiogenesis](#)

Ventricular Zone

Definition

The ventricular zone is the germinal zone located at the walls surrounding the ventricles where the majority of neurones of higher brain areas are generated. From there, neurones migrate to their final target regions using intrinsic or extrinsic cues and/or radial glia.

► [Neural Development](#)

► [Neurons](#)

Very Low Density Lipoprotein

Definition

Very low density lipoprotein VLDL is a lipoprotein substance (complex of fat and protein) which serves as a carrier for cholesterol and triglycerides in the

bloodstream. Increased levels of VLDL are associated with atherosclerosis and coronary artery disease.

► [Familial combined hyperlipidemia](#)

► [hypertriglyceridemia](#)

Very-Long-Chain Fatty Acids

Definition

Very-long-chain fatty acids (VLCFA) are fatty acids with a chain length of 22 carbon atoms or more.

► [Peroxisomal Disorders](#)

Vesicular Traffic

RAINER DUDEN

School of Biological Sciences, Royal Holloway
University of London, Egham, Surrey, UK
rainer.duden@rhul.ac.uk

Definition

Eukaryotic cells are elaborately subdivided into functionally distinct, membrane-bounded compartments, which together form a sophisticated endomembrane system. Each compartment or organelle contains its characteristic set of enzymes and other resident proteins with specialized functions and complex distribution systems connect compartments with each other. Within the endomembrane system, the ► [secretory pathway](#) delivers newly synthesized proteins, carbohydrates and lipids to the outside of the cell or to other, intracellular final destinations. The ► [endocytic pathway](#) takes up macromolecules into the cell and ensures their delivery to endosomes and lysosomes. Despite a large flux of material through these compartments, each compartment maintains its unique biochemical make-up of resident proteins at steady state. This is made possible by highly regulated mechanisms for the transfer of material between compartments, which ensure the proper specificity and directionality of the secretory and endocytotic processes. In the past two decades a large number of evolutionarily conserved components of these vesicular transport pathways have been identified using biochemical and genetic approaches, in mammalian cells and in the yeast model *Saccharomyces cerevisiae* (budding yeast). The majority of these proteins participate in transport in the context of

macromolecular protein complexes, sometimes also called “protein machines”.

Characteristics

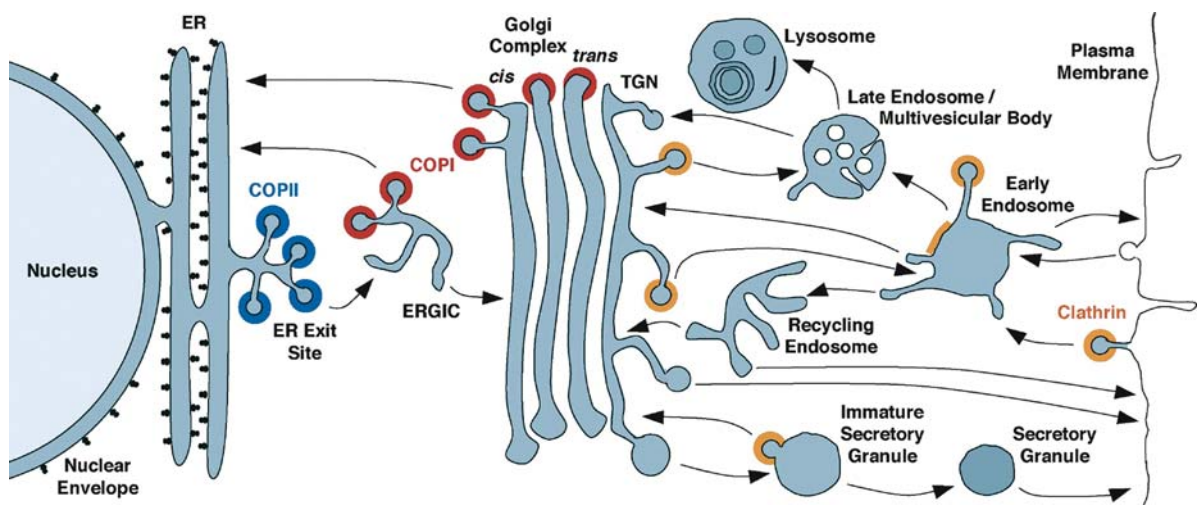
Traffic in between membrane-bounded compartments is mediated by vesicular and tubular carriers, which are shaped by cytoplasmic ►coat proteins (Fig. 1) (1). Coat proteins form an electron-dense outer layer of vesicles that can be visualized under the electron microscope. Cytoplasmic coat proteins perform the dual, essential tasks of providing the mechanical force for deforming the lipid bilayer of appropriate “donor” membranes into buds and vesicles and selecting appropriate cargo proteins during budding. Thus, protein sorting and vesicle budding are functionally integrated. Different intracellular transport steps utilize distinct sets of coats and several sets of coat proteins have been well characterized, including COP I, COP II, ►clathrin, the ►adaptor complexes AP-1, AP-2, AP-3 and AP-4 and ►GGAs. Binding of coat proteins to membranes is highly compartment specific. All vesicular protein transport involves the selective recruitment of cargo into the vesicles, controlled formation of the vesicle from membranes of a specific compartment, partial uncoating and transport to the target membrane/organelle and binding to and fusion with the target membrane (1).

Upon arrival of transport vesicles at their appropriate target membranes, and preceding membrane fusion, they are “tethered” by specific tethering complexes, some of which have been characterized in molecular detail (2). Lastly, the specificity of intracellular membrane fusion is governed by ►SNARE proteins, which have compartment specific patterns of localization. Pairing of cognate v- and t-SNAREs between two opposing lipid bilayers drives spontaneous membrane fusion and confers specificity to intracellular membrane trafficking events (3). The incorporation of v-SNAREs into the budding vesicle through interactions with coat proteins programs the vesicle for fusion at the appropriate target membrane. Fusion of the vesicle with the target membrane leads to delivery of cargo molecules and thus completes a round of transport (Fig. 2) (1).

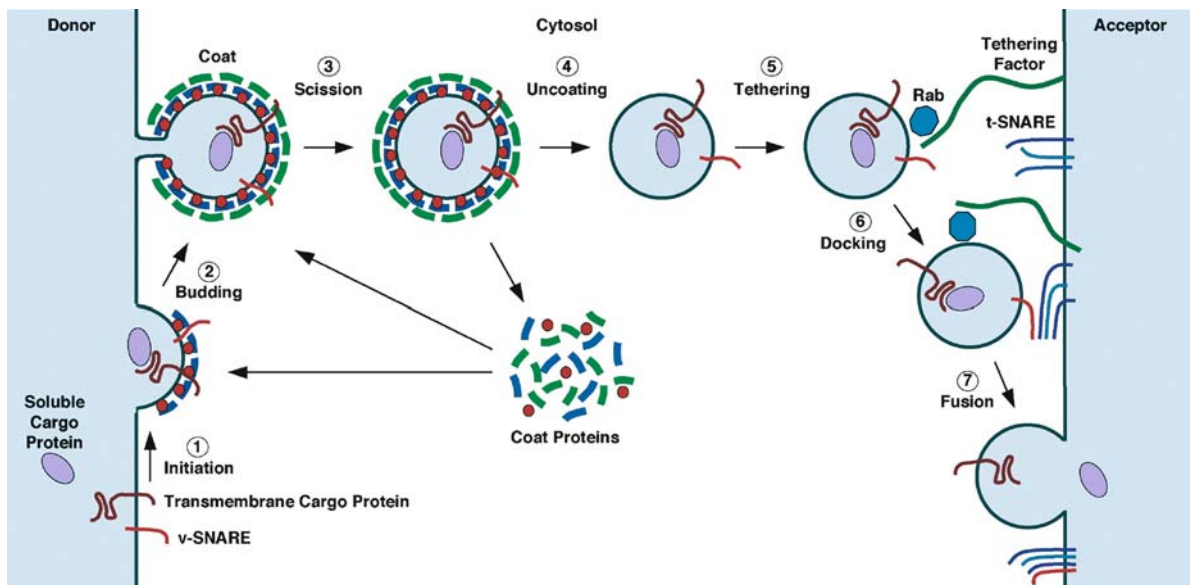
Several distinct classes of coated vesicles and the coat proteins forming them, have been characterized in molecular detail (4), and are described below. All these coat proteins are organized as multi-subunit complexes on “donor” membranes and their individual subunits and interactions are conserved in evolution.

COP II-Coated Vesicles

COP II-coated vesicles (size 60–70 nm) are required for selective export of newly synthesized proteins from the



Vesicular Traffic. 1 Intracellular Transport Pathways. The scheme depicts the compartments of the secretory, lysosomal/vacuolar and endocytic pathways. Transport steps are indicated by arrows. Colours indicate the known or presumed locations of the coat proteins COPII (blue), COPI (red), and clathrin (orange). Clathrin coats are heterogeneous and contain different adaptor and accessory proteins at different membranes. Only the function of COPII in ER export and of plasma membrane-associated clathrin in endocytosis are known with certainty. Less well understood are the exact functions of COPI at the VTCs and Golgi complex and of clathrin at the TGN, early endosomes and immature secretory granules. There are several vesicular transport pathways within the cell for which no coat proteins have been described so far. Therefore the identification of further, novel vesicle coats in the near future is likely. Reproduced with copyright permission from Bonifacino and Glick (2004).



Vesicular Traffic. 2 Steps of Vesicle Budding and Fusion. (1) Initiation of coat assembly. The membrane-proximal coat components (blue) are recruited to the donor compartment by binding to a membrane-associated GTPase (red) and/or to a specific phosphoinositide. Transmembrane cargo proteins and SNAREs begin to gather at the assembling coat. (2) Budding. The membrane-distal coat components (green) are added and polymerize into a mesh-like structure. Cargo becomes concentrated and membrane curvature increases. (3) Scission. The neck between the vesicle and the donor compartment is severed either by direct action of the coat or by accessory proteins. (4) Uncoating. The vesicle loses its coat due to various events including inactivation of the small GTPase, phosphoinositide hydrolysis and the action of uncoating enzymes. Cytosolic coat proteins are then recycled for additional rounds of vesicle budding. (5) Tethering. The “naked” vesicle moves to the acceptor compartment, possibly guided by the cytoskeleton, and becomes tethered to the acceptor compartment by the combination of a GTP bound Rab and a tethering factor. (6) Docking. The v- and t-SNAREs assemble into a four-helix bundle. (7) This “trans-SNARE complex” promotes fusion of the vesicle and acceptor lipid bilayers. Cargo is transferred to the acceptor compartment, and the SNAREs are recycled. Reproduced with copyright permission from Bonifacino and Glick (2004).

► **endoplasmic reticulum (ER)**. The minimal machinery to bud ► **COP II vesicles** from ER membranes or liposomes comprises the small GTPase ► **Sar1p** and two hetero-dimeric protein complexes, Sec23/24p and Sec13/31p. All components involved in COP II vesicle formation are essential for cell viability in yeast. COP II vesicles produced *in vitro* from membranes in the presence of the minimal machinery are capable of packaging a large set of cargo proteins, including the SNARE proteins necessary for their targeting and fusion with membranes of the ► **Golgi complex** (4, 5). The dynamics of COP II-coated ER exit sites have been studied by time-lapse fluorescence microscopy of living cells, using functional chimeras of COP II subunits with green fluorescent protein. The majority of COP II labelling is in stable ‘spots’ tightly associated with ER membranes. Secretory cargo segregates from these sites and is transported to the Golgi complex without COP II proteins being associated with the moving structures. The current model is that COP II vesicles rapidly uncoat after budding, allowing them to

fuse with each other and generate so-called VTC membrane structures (VTC = vesicular-tubular clusters, also named ERGIC (ER/Golgi intermediate compartment, or just ‘intermediate compartment’) (Fig. 1). COP II coat proteins do cycle rapidly on and off the ER membrane, but they get recruited back to the same domains, thus defining kinetically stable ER exit sites (5).

COP I-Coated Vesicles

COP I-coated vesicles (size 70–80 nm) mediate an essential and conserved retrograde transport pathway that selectively recycles proteins from the *cis*-Golgi complex back to the ER (Fig. 1). Importantly, ► **COP I vesicles** retrieve the v-SNAREs involved in COP II vesicle consumption. Furthermore, COP I vesicles are instrumental in the retrieval of misfolded proteins that escape the ER quality control mechanism and of ER resident proteins that leak out of the ER. Cargo packaging into COP I vesicles is based on specific trafficking motifs, e.g. the di-lysine motif (KKXX) or

the KDEL motif, which is present on the ER luminal chaperones BiP and protein disulphide isomerase. Escaped, misfolded proteins can be retrieved to the ER in complex with BiP. This aspect of COP I function is therefore an extended part of the ER quality control system. Additionally, COP I coat proteins have important and complex functions in intra-Golgi trafficking and in maintaining the normal structure of the mammalian interphase Golgi complex. The minimal machinery to form COP I-coated vesicles from Golgi membranes or liposomes with a defined composition comprises the small GTPase **▶ARF** in its GTP-bound, active form and coatomer, a stable heptameric protein complex containing α -, β -, β' -, γ -, δ -, ϵ -, and ζ -COPs (4, 5). Conserved GTPase-activating (ARFGAP) proteins are essential to accomplish GTP hydrolysis on ARF and thus help achieve vesicle uncoating, but these proteins are also important for cargo selection in ways so far incompletely understood (5).

Clathrin Coated Vesicles, Adaptor Complexes, GGAs

The clathrin coat was the first vesicle coat to be discovered and is still the best understood (6). The coat protein clathrin provides the mechanical scaffold to deform the **▶plasma membrane** or membranes on the *trans*- or exit side of the Golgi complex into clathrin coated vesicles. Endocytosis mediated by clathrin coated vesicles plays a central role in regulating the level of integral membrane proteins in the plasma membrane, including signalling and nutrient receptors, cell surface markers and adhesion proteins, ion channels and pumps. In many cells, clathrin dependent receptor mediated endocytosis is the major pathway for internalization of materials from the outside world, but other pathways for internalization of materials exist (reviewed in 8).

Cargo inclusion into clathrin coated vesicles is mediated by adaptor proteins that can associate both with trafficking motifs in the cytoplasmic domains of cargo proteins and with clathrin (6, 7). Classically, four related hetero-tetrameric AP complexes have been described, AP-1, AP-2, AP-3 and AP-4. Each of these complexes is composed of two large subunits (one each of $\gamma/\alpha/\delta/\epsilon$ and $\beta 1$ -4, respectively; 90–130 kD), one medium adaptin ($\mu 1$ -4; ~50 kD) and one small adaptin ($\sigma 1$ -4; 20 kD). Only AP-1 and AP-2 associate with clathrin. AP-1 is present on vesicles budding from the *trans*-side of the Golgi complex to effect transport of specific cargo to endosomal and lysosomal membranes and AP-2 associates with endocytotic vesicles budding from the plasma membrane. Gene knockouts of subunits of AP-1 and AP-2 in mice are embryonic lethal. GGAs comprise another family of proteins that associate with clathrin and ARF and are involved in trafficking of specific cargo in between the Golgi complex and endosomes. AP-3 mediates selective

transport to lysosomes and lysosome-related organelles; its associated coat protein is so far unknown. Mutations in AP-3 subunits result in viable mice with coat colour defects and bleeding disorders and in *Drosophila* result in eye pigmentation defects.

SNAREs

SNARE proteins ensure the specificity of vesicle docking and fusion with the appropriate target membrane. SNAREs constitute a superfamily of small, coiled-coil transmembrane proteins present on vesicles and target organelles (named v- and t-SNAREs) that act as molecular “address tags” to ensure this specificity. During membrane fusion, four α -helices from SNAREs on the vesicle and target membranes come together to form a stable, parallel four-helix bundle. A conserved domain of ~60 amino acids (the SNARE motif) mediates this association, which helps to overcome the energetic barrier to lipid bilayer fusion. After membrane fusion SNARE complexes need to be disassembled in an ATP-dependent reaction mediated by **▶NSF** (3). Also, since the v-SNAREs are now located in the wrong compartment they need to be recycled to the “donor” compartment to be made available for the next transport round. Therefore, all sustained trafficking in any of the vesicular transport steps in the secretory or endocytic pathways in one direction is coupled to recycling pathways that mediate retrieval of SNAREs (and other escaped proteins) in the reverse direction.

Molecular Interactions

Different coat proteins get recruited to the cytosolic face of specific “donor” membranes through molecular interactions with specific lipids or organelle resident proteins, or with sorting motifs present in the cytoplasmic tails of cargo proteins, or through a combination of these binding determinants. Below the molecular constituents of the best well-known vesicle coats and their interactions are discussed (2, 3, 4, 7).

COP II-Coated Vesicles

COP II vesicle budding is initiated when an ER-localized transmembrane protein, Sec12p, mediates the GTP for GDP exchange on Sar1p. GTP-loaded Sar1p tightly binds to ER membranes and recruits the Sec23/24 complex and subsequently the Sec13/31p complex, which induces coat polymerization and membrane deformation into buds and vesicles. Sar1p and Sec23/24p together can bind to cytoplasmic domains of cargo proteins, providing a ‘primer’ for COP II vesicle formation. These direct interactions of COP II proteins with cargo during budding are thought to increase efficacy and fidelity of cargo sorting into COP II vesicles. By mechanisms so far unknown, misfolded

proteins present in the ER and ER-resident proteins are excluded from the vesicles. After budding, Sec23p acts as a Sar1p-specific GTPase activating protein (GAP), helping Sar1p to hydrolyse its GTP, which in turn triggers vesicle uncoating. The Sec13/31p complex, once recruited to the membrane and having induced membrane curvature, leads to increased Sar1p-GTPase activity mediated through the action of Sec23p. This built-in GTPase-activating protein makes the COP II coat intrinsically unstable. Non-hydrolysable analogues of GTP have therefore been used to isolate and characterize COP II-coated vesicles (4, 5).

X-ray crystallography, deep-etch rotary shadowing and electron microscopy have recently been used to visualize the structure of COP II subunits and the surface structure of COP II-coated vesicles. These data demonstrated that the Sec23/24p complex resembles a bow tie and the Sec13/31p complex has a flexible structure of 24–30 nm in length comprising a terminal bilobed globular structure bordering a central rod. A surface view of vesicles by electron microscopy revealed a coat built with polygonal units (5).

The Sec24p subunit recruits cargo into the vesicles, since the Sec23/24p heterodimer, and in some cases Sec24p alone, can bind to cytoplasmic domains of a variety of COP II-vesicle cargo molecules. Yeast cells possess three Sec24p homologs and higher eukaryotes express at least four isoforms. This variation in the Sec24p subunit provides COP II vesicles with a broader specificity for cargo selection, as well as greater flexibility in vesicle size. The specificity of the cargo coat protein interaction is mediated by sorting signals present in the cytoplasmic domains of cargo proteins to which the Sec23/24 proteins bind. Two such ER export signals have been well characterized so far, namely a di-phenylalanine motif and a di-acidic sorting signal Asp-X-Glu (D-X-E, where X is any amino acid). Soluble cargo may get incorporated into COP II vesicles either by default ('bulk flow') or through selective interactions of the coat with export receptors that bind this cargo.

Clathrin and Adaptor Complexes

The basic assembly unit of the clathrin coat is the clathrin triskelion, composed of three clathrin heavy chains and three clathrin light chains. Clathrin assembles into hexagonal lattices on membranes of the plasma membrane and on membranes on the *trans*- or exit side of the Golgi complex. During the process of vesicle formation, the clathrin arrangement changes into a basket made out of pentagons (6). Disassembly of the clathrin coat after budding involves the HSC70 (heat-shock-cognate 70) ATPase and the co-chaperone auxilin, which *in vitro* bind to the clathrin light chains. While clathrin provides the mechanical scaffold to deform the plasma membrane or *trans*-Golgi

membranes into coated buds and vesicles, adaptor proteins mediate protein cargo selection. Different adaptors, e.g. AP-1, AP-2, GGAs and PACS-1, have binding specificity for distinct types of cargo proteins and sorting motifs and also associate with clathrin. The complete structures of AP-2 and clathrin have been solved by X-ray crystallography. X-ray crystallography has also allowed understanding at atomic resolution of how adaptors and GGAs bind their different cargo proteins through specific trafficking motifs. There is an ever-growing number of other cargo-specific adaptors (7). In general, adaptors team up with cytoplasmic coat proteins to effect the enrichment of cargo into transport vesicles. Whereas AP-1 and AP-2 cooperate with clathrin, AP-3 and AP-4 do not seem to associate with clathrin and their binding partners are unknown.

Regulatory Mechanisms

Our knowledge of the regulation of vesicular trafficking events in time and space is sketchy at best. It is clear, however, that these processes are tightly regulated, since e.g. the size and molecular composition of intracellular membrane bounded compartments stay constant at steady state despite an enormous flux of proteins and lipids through them. Strikingly, a fibroblast cell under typical conditions will internalize an amount of membrane equivalent to the whole surface area of the cell in just one hour. For the regulation of vesicle coat formation, small GTP-binding proteins of the ras-superfamily and their regulators, i.e. guanine nucleotide exchange factors or GEFs and GTPase-activating proteins or GAPs, hold the key to understanding how cargo selection is coupled to vesicle formation. These proteins function as "molecular switches" whose conformational changes are regulated by the cycle of GDP/GTP exchange and GTP hydrolysis. Similarly, the proper and efficient targeting of vesicles to their destinations, i.e. their "tethering", requires proteins from the large and diverse Rab/Ypt family of small GTPases, which provide another example of how a GTP/GDP-nucleotide dependent conformational switch mechanism is used to regulate cellular events (2). In the case of clathrin/AP-2-coated vesicles in endocytosis, phosphorylation on a crucial residue on AP-2 (threonine 156 on the μ 2 subunit) and binding of specific phosphoinositides crucially regulate binding of cargo and vesicle formation. A large number of regulatory proteins are known to bind the "ear" domains of the large subunits of adaptor complexes, and are likely to fine-tune aspects of cargo selection, membrane binding and vesicle formation. In the case of COP II and COP I vesicles only a few such studies have been published. Active research in many laboratories is currently addressing these questions.

References

1. Bonifacino JS, Glick BS (2004) The mechanisms of vesicle budding and fusion. *Cell* 116:153–166
2. Whyte JRC, Munro S (2002) Vesicle tethering complexes in membrane traffic. *J Cell Sci* 115:2627–2637
3. Ungar D, Hughson FM (2003) SNARE protein structure and function. *Annu Rev Cell Dev Biol* 19:493–517
4. Kirchhausen T (2000) Three ways to make a vesicle. *Nat Rev Mol Cell Biol* 1:187–198
5. Duden R (2003) ER-to-Golgi Transport: COP I & COP II function. *Mol Membr Biol* 20:197–207
6. Brodsky FM, Chen C-Y, Knuehl C et al (2001) Biological Basket weaving: formation and function of clathrin – coated vesicles. *Annu Rev Cell Dev Biol* 17:517–568
7. Robinson MS (2004) Adaptable adaptors for coated vesicles. *Trends Cell Biol* 14:167–174
8. Conner SD, Schmid SL (2003) Regulated portals of entry into the cell. *Nature* 422:37–44

VH

Definition

VH denotes the variable domain of the heavy-chain of conventional antibodies of all vertebrates.

- [Camel as a Model for Functional Genomics](#)
- [Immunoglobulin](#)

VHH

Definition

VHH stands for the variable domain of the heavy-chain antibodies of camelids. This domain is responsible for antigen binding.

- [Camel as a Model for Functional Genomics](#)

VHL-(von Hippel-Lindau) Disease

Definition

VHL disease is an inherited human angiomatous cancer syndrome with an autosomal-dominantly transmitted pre-disposition to develop renal cell carcinomas, haemangioblastomas and pheochromocytomas associated with constitutive upregulation of HIF- α due to a loss of function in pVHL.

- [Hypoxia Inducible Factors](#)

VH-VL Pair

Definition

VH-VL pair designates the paired VH and VL domains of conventional antibodies of all vertebrates. This pair forms the antigen binding fragment.

- [Camel as a Model for Functional Genomics](#)
- [Antibody](#)

Villin

Definition

Villin is a cytoskeletal protein of microvilli of epithelial cell brush border. It belongs to the gelsolin/villin group of actin binding proteins, which are characterized by the possession of three or more gelsolin-like domains.

- [Microvilli](#)

Vinblastine/Vincristine

Definition

Vinblastine and vincristine are alkaloids derived from the periwinkle plant, *Vinca rosea*. These, and some semisynthetic derivatives, can block mitosis. They bind with 1:1 stoichiometry to β -tubulin in disassembled tubulin heterodimers that can no longer polymerize to microtubules. During mitosis the spindle cannot be formed. Chromosomes are no longer separated but dispersed or clumped in the cytoplasm. The cells undergo apoptosis. Vinca alkaloids are clinically used for cancer treatment.

- [Cytoskeleton](#)
- [Multi-drug Resistance](#)

Vinca Alkaloids

- [Vinblastine/Vincristine](#)

Vincristine

► Vinblastine/Vincristine

selection. Fitness is a Darwinian concept that has been applied to explain the effects of genetic variation and environmental selection with regard to the evolution of rapidly replicating viral populations.

► Reverse Transcriptase

Vinculin

Definition

Vinculin is a 117 kDa protein localized at the cytoplasmic side of focal contacts or adhesion plaques. The globular head region of vinculin contains the binding sites for talin and alpha-actinin, whereas its rod-like tail domain contains the binding sites for F-actin and paxillin. Vinculin is involved in cell adhesion through attachment of the actin-based micro-filaments to the plasma membrane.

► Cytoskeleton

► Signal Transduction, Integrin-Mediated Pathway

► Focal Complexes/Focal Contacts

Viral Fitness

Definition

Viral Fitness is a term used to describe the relative ability of a viral variant to adapt and replicate competitively in a population of viruses undergoing

Viral Oncogenesis

RALPH GRASSMANN, BERNHARD FLECKENSTEIN
Institute of Clinical and Molecular Virology, Erlangen, Germany
grassmann@viro.med.uni-erlangen.de

Definition

Viral oncogenesis can be defined as the feature of ► tumor viruses that induces benign or malignant proliferation of infected cells.

Characteristics

Tumor viruses constitute a heterogeneous group of viruses, which play a causative role in naturally occurring malignancies, benign tumors and tumors of experimentally infected laboratory animals. Although they are members of different and unrelated systematic entities, they have several features in common. They mostly encode growth-stimulating proteins, contain DNA as genetic material, and establish long lasting persistence in the host (Table 1). The last frequently relies on episomal replication of circular viral genomes

Viral Oncogenesis. Table 1 Oncogenic viruses

Systematic group	Virus oncogenic in humans	Virus oncogenic in animals (examples)	Nucleic acid in virus particle	Nucleic acid in host cell/ malignancy
γ-Herpesviruses	Epstein-Barr virus (EBV) Human herpesvirus 8 (HHV-8)	Herpesvirus saimiri	DNA linear (110-180kb)	DNA circular episomal
Adenoviruses	-	Ad12	DNA linear	DNA integrated
Polyomaviruses	-	SV-40 Polyoma virus	DNA circular	DNA circular/integrated
Papillomaviruses	Human papillomavirus 16, 18 (HPV16,18 and others)	Bovine papillomavirus (BPV-1)	DNA circular	circular/integrated
Hepadnaviruses	Hepatitis B virus (HBV)	Woodchuck hepatitis virus (WHV)	DNA circular	DNA circular/integrated
Retroviruses	Human T-cell leukemia virus (HTLV-1)	Avian leukemia virus (ALV) Rous sarcoma virus (RSV)	RNA	DNA integrated

with a viral origin of replication or the insertion of the genome into the host chromosome.

Most oncogenic viruses contain DNA genomes that are widely different in size and structure. Among these are hepadnaviruses, ►[papillomaviruses](#), polyomaviruses, adenoviruses, and γ -herpesviruses. In contrast, ►[retroviruses](#) contain RNA as genomic material, which upon infection of the host cell is reverse transcribed into DNA and inserted into the host genome by means of a viral enzyme (integrase). The most frequent oncogenic viruses are DNA viruses with relative small genomes. They replicate by entering host cells and programming the cellular DNA synthesis machinery to produce copies of the invading viral genome. Since these viruses rely on cellular enzymes for DNA polymerization, which are only produced during the S phase of the cell cycle, these viruses stimulate cells to enter the cell cycle.

Virus infections that cause complete lysis of the host cell and efficient production of progeny viruses cannot contribute to tumor development. Thus, induction of malignant tumors is frequently the result of abortive infections and long lasting virus persistence in the host. Abortive infections are caused by viruses with defects in essential genes or by infection of host cells that do not support the full cycle of viral replication. The viral DNA can be randomly integrated into the host chromosomes. This is frequently accompanied by deletions and rearrangements of viral and cellular sequences. The chromosomal integration results in the stimulation of cell growth by continuous synthesis of viral proteins

Only a few tumor viruses of animals induce malignant tumors within a few days or weeks after infection. Most viruses including all human oncogenic viruses cause malignancies after long latency periods of many years. This hints at additional changes that the cells have to undergo before the malignant diseases develop. The long-term persistence of many tumor viruses results in the benign polyclonal stimulation of infected cells due to expression of viral growth promoting proteins. These cells may acquire additional mutations, which result in the promotion of rare infected clones into malignant cancer. Alternatively, the viral interference with cellular pathways of apoptosis may result in the preservation of cells carrying oncogenic mutations that would otherwise be eliminated.

About 20% of all human malignancies are the late sequels of virus infections. World wide about 450,000 cases of cervical carcinoma develop annually; these are caused by human papillomaviruses (HPV). Papillomaviruses constitute a large systematic group characterized by small icosahedral, non-enveloped capsids, circular doubled-stranded DNA genomes and a strict tropism for epithelia for their replication. The steps of viral replication are closely linked to the differentiation

of the keratinocytes. Depending on the HPV type, the productive infection results in different types of benign tumors of skin (wart, papilloma) or genital mucosa (condyloma). These lesions are characterized by a thickening of the epithelium. This is caused by increased rates of mitosis in keratinocytes mediated by the activities of the viral oncogenes E6 and E7. Depending on the presence of high-risk HPV types and chromosomal integration of viral genomes, these lesions can convert to genital cancer. High-risk type papillomaviruses are HPV16, HPV18, 45, 56. In addition malignant cancer of the skin is also frequently found to be associated with HPV, particularly under conditions of immune suppression.

Various types of lymphoid malignancies and nasopharyngeal carcinoma are causally linked to ►[Epstein-Barr virus](#) (EBV) infections. The gamma-herpesvirus infects and persists in B-lymphocytes, which are stimulated to growth by the expression of viral latency-associated gene products. Like persistently infected cells, all EBV-associated malignancies contain the viral episome and express various numbers of the EBV latency-associated gene products. A benign lymphoproliferation of EBV infected B-lymphocytes and cognate cytotoxic T-lymphocytes are hallmarks of infectious mononucleosis, a self-limiting disease caused by the virus. In immunocompromised individuals the lymphoproliferative-stimulated cells grow out to become malignant polyclonal B-cell lymphomas. In addition, EBV is strongly associated with the endemic form of Burkitt's lymphoma, which is prevalent among children in tropical Africa, with Hodgkin's lymphoma, and with T-cell lymphoma.

Kaposi's sarcoma, and two rare lymphoproliferative diseases (Castelman's disease and body cavity based lymphoma) are etiologically linked to infection with the ►[human herpesvirus type 8](#) (HHV-8). This gamma-herpesvirus establishes life-long persistence in lymphoid cells and expresses growth-stimulating, latency-associated gene products. Kaposi's sarcoma is a multifocal cancer of the skin, which develops from endothelia. Most of its forms are linked to severely immune-compromised patients, such as AIDS patients or transplant recipients. The viral genome and the expression of latency-associated gene products are present within the malignant spindle cell.

Adult T-cell leukemia/lymphoma (ATLL) is an aggressive malignancy of T-helper lymphocytes, which is caused by the ►[human T-cell leukemia virus type 1](#) (HTLV-1), a member of the retrovirus family. The leukemic cells are clonal according to retroviral genome integration site. The disease develops from infected T-lymphocytes in adults after decades of viral persistence. The virus persistence is associated with polyclonal non-malignant expansion of infected T-cells, which is probably due to the expression of the

growth stimulating viral Tax protein. Tax is a multi-functional oncoprotein, which is capable of immortalizing human lymphocytes in culture and inducing leukemia in transgenic mice.

Hepatocellular carcinoma is the frequent consequence of chronic ►hepatitis B virus infection of the liver. The virus belongs to the hepadna virus family; the enveloped particles contain small circular DNA genomes with four major genes. At least one of its proteins, the X protein, has transforming features in tissue culture and in transgenic mice. Factors for the development of the disease are dysregulation of cellular signaling by virus proteins, integration of the viral genome into the chromosome and the continued destruction and regeneration of liver parenchyma cells. The latter is caused by the cytotoxic effect of the virus and, more importantly, by the host immune response. Broad vaccination against hepatitis B virus has significantly lowered the prevalence of hepatitis B. This provides the first example of a highly efficient prophylaxis against virus associated tumor forms in the general population.

Molecular Interactions and Regulatory Mechanisms

As other tumors, virally induced tumors basically rely on a defect of growth and proliferation control and thus on a deregulation of ►proto-oncogenes and ►tumor suppressors. Tumor viruses apply several mechanisms to achieve this. Animal retroviruses and herpesviruses can contain ►oncogenes within their genomes, which are expressed as a consequence of the persisting infection. These oncogenes (v-onc) are mutant variants of cellular proto-oncogenes (c-onc), from which they

are derived (Table 2). The viral oncogenes constitutively deliver growth signals since they are either (1) over-expressed or (2) produced in an inappropriate cell or (3) devoid of negative control domains. Such oncogene transducing retroviruses as the Rous sarcoma virus, which contains the v-src oncogene, can efficiently and quickly induce tumors in laboratory animals. Moreover, they are capable of inducing measurable changes in growth behavior in cultured cells (transformation). These v-onc genes are derived from growth factors (v-sis), growth factor receptors (v-erbB), intracellular cytoplasmic signal proteins (v-src), transcription factors or cell-cycle regulators.

A dysregulation of cellular c-onc genes can also be induced by viral gene products, which act as transcriptional transactivators or by the insertion of viral transcriptional control sequences (enhancer, promoter, polyA signal) into or close to c-onc genes. The latter phenomenon is designated as ►insertional activation and is frequently accompanied by genomic mutagenesis. Insertional c-onc activation is the most frequent form of tumor induction by murine and avian retroviruses such as murine leukemia virus. Since the integration of retroviral sequences into chromosomal sequences occurs at random, hitting an oncogene is a rare event; this explains the long latency of tumors and their monoclonality.

Many viral oncogenes have no obvious homology to cellular c-onc proteins and thus constitute genuine viral functions (Table 3). These proteins exert their growth promoting activities by binding to cellular growth regulators and in this way induce cellular signaling or prevent its negative regulation. Cellular targets can be components of receptors, the src and related tyrosine

Viral Oncogenesis. Table 2 Cell-derived viral oncogenes

Viral oncogen (v-onc)	Cellular homologues protooncogen (c-onc)	Function of protooncogen	Virus (family)
v-sis	Platelet derived growth factor (PDGF)	Growth factor	Simian sarcoma virus (retrovirus)
v-erb B	Epidermal growth factor receptor	Growth factor receptor	Avian erythroblastosis virus (retrovirus)
v-src	c-src	Non-receptor tyrosine kinase	Rous sarcoma virus (retrovirus)
v-myc	c-myc	Transcription factor	Avian myelocytoma virus (retrovirus)
v-cyclin	cyclin D	Cell cycle regulator	Human herpesvirus 8 Herpesvirus saimiri
v-bcl -2	c-bcl-2	Apoptosis inhibitor	Human herpesvirus 8 Herpesvirus saimiri

Virtual Screening. Table 3 Cellular targets of viral transforming proteins (examples)

Cellular target (function)	Viral Oncogen	Virus
Ikk γ (NF- κ B signalling)	Tax	Human T-cell leukemia virus type 1
TRAF (tumor necrosis factor signalling)	LMP-1	Epstein-Barr virus
lck, src (non-receptor tyrosin kinase oncogenes, growth factor signal transduction)	Tip	Herpesvirus saimiri C488
p53 (tumorsuppressor)	Middle T-antigen	Polyomavirus
	Large T-antigen	SV40
	E6	Human papillomavirus 16
	E1B	Adenovirus Ad12
Rb (tumorsuppressor; G1 phase control within cell cycle)	E7	Human papillomavirus 16
	E1A	Adenovirus Ad12
	Large T-antigen	SV40
	Tax	Human T-cell leukemia virus type 1
CDK 4 (cyclin dependent kinase; G1 phase control within cell cycle)		

kinases, ►NF- κ B and other transcription factors or cell-cycle kinases. Many DNA viruses encode proteins, which inactivate cellular tumor suppressors by direct binding. Most frequent targets are the ►Rbprotein and the ►p53 tumor suppressor protein. The viral proteins directly stimulate the cell-cycle of infected cells in the G1 phase and result in the entrance of the cell into the S phase. These proteins are frequently multifunctional, i.e. they have many different cellular targets. In addition to their growth stimulating capacity some tumor viruses have acquired functions that interfere with induction of ►apoptosis, ►DNA repair and immune evasion, which also contribute to the induction of malignant growth.

Further reading

1. Bouchard MJ, Schneider RJ (2004) The enigmatic X gene of hepatitis B virus. J Virol 78:12725–12734
2. Flint SJ, Enquist LW et al (2004) Principles of virology: molecular biology, pathogenesis, and control of animal viruses. ASM Press Washington DC
3. Grassmann R, Fleckenstein B, Desrosiers RC (1994) Viral transformation of human T lymphocytes. Advances in Cancer Research 63:211–241
4. Jeang KT (2001) Functional activities of the human T-cell leukemia virus type I Tax oncoprotein: cellular signaling through NF- κ B. Cytokine Growth Factor Rev 12:207–217
5. Munger K, Baldwin A, Edwards KM et al (2004) Mechanisms of human papillomavirus-induced oncogenesis. J Virol 78:11451–11460

6. Neipel F, Albrecht J-C, Fleckenstein B (1997) Cell homologous genes in the Kaposi's sarcoma-associated rhadinovirus human herpesvirus 8: determinants of its pathogenicity? J Virol 71:4187–4192
7. Nevins JR (2001) Cell transformation by Viruses. In: Knipe DM, Howley PM (eds) Fields Virology, Lippincott, Williams and Wilkins, Philadelphia, pp 245–283
8. Young LS, Rickinson AB (2004) Epstein-Barr virus: 40 years on. Nat Rev Cancer 4:757–768

Virion

Definition

The virion is the extracellular, particulate form of a virus.

►Retroviruses

Virtual Screening

RONALD KÜHNE

Research Institute for Molecular Pharmacology, Berlin, Germany
kuehne@fmp-berlin.de

Definition

Virtual screening (VS, often also referred to as *in silico* screening or protein based database screening)

describes the process of reducing a library of available and/or virtual compounds to a limited number of potentially bioactive structures for a target or a target family using computational techniques (1). VS comprehends a broad spectrum of methods. These include computational approaches to filter compound libraries according to their structural diversity, drug likeness, ADME properties (ADME: absorption, distribution, metabolism, excretion), presence of 2D- or 3D-pharmacophores and binding capacity for a target protein calculated using automated docking procedures. The aim of VS consists of the rationalization of the drug discovery process by prioritizing compounds for screening and improving hit rates by using computationally filtered compound libraries.

Description

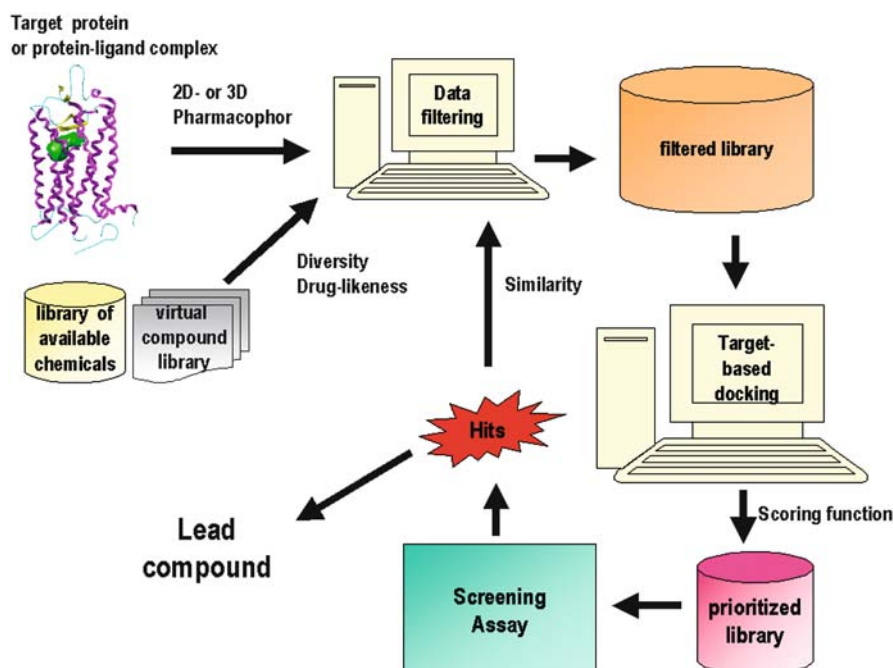
The broad application of new technologies like combinatorial chemistry (2) and high throughput screening (3) in drug discovery allows the synthesis and testing of hundreds of thousands of compounds. Additionally, the number of potential therapeutic targets emerging from the field of functional genomics is growing dramatically. All these factors increase the pressure to develop a valuable VS strategy for prioritizing screening compounds. The recent advances and developments in automated computational methods for VS have allowed its application to different

stages of the drug discovery process, namely hit discovery and lead optimization (Fig. 1). VS includes a broad spectrum of different computational methods. These consist of library filtering methods, checking compounds according to specific properties or occurrence of pharmacophoric patterns and docking approaches used to sort out all compounds within a library having no chance of binding into a predefined binding site. The basic idea of all of these approaches is the similarity property principle, which means that structurally and property related compounds should demonstrate similar biological activities.

Filtering of Compound Libraries

Database filtering methods are mainly applied to more general libraries of chemical compounds. They are used at an early stage in a project, where only structural information on active compounds is available. The aim of these methods is the reduction of the available chemical space in libraries containing compounds with the desired properties, high structural diversity, certain structural scaffolds or pre-defined pharmacophoric groups.

An important step in database filtering is related to the calculation of compound descriptors. Compound descriptors can be divided into structural and property descriptors. Structural descriptors depict the 2D or 3D structure of chemical entities. Commonly used 2D structural descriptors are fingerprints denoting the



Virtual Screening. Figure 1 The role of virtual screening tools in the drug discovery process. By using data filtering methods, virtual and available chemical libraries can be reduced according to different desired properties. The reduced libraries are docked into the target binding site in order to prioritize compounds for synthesis and screening.

presence or absence of certain substructures and a wide variety of descriptors based on a graphic representation of the chemical structural formula. 3D descriptors depend on the calculated molecular conformation. Commonly used 3D descriptors are accessible surface area and molecular volume. It was found that 2D descriptors were superior to 3D for clustering compounds into groups displaying similar physicochemical and biological activities (4). Because 2D methods are faster than 3D approaches, they are often used to select compounds from a general library.

Property descriptors are mainly applied to select compounds according to their lipophilic, electronic and steric properties. These properties are important in analyzing the drug likeness of compounds. A frequently used filter to evaluate the drug likeness of compounds was developed by Lipinski (5). The so-called 'rule of five' was derived to predict the bioavailability of a compound that can be considered as the major feature for drug likeness. If the number of hydrogen bond donors is <5 , of hydrogen bond acceptors <10 , the relative molecular weight <500 and the lipophilicity expressed as $\log P < 5$, the compound is probably orally bioavailable. Different types of descriptors can be combined using a suitable distance function to analyze the structural diversity of compound databases. A number of distance functions are available including the well-known Euclidian metrics for real valued descriptors and the Tanimoto coefficient for binary descriptors like fingerprints. In general, large descriptor sets are necessary to provide adequate structural diversity within a compound library. However, the more descriptors used for a diversity study, the greater the probability that they are correlated. In order to minimize this redundancy, approaches like cluster analysis or principal component analysis are used to reduce the dimensionality of the problem and to allow a good graphical representation of structure and property based similarities within a compound library.

Because a drug interacts with its target protein in 3D space, 2D descriptors are unsuitable to describe the entire protein-ligand interaction. Therefore it is often desirable to enrich a library for compounds that share common 3-dimensional structural elements with known active ligands. These elements are commonly called **▶ pharmacophores**. A typical pharmacophore represents a set of 3 or 4 pharmacophoric centers forming a triangle or tetrahedron and can be formulated as a 3D query to extract molecules containing the pharmacophoric definition. It becomes clear that the higher the complexity of pharmacophoric definition the better is the ability to separate inactive and active compounds, but the calculation time increases accordingly. However, this approach underlines some important limitations that are related to the large number of possible solutions and the large number of possible low energy conformations. An

often-used simplification of this approach consists of encoding the 3D pharmacophore in the form of long binary vectors called pharmacophoric fingerprints. This procedure reduces the computational time drastically, but is really dependent on the quality of the conformational sampling that is a prerequisite for fingerprint definition. Therefore attempts have been made to overcome this problem by using substituent based pharmacophoric descriptors such as the topomerically aligned conformer fields (6).

Docking Approaches

A basic problem of all similarity-based searches is that these approaches cannot be used with targets for which no active compounds are known. Further, 2D and 3D searches have a tendency to enrich structures that are relatively closely related to the known active leads, thus limiting the chance of finding new patentable scaffolds. A possibility for overcoming this limitation is to model the structures of protein-ligand complexes using fast small molecule docking algorithms. These algorithms must be able to recognize the key interactions involved in ligand binding and calculate reasonable 3D structures and binding affinities with the aim of prioritizing chemical entities. On the other hand, a protein-ligand interaction often induces conformational changes in both binding site and ligand. In principle such conformational flexibility should be considered but this is computationally extensive. Therefore all available docking algorithms have implemented simplifications to reduce the complexity of this task. The highest level of simplification is to calculate a rigid docking in which both the binding site and the ligand are held to be conformationally rigid. In that case the docking problem is reduced to finding a match between structural features of the binding site and the ligand. In this approach a distance compatibility graph is constructed in order to get reliable matches. Each pair of matching features between the site and the ligand represents a node within the graph. If two pairs of matching features have the same distance within a certain tolerance, the two nodes are connected by an edge. A 3-dimensional match between the site and the ligand is equivalent to a fully connected subgraph, which is called in graph theory a clique. The search for the optimal docking orientation is then reduced to the finding of the largest connected subgraph. In principle, application of this algorithm is not limited to rigid body docking. The inclusion of conformational flexibility of ligands can be done in several ways. The most obvious way is to individually dock every reasonable ligand conformation. In that case all low energy conformations for each ligand in the library must be calculated and stored in a database. Such pre-calculated databases of conformers can then be used in rigid docking programs. The result depends strongly

on the quality of the conformational analysis and the energy cutoff used to decide which conformers are accepted for docking.

Another approach is based on cutting the rotatable bonds of the ligand and dividing the ligand into a set of rigid fragments which then can be docked individually into the binding site by using rigid docking algorithms. In this so-called incremental construction approach, a core fragment is first docked and the residual fragments are added in an energetically favorable orientation. The first flexible docking program based on the incremental approach was developed by Kuntz (7). This approach was further developed and automated in the last releases of the program DOCK. The program FlexX (8) also utilizes the incremental construction approach. FlexX differs significantly from DOCK in the method used for placement of the core fragment. FlexX defines interaction sites for each possible interacting group of the binding site and the ligand. A pre-defined interaction type such as hydrogen bond donor, hydrogen bond acceptor etc., which is geometrically represented by an interaction center and a spherical surface, is assigned to each interaction site. The orientation of the core fragment is done by using an algorithm that matches each triangle of interaction sites in one object to the corresponding triangles in the other object. For each match a transformation superimposing the objects is determined. Once all possible transformations have been calculated, they are checked to remove all those that collide with the binding site. In a last step, all placements are clustered according to the RMS deviation using a hierarchical clustering algorithm.

The second type of approach to include flexibility is summarized under the generic term stochastic methods. Docking is a global optimization problem in which the best position of a ligand within a binding site has to be found. Because such an optimization cannot be resolved by using a fast algorithm, several efficient stochastic techniques were developed. One of these techniques is the Monte Carlo based optimization which is implemented in docking programs like AUTODOCK (9) and PRODOCK (10). Another widely used stochastic technique for optimization is the genetic algorithm. In a genetic algorithm, the state of a system is described as a binary string, which is called a chromosome. There are two ways to produce the next generation. In 'sexual' reproduction the two binary strings of the parent generation exchange bits to produce the 'children' state. 'Asexual' reproduction consists of point mutations in the parent chromosome. After each reproduction cycle, the binary strings are transformed back to the system state, which is then evaluated by an appropriate fitness function. System states with higher fitness have than a better chance to take part in the next reproduction cycle. The most widely used program based on genetic algorithm is GOLD (11).

Scoring Function

Another problem, which is in common to all docking approaches, is the evaluation and ranking of obtained solutions. The ideal scoring function is the binding free energy between the protein and the ligand. Methods that are able to calculate this very complex parameter with sufficient accuracy are only applicable for small systems. Therefore, simplifications must be made to prioritize a large number of binding modes for a large number of screening compounds. A widely used strategy to develop scoring functions is to analyze protein ligand complexes with known binding free energies with the aim of creating simple models that reproduce the experimental data. These models try to find incremental contributions for certain interaction types like hydrogen bonds, ion-pair interactions, hydrophobic interactions and aromatic stacking. Another possibility is calculation of the statistical preference of atom pair interactions. These scoring functions are based on the analysis of pair distribution functions that describe how much an atom pair prefers a certain distance compared to randomly distributed atom positions. In contrast to the former scoring functions, these pair distribution functions cannot distinguish between directional and undirectional interactions and often have a poor performance if directional interactions like hydrogen bonds dominate the binding free energy.

It has been shown that no scoring function is superior to all others. Hence a set of scoring functions is used in order to find compounds that consistently score highly. This so-called consensus scoring strategy, in which only compounds that are common in the top lists of different scoring functions are selected, can reduce the number of false hits in virtual screening.

► [Molecular Docking](#)

► [Molecular Dynamics Simulations in Drug Design](#)

References

1. Walters PA, Stahl MT, Murcko MA (1998) Virtual screening – an overview. *Drug Disc Today* 3:160–178
2. Geysen HM, Schoenen F, Wagner D et al (2003) Combinatorial Compound libraries: an ongoing challenge. *Nat Rev Drug Disc* 2:222–230
3. Sundberg SA (2000) High-throughput and ultra-high-throughput screening: solution and cell-based approaches. *Curr Opin Biotechnol* 11:47–53
4. Brown RD, Martin YC (1997) The information content of 2D and 3D structural descriptors relevant to ligand-receptor binding. *J Chem Inf Comput Sci* 37:1–9
5. Lipinski CA, Lombardo F, Dominy BW et al (1997) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev* 23:3–25
6. Cramer RD, Clark RD, Patterson DE et al (1996) Bioisosterism as a molecular diversity descriptor: steric fields of single topomeric conformers. *J Med Chem* 39:3060–3069

7. Leach AR, Kuntz ID (1992) Conformation analysis of flexible ligands in macromolecular receptor sites. *J Comput Chem* 15:411–428
8. Kramer B, Rarey M, Lengauer T (1999) Evaluation of the FlexX incremental construction algorithm for protein-ligand docking. *Proteins* 37:228–241
9. Morris GM, Goodsell DS, Huey R et al (1996) Distributed automated docking of flexible ligands to proteins: parallel applications of AUTODOCK 2.4. *J. Comput Aided Mol Design* 10:293–304
10. Trosset JY, Scheraga HA (1999) PRODOCK: Software package for protein modeling and docking. *J Comp Chem* 20:412–427
11. Jones G, Willett P, Glenn RC et al (1997) Development and validation of a genetic algorithm for flexible docking. *J Mol Biol* 267:727–748

Virulent / Virulence

Definition

Virulent designates the degree of a aggressiveness of microorganisms in a macroorganism whereas pathogenicity refers to the capability of microorganisms to cause a pathological state (disease).

► [Proteomics in Human-Pathogen Interactions](#)

Viscoelasticity

Definition

Viscoelasticity is a rheological parameter that describes the flow properties of complex fluids like blood or the cytoplasm. Viscoelasticity combines viscosity and elasticity. Viscosity is characterized by the energy lost during flow, primarily due to sliding and deformation. Elasticity is related to the energy stored during the flow due to orientation and deformation. Actin filaments cause a high viscosity in the cytoplasm, which allows a steady flow during slow migration. Actin at the crosslinking of proteins adds an elastic modulus because they guarantee a rubber-like resistance against sudden perturbation.

► [Actin Cytoskeleton](#)

Vitiligo

Definition

Vitiligo is an acquired, sometimes familial, depigmentary disorder of the skin and hair that results from selective destruction of melanocytes that produces white patches.

Usually both sides of the body are affected. Common areas of involvement are the face, lips, hands, arms, legs, and genital areas. People with vitiligo have an increased risk to get skin cancer. The pathogenesis of vitiligo is complex, with both genetic and non-genetic factors contributing to the development of the disease. Genes that contribute to susceptibility to vitiligo are still unknown. The inherited nature of vitiligo and its frequent association with autoimmune diseases has prompted studies on the association between HLA and vitiligo.

► [Heritable Skin Disorders](#)

Vitreous Ice

Definition

Vitreous ice refers to ice lacking hexagonal or cubic crystal order, which is generated by shock-freezing water solutions.

► [Cryo-Electron Microscopy: Single-Particle Reconstruction](#)

VL

Definition

VL defines the variable domain of the light-chain of conventional ► [antibodies](#) of all vertebrates.

► [Camel as a Model for Functional Genomics](#)

VLCFA

► [Very-Long-Chain Fatty Acids](#)

VLDL

► [Very Low Density Lipoprotein](#)

Voltage-Clamp Mode

Definition

Voltage-clamp mode is the commonly used recording mode in the patch clamp analysis. The membrane

potential of a cell is set to a certain voltage (clamp), while ion currents through the cell membrane are measured.

► [Patch Clamping](#)

von Hippel-Lindau Disease

► [VHL-\(von Hippel-Lindau\) Disease](#)

von Willebrand Disease

Definition

The von Willebrand disease is an inherited, autosomal dominant bleeding disorder caused by a defective or deficient von Willebrand factor, a protein involved in normal blood clotting. It is due to a deficiency or anomaly of an attribute of factor VIII, which is necessary for the adhesion of platelets to vascular elements. Symptoms include nosebleeds, bleeding gums, heavy menstrual bleeding, bruising and skin rashes.

► [Hereditary Hemostatic Defects and Recombinant Proteins for Treatment](#)

von Willebrand Factor

Definition

The von Willebrand Factor (vWF) is a heterogeneous multimeric plasma glycoprotein produced by megakaryocytes and endothelial cells. vWF is found in platelets and the subendothelium, and it facilitates platelet adhesion through an interaction with Factor VIII.

v-SNARE

Definition

v-SNARE denotes ► [SNARE proteins](#) present on the vesicle membrane.

► [Cell-Polarity](#)

► [Vesicular Traffic](#)

► [Protein and Membrane Transport in Eukaryotic Cells](#)

vWF

► [von Willebrand Factor](#)

Walker A and B Motif

Definition

Walker A and B motif are conserved peptide sequences within the nucleotide binding fold. Highly conserved amino acids are glycines for Walker A and aspartates for Walker B.

► [High-HDL Syndrome](#)

Wavelength of Light

Definition

Wavelength of light defines the distance in nanometers between the nodes in a wave of light. The wavelength is inversely proportional to the energy of the light.

► [Fluorescence Microscopy: Spectral Imaging vs. Filter-Based Imaging](#)

WASp

Definition

WASp (Wiskott Aldrich Syndrome Protein) is a protein that is mutated in patients suffering from Wiskott-Aldrich Syndrome, which is an X-linked recessive primary immune deficiency disease. It leads to disorganised F-actin in T-cells that results in a disturbed polarisation. Overexpression of this protein increases the F-actin content. WASp is an established effector of cdc42, which is itself a regulator of cell polarity.

► [Actin Cytoskeleton](#)

► [Microvilli](#)

Wegener's Granulomatosis

Definition

Wegener's granulomatosis is a systemic necrotising inflammation of small vessels (vasculitis) that leads to severe impairment of affected organ systems.

Weighted Back-Projection

Definition

Weighted Back-Projection is an algorithm commonly used in 3-D imaging that allows a 3-D volume to be reconstructed from projection views.

► [Electron Tomography](#)

Watson-Crick DNA

Definition

Watson-Crick DNA describes the DNA duplex (double helix) comprising of only canonical base pairing, as first described by Watson and Crick in 1953.

► [DNA Structure](#)

► [Thermodynamic Properties of DNA](#)

Werner's Syndrome

Definition

Werner's syndrome (WS) is an inherited autosomal recessive disease that has symptoms that resemble premature aging. WS patients prematurely develop

many age related diseases including arteriosclerosis, malignant neoplasms, Type II diabetes mellitus, osteoporosis, ocular cataracts, early graying, loss of hair, skin atrophy and aged appearance. The mutated gene, *WRN*, encodes a member of the RecQ helicase family, which is unique in featuring exonuclease activity. WRN interacts with several proteins involved in DNA metabolism.

► [Molecular Aging Research](#)

► [Proteomics in Ageing](#)

Western Blotting

Definition

Western blotting refers to a procedure in which proteins separated by electrophoresis in ► [polyacrylamide gels](#) are transferred (blotted) onto nitrocellulose or nylon membranes and identified by the formation of specific complexes with antibodies. The protein specific antibody is subsequently detected with a secondary protein bearing a label which results in the detectable signal.

► [Immunological Methods, the Use of Antibodies to Discover the Function of Novel Gene Products](#)

► [Protein Microarrays as Tool for Protein Profiling and Identification of Novel Protein-Protein Interactions](#)

► [Recombinant Protein Expression in Bacteria](#)

White Sponge Naevus

Definition

White sponge naevus is an autosomal dominant disorder that is characterized by thickening and fragility of mucous epithelia.

► [Heritable Skin Disorders](#)

Whole Mount ISH

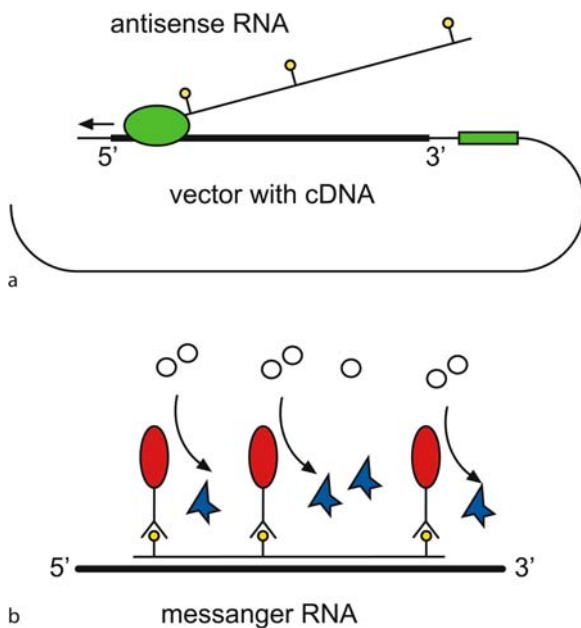
BERNHARD G. HERRMANN, LORENZ NEIDHARDT
Department of Developmental Genetics, Max Planck
Institute for Molecular Genetics, Berlin, Germany
herrmann@molgen.mpg.de
neidhardt@molgen.mpg.de

Synonyms

WISH

Definition

Whole mount *in situ* hybridisation is used to determine the RNA expression pattern of a gene in the context of a whole embryo or embryo piece/organ. The technique allows spatial assignment of the transcriptional activity of a particular gene to cell type, tissue or organ at high resolution. Whole mount ISH is routinely used on embryos of vertebrate (e.g. mouse, chick, fish, frog) and invertebrate (e.g. fruit fly, ascidia, *Hydra*) species. The principal behind whole mount ISH is the specific annealing of labelled nucleic acid probes to complementary RNA sequences in fixed tissue, followed by detection of the probe (Fig. 1). For probes, labelled DNA or RNA can be utilized; RNA probes are more common. Antisense RNA probes (which are complementary to the mRNA) can readily be produced *in vitro* by run-off transcription from cDNA fragments cloned into vectors containing promoters for RNA polymerases adjacent to the cDNA insert. DNA labelling is achieved by nick-translation, random priming or PCR amplification. Probes are labelled by synthesis of DNA or RNA in the presence of a modified nucleotide coupled through a spacer to a hapten group. Commonly used non-radioactive labels are digoxigenin, fluorescein and biotin. Detection of these haptens is achieved *via* enzymes coupled to antibodies directed against the hapten group or to avidin or streptavidin, which form stable complexes with biotin. Commonly used enzymes are alkaline phosphatase (AP) and peroxidase (POD), usually used in combination with the substrates 5-bromo-4-chloro-3-indolylphosphate/nitro blue tetrazolium (BCIP/NBT) or diaminobenzidine (DAB)/imidazole, respectively. The enzymatic reactions produce coloured precipitates (blue and brown, respectively), which are insoluble in water and stable. Alternatively, fluorochromes can be used as labels, either directly coupled to the nucleoside triphosphate used for probe synthesis (as is the case for fluorescein) or coupled to the antibody directed against the hapten of choice or avidin/streptavidin. However, in general, fluorochrome labels are not sufficiently sensitive for mRNA detection in whole mount ISH. At present, parallel detection of two mRNA species or one mRNA and one protein are feasible by using non-fluorescent labelling. In the near future, nanocrystals/quantum dots may overcome the sensitivity issue and allow parallel detection of several transcripts in the same embryo. Detection of mRNA may then be combined with visualization of proteins (by antibody staining) and reporters (fluorescent protein such as eGFP, YFP etc.) in the same specimen.



Whole Mount ISH. Figure 1 Schematic presentation of RNA probe preparation *in vitro* and mRNA detection *in situ*. (a) Antisense RNA is synthesized in the presence of hapten-linked nucleotide triphosphates by RNA polymerase *in vitro* from a linearised plasmid containing the cDNA fragment (bold bar) and a RNA polymerase promoter (green box) close to the 3'-end of the fragment. (b) After hybridization of the labelled antisense RNA to the corresponding messenger RNA *in situ*, an antibody, which is coupled to an enzyme (red) is bound to the hapten group. The enzyme catalyzes conversion of a soluble substrate to a colored precipitate, thereby marking the cells and tissues in which the gene of interest is expressed.

Characteristics

Whole mount ISH was first described for embryos of the fruit fly *Drosophila melanogaster* (1), followed by application to *Xenopus* (2) and mouse embryos (3, 4). Since then these protocols have been adapted to other organisms. To date whole mount ISH can be applied to embryos of basically every model organism of developmental biology including plants (Fig. 2). Whole mount ISH has become the major technique for gene expression analysis in embryos. In a variety of model organisms it is an essential tool for the characterisation of mutants and manipulated embryos, thus promoting our understanding of pattern formation and organogenesis. The success of whole mount ISH as a tool for gene expression analysis is based on the fact that genes, once available in form of cDNA clones, can immediately be subjected to expression analysis in the

organism. In contrast, detection of proteins encoded by the genes of interest requires very time consuming schemes of antibody production. Genes are readily isolated from cDNA libraries, either by random picking or through screening or selection procedures. Gene catalogues and collections holding full-length cDNA clones of all genes encoded by the genome are being established for a variety of organisms.

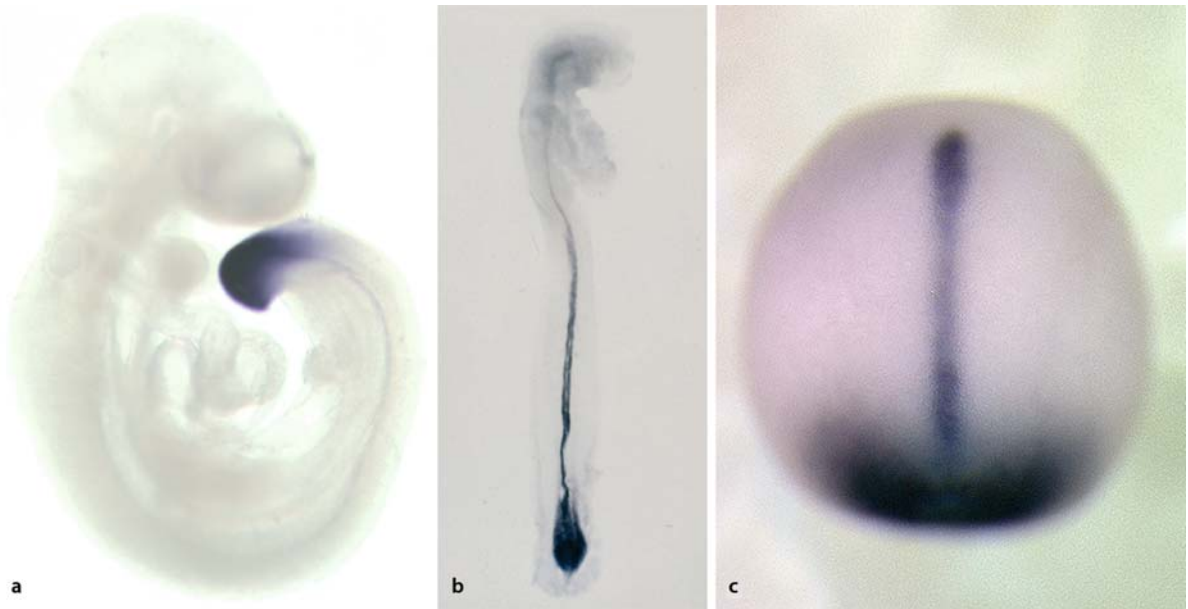
Microarrays of gene collections provide ideal tools for expression profiling and description of the transcriptome of particular cell types, tissues or organs and for analysis of alterations in the transcriptome caused by e.g. inducing agents, loss of gene function, pathogenic or other triggers. However, such data do not provide cellular resolution if complex tissues or whole organisms are analysed. Whole mount ISH can be employed to complement such data sets by providing high-resolution expression data.

Whole mount ISH is usually, where applicable, also favoured over ISH on embryo sections because the procedure takes much less effort, while at the same time the result can more readily be interpreted since the pattern is viewed in the context of the whole organism which provides morphological landmarks. In contrast, three-dimensional expression patterns are often difficult to reconstruct from the signals seen on sections.

High-Throughput Whole Mount ISH as Tool for Systematic Analysis of Embryonic Pattern Formation and Organogenesis

The expression analysis of a considerable number of genes involved in the control of embryonic processes in invertebrate and vertebrate species by ISH has shown that genes controlling tissue differentiation and organogenesis very often show a temporally and/or spatially restricted expression pattern. For example, all known members of the Hox, Pax and T-box gene families, prominent developmental control factors, show a restricted expression. Tissue restricted expression of control factors usually precedes morphological changes in the embryo. Thus, the molecular anatomy prefigures the morphology of the embryo. In consequence, if the molecular anatomy of the organism could be deciphered, this information should greatly facilitate unravelling the regulatory networks controlling differentiation processes in molecular detail. The molecular anatomy will become available by a detailed description of the expression patterns of all genes encoded in the genome.

The large number of genes envisioned a decade ago to be encoded by vertebrate species could not be analysed by the "classical" whole mount ISH techniques, which allowed parallel analysis of only a few genes at a time. Therefore, the technique had to be adapted to high-throughput scale, allowing parallel processing of up to



Whole Mount ISH. Figure 2 Expression analysis of *Brachyury* by whole mount *in situ* hybridization. (a) E9.5 mouse embryo, (b) chick embryo (HH stage 14, kindly provided by Susan Mackem (5)), (c) *Xenopus laevis* embryo (from (6)).

96 genes at a time. To achieve that, hands-on time for processing of individual genes needed to be minimized. Briefly, in this procedure, for probe preparation, cDNA inserts flanked by RNA-polymerase promoters are amplified by PCR and directly used for antisense RNA transcription. Probe preparation, *in situ* hybridisation, washes and staining are carried out in a 96-well format device. Large numbers of genes can be assayed with relatively small effort using such technology. However, it should be stressed here that, rather than the embryo preparation and WISH procedure itself, the evaluation and documentation of the results obtained is limiting for the progress of such a project.

Interestingly, high-throughput whole mount ISH was first carried out on species not or hardly amenable to large-scale mutant screening for recessive developmental phenotypes, *Xenopus laevis* and the mouse (6, 7, 8). In these systems, genes controlling embryonic development have been isolated traditionally on the basis of sequence conservation to genes of known importance in *Drosophila* development, using filter hybridisation technology, or through functional screening.

High-throughput whole mount ISH has been introduced more recently to identify potential control factors solely on the basis of a restricted expression pattern. The latter, complemented by hints on the molecular function derived from sequence information, can suffice as an indication of a possible regulatory role in a particular differentiation process.

Various methods are available for testing the functional role of a gene selected on the basis of these criteria. (Ectopic) over-expression or mis-expression of wild type or dominant (active or negative) gene products often provides important insight into the role of a gene. This approach is the prime methodology for functional gene analysis in *Xenopus laevis*. It has also been widely applied to zebrafish embryos and, to a minor degree, to the chick and the mouse. It has proven to be very successful with respect to its impact in understanding control mechanisms of embryonic development. In mouse, gene targeting by [homologous recombination](#) ([“knock-out” mutagenesis](#)) in embryonic stem cells has been utilized primarily as tool for functional analysis. More recently, “lack of function” [mutagenesis](#), based on [RNAi](#) or morpholino ([Morpholino Oligonucleotides](#), [Functional Genomics by Gene ‘Knock-down’](#)) technology, has become available and is widely applied. With this technology the protein level produced from a particular gene can be strongly down-regulated, allowing the generation of phenocopies of loss of function mutations. Therefore, adequate tools promoting the functional analysis of genes identified by virtue of a restricted expression pattern or through functional screening are available. The “classical” random mutagenesis approach for the identification of genes essential for embryonic development, requiring time and labour intensive positional cloning of the mutated genes, has therefore become dispensable.

In addition, disadvantages of the mutagenesis approach are avoided. In vertebrates, a high degree of (partial) functional redundancy has been observed. This is due to gene amplifications having occurred during evolution. Often, important gene functions are provided by two or several members of a gene family co-expressed in the same cells. The consequence is that a loss of function mutation in one gene is often complemented by the wild type function of another co-expressed family member and a mutant phenotype is not obtained. Such a gene would not be detected by the mutagenesis approach identifying only essential genes. Being “non-essential” is often falsely interpreted as not having functional importance, which is of course erroneous.

Dominant mutant gene products often reveal phenotypes providing important insight into the embryonic role of a gene, which is not apparent when only the lack of a recessive mutant phenotype is considered. Therefore, it is important to identify all genes expressed in patterning and differentiation processes, not only “essential” genes showing loss of function mutant phenotypes.

Another disadvantage of random mutagenesis as a tool for the identification of genes controlling embryonic development is the fact that the same signal cascades and control factors are utilized in multiple processes and cell types of an organism. Thus, mutations in genes involved in processes taking place later in development (e.g. in organogenesis) may escape detection if the mutations result in early lethal phenotypes or in morphological changes affecting the formation of late appearing structures.

In recent years, large-scale gene expression analysis has also been applied to the zebrafish, medaka and *Drosophila melanogaster* (9, 10, 11). Large fractions of the corresponding genomes have been screened already and for some of these systems saturation may be achieved soon. As a result, the molecular anatomy of the embryo will emerge. This will allow reconstruction of the development of pattern, tissues and organs on a molecular scale in the computer. In combination with the genome sequence, the coding sequences of all genes encoded in the genome and the molecular function of all these genes, computational simulation of the generation of complex organisms can be initiated.

Gene Expression Data Derived by Whole Mount ISH Can Be Utilized in Multiple Ways

Gene expression analysis *per se* provides an assignment of gene functions to the tissues in which they act. This basic functional information on a gene can be utilized in several ways. Genes can be grouped according to identity of the expression pattern in several processes/cell types. Such so-called “synexpression groups” (12) often identify genes acting in the same signal pathways. Genes can also be grouped with

respect to their expression in a particular cell type or organ. This identifies clusters of genes involved in particular differentiation processes. The latter clustering is highly stage dependent since cellular readouts can rapidly change during development. All this information, in combination with predicted protein functions, can be used to make predictions with respect to the possible role of a gene or set of genes, which can be tested by functional studies.

Expression data can also be utilized to identify candidates for mutant genes causing inherited diseases. Candidates can be identified on the basis of a match between the expression pattern of a gene and the mutant phenotype, in combination with co-localisation in the genome.

In combination with genome sequence data, gene expression data can be utilized to extract cis-acting control elements likely to regulate the expression of a synexpression group or a gene cluster in particular tissues. The latter data can again be related to the expression patterns of transcriptional control factors likely to bind to such control elements. Such bio-informatic evaluation of gene sequence, function, expression and control data can be utilized to predict regulatory networks (the “regulome”) controlling embryonic processes. In addition, regulonome data on various organisms may allow unravelling of the evolutionary mechanisms promoting the phylogeny of higher vertebrates (including ourselves), on a genomic level.

Large-Scale Mutant Screening Using Whole Mount ISH

More recently whole mount ISH has also been applied as monitoring system in large-scale mutagenesis projects, mainly in the zebrafish (*Danio rerio*) system. Chemical random mutagenesis has been utilized to generate mutations in virtually any gene encoded by the zebrafish genome. Complex breeding schemes are utilized to obtain homozygosity at the mutated gene loci. In the “classical” approach, the identification of a homozygous phenotype relied exclusively on morphological criteria. However, subtle phenotypes can readily be missed by this approach especially if small organ anlagen are affected. To overcome this limitation, the expression of genes marking the tissue or organ of interest is monitored using whole mount ISH. This requires the expression analysis of hundreds to thousands of offspring derived from intercrosses of heterozygous mutant carriers. Therefore, exhaustive mutant screens in zebrafish require high-throughput whole mount ISH technology.

Practical Considerations

Whole mount ISH is limited by several factors, in particular the size and translucency of the object and the expression level of the gene to be assayed.

The size limit is especially critical for mouse embryos, which can be reliably assayed by whole mount ISH only up to 11.5 days post fertilization. Older embryos get too large, preventing probe access to inner organs. The size limit also applies to chick embryos. Embryonic stages of lower vertebrates or invertebrates undergoing larval development do not usually face a size limitation. Embryos beyond the size limit must be analysed by conventional ISH on tissue sections.

In our experience, a significant proportion of genes cannot be detected by whole mount ISH with current technology. To date it is not clear, if some of these genes play a role in differentiation processes.

Clinical Relevance

Whole mount ISH serves as standard tool in the functional analysis of genes involved in tissue differentiation and organogenesis in the embryo. Many human syndromes and diseases derive from aberrant gene functions during embryonic development and pathologies of diseases often arise from malfunction of regulatory mechanisms and signalling cascades acting in the adult as well as in the embryo. Moreover, regeneration processes in the adult also reemploy embryonic mechanisms. Therefore, insight into molecular mechanisms controlling embryonic development becomes more and more important for understanding the triggers of clinical syndromes and pathologies as well as regeneration processes.

References

1. Tautz D, Pfeifle C (1989) A non-radioactive *in situ* hybridization method for the localization of specific RNAs in *Drosophila* embryos reveals translational control of the segmentation gene hunchback. *Chromosoma* 98:81–85
2. Hemmati-Brivanlou A, Frank D, Bolce ME et al (1990) Localization of specific mRNAs in *Xenopus* embryos by whole-mount *in situ* hybridization. *Development* 110:325–330
3. Herrmann BG (1991) Expression pattern of the Brachyury gene in whole-mount TWis/TWis mutant embryos. *Development* 113:913–917
4. Wilkinson DG (1992) Whole mount *in situ* hybridization of vertebrate embryos. In: Wilkinson DG (ed) *The Practical Approach Series. In situ hybridization: a practical approach*. Oxford University Press, New York, pp 75–83
5. Knezevic V, De Santo R, Mackem S (1997) Two novel chick T-box genes related to mouse Brachyury are expressed in different, non-overlapping mesodermal domains during gastrulation. *Development* 124:411–419
6. Gawantka V, Pollet N, Delius H et al (1998) Gene expression screening in *Xenopus* identifies molecular pathways, predicts gene function and provides a global view of embryonic patterning. *Mech Dev* 77:95–141
7. Neidhardt L, Gasca S, Wertz K et al (2000) Large-scale screen for genes controlling mammalian embryogenesis, using high-throughput gene expression analysis in mouse embryos. *Mech Dev* 98:77–94
8. Wertz K, Herrmann BG (2000) Large-scale screen for genes involved in gonad development. *Mech Dev* 98:51–70
9. Kudoh T, Tsang M, Hukriede NA et al (2001) A gene expression screen in zebrafish embryogenesis. *Genome Research* 11:1979–1987
10. Henrich T, Ramialison M, Quiring R et al (2003) Medaka Expression Pattern Database (MEPD): a Medaka gene expression pattern database. *Nucleic Acids Research* 31:72–74
11. Tomancak P, Beaton A, Weiszmarm R et al (2002) Systematic determination of patterns of gene expression during *Drosophila* embryogenesis. *Genome Biology* 3: RESEARCH0088.0081–0088.0014
12. Niehrs C, Pollet N (1999) Synexpression groups in eukaryotes. *Nature* 402:483–487

Wide-Field Fluorescence Microscopy

Definition

Wide-field fluorescence microscopy is a commonly used technique to obtain both topographical and dynamic information from biological samples, such as cell cultures or tissue sections. The sample, in which specific structures have been labelled by fluorescent dyes, is most usually illuminated with a mercury lamp, giving out pure white light. Optical filters (excitation filters), or a monochromator, are used to select the optimal wavelength of excitation light for the used dye. Excitation light is directed to the sample via a dichroic mirror (a mirror that reflects some wavelengths but is transparent to others). The fluorescent light emitted from the sample is selected by further optical filters (emission filters, cut-off filters; e.g. longpass or bandpass filter) and then detected by a camera.

Wild Type

Definition

Wildtype is the genotype of an organism that is found in nature, often also genotypes that serve as parental lines or strains for gene transformation or mutation experiments.

► Mass Spectrometry: Quantitation

► Transgenic and Knock-out Animals

William-Beuren Syndrome

Definition

A well defined disorder (incidence 1:10000 newborns) caused by chromosomal microdeletion of the elastin gene at 7q11.23. It is characterized by short stature, facial dysmorphism (periorbital fullness, strabismus, stellate iris, long philtrum, thick lips, and large open-held mouth), supravalvular aortic stenosis or other cardiovascular defects, variable hypercalcemia often manifesting under stress (accident operation), and mild mental retardation. There is also a hoarse voice; loquacious behaviour ("cocktail man"), pronounced sensitivity to sound and hypodontia.

► [Microdeletion Syndromes](#)

Winged Helix

Definition

Winged helix refers to the 100 amino acid DNA binding domain that is comprised of three alpha-helices and two loops, which form characteristic "wings" that give this domain its name.

► [Winged Helix Transcription Factors](#)

Winged Helix Transcription Factors

PATRICIA A. LABOSKY, KLAUS H. KAESTNER
Department of Cell and Developmental Biology,
University of Pennsylvania, PA, Philadelphia, USA
plabosky@mail.med.upenn.edu
kaestner@mail.med.upenn.edu

Synonyms

Forkhead genes or proteins; *Fox* genes or proteins

Definition

The *forkhead* genes form a large family of transcriptional regulators with approximately 40 members identified in mouse and humans. The name "*forkhead*" derives from the *Drosophila fork head* mutant, which led to the discovery of the first member of this gene family in 1989 (1). Flies homozygous for the *fork head* mutation exhibit homeotic transformations of portions

of the gut as the foregut and hindgut are replaced by ectopic head structures. Forkhead proteins are involved in embryonic development, tumorigenesis and tissue-specific gene expression in all metazoans. Members of this gene family share a common highly conserved DNA binding motif, the ► [winged helix](#) DNA-binding domain consisting of 100 amino acids (2) (see below). Historically, members of this gene family have had a variety of various names reflecting their discovery but in 2000 the nomenclature was standardized using "Fox" for forkhead box as the unifying term for all the *forkhead* genes from vertebrates (3). The ► [Fox gene](#) tree can be accessed at: <http://www.biology.pomona.edu/fox.html>. Within the Fox family there are classes denoted from "A" to "Q" based on sequence similarity. Different numbers of genes are found in the various classes and are indicated by numbers. For example, the largest class consisting of the *Foxd* genes has six members, *Foxd1-Foxd6*, while some classes have only one member; *Foxg*, *Foxh*, *Foxi*, *Foxm*, *Foxp* and *Foxq*.

Characteristics

Evolution

Winged helix genes are found in all animal phyla and fungi with fewer members in the lower phyla (four in yeast) and more in the upper phyla (16 in *Drosophila* and 42 in humans), suggesting that as the body plan became more complex, the *Fox* gene family was expanded (4). Interestingly, no ► [orthologs](#) for the *Fox* subclasses E, H, I, J, M and Q1 have been found in *Drosophila* or *C. elegans*, suggesting that these subclasses have evolved in the deuterostome lineage. Much useful information about the Fox genes has come from model organisms. One example is the *Caenorhabditis elegans* gene DAF-16, which is an ortholog of the *FOXO* group in humans. The discovery of DAF-16 as a gene involved in regulating the rate of ageing in *C. elegans* in response to insulin/IGF-1 signaling prompted the reevaluation of this pathway in mammals and led to the discovery that the mammalian *FOXO* genes function in the regulation of genes encoding metabolic enzymes in response to insulin (5). Thus, while the output of the signaling pathway, i.e. ageing or glucose metabolism differs between worms and humans, the underlying regulatory cascade is conserved.

Structure

The structure of the winged helix DNA binding domain defines the family of forkhead proteins. The crystal structure of Foxa3 (Hnf γ) bound to DNA was solved in 1993 and revealed that the DNA binding motif formed three α -helices with a helix-turn-helix structure (6). This core is flanked by two larger loops or "wings". The conserved winged helix binding motif is shared by

other DNA binding proteins (LexA, E2F, DP, RFX and ADAR1) that are not termed Forkhead proteins. Winged Helix proteins bind DNA as monomers with the recognition helix (helix 3) binding to a seven base pair core (RYMAATA R=A or G, Y=C or T and M=A or C). Upon binding of the forkhead protein the DNA is bent, effectively opening the chromatin structure to a more active configuration and allowing other transcription factors access to nearby cis-regulatory elements. At the albumin enhancer, which is currently the best-studied model of Fox transcriptional regulation, Foxa proteins can open chromatin by bending DNA and altering nucleosome position independently of histone acetylation.

Transcriptional and Metabolic Regulation

Most winged helix genes are activators of transcription but several have repressor function (Foxc2, d2, d3 and g1) and one *C. elegans* winged helix protein (LIN-31) can act either as a repressor or as an activator depending on the phosphorylation state of the protein. An interesting mechanism governs the activity of the Foxo proteins. Although normally nuclear, these proteins are targets for phosphorylation, which leads to nuclear exclusion and loss of DNA binding. In addition, Foxo proteins act as repressors or activators of transcription, depending on the cellular context. In hepatocytes, Foxo proteins activate several genes involved in gluconeogenesis in the fasted state. However, when insulin levels rise in response to feeding, binding of insulin to its receptor on the hepatocyte surface initiates a signaling cascade that leads to phosphorylation and nuclear exclusion of the Foxo proteins. In pancreatic β -cells, however, Foxo proteins occupying the enhancer of the *Pdx1* gene prevent its activation in the basal state. After insulin binding initiates the phosphorylation and removal of the Foxo proteins, the Foxa proteins are able to bind to the same sites and initiate activation of transcription. Thus in the case of the β -cell, Foxo occupancy represses transcription by competing with the stronger transcriptional activator, Foxa.

A yet unexplained puzzle is how certain Fox genes can exert multiple functions in different tissues. The best example for this is the case of the *Foxa2* gene, which has been studied through both transgenic over-expression and complete and tissue-specific gene ablation studies in mice in great detail. Thus it was demonstrated that *Foxa2* can induce re-specification of the dorsal neural tube into a ventral, floor-plate-like structure and that germ-line deletion of *Foxa2* leads to severe defects in axial mesoderm development with the absence of node and notochord (7, 8). In addition to these early developmental phenotypes, tissue-specific deletion has shown that *Foxa2* is also required for the terminal differentiation of pancreatic α - and β -cells, the

regulation of glucose homeostasis and the proper development of the gastrointestinal tract. Clearly, the molecular targets of *Foxa2* must differ between all these different cell types, but at present no explanation exists as to how this is achieved.

Mechanisms of Transcriptional Regulation

Not much is known about co-factors for winged helix proteins but several (Foxa2, g1 and d3) bind to *Groucho* co-repressor proteins. Foxg1 recruits the basic helix-loop-helix protein Hes1 in a complex with Groucho to then recruit a histone deacetylase to repress transcription.

Changing the intracellular concentration of specific transcription factors is a mechanism for modulating the expression of downstream genes and the precise level of some winged helix proteins can affect cellular function. This is the case for *Foxa2*; *Foxa2*^{+/-} mice are not viable on certain genetic backgrounds and have pronounced phenotypes on others. The binding of the *C. elegans* orthologue of Foxa (PHA-4) to target genes occurs with different affinities causing PHA-4 to act in a concentration dependent manner. The same variation in phenotype with gene dosage was shown for *Foxf1* mice with respect to lung repair and *FOXC1* over-expression in humans and/or decreased gene dosage in mice and humans causes eye defects.

Interactions with Signal Transduction Pathways

Winged helix proteins are downstream components of several signal transduction cascades demonstrating a universal involvement in cellular responses. Foxh1 acts downstream of TGF β -like signals (activin) by binding directly to phosphorylated Smad protein and target gene regulatory regions. Conversely, Foxg1 can inhibit signals from TGF β -like proteins (Bmp4) presumably by binding Groucho and Hes as mentioned above. *Foxl1*^{-/-} mice display overproliferation of cells in the intestinal epithelium that is correlated with activation of the Wnt signaling pathway. Foxd3 regulates a secreted signal required for the proliferation of stem cells in the early mouse embryo. Sonic hedgehog (\blacktriangleright Shh) from the notochord induces the expression of Foxa2 in the neural tube. Foxa2 then maintains the expression of Shh in the neural tube by directly binding to a Shh enhancer. The *C. elegans* forkhead protein LIN-31 acts downstream of an EGF-related ligand. In zebrafish, Foxi1 is required for cells to respond to FGF signaling in patterning the developing ear and jaws. The Foxo proteins are phosphorylated by protein kinase B (PKB or AKT) as described above in response to signaling via insulin and insulin-like ligands. The molecular mechanism of most of these interactions is yet to be determined and it is not clear if all of them are direct.

Winged Helix Transcription Factors. Table 1 Characteristic features of the mammalian Fox genes

Gene Name	Synonym (rodent and human only)	Chromosome (human)	Expression Domain	Phenotype
Foxa1	HNF-3 α	14q12-13	Notochord, floorplate, gut, lung, liver, pancreas	Hypoglycemia, postnatal lethality (recessive)
Foxa2	HNF-3 β	20p11	Notochord, floorplate, gut, lung, liver, pancreas	Deficient axial mesoderm, embryonic lethal (recessive)
Foxa3	HNF-3 γ	19q13.2-4	gut, liver, pancreas	Fasting hypoglycemia (recessive), male sub-fertility (dominant)
Foxb1	fkh5, mf3, TWH, hfh5	15q21-26	Central nervous system, mammary glands	Growth retardation, lactation defects.
Foxb2	fkh4		Central nervous system	
Foxc1	Fkh1, MF1, FREAC3, FKHL7	6p25	Prechondrogenic mesenchyme, periocular mesenchyme, meninges, endothelial cells, and kidney	Congenital hydrocephalus (mouse), cardiac, skeletal, ocular (recessive); cardiac, glaucoma (dominant)
Foxc2	MFH1	16q24	Sclerotome of somite, vertebrae, head mesenchyme, kidney	Ocular, cardiac, renal (recessive); lymphoedema-distichiasis (dominant), craniofacial and vertebral column defects (low penetrance)
Foxd1	FREAC4, BF2, Hfh10	5q12-13	Stromal cells of the kidney	Kidney defects
Foxd2	Mf2, FRKHDB, FREAC9	1p32-34	Kidney, somite, neural crest	Incomplete penetrance (~40%) of hydrourter, hydronephrosis, hypoplastic kidneys, and short ureters
Foxd3	Genesis, HFH2	1p31-32	ES cells, early preimplantation embryo, neural crest	Peri-implantation lethal, required for embryonic stem cells
Foxd4	FREAC5, HFH6, Fkh2	9q21	Anterior central nervous system, developing gut, red nuclei of the midbrain	
Foxe1	FKHL15, TITF2	9q22	Thyroid, craniopharyngeal ectoderm	Thyroid agenesis, cleft palate (recessive)
Foxe2	HFKH4	22q13-qter		
Foxe3	FREAC8, HFH7, dyl	1p32	Undifferentiated lens	Ocular anterior segment anomalies, cataract
Foxf1	FREAC1, HFH8	16q24	Brain and head mesenchyme	Defects in gall bladder development and lung defects (dominant); extra-embryonic and lateral plate mesoderm (recessive)
Foxf2	FREAC2, lun	6p25	Tongue, mesenchyme adjacent to the epithelium in alimentary, respiratory, and urinary tracts	Abnormal development of secondary palate (recessive)
Foxg1	BF1, HBF1	14q13	Telencephalic neuroepithelium, anterior optic vesicle	Hypoplasia of the cerebral hemispheres, eye development
Foxh1	FAST1, FAST2	8q24	Epiblast of early embryo	Lethal around gastrulation, specifies anterior primitive streak

Winged Helix Transcription Factors. Table 1 Characteristic features of the mammalian Fox genes (Continued)

Gene Name	Synonym (rodent and human only)	Chromosome (human)	Expression Domain	Phenotype
Foxi1	FREAC6, HFH5	5q35	Kidney, otic vesicle	Ear malformation, vestibular dysfunction and hearing impairment
Foxj1	HFH4	17q22-25	Ciliated cells of the respiratory tract, oviduct and ependyma	Defective ciliogenesis in airway epithelial cells and left-right axis defects
Foxj2	Fhx	12p13	Testis, ovary, early embryo, colon, nerve fibers of the GI tract	
Foxj3		1pter-q31.3	Neuroectoderm, neural crest, myotome	
Foxk1a	MNF	17q25	Muscle satellite cells	Growth retardation and severe impairment in skeletal muscle regeneration
Foxk1b		14		
Foxl1	fkh6, FREAC7	16q24	Lateral plate mesoderm, gut mesenchyme	Abnormal gut architecture
Foxl2	Pfrk	3q23	Ventral part of developing pituitary	
Foxm1	Trident, INS1, HFH11	12p13	Ubiquitous	Liver specific deletion results in defects in liver regeneration
Foxn1	WHN, whn, Foxr1	17q11-12	Thymus and skin	Severe immunodeficiency, absence of hair (recessive)
Foxn2	HTLF, Fkh19	2p16-22	Embryonic craniofacial, limb, central nervous system and somites	
Foxn3	CHES1	14q24		
Foxn4		12q24	Retina, central nervous system	Loss of amacrine and horizontal cells by retinal progenitors
Foxn6	Foxr2	Xp11.21	Breast cancers	
Foxo1	Afxh, FKHR, Fkhr1, Foxo1a	13q14	Muscle, adipose tissue and liver	Regulates cell cycle progress and apoptosis, required for fusion of primary myoblasts, adipocyte differentiation and angiogenesis; mediates insulin signalling
Foxo2	AF	6q21		
Foxo3a	FKHRL1	6q21	ubiquitous	Survival of hematopoietic progenitors and age dependent infertility
Foxo3b	FKHRL1P1	17q11	ubiquitous	
Foxo4	AFX, <u>Mlt7</u>	Xq13	ubiquitous	No obvious phenotype
Foxo6		1p34.1	Embryonic and adult brain	
Foxp1	QRF1	3p13	Lung, intestine, central nervous system, diffuse large B-cell lymphoma	Cardiac defects
Foxp2		7q31	Central nervous system, lung, intestine, stomach	Speech and language disorder (dominant)

Winged Helix Transcription Factors. Table 1 Characteristic features of the mammalian Fox genes (Continued)

Gene Name	Synonym (rodent and human only)	Chromosome (human)	Expression Domain	Phenotype
Foxp3	JM2, scurfin, scurfy, sf	Xp11.23	Thymus	Neonatal diabetes, enteropathy and endocrinopathy, abnormal T cell development, enlarged spleen and lymph nodes, aggressive lymphoproliferative autoimmunity
Foxp4		6 (p21.1)	Intestine, lung heart, brain, liver, kidney, and testis	Cardiac defects (duplication of heart)
Foxq1	HFH1, HFH1L, Satin	6 (p25)	Epidermis, hair follicle	Aberrant differentiation of hair shaft

Clinical Relevance

Cell Proliferation and Cancer

There are multiple examples of *winged helix* genes controlling proliferation of the cells in which they are expressed. *Foxg1* was shown to control proliferation of cells in the telencephalon while *Foxb1* controls the proliferation of cells in the mammillothalamic tract. *Foxm1* is expressed highly in cycling cells, is phosphorylated during mitosis and a liver-specific deletion of *Foxm1* revealed that this gene is critical for regulating expression of cell-cycle genes required for hepatocyte proliferation. Normal control of cell-cycle exit and entry involves protein kinase B-regulated *forkhead* transcription factors. Two *Fox* genes have been linked to stem cell function; *Foxk1* is expressed in muscle satellite cells and is required for muscle regeneration while *Foxd3* is expressed in embryonic stem cells and is required for their establishment and maintenance *in vitro*.

The misregulation of some *Fox* genes results in abnormal cell proliferation and disease. Chromosomal translocation of several different *Fox* genes can cause aberrant cell growth resulting in tumors. Alveolar rhabdomyosarcomas can be caused *via* the fusion of either the *PAX3* or *PAX7* DNA binding domain with the transactivation domain of *FOXO1* and acute lymphoid leukemia can be caused by the fusion of the DNA binding domain of MLL with the transactivation domain of *FOXO3* and *FOXO4* (among others). Ectopic expression of *Fox* genes is correlated with some cancer tumors. *FOXA1* is overexpressed in esophageal and lung adenocarcinomas and *FOXD3* expression is up-regulated in a number of cancer cell lines. In addition, *FOXMI*, noted above in hepatocyte proliferation, is also a downstream target of Gli1 in basal cell carcinomas.

Human Disease and Behavior

The *Fox* genes have been shown to play important roles in a diverse range of developmental processes and human

diseases, which are summarized in Table 1. Space does not permit a detailed description of all of them, however, the case of *FOXP2* provides a particularly striking example. In a large three-generation pedigree, a severe deficit in language processing and grammatical skills was found to be inherited as an autosomal dominant trait. This disease was traced to a mutation in the *forkhead* gene *FOXP2* and is probably the result of subtle deficits in neural or intellectual development (9). *FOXP2* represents the first gene to be shown to be involved in speech and neural development. Further analysis revealed a difference of just two amino acids between the human and primate *FOXP2* sequences. One of these changes generates a new phosphorylation site that was fixed in the human genome approximately 200,000 years ago, i.e. close to the time that anatomically modern humans appeared. This suggests the intriguing possibility that this change in the ancestral *FOXP2* gene contributed to the alterations in neural development that allowed for the acquisition of speech during human evolution.

Several *Fox* genes, *FOXC1*, *FOXC2*, *FOXE3* and *FOXL2*, have been implicated in eye development with mutations causing varied defects in the eye chamber, the eyelids and the eyelashes (4). Some human *Fox* genes are also linked to immune deficiencies, *FOXE1*, *FOXN1* and *FOXP3*. In summary, the *Fox* genes control a remarkable spectrum of developmental processes and contribute to the whole range of human diseases.

► [Wnt/β-catenin Signaling Pathway](#)

► [Hedgehog Signalling](#)

References

1. Weigel D, Jürgens G, Küttner F et al (1989) The homeotic gene fork head encodes a nuclear protein and is expressed in the terminal regions of the *Drosophila* embryo. *Cell* 57:645–65

2. Weigel D, Jackle H (1990) The fork head domain: a novel DNA binding motif of eukaryotic transcription factors? *Cell* 63:455–456
3. Kaestner KH, Knöchel W, Martinez DE (2000) Unified nomenclature for the winged helix/forkhead transcription factors. *Genes Dev* 14:142–146
4. Lehmann OJ, Sowden JC, Carlsson P et al (2003) Fox's in development and disease. *Trends Genet* 19:339–344
5. Ogg S, Paradis S, Gottlieb S et al (1997) The Fork head transcription factor DAF-16 transduces insulin-like metabolic and longevity signals in *C. elegans*. *Nature* 389:994–999
6. Clark KL, Halay ED, Lai E et al (1993) Co-crystal structure of the HNF-3/fork head DNA-recognition motif resembles histone H5. *Nature* 364:412–420
7. Ang SL, Rossant J (1994) HNF-3 beta is essential for node and notochord formation in mouse development. *Cell* 78:561–574
8. Weinstein DC, Ruiz i Altaba A, Chen WS et al (1994) The winged-helix transcription factor HNF-3 beta is required for notochord development in the mouse embryo. *Cell* 78:575–588
9. Lai CS, Fisher SE, Hurst JA et al (2001) A forkhead-domain gene is mutated in a severe speech and language disorder. *Nature* 413:519–23

WISH

- Whole Mount ISH

Wiskott Aldrich Syndrome Protein

- WASp

Withdrawal

Definition

After cessation of drug use (withdrawal may occur as early as a few hours after the last administration) the addict usually undergoes withdrawal, which is often associated with very aversive physical symptoms (major symptoms peak between 48 and 72 hours after the last dose and subside after about a week). Most importantly, after the physical withdrawal symptoms have disappeared and the drug addict is detoxified, he/she is still an addict.

- Addiction, Molecular Biology

WNK

Definition

With no lysine kinases (WNKs) are serine/threonine protein kinases involved in regulating human blood pressure and electrolyte balance. Mutations in two members of the WNK family, WNK1 and WNK4, cause familial hyperkalemic hypertension.

- Receptor Serine/Threonine Kinases

Wnt/ β -catenin Signaling Pathway

AJAMETE KAYKAS, RANDALL T. MOON

Department of Pharmacology, Howard Hughes Medical Institute, University of Washington School of Medicine, Seattle, WA, USA
rtmoon@u.washington.edu

Definition

Wnts are secreted ligands that activate multiple signal transduction pathways. The most rigorously studied ► Wnt pathway is initiated by Wnt-mediated activation of its cognate cell surface receptors, which induces a core set of cytosolic signaling molecules leading to the stabilization of ► β -catenin. Stabilized β -catenin participates in a bipartite transcription factor complex with the ► TCF family of transcription factors. This bipartite transcription factor activates a set of specific genes in the nucleus of target cells. A large base of genetic evidence from *Hydra* to man supports the interactions and regulations of the multiple proteins that transduce the Wnt signal. Further, the plethora of molecular interactions of the pathway are supported by cell biological and biochemical experiments.

The Wnt/ β -catenin signaling pathway is one of a few central signaling pathways that regulate many aspects of embryonic development. The major cellular outcomes of activation of the Wnt/ β -catenin pathway are altered gene expression, which in turn modulates cell fate, proliferation and apoptosis. Wnt/ β -catenin signaling plays a role in cardiac differentiation, neural patterning and development of multiple other tissues such as kidney and the female reproductive tract. The pathway also plays a role in ► adipogenesis, ► hematopoiesis and skin formation, by maintaining and inducing differentiation of stem cells. Aberrant regulation of the Wnt/ β -catenin pathway is implicated in multiple human diseases and cancers. Hence, the

Wnt/ β -catenin pathway is medically important and a major target for drug development.

Characteristics

The Wnt/ β -catenin Pathway Functions in Development of Multiple Cell Types and Tissues

It is well established that Wnt/ β -catenin signaling plays a role in dorsal ventral specification in early *Xenopus* embryos. The pathway has subsequently been shown to play a role in formation of multiple tissues and cell types. The Wnt/ β -catenin pathway suppresses cardiac development and mice that have been engineered to conditionally delete β -catenin develop multiple hearts. Multiple Wnts and their effector signaling proteins are expressed throughout the developing nervous system, where they control many aspects of neural development. In support of this role, mice deficient for Wnt-1 fail to develop a midbrain and cerebellum and are deficient in neural crest derivatives. Wnts are also important for formation of other organs such as kidney. Wnt-4 is required for the development of kidney and mice deficient for it fail to have correct kidney tubule formation. Wnts are also important in the development of the female reproductive organs. Mice that are deficient in WNT-7a are infertile because of abnormal development of their oviduct and uterus.

The Wnt/ β -catenin pathway is involved in maintaining and specifying multiple types of stem cells and subsequently plays a role in adipogenesis, hematopoiesis and skin cell formation. The Wnt signaling pathway maintains preadipocytes in an undifferentiated state by inhibition of adipogenic transcription factors. Inhibition of Wnt signaling in preadipocytes by expression of axin or dominant negative TCF leads to formation of adipocytes. Wnts also regulate hematopoiesis. Wnts are expressed in very specific domains in the embryonic mesoderm and act as hematopoietic growth factors. In addition, TCFs are required for T and B-lymphocyte development in the mouse. Components of the Wnt/ β -catenin signaling pathway control specification of skin cell fates into keratinocytes, sebaceous glands and hair follicles. Inhibition of signaling leads to epidermal cell fates, while activation of signaling leads to formation of hair cells. Wnts are probably involved in maintaining many types of stem cells that have yet to be appreciated.

The Wnt/ β -catenin Pathway is Conserved in Evolution

One of the strengths in support of the models of the Wnt/ β -catenin pathway is its conservation in metazoans. Its conservation has ultimately allowed rigorous genetic studies to be conducted in *Drosophila*, *C. elegans*, zebrafish and mouse (5). All of these organisms have a core Wnt/ β -catenin signaling pathway made up of Wnts, \blacktriangleright Frizzleds, \blacktriangleright Dishevelled,

\blacktriangleright GSK3, β -catenin and TCFs. In addition, the general function of the pathway is conserved as it regulates similar phenomena in all of these organisms. Simple metazoans such as *Hydra* also have an intact Wnt/ β -catenin pathway, which has all of the core components of the pathway and is involved in specifying the formation of the head and foot of the organism. There is no evidence that yeast or bacteria have a Wnt/ β -catenin signaling pathway. However, the pathway has been reconstituted in yeast. Even organisms as evolutionarily primitive as the slime mold *Dictyostelium* have some of the components of the Wnt/ β -catenin pathway. Specifically, *Dictyostelium* has receptors similar to Frizzleds, \blacktriangleright CARs, which activate a GSK3 homologue. The Wnt/ β -catenin pathway is similar in many ways to the \blacktriangleright hedgehog (HH)/ \blacktriangleright Gli signaling pathway and many components of the pathways are shared (4). Both Wnts and HH are secreted lipid-modified ligands that bind and activate cell surface receptors. \blacktriangleright Smoothed (Smo), a receptor component of the HH pathway is the closest known relative to Frizzled. While Smoothed does not bind Wnt nor has its ligand been identified, Smo shares the overall structure of Frizzled, even the domain that is known to bind Wnt. In addition, both Frizzled and Smoothed have structures that are reminiscent of \blacktriangleright G-protein coupled receptors (GPCRs). However, they share little sequence homology with any other families of GPCRs. Another similarity lies in \blacktriangleright megalyn, a component of the HH signaling complex, which shares homology to \blacktriangleright LRP5/6. In addition, both the HH and Wnt pathways use GSK3 as a core component of their respective pathways. In *Drosophila* Shaggy, the GSK3 orthologue inhibits HH signaling by phosphorylating Cubitus interruptus (Ci), to generate a truncated repressor, which inhibits HH's induction of transcription.

Aberrant Regulation of the Wnt/ β -Catenin Pathway Leads to Human Disease

Wnt/ β -catenin is very important for embryonic development and when mutated leads to multiple diseases. Wnt-1 was first identified in vertebrates because it was transactivated by mouse mammary tumor virus, which integrated near it and induced ectopic expression leading to mammary tumorigenesis. Other components of the pathway are linked to cancer. \blacktriangleright APC, which is an inhibitor of β -catenin, is frequently mutated in colon cancer (6). The mutant form of APC does not bind β -catenin or participate in the " \blacktriangleright destruction complex". Hence, β -catenin is stabilized and can transactivate genes that lead to cell proliferation. \blacktriangleright Axin is another component of the β -catenin "destruction complex" that is also mutated in human cancers such as \blacktriangleright HCC; again the mutation in axin leads to β -catenin stabilization and activation of transcription. β -catenin is itself found to be mutant in some forms of HCC (6). The mutants of

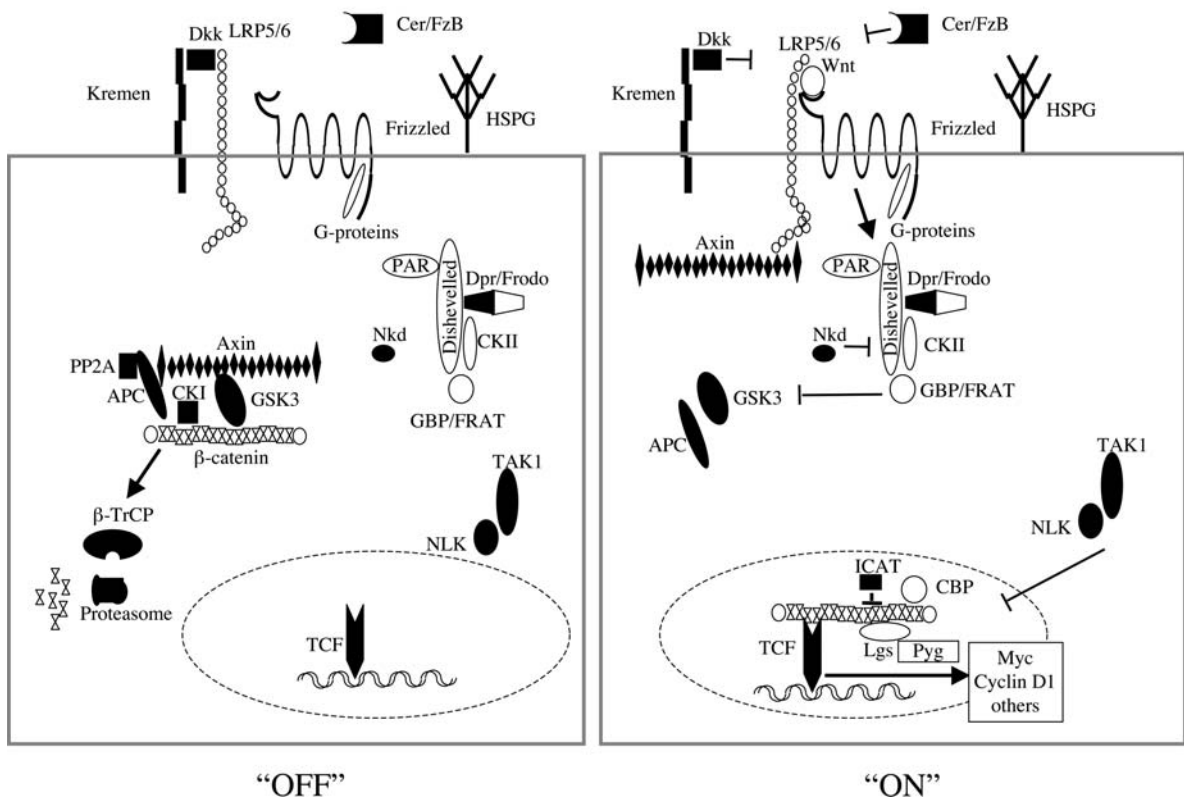
β -catenin do not interact with the “destruction complex”, are stabilized and activate genes that promote cell growth.

In addition to cancer, the Wnt/ β -catenin pathway is implicated in many other diseases. Many of the components of the pathway such as LRP5 and GSK3 are known to be involved in the pathogenesis of [▶ Alzheimer’s disease](#). Although definitive proof is yet to show a direct link between aberrant regulation of the Wnt pathway and Alzheimer’s disease there are many striking correlations (1). Interestingly, the Wnt/ β -catenin pathway is also involved in bone formation. Loss-of-function mutations in the gene LRP5 have been shown to cause osteoporosis-pseudoglioma or low bone mass (2). Conversely, gain of function mutations in LRP5 causes high bone mass traits in humans. The Wnt/ β -catenin pathway is also implicated in [▶ angiogenesis](#). Frizzled-4, one of the human receptors for Wnts, is mutated in FEVR (7), a rare dominant disease of the retina, which is due to defects in angiogenesis. However, the full implication of this and the role of the Wnt/ β -catenin pathway in angiogenesis have not been fully explored.

Molecular Interactions

The Wnt/ β -catenin pathway is more of a web than a linear pathway. Some of the relevant proteins that make up the signaling pathway are shown in Fig. 1. Many additional proteins and contacts are not shown. In addition, many of the proteins in the pathway make contacts with proteins that are in different signaling pathways. Therefore, there is much scope for cross talk between the Wnt/ β -catenin pathway and other signaling pathways, adding a complex layer of contacts and modes of regulation. A more detailed up to date description of the pathway can be found at STKE (3). The Wnt/ β -catenin pathway is initiated by Wnt binding to the serpentine Frizzled receptor and forming a ternary complex with LRP5/6 at the plasma membrane. Wnt also makes contacts with [▶ secreted Frizzled related proteins \(FzBs\)](#), [▶ Cerberus \(Cer\)](#), and [▶ heparan sulfate proteoglycans \(HSPG\)](#), like syndecan-1 or Knypek. LRP6 binds to [▶ Dickkopf \(Dkk\)](#), which interacts with [▶ Kremen](#). The activated Frizzled-LRP6 complex activates Dishevelled (Dvl) possibly *via* G-proteins. Dvl interacts with multiple proteins, such as [▶ Dapper \(Dpr\) /Frodo](#), [▶ Naked Cuticle \(Nkd\)](#) and

Wnt/ β -catenin Signaling in the Absence and Presence of Wnt



Wnt/ β -catenin Signaling Pathway. Figure 1

►**PAR** that regulate its function. Activated Dvl inhibits GSK3 possibly *via* its interaction with ►**GBP/FRAT**. GSK3 is found in the β -catenin “destruction complex” with the scaffolding proteins axin and APC. β -catenin interacts with ► **β -TrCP** and is degraded by the proteasome. In the nucleus β -catenin interacts with ►**Legless (Lgs)**, ►**Pygopus (Pgy)**, ►**ICAT**, and ►**CBP**, which regulate its function. TCF interacts with and is regulated by ►**NLK**, which forms a complex with ►**TAK1**.

Regulatory Mechanisms

Regulation at Membrane and in the Extracellular Lumen

Wnts are secreted outside cells and their access to their receptors is regulated by FzBs, which resemble the Wnt binding domain of Frizzled. In addition, Cerberus also binds and inhibits Wnts. Both of these inhibitors bind and sequester Wnts from binding Frizzled. The access of Wnts to the co-receptors LRP5/6 is also regulated. The presentation of LRP6 to Wnts at the plasma membrane is inhibited by Dkk and Kremen, which cause LRP6 to be internalized and cleared from the plasma membrane.

Regulation of Disheveled in the Cytosol

Disheveled is a modular molecule that functions downstream of Frizzled in the Wnt/ β -catenin pathway. Activation of Dvl is central to activation of signaling and therefore Dvl activity is regulated by multiple interacting proteins (8). Although it is not known how Frizzled activates Dvl, its phosphorylation status correlates with its activation. The kinases PAR-1 and ►**casein kinase II (CKII)** are known to participate in phosphorylating Dvl and regulate its subsequent activation. Dvl is also regulated by the Dpr/Frigo family of proteins, which interact with Dvl and can both activate and inhibit its activity. The EF domain containing protein, Nkd, is induced by Wnt and participates in a feed back loop to dampen the activity of Dvl.

Regulation of β -Catenin Stability

Activated Disheveled inhibits GSK3 from phosphorylating β -catenin, possibly by interaction with GBP/FRAT. GSK3 is found in a complex of proteins required for phosphorylation of β -catenin, which include axin and APC. This complex of proteins that regulates the destruction of β -catenin is referred to as the “destruction complex”. β -catenin phosphorylation is also regulated by ►**casein kinase I (CKI)** and the phosphatase ►**PP2A**. Phosphorylated β -catenin is recognized by the ubiquitin ligase β -TrCP and targeted for degradation by the proteasome.

Regulation of Transcription in the Nucleus

Once in the nucleus β -catenin forms a bipartite transcription complex with the TCF family of transcription factors. β -catenin activation of transcription is facilitated by interaction with the co-activator CBP. In addition β -catenin’s interaction with Legless, which recruits Pygopus, is required for its activation of genes in the nucleus. β -catenin’s transcriptional activation is antagonized by its interaction with ICAT, which competes for binding TCF.

References

1. De Ferrari GV, Inestrosa NC (2000) Wnt signaling function in Alzheimer’s disease. *Brain Res Brain Res Rev* 33:1–12
2. Gong Y, Slee RB et al (2001) LDL receptor-related protein 5 (LRP5) affects bone accrual and eye development. *Cell* 107:513–23
3. <http://stke.sciencemag.org>
4. Kalderon D (2002) Similarities between the Hedgehog and Wnt signaling pathways. *Trends Cell Biol* 12:523–531
5. Moon RT, Bowerman B et al (2002) The promise and perils of Wnt signaling through beta-catenin. *Science* 296:1644–1646
6. Polakis P (2000) Wnt signaling and cancer. *Genes Dev* 14:1837–1851
7. Robitaille J, Mac Donald ML et al (2002) Mutant frizzled-4 disrupts retinal angiogenesis in familial exudative vitreoretinopathy. *Nat Genet* 32:326–330
8. Wharton KA (2003) ‘Runnin’ with the dvl: proteins that associate with dsh/dvl and their significance to Wnt signal transduction. *Dev Biol* 253:1–17

Wnt/Wg

Definition

Wnt/Wg are components of a signaling pathway in vertebrates (Wnt) and in drosophila (Wg, wingless) that are important in regulating pattern development during embryogenesis. In vertebrates, Wnts include at least 18 members. Not all Wnts activate the Wnt/ β -catenin pathway.

►**Axis Formation – Formation and Function of the Dorsal Organizer**

►**Desmosomes**

►**Wnt/Beta-Catenin Signaling Pathway**

Wolffian Duct

Definition

Wolffian duct refers to a duct that originates from the mesonephros, which gives rise to the ureteric bud of the

metanephric kidney. The duct persists in males to contribute to parts of the gonad, but regresses in females.

► [Kidney](#)

Wolf-Hirschhorn Syndrome

Definition

Wolf-Hirschhorn syndrome is a widely known disorder (1:30,000 newborns) that is caused by a deletion at chromosome 4p16.3, and is characterized by severe pre- and postnatal growth retardation, facial abnormalities, heart defects, severe mental deficiency, and high lethality in infancy. Expression is variable and the mild forms, which have been referred to as Pitt-Rogers-Danks syndrome, allow survival into adulthood.

► [Microdeletion Syndromes](#)

Wound Healing

SUSANNE BRAUN, SABINE WERNER

Department of Biology, Institute of Cell Biology, ETH Zurich, Zurich, Switzerland

sabine.werner@cell.biol.ethz.ch

Definition

Wound healing is a complex and dynamic process aimed at restoring the integrity of injured skin. This repair process involves blood clotting, inflammation, new tissue formation and tissue remodeling. Wounds can be of variable size and depth. Partial-thickness wounds are characterized by a loss of skin that is limited to the ► [epidermis](#) and potentially parts of the ► [dermis](#). These wounds heal predominantly by reepithelialization (proliferation and migration of epithelial cells) and the injured tissue fully regenerates if the wound is restricted to the epidermis. Full-thickness wounds involve damage to the epidermis, dermis and subcutaneous tissue. These wounds heal by granulation (replacement of the lost dermal tissue by new cells and matrix), contraction and reepithelialization. The final result in adult mammals is not perfectly regenerated skin but a ► [scar](#).

Characteristics

Most wounds affect both layers of the skin, the epidermis and the dermis. Immediately after lesion

the repair process is initiated and proceeds in three phases – inflammation, new tissue formation, remodeling – that overlap in time (Fig. 1).

Inflammation

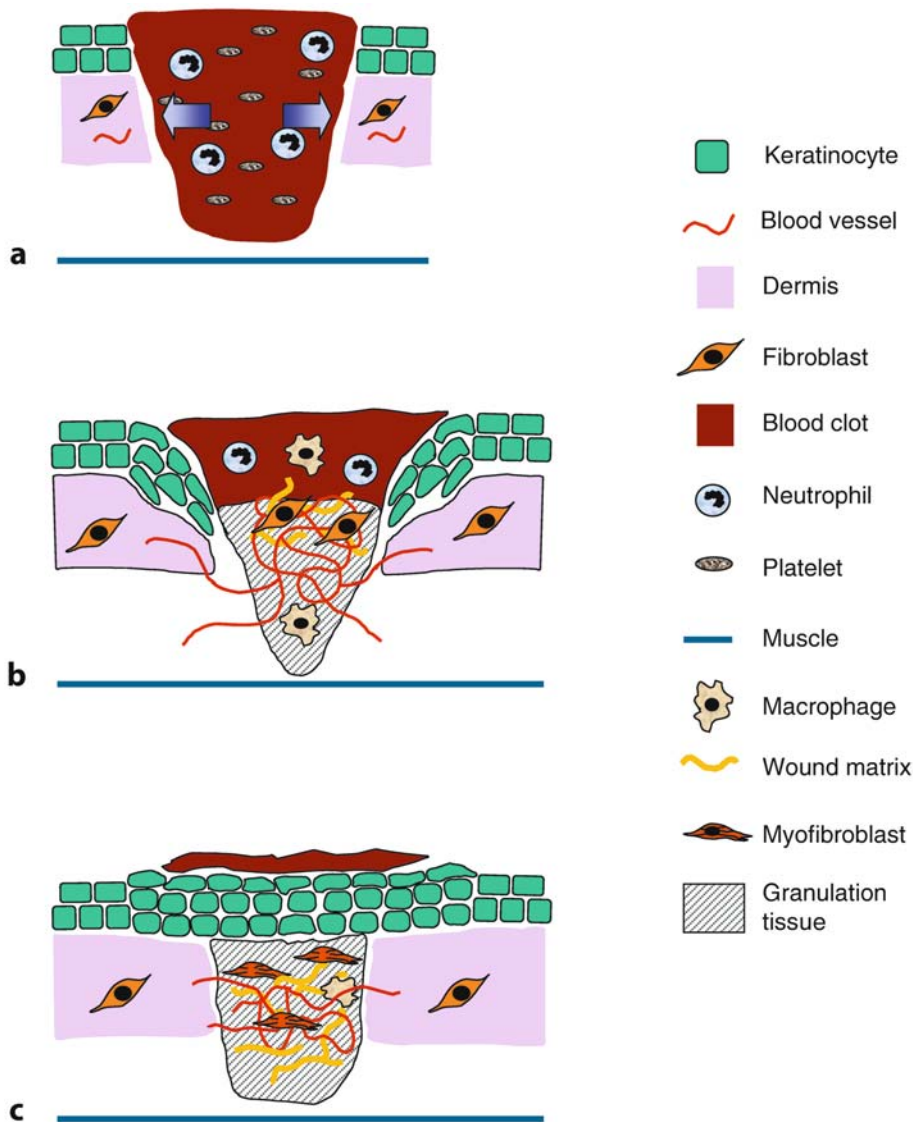
Upon skin injury, blood vessels are disrupted and blood constituents extravasate. This leads to the formation of the blood clot, which is composed of cross-linked fibrin and of ► [extracellular matrix](#) (ECM) proteins such as fibronectin, vitronectin and thrombospondin. The blood clot plugs the wound, provides a barrier against microorganisms and serves as a provisional matrix for invading cells. Furthermore, the repair process is initiated by the release of various ► [growth factors](#) and ► [cytokines](#) from the serum and degranulating platelets. Among them are chemotactic factors, which attract ► [neutrophils](#), lymphocytes and monocytes to the wound site. These inflammatory cells in turn produce and secrete reactive oxygen species and proteinases that play an important role in the defence against contaminating bacteria.

New Tissue Formation

In addition to their role in inflammation, neutrophils and ► [macrophages](#) are also an important source of growth factors and cytokines that initiate the second phase of wound repair, the formation of new tissue. ► [Keratinocytes](#) at the wound margin are stimulated to migrate and to proliferate, leading to reepithelialization of the wound. To restore the injured dermis, the formation of new stroma, called ► [granulation tissue](#), is initiated a few days after injury. ► [Fibroblasts](#) at the wound edge start to proliferate and migrate into the wound space, where they deposit large amounts of extracellular matrix. At later stages of wound repair, a proportion of the fibroblasts transform into myofibroblasts, which can generate contractile forces, leading to wound contraction. ► [Endothelial cells](#) migrate into the wound space where they proliferate and form new capillaries. Nerve sprouting occurs at the wound edge, leading to innervation of the new tissue.

Remodeling

Finally, the transition from granulation tissue to mature scar tissue takes place during the remodeling phase. This process is characterized by reduction of the cellularity within the granulation tissue due to apoptosis, continued collagen synthesis and degradation at a low rate and formation of larger and more cross-linked collagen bundles. The wound gains more tensile strength but never reaches that of uninjured skin. Furthermore, scar tissue lacks appendages, including hair follicles, sebaceous glands and sweat glands. Scarring can also be excessive, leading to hypertrophic scars and keloids. Interestingly, wound healing in



Wound Healing. Figure 1 Schematic representation of different stages of wound repair. (a) 12–24 hours after injury the wounded area is filled with a blood clot. Neutrophils have invaded into the clot. (b) At days 3–7 after injury, the majority of neutrophils have undergone apoptosis. Instead, macrophages are abundant in the wound tissue at this stage of repair. Endothelial cells migrate into the clot; they proliferate and form new blood vessels. Fibroblasts migrate into the wound tissue, where they proliferate and deposit extracellular matrix. The new tissue is called granulation tissue. Keratinocytes proliferate at the wound edge and migrate down the injured dermis and above the provisional matrix. (c) 1–2 weeks after injury the wound is completely filled with granulation tissue. Fibroblasts have transformed into myofibroblasts, leading to wound contraction and collagen deposition. The wound is completely covered with a neoepidermis. Reprinted with permission from Ref. 2, copyright American Physiological Society.

mammalian embryos up to the beginning of the third trimester occurs without scarring.

Molecular Interactions

Wound healing is a highly orchestrated process that is largely dependent on a network of interactions among different cells. In cutaneous wound repair, keratinocytes, fibroblasts, endothelial cells and inflammatory cells

communicate with each other through cell-cell and cell-matrix contacts and through cytokines and growth factors. These signals are interpreted by the target cells within the context of their surrounding extracellular matrix. Thus, wound repair requires continuous interactions among cells, cytokines and matrix. In the following part, a few examples are chosen to illustrate basic principles of these interactions.

Growth Factors and Cytokines

During wound healing large amounts of growth factors and cytokines are released from different sources. The first wave of these factors reaches the wound immediately after injury and originates from the serum of injured blood vessels and from degranulating platelets. These growth factors and cytokines attract inflammatory cells that in turn secrete new factors. In addition, they act on resident cells at the wound margin and initiate many different cellular processes, e.g. proliferation, migration, differentiation as well as production of extracellular matrix, proteinases and additional cytokines.

As an example of a growth factor involved in wound reepithelialization, the role of keratinocyte growth factor (KGF) will be described in more detail. KGF belongs to the fibroblast growth factor (FGF) family and was therefore designated FGF7. In contrast to other FGF family members, which are broad-spectrum mitogens, KGF seems to act specifically on epithelial cells. In addition to its mitogenic effect, it regulates migration and differentiation of its target cells and protects them from cell death under stress conditions. Upon skin injury KGF expression is strongly induced in dermal fibroblasts and in intraepidermal $\gamma\delta$ T cells, whereas the receptor is only produced by keratinocytes. These results strongly suggest that dermally derived KGF stimulates wound reepithelialization in a **▶paracrine** manner. But what are the factors that can induce KGF expression? Local hemorrhage causes extravasation of thrombocytes that subsequently release, among other growth factors, platelet-derived growth factor and epidermal growth factor. The latter are potent stimulators of KGF expression in fibroblasts. In addition, growth factors and cytokines, in particular platelet-derived growth factor, transforming growth factor α , tumor necrosis factor α and interleukin-1, derived from inflammatory cells can enhance KGF expression in fibroblasts.

Another process that is strongly influenced by growth factors is **▶angiogenesis**, the generation of new capillaries from the pre-existing vasculature. Angiogenesis is especially important for wound healing, since a rich blood supply is vital to sustain newly formed tissue. Of particular importance for wound angiogenesis are FGF2 and members of the vascular endothelial growth factor (VEGF) family. After cutaneous injury, the expression of VEGF-A is strongly upregulated in keratinocytes and macrophages at the wound site. This factor acts in a paracrine manner on endothelial cells, which express the appropriate receptors. The impact of VEGF-A on angiogenesis and wound repair has been shown in several studies. For example, treatment of pig wounds with neutralizing VEGF-A antibodies caused a severe reduction in

wound angiogenesis and granulation tissue formation. Similar defects were observed when a mutant VEGF receptor, which blocks VEGF-A signaling, was over-expressed in mouse skin by viral vector delivery.

Cell-Matrix Interactions

The extracellular matrix (ECM) is comprised of a large number of components, including different collagens, fibronectin and proteoglycans. These molecules form a network that imposes structure and is involved in cell regulatory functions. A number of ECM receptors that mediate cell-matrix interactions are expressed on cell surfaces. Among them, members of the integrin family are important regulators of wound healing. Integrins are heterodimeric transmembrane proteins consisting of an α and a β subunit. Many different α and β subunits are known, and the combination of the various subunits defines the ECM proteins that are bound. In the wound repair process, fibroblast integrins appear to be involved in wound contraction, keratinocyte integrins in reepithelialization, and endothelial cell integrins in angiogenesis.

In the intact epidermis, the major integrins are $\alpha2\beta1$, $\alpha3\beta1$, $\alpha9\beta1$ and $\alpha6\beta4$ and their expression is strictly confined to the basal layer of keratinocytes. After wounding, keratinocytes at the wound margin change their integrin expression pattern, characterized by suprabasal integrin expression and induction of $\alpha5\beta1$, $\alpha v\beta5$, $\alpha v\beta6$ integrins. This set of integrins is necessary for keratinocyte migration, and thus for reepithelialization, because they can recognize proteins of the dermal and provisional matrix.

Integrin $\alpha v\beta3$, the receptor for fibrin and fibronectin, is expressed on blood vessels invading the blood clot and wound granulation tissue, but not in normal skin. Recent investigations have shown that the wound ECM can regulate angiogenesis in part by modulating the cells' integrin receptor expression.

The change of the integrin expression profile is also important for the behavior of fibroblasts. Early after wounding, these cells acquire a migratory phenotype. Like migrating keratinocytes, the expression of a specific set of integrins is necessary for fibroblasts to bind to ECM molecules of the provisional matrix. At later stages of healing, fibroblasts begin to differentiate into myofibroblasts, which lead to wound contraction. This process depends strongly on fibroblast attachment to the collagen matrix through the $\alpha2\beta1$ integrin receptors.

Cell-Cell Interactions

The formation of surface epithelium to close the wound is precisely coordinated with the repair of the underlying dermis. This synchrony is the key to preventing

either insufficient or excessive wound repair. To achieve this aim, communication between the two layers, epidermis and dermis, is very important. For example, keratinocyte proliferation and differentiation may be regulated by a double paracrine loop of cytokines as suggested by *in vitro* studies. After wounding, keratinocytes of the epidermis synthesize and secrete interleukin 1 α and interleukin 1 β that act in a paracrine manner on the fibroblasts in the dermis. In response to this stimulus, fibroblasts release KGF, granulocyte-macrophage colony stimulating factor and additional, yet unknown factors. In a second paracrine action these cytokines stimulate keratinocyte proliferation and differentiation.

Regulatory Mechanisms

Cutaneous wound repair is well described at the histological level but the genes that orchestrate this process have been only partially defined. Recently, several groups have identified genes that are up- or down-regulated after skin injury. These genes are likely to have a functional role in the repair process, but this has only been proven for a few selected genes. During the past few years, the development of technologies for the generation of genetically modified mice has allowed elucidation of the function of various genes in the healing process. These technologies allow gain of function experiments (over-expression of genes) as well as loss of function studies (e.g. gene knockouts). Most importantly, spatial and temporal control of gene ablation or over-expression, using more specialized technologies (inducible and/or tissue-specific over-expression or ablation) makes it possible to determine the functions of genes formerly precluded due to embryonic lethality. A large number of viable genetically modified animals are now available that can be used to elucidate the role of the deleted, mutated or over-expressed genes in the repair of different tissues and organs. The first example of a genetic approach to the study of wound repair was the analysis of KGF receptor function in wound healing. ► **Transgenic mice** were generated that express a dominant negative KGF receptor mutant specifically in the epidermis. Therefore, keratinocytes are unable to respond to the KGF stimulus. These mice have severely impaired healing,

with reduced proliferation of wound edge keratinocytes and substantially delayed reepithelialization. Thus, this wound healing study revealed that KGF receptor signaling is essential for wound reepithelialization. The last few years have seen an exponential growth in the number of ► **transgenic and knockout mice** used for wound healing studies, and these experiments have provided exciting, and often unexpected, results concerning the function of genes in the wound healing process *in vivo* (see <http://icbxs.ethz.ch/members/grose/woundtransgenic/home.html>).

References

1. Clark RAF (1996) Wound repair. Overview and general considerations. In: Clark RAF (ed) *The Molecular and Cellular Biology of Wound Repair*, 2nd ed. Plenum Press, New York, pp 3–50
2. Grose R, Werner S (2003) Regulation of wound healing by growth factors and cytokines. *Physiol Rev* 83:835–870
3. Martin P (1997) Wound healing – aiming for perfect skin regeneration. *Science* 276:75–81

WSC Domain

Definition

The term WSC domain comes from *S. cerevisiae* WSC1 (cell-wall integrity and stress-response component 1). WSC domains contain conserved cysteine residues and might be involved in carbohydrate binding.

► **Polycystic Kidney Disease, Autosomal Dominant**

WT

► **Wild Type**

X Chromosome

Definition

Women carry two X while men carry an X and a Y chromosome. X and Y collectively are called sex chromosomes.

- ▶ Fragile X Syndrome
- ▶ Repeat Expansion Diseases

Type III hyperlipoproteinemia and some forms of apoA-I deficiency.

- ▶ HDL-Dyslipidemia

XALD

- ▶ X-Linked Adrenoleukodystrophy

Xanthelasmata

Definition

Xanthelasmata are tiny (1–2 mm) yellowish plaques that are slightly raised on the skin's surface of the upper or lower eyelids. They are a distinguished form of xanthomas by their specific localization.

- ▶ Familial Hypercholesterolemia

Xanthophore

Definition

Xanthophore is a neural crest-derived cell that is pigmented yellow.

- ▶ Mutagenesis Approaches in Medaka

X-Chromosome Inactivation

BRYAN K. SUN, JEANNIE T. LEE

Howard Hughes Medical Institute, Department of Molecular Biology, Massachusetts General Hospital, Department of Genetics, Harvard Medical School, Boston, MA, USA

bryansun@post.harvard.edu

lee@molbio.mgh.harvard.edu

Synonyms

Lyonization; X-chromosome silencing; Sex chromosome silencing

Xanthoma

Definition

Xanthoma originates from a lesion of foam cell (cholesterol) accumulation in skin or tendon. Skin xanthomas in eyelids are called xanthelasma. Xanthoma is seen in familial hypercholesterolemia, cerebrotendinous xanthomatosis sitosterolemia (phytosterolemia),

Definitions

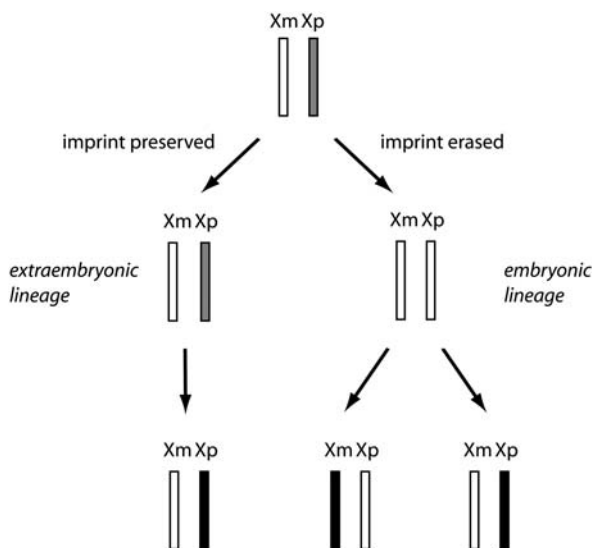
X-chromosome inactivation (XCI) is the process by which one of two X-chromosomes is transcriptionally silenced in female mammals, equalizing the dose of X-linked genes in females (XX) and males (XY). XCI takes place early in embryonic development when each cell chooses to inactivate one of the X-chromosomes. After the choice has been made between the two, all daughter cells maintain the same inactivated X (Xi). As a result, all females are ▶ **mosaics**, composed of two

populations of cells that express the genes from one or the other X-chromosome. The process of X-chromosome inactivation is unique to mammals, but is only one of several known mechanisms of [dosage compensation](#) found in organisms such as the mouse, fruit fly and nematode worm. One intriguing characteristic that is common to dosage compensation in mammals and fruit flies is the essential role of [noncoding RNAs](#) in the process. From a medical perspective, the mechanism and pattern of XCI has important effects on the phenotype of X-chromosome [aneuploidies](#) and X-linked genetic diseases. Emerging evidence suggests that XCI also has implications in cancer and other areas of medicine as well.

Characteristics

Forms of X-Chromosome Inactivation

Two forms of X-chromosome inactivation have been characterized in mammals, imprinted X-inactivation and random X-inactivation (Fig. 1). From an evolutionary standpoint, the ancestral form was probably imprinted X-inactivation, in which the father's X-chromosome is always silenced and the mother's X-chromosome remains active. To distinguish the two



X-Chromosome Inactivation. Figure 1 Imprinted and random X-inactivation. After early cleavage stages, cells of the zygote separate into extraembryonic and embryonic lineages. Cells of the extraembryonic lineage display imprinted X-inactivation. A parental imprint protects the maternal X chromosome (Xm) from inactivation and/or designates the paternal X (Xp) for silencing. In cells of the embryonic lineage, the imprint is erased, allowing each female cell to randomly inactivate either the maternal X (Xm) or paternal X (Xp). Transcriptionally active chromosomes are depicted in white, transcriptionally inactive chromosomes in black.

X- chromosomes from each other, a parental “imprint” provides a mark that assures the maternal X is protected and the paternal X becomes designated for silencing. This type of XCI is observed today in marsupials, in the extraembryonic tissues of eutherian mammals such as the mouse, and possibly the human as well.

By contrast, the embryonic cells of eutherian mammals inactivate either the maternal or paternal X-chromosome and the choice between the two is random. This form is known as random X-inactivation, and occurs when the parental imprint is erased around the time of implantation, equalizing the chance for the cell to choose either the maternal or paternal X to be active. The [stochastic](#) choice between the two chromosomes suggests that each female cell has a mechanism that senses the presence of an extra X-chromosome and is able to precisely designate one for inactivation while keeping the other one activated.

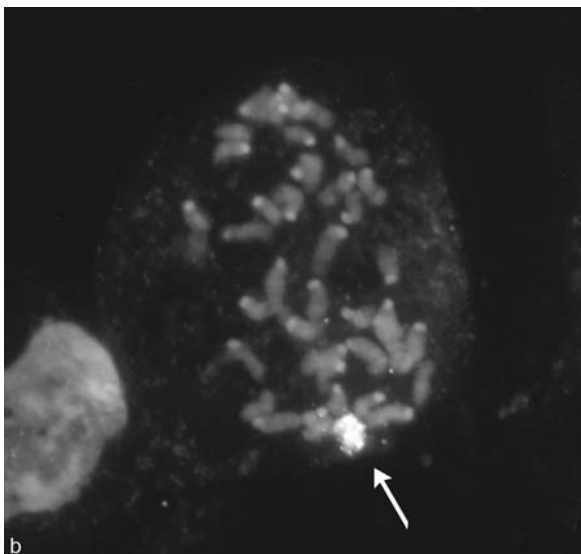
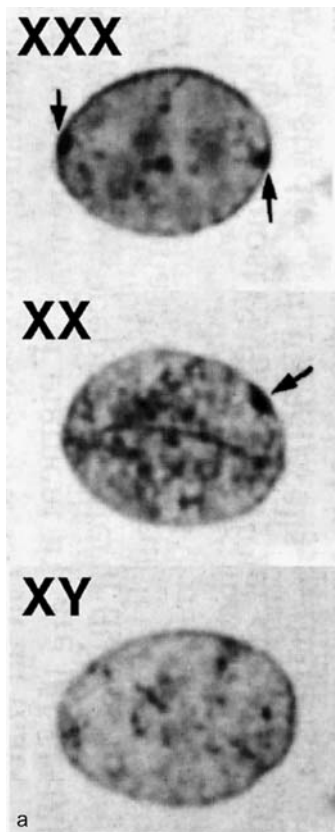
Since most human disease involves somatic cells and tissue, this form of X-inactivation is the most relevant mechanism when considering the impact of XCI on health and disease processes. Imprinted XCI, however, may prove to be important to issues of fertility and pregnancy loss, due to its possible role in placental (extraembryonic) tissues.

Development and Physiology

Recent evidence indicates that X-inactivation may be established much earlier than previously thought. One study proposes that imprinted XCI patterns are initiated in the paternal germline, where meiotic sex chromosome inactivation (MSCI) occurs, and maintained through conception. This imprinted pattern of XCI is retained in extraembryonic tissues but “reset” in the embryonic lineage by reactivation of the paternal X, allowing random XCI to occur.

Random X-inactivation begins when the embryo proper consists of only a few dozen cells. The process can be conceptualized in several steps. First, the cell determines the number of X-chromosomes relative to autosomes and permits only one active X for every two autosomes (counting). In the usual case of diploid female cells, two X-chromosomes are counted and one is kept active. Next, the cell randomly designates one of the two X- chromosomes to be inactivated (choice). Once chosen, the inactive X remains silenced in all future daughter cells (maintenance of inactivation). On a cellular level, the inactivated X acquires characteristics of heterochromatin including extensive [CpG methylation](#), hypoacetylated histones and late replication during the cell-cycle. During interphase of human cells, the inactivated X can be seen as a compact element at the periphery of the nucleus known as the [Barr body](#) (Fig. 2a).

Since the inactivation of X-chromosomes occurs so early in embryogenesis, patterns of cell growth and



X-Chromosome Inactivation. Figure 2 Cytological characteristics of X-chromosome inactivation. (a) Cresyl violet stain shows Barr bodies in human cheek cells. (from Development, X-chromosome inactivation, JT Lee, 1999, Springer-Verlag, Berlin, p. 407; Moore and Barr, 1955, *Lancet*, vol 2, p. 57) (b) Fluorescence *in situ* hybridization with an *Xist* probe reveals an *Xist* “cloud” coating the inactive X chromosome (arrow) in a

migration determine whether body tissues consist of cells of a polyclonal mixture of active X-chromosomes or are derived from a monoclonal set of cells with the same active X. In some cases, the compositional patterns tend to be consistent for particular organs or tissues. Some studies suggest that intestinal crypts and thyroid nodules are often monoclonal tissues because they are presumably descendants of a single stem cell. Conversely, intestinal villi and muscle fibers are believed to be polyclonal because they tend to consist of cells derived from a mixture of precursor cells. The monoclonal or polyclonal composition of tissues is an important consideration when a female carries a heterozygous X-linked mutation or when evaluating a tumor that originates from one of these tissues.

Molecular Mechanisms of X-Inactivation

Progress in understanding the molecular mechanisms of X-inactivation has occurred largely in mouse embryonic stem (ES) cells, which recapitulate the process of XCI when induced to differentiate in culture. Using this model system, important genetic regions and elements have been identified that are central to the control and execution of X-inactivation.

The initiation of X-inactivation and subsequent spread of chromosomal silencing *in cis* depends on a region of the X-chromosome known as the X-inactivation center (Xic) (1). The best-characterized element of the Xic is the X-inactive specific transcript gene (*Xist*). *Xist* is a 17-kb, non-translated gene that is necessary and sufficient to initiate the process of X-inactivation. In pre-inactivated female cells, *Xist* RNA is transcribed at very low levels from both X-chromosomes. At the onset of XCI, *Xist* is upregulated on the future inactive X and spreads along the length of the chromosome (Fig. 2b). Shortly afterwards, *Xist* expression from the future active X is shut off. Although detailed mechanisms are not fully understood, *Xist* RNA contains functional domains that allow it to remain associated with its chromosome of origin and to mediate transcriptional silencing by interacting with other cellular factors (2).

Since the up-regulation of *Xist* is central to the initiation of X-inactivation, a major research focus has been to identify factors that regulate *Xist* and therefore determine the mechanism of choice. One of these factors is *Tsix*, a gene that originates approximately 12-kb downstream from the 3' end of *Xist*. Like *Xist*, *Tsix* is a non-translated RNA gene that is spliced,

metaphase spread of female mouse fibroblasts (Photograph JT Lee from Development, X-chromosome inactivation, JT Lee, 1999, Springer-Verlag, Berlin, p. 413).

polyadenylated and retained exclusively in the nucleus. The *Tsix* locus is oriented antisense to *Xist* and produces a transcript that overlaps entirely with that of *Xist*. In addition, *Tsix* is expressed in a reverse dynamic pattern to *Xist*, suggesting that it may function as a negative regulator (Fig. 3). Genetic manipulations that abolish *Tsix* expression on one of two X-chromosomes consistently cause the *Tsix*-deleted chromosome to be inactivated, (i.e., *Xist* is up-regulated on that chromosome). Although the antisense orientation of *Tsix* suggests the possibility of a base-pairing mechanism of antagonism, the exact way in which *Tsix* blocks *Xist* expression is still unknown.

While noncoding *XIST/Xist* transcripts have been identified in both human and mouse, *Tsix* has been characterized mostly in the mouse ES cell system. Recent studies of human cells have identified human *TSIX* antisense transcripts as well, suggesting a potential conserved role for this gene. However, since the genetic region near *XIST/Xist* and *TSIX/Tsix* is not highly conserved between the two species, some uncertainty exists as to whether the sense-antisense mechanisms observed in the mouse can be reliably extrapolated to humans and future work will clarify these issues.

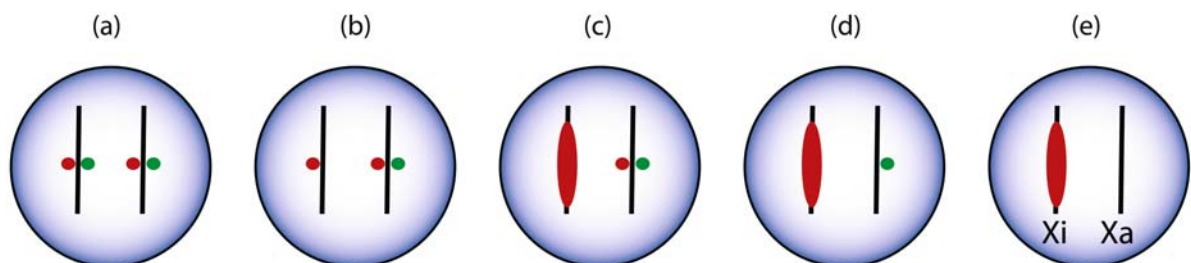
In addition to the sense-antisense pair *Xist/Tsix*, the X-inactivation center in mouse also contains other elements that appear to regulate aspects of X-chromosome inactivation. One of these is DXPas34, a CpG-rich region containing 34-mer repeats that lies very near the start site of *Tsix*. The methylation patterns of DXPas34 correlate with the active or inactive state of the X-chromosome, indicating that it may have some relationship to X-chromosome choice and [▶imprinting](#). Other elements that affect the probability of X-chromosome choice have been identified further

upstream from the *Tsix* promoter. These candidate elements are particularly interesting because they may represent the X-controlling element (*Xce*), a modifier whose existence has been inferred through genetic studies in mice. In mice, different alleles of the *Xce* influence the relative probability that the *cis* chromosome will be chosen for inactivation. In a similar way, an *Xce*-like effect has been observed in humans which modifies the probability that a given X-chromosome is active, a phenomenon which has important implications to female carriers of X-linked mutations (discussed further below).

Escape from X-Inactivation

Although X-inactivation silences a vast number of genes along the X-chromosome, a significant number of genes escape inactivation and are active on both X-chromosomes in females. Many of these genes reside in an area found on the distal tips of both X- and Y-chromosomes known as the [▶pseudoautosomal region](#). During meiosis, the pseudoautosomal regions of the X- and Y-chromosome pair and participate in crossing-over, exchanging genetic material just as paired autosomes do. The similar sequences of the pseudoautosomal region contain genes that are actively transcribed on both X- and Y-chromosomes, such that all females (on both Xs) and males (on the X and Y) express two active copies of these genes.

Two other groups of genes also escape inactivation on the X-chromosome. One group includes genes that, like those in the pseudoautosomal region, contain active copies on the Y-chromosome as well. Unlike the pseudoautosomal region, however, these genes do not participate in crossing over, but otherwise are always active in two copies in each cell. The second group of genes consists of genes that are expressed from the



X-Chromosome Inactivation. Figure 3 Dynamic relationship of *Xist* and *Tsix* during X-chromosome inactivation. A schematic diagram representing fluorescence *in situ* hybridization profiles of *Xist* and *Tsix* during the process of X-inactivation in female mouse embryonic stem cells. (a) Prior to inactivation, *Xist* (red) and *Tsix* (green) signals can be seen as small pinpoints originating from the X-inactivation center (Xic). (b) On one of the two X chromosomes, *Tsix* pinpoint expression disappears. (c) Subsequently, *Xist* is up-regulated on this chromosome and spreads to coat the chromosome *in cis*, which eventually causes the transcriptional silencing of this designated inactive X. (d) On the other chromosome, *Xist* expression is turned off while *Tsix* persists. (e) Finally, *Tsix* is turned off on the active X chromosome, leaving an *Xist* “cloud” over the inactive X.

X-chromosome only and do not have copies on the Y. Genes in this group are expressed at higher levels in females than males, since the male Y-chromosome does not have copies of the gene. The clinical impact of the over-expression of this class of genes is unknown.

Clinical Relevance

Sex Chromosome Aneuploidies and Structural Abnormalities

In contrast to autosomal aneuploidies, which are invariably lethal or severely disabling, X-chromosome aneuploidies occur frequently (1 in 500 births) and are better tolerated in humans. Two common sex chromosome aneuploidies are XO (Turner's syndrome females) and XXY (Klinefelter's syndrome males). All these individuals, as well as the less common individuals with other karyotypes (XXXY, XXYY) are viable since the X-inactivation mechanism is able to count the X-chromosomes and inactivate all but one. Generally, individuals with sex chromosome aneuploidies have moderate behavioral and developmental abnormalities, though the extent and severity can vary widely between individuals. Patients with Turner's syndrome usually have normal intelligence but often have streak ovaries and are infertile. Individuals with Klinefelter's syndrome are also infertile and frequently have learning difficulties and poor psychosocial skills. Females with trisomy X (XXX), by contrast, are usually fertile, rarely display behavioral abnormalities and may often go undetected by clinicians.

One hypothesis suggests that the genes that escape X-inactivation may contribute to the phenotype of individuals with sex chromosome aneuploidies. Turner's syndrome patients lack the dose of escape genes normally provided by the second X-chromosome, making them haploinsufficient for this set of genes. Conversely, even though all but one X-chromosome is inactivated in individuals with supernumerary X-chromosomes, escape genes on the extra inactivated X(s) are still active and provide an abnormal additional dose of these genes. Consistent with this hypothesis, individuals with progressively more Xs tend to have increasingly severe behavioral and developmental phenotypes; pentasomy X individuals (XXXXX), for example, usually present with severe developmental retardation and many physical defects, despite grossly inactivating four of the five X chromosomes. For the most part, it is unknown whether the mis-dosage of a particular gene or subset of genes is responsible for the phenotypes seen in these aneuploidies or whether the generalized mis-dosing of many genes is intrinsically harmful. In Turner's syndrome, at least, haploinsufficiency of the *SHOX* gene (short stature, X chromosome) has been directly linked with the short stature and skeletal anomalies found in these patients. As more detailed and efficient DNA mapping and expression

analysis methods become available, it is possible that future research will be able to associate specific genes to the phenotypic traits of patients with X-chromosome aneuploidies.

In addition to numerical abnormalities, various X-chromosome structural abnormalities have also been observed, including deletions, duplications, X;autosome [translocations](#) and ring X-chromosomes. As is the case with aneuploidies, certain X-chromosome structural abnormalities are better tolerated than analogous defects on autosomes because X-chromosome inactivation minimizes the cellular and clinical impact of the anomaly. Individuals harboring an X-chromosome deletion or duplication tend to have cells that have preferentially inactivated the abnormal chromosome, leaving the normal X-chromosome active. This is probably due to a process of cell selection, where cells that choose to keep the defective X-chromosome active are less viable than the cells that choose to keep the normal X-chromosome active. The same process of selection affects individuals with balanced X;autosome translocations, who preferentially inactivate the intact X-chromosome. Since the translocated portion of the X-chromosome can induce gene silencing on the attached autosome, cells which choose to silence the translocated X will also inactivate important autosomal genes and will be at a disadvantage compared to cells expressing a normal complement of autosome and sex chromosome genes. One other rare structural abnormality is the ring X-chromosome, which usually results after X-chromosome breaks and deletions are followed by circularization of the chromosome. Various sizes of ring Xs have been observed, depending on the amount of the chromosome that has been deleted. In addition to the number of genes that have been lost, an important factor in the severity of ring X phenotypes is whether or not the ring still contains the XIST region. Ring X-chromosomes that have lost the XIST region will generally not be inactivated, hence providing an inappropriate extra dose of X-linked genes and resulting in severe phenotypes. Patients with rings that have preserved the XIST region can have a phenotype that is normal, although this depends greatly on the extent of the deletion.

Manifestation of X-Linked Genetic Diseases

Random X-chromosome inactivation causes females to be mosaic with respect to X-linked genes, resulting in a difference in how recessive X-linked mutations are manifested in the two sexes (3). Almost all X-linked mutations are revealed in males regardless of whether the mutation is dominant or recessive. In heterozygote females, however, roughly half of all somatic cells will express the mutant gene while the other half will express wild type copies. Immunostaining for

dystrophin in females with Duchenne muscular dystrophy (DMD), in which the dystrophin gene is mutated, shows a mosaic pattern where patches of muscle fibers express normal dystrophin and other patches do not. In many cases, such as in DMD and hemophilia A, heterozygote female carriers can therefore be clinically unaffected because the single wild-type copy compensates for the mutant gene.

Although unaffected or mildly affected heterozygotes represent the majority of cases in X-linked recessive diseases, variations in X-chromosome inactivation can influence the clinical picture. In a purely random system, each female would have cells expressing the two X-chromosomes in a 50:50 ratio. In reality, this ratio sometimes favors one chromosome, resulting in ►skewed X-inactivation (80:20 instead of 50:50, for instance). One way in which this occurs is by pure chance. Since X-inactivation is a random process that occurs when the embryo contains relatively few cells, the total percentage of cells inactivating one or the other X-chromosome can vary, leading to skewed X inactivation in that individual as a whole. Consequently, a heterozygote in whom a mutant disease allele resides on the predominantly active X can manifest characteristics of a normally recessive disease. Another way in which skewed X-inactivation is thought to occur is by the presence of genetic elements that predispose one X-chromosome for inactivation over the other. Several families have been identified where skewed X-inactivation occurs in clusters, implicating the existence of human Xce-like genetic modifiers or alleles that cause different predispositions for inactivation. A strong modifier allele makes the *cis* X-chromosome less likely to be inactivated and therefore causes all genes on that chromosome to be preferentially expressed in the organism as a whole. If a genetic mutation resides on an X-chromosome with a strong modifier, then this mutation will be preferentially expressed, leading to disease phenotypes in heterozygotes despite the presence of the normally protective wild type copy. Whether occurring by random chance or by genetic predisposition, skewed X-inactivation has been reported to cause disease in female carriers with DMD, fragile X syndrome, hemophilia, Wiskott-Aldrich syndrome, color blindness and several other X-linked disorders.

XCI and Assessing Tumor Clonality

Establishing the clonality of tumors is a major interest for cancer biologists and clinical oncologists. The phenomenon of XCI has been used for assessing clonality since the process stably “marks” each cell and all its daughter cells with one of the two X-chromosomes. By identifying differences between the two X-chromosomes and determining whether a tumor cell population has one active X or a mixture of

the two, one can infer the clonality of the population. Historically, variant isotypes of proteins were identified for assessing clonality, but recently the more-common differences in DNA and RNA have been used instead. Some examples of X-linked genes that have proven useful include the glucose-6-phosphate dehydrogenase (G6PD), hypoxanthine phosphoribosyltransferase (HPRT) and human androgen receptor (HUMARA) genes, in which common polymorphisms and/or methylation differences have been identified. Although precise methods of assaying polymorphisms and methylation differences are numerous and varied, the principle is the same. If tumor cells express a gene that can be uniquely traced to only one X-chromosome, one can infer that the tumor population arose from a single progenitor. Using XCI as a marker for clonality has several limitations and drawbacks (it works only in females and is subject to confounding factors such as the natural X-inactivation patterns in tissues) but it has been a valuable tool in conjunction with other methods.

XCI and Cancer

Long-standing evidence has connected inappropriate X-chromosome regulation with cancers, and recent work continues to strengthen this link. Historically, breast cancer tumor cells that lacked Barr body staining by cytology were associated with a poor prognosis in patients. Recently, a gene expression profile of recurrent ovarian tumors found that *XIST* expression was consistently down-regulated and that this down-regulation correlated with insensitivity to taxol treatment. The study suggests that *XIST* may be a potential marker for chemotherapeutic response in ovarian cancer. Taken a step further, another study found that the *BRCA1* gene, a breast and ovarian tumor suppressor, supports the localization of *XIST* RNA on the inactive X-chromosome. Although this provides no direct causal link, it raises the possibility that *XIST* may not only serve as a marker for cancer and cancer therapy, but may ultimately have some role in tumorigenesis. Given the importance of gene regulation in cancer and the ability for XCI and *XIST* to affect gene expression on a wide scale, it is likely that future research will continue to discover connections between X-chromosome inactivation, cancer and other genetically related health conditions.

► CpG Islands

► Repetitive DNA

References

1. Avner P, Heard E (2001) X-chromosome inactivation: counting, choice, and initiation. *Nat Rev Genet* 2:59–67
2. Brockdorff N (2002) X-chromosome inactivation: closing in on proteins that bind Xist RNA. *Trends Genet* 8:352–358

3. Lyon MF (2002) X-chromosome inactivation and human genetic disease. *Acta Paediatr Suppl* 439:107–112

²Molecular, Cellular and Developmental Biology, University of Colorado, Boulder, CO, USA
klym@spot.colorado.edu

X-Chromosome Silencing

► X-Chromosome Inactivation

Xenobiotic Substances

Definition

Xenobiotic substances are synthetic substances or drugs, which cannot be synthesised by an organism.

► Metabolomics

Definition

The human genome project and subsequent efforts have driven technological advances and led to an exponential increase in sequence data acquisition. New bioinformatic tools allow researchers to query and analyze sequence data as never before. Yet, the functional meaning of sequence and expression information, how genotype maps to phenotype (and *vice versa*), and the dynamic relationship between gene expression and cellular behavior, tissue function, developmental processes and homeostatic and adaptive mechanisms, has lagged behind. There is an urgent need for the integration of high-throughput ► [genomics](#) methods and experimentally accessible model systems, in order to answer central questions of biological behavior. An obvious choice for a biological system in which to reconcile genomics and functions is the clawed frog ► *Xenopus laevis*. We discuss the advantages of *Xenopus* as an experimental model, and its potential for making significant contributions in the area of ► [functional genomics](#).

XenoMouse

Definition

The XenoMouse is a transgenic mouse strain which does not produce murine but human antibodies. This is possible because the mouse immunoglobulin gene locus has been knocked out and a complete human immunoglobulin gene locus was introduced by transgenesis (humanizing).

► Monoclonal Antibodies

► Transgenic and Knockout Animals

Characteristics

Introduction

Xenopus' Advantages

For over fifty years the African clawed frog *Xenopus laevis* has served as a model system for biologists seeking to understand the basic processes underlying animal development. The ease of maintaining these frogs in captivity, the hundreds of eggs that can be produced and externally fertilized at one time, and the ability to readily manipulate the large oocytes and embryos have made them a favorite of classical and molecular embryologists for mechanistic studies of patterning, induction and morphogenesis.

The Genus *Xenopus* is in the Order Anura, Family Pipidae (http://www.columbia.edu/itc/cerc/danoff-burg/invasion_bio/inv_spp_summ/xenopus_laevis.htm); they are the only frogs with clawed toes. They are endogenous to cooler regions of sub-Saharan Africa, and a number of species have been described (<http://clawedfrogs.tripod.com/acf/id23.html>). Fully aquatic, they are typically found in muddy ponds. While they lack both a tongue and teeth they are voracious predators. Several *Xenopus* species have been studied in the laboratory, with *Xenopus laevis* being by far the most common. Recently, a second *Xenopus* species, *tropicalis* (also known as *Silurana tropicalis*; Fig. 1), has been adopted as an experimental

Xenopus as a Model Organism for Functional Genomics: Rich History, Promising Future

MATTHEW W. GROW¹, MICHAEL W. KLYMKOWSKY²

¹Biochemistry and Molecular Biology, Indiana University School of Medicine, Biotechnology Research and Training Center, Indianapolis, IN, USA



***Xenopus* as a Model Organism for Functional Genomics: Rich History, Promising Future.**

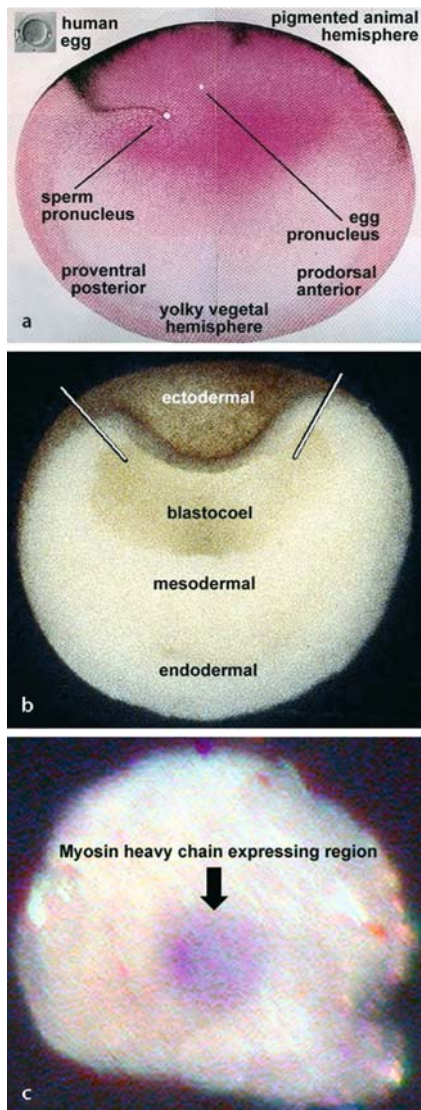
Figure 1 A sexually mature *X. tropicalis* female, shown here next to a US quarter for size comparison. *X. laevis* females differ little in appearance, but grow to as much as 23 times larger.

organism, primarily due to its genomic simplicity when compared to *X. laevis* (see below). For the remainder of this discussion, we will use the term *Xenopus* to refer to both species collectively. The advantages of using *Xenopus* as an experimental model organism are outlined below:

- **Ease of maintaining large breeding colonies:** Being completely aquatic, these animals may be housed in many types of aquaria. Several commercial, recirculating “rack” systems are available. *Xenopus* are also extremely resistant to disease, provided proper care is taken.
- **Large clutch size:** A mature *Xenopus* female can produce hundreds of eggs several times in one day (and can repeat this process after a recovery period of ~6 weeks). This allows for the collection of large numbers of embryos, which benefits any high-throughput study, and provides statistically significant numbers for experiments.
- **Large oocyte size:** *X. laevis* oocytes and eggs, and extracts prepared from them, have long been used for *in vitro* biochemical and subcellular studies. A particularly attractive feature of the late stage oocyte, is the ability to manually isolate the nucleus (►[germinal vesicle](#)), which facilitates subcellular localization studies.
- **Large embryo size:** A mature *X. laevis* egg is ~1.2 mm in diameter,

while *X. tropicalis* eggs are ~400 µm in diameter. The large *X. laevis* egg size facilitates microinjection, microdissection and complex tissue recombination studies (Fig. 2). (To view examples of microsurgery using a ‘gastromaster’ visit (►<http://www.gastromaster.com/videos.html>)). There is no net growth of the embryo until ~5 days after fertilization, when feeding begins. Gastrulation takes place ~9 hours after fertilization (at 23°C); neurulation begins in earnest at stage 12.5/13, ~14 hours after fertilization; neural crest is well established and neural tube closure is complete by stage 21 (~23 hours after fertilization). Around this same stage, a number of other organ systems are forming, and by stage 30 (~36 hours after fertilization) organogenesis is well underway (For stages and timings see – ►<http://www.xenbase.org/atlas/NF/NF1-10.html> and related pages). Cells are therefore relatively large during the key morphogenic processes of gastrulation, neurulation and neural crest migration, facilitating the visualization of cell behavior.

- **Externally fertilized embryos:** Fertilization is typically carried out artificially by adding macerated testis to expressed eggs. This allows fertilization to occur reasonably synchronously (within a 5–10 min window). It is therefore straightforward to generate large numbers of stage-matched embryos.
- **Externally-developing embryos:** This feature enables researchers to observe the embryos’ progression through all developmental stages.
- **Well-characterized embryonic development:** Intense efforts have led to detailed fate maps for *X. laevis* (1,2) (Forward and reverse fate maps ►<http://www.xenbase.org/atlas/xenbasefate.html>). This, and a cumulative knowledge base from decades of study, can be used to target micromanipulations (microinjections, dissections, etc.) to specific regions or tissues of the embryo. There is a critical mass of researchers working in the system, actively producing both reagents and observations that spur on new projects and discoveries. It is very likely that the bulk of the observations made in *X. laevis* will hold true for *X. tropicalis* as well.
- **The animal cap system:** Unique to *X. laevis* is the animal cap assay (7,10). In *X. laevis* zygotic transcription does not begin in earnest until stage 8.5, the midblastula transition (~4000-5000 cells). At this point, two features of the embryo combine to provide a unique experimental opportunity. First, a large fluid filled space, the blastocoel, acts to minimize cellular interactions between the animal pole ectoderm and the vegetal endodermal cells. The cells that constitute the “roof” of the *X. laevis* blastocoel can be excised from the



Xenopus as a Model Organism for Functional Genomics: Rich History, Promising Future.

Figure 2 Compared to the $\sim 70 \mu\text{m}$ diameter human egg (inset panel a), with a volume of $\sim 2 \times 10^5 \mu\text{m}^3$, the $\sim 1200 \mu\text{m}$ diameter *X. laevis* egg (a) has a volume of $\sim 9 \times 10^{10} \mu\text{m}^3$. Following sperm entry, male and female pronuclei migrate toward the center of the animal hemisphere and fuse. A microtubule system, nucleated by the sperm's centrosome, drives a cortical rotation; this rotation breaks the egg's cylindrical symmetry and presages the dorsal-anterior and ventral-posterior axes of the embryo. (b) A series of rapid cleavage divisions transforms the fertilized egg into a blastula. As early as the second division a space, the blastocoel, appears and separates the animal ectoderm from the equatorial mesoderm. During normal development inductive signals are relayed from vegetal endoderm to mesoderm and then to ectodermal tissues. At stage 8/9 the animal pole region (marked by white lines) can be readily dissected away. At this point, inductive signals that

embryo and cultured separately (animal cap). If excised prior to the arrival of inductive signals from the mesoderm, which acts as a relay station for signals that originate in endodermal cells, the piece of blastocoel roof heals to form a spherical structure and proceeds to form atypical epidermis. These cells possess the ability, however, to respond to a number of inductive signals, signals that can be supplied by the experimenter. These signals can drive the animal cap to differentiate into a wide variety of tissues, including blood, muscle, nerve, cartilage, pronephros, pancreas and heart (7,8). This multi-potent, homogeneous cell culture offers investigators an *in vivo* "test bed" for studying the regulatory networks that control the patterning and differentiation of many different embryonic tissues, particularly when coupled with the use of hormone-regulated proteins (see below). At the same time, the animal cap has been freed from many of the dynamic interactions that characterize the developing embryo. Specific inductive signals can be applied at specific times to the system, and their effects systematically analyzed. The large size of the animal cap offers the possibility of analyzing individual caps, which should reveal much of the homeostatic mechanisms and regulatory dynamics. From an experimental perspective, an important practical feature of the animal cap system is that molecular markers of mesodermal or endodermal differentiation are generally completely absent, so that their induction is readily apparent even using simple (rather than quantitative) RT-PCR methods.

- **Hormone regulated proteins:**

The technique to generate hormone-regulated versions, particularly of transcription factors, has proven extremely useful in *Xenopus*, particular when used in combination with the animal cap system (Sive et al. 2000). The most common approach is to generate a plasmid in which the region encoding the hormone-binding domain of the human glucocorticoid receptor is appended to either the 5' or 3' end of the target protein's coding sequence. In a typical experiment, RNA encoding this chimeric

originate in the vegetal endoderm and propagate through the mesodermal region have not yet reached the animal pole ectoderm. Such "animal caps" differentiate into atypical epidermis (ectoderm), but can be driven toward various fates through manipulation of signaling and gene regulatory pathways (7,8). (c) An animal cap generated from a fertilized egg injected with an RNA encoding the HMG-box transcription factor SOX7; the cap was fixed and stained *in situ* for myosin heavy chain a RNA, a marker of cardiogenesis (9). Panel A is adapted from Hausen & Riebesell (10).

polypeptide is injected into fertilized eggs. While the polypeptide is rapidly synthesized, it is inactive in the absence of the hormone dexamethasone. Addition of dexamethasone rapidly activates the polypeptide. In such experiments, animal caps are generated from RNA injected fertilized eggs. They can then be exposed to hormone in the presence or absence of a protein synthesis inhibitor (we find that emetine is more useful than cycloheximide, since it produce fewer non-specific effects) (see 9). This makes it possible to identify “immediate early” and possibly direct targets of transcription factors. In the context of the intact embryo, it is possible to use such hormone-regulated constructs to circumvent early effects and focus on later stages of development.

- **Transgenesis:**

The techniques for generating transgenic *Xenopus* are now well established (3), and the number of characterized promoters available is increasing rapidly. Stable transgenic lines are being produced, enhancer-trapping screens are underway and transposon-mediated transgenesis methods are being tested.

- **Genetic Studies in *X. tropicalis*:**

X. laevis is effectively ▶ **allotetraploid**. Some alleles are non-functional or distinct from a functional perspective. In contrast, *X. tropicalis* has a typical diploid genome and presents the opportunity to conduct traditional genetic and molecular genetic experiments (4). That said, the value of traditional genetic approaches in *Xenopus* remains to be established, particularly given the relative ease of transferring genetic insights from other well established genetic model systems (e.g. *Dario rerio* – zebrafish, *Drosophila melanogaster* and *Caenorhabditis elegans*) to the more experimentally manipulable *X. laevis* system.

- **Gene screens**

A particularly powerful screening approach, made possible by the inherent and visible dorsoventral asymmetry of the early *Xenopus* embryo, is the screening of pools of cDNAs via injection and sib selection (see 12 et al. 2000, and references therein). This method has proven extremely useful in identifying previously unknown genes, and previously unknown functions for known genes.

- **Rapidly accumulating EST and genomic *Xenopus* sequence:**

Of the many chordates studied, the *Xenopus laevis* transcriptome is among the most frequently sequenced. As of April 2005, Unigene build #63 (▶ <http://www.ncbi.nlm.nih.gov/UniGene/UGOrg.cgi?TAXID=8355>) contained 12,475 mRNAs and 389,701 total sequences in clusters. Additionally, the DOE Joint Genome Institute has initiated an *X. tropicalis* genome sequencing project, with 6 to

8-fold coverage expected before the end of 2005; assembly version 3.0 is currently available (<http://genome.jgi-psf.org/Xentr3/Xentr3.home.html>).

- **A growing array of molecular tools**

In addition to well-characterized promoter sequences, there are a growing number of characterized genes, in situ probes and antibodies that serve as markers of gene and protein expression and cellular differentiation (Fig. 3). Confocal imaging of organogenesis is straightforward (13). It is also worth noting that *X. laevis* and *X. tropicalis* are close enough evolutionarily that probes derived from one generally function in the other (5, 6). Furthermore, the growing collection of sequence data allows researchers to readily clone *Xenopus* orthologs of genes implicated in specific processes in other systems. Additional molecular tools for imaging, gene ▶ **knockdowns**, etc. are making functional genomic studies in *Xenopus* increasingly straightforward (see below).

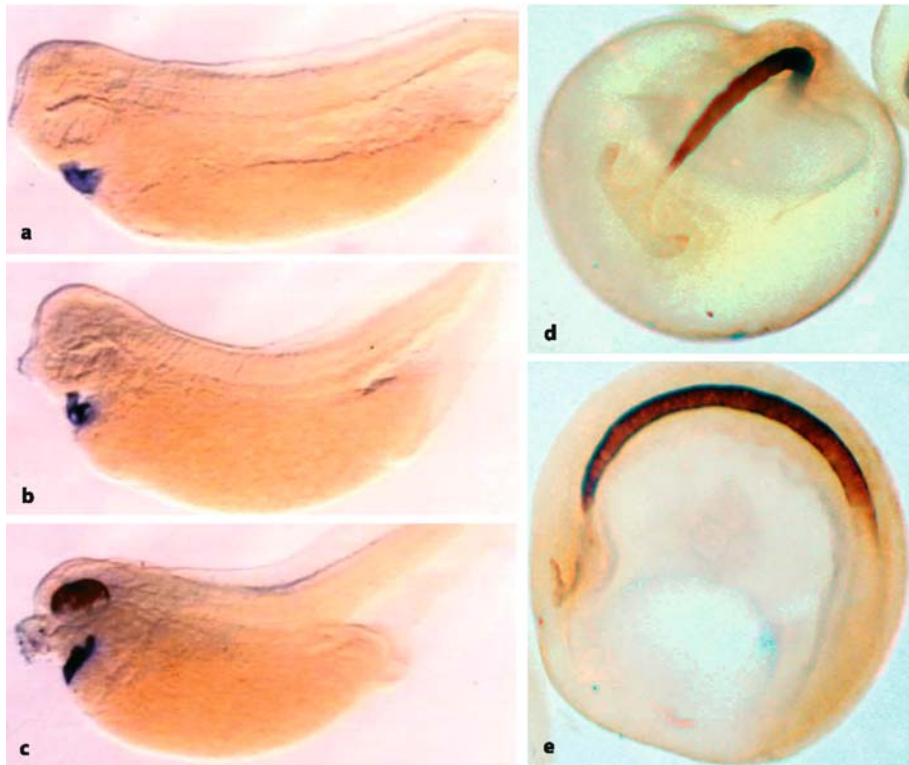
Functional Studies in *Xenopus*: Methods and Resources

▶ Misexpression/▶ Over-Expression

The standard approach to study gene function is to disrupt the normal pattern of gene expression. This can involve expression of the gene product in the wrong place, at the wrong time, or at the wrong level, expressing an altered form of the gene product or suppressing to various extents the expression of the gene. In *X. laevis* over-expression and misexpression studies have been used quite successfully to analyze a number of key developmental and differentiation events. RNA or DNA can be injected into the fertilized egg or specific blastomeres in the early embryo or introduced *via* transfection or transgenesis. Specific regions of an injected embryo can be transplanted into homologous or ectopic regions of uninjected embryos and *vice versa*. Transplanted cells can be followed through the use of various lineage tracers, and conditional expression can be achieved by using tissue-specific promoters and inducible chimeric constructs, such as glucocorticoid receptor fusions (see above). Transgenic *Xenopus* lines have been made that allow conditional expression using the Gal4-UAS system (14).

Knockdown/Inhibition Studies

While targeted gene knockouts are not yet possible in *Xenopus*, analogous methods have been used quite successfully. Conventional and modified (morpholino (morpholino (▶ **Morpholino Oligonucleotides, Functional Genomics by Gene ‘Knock-down’**)- or phosphorothioate-type) antisense reagents can dramatically reduce the translation or splicing of the targeted mRNA. (11, 15). This makes *X. laevis* a particularly valuable model system for the study of maternally supplied gene products. It is also possible to perturb



Xenopus as a Model Organism for Functional Genomics: Rich History, Promising Future. Figure 3 Many molecular markers and characterized genes are available for *Xenopus*. In the left panel, whole mount *in situ* hybridizations were used to visualize *cardiac troponin I*; this gene is expressed in myocardium of the heart. From top to bottom, the embryos are at stage 31, 35, and 40 of development. In the right panel, embryos have been stained in whole-mount with the MZ15 monoclonal antibody directed against a keratan sulfate, a component of the notochord.

normal gene product function *via* the expression of dominant negative versions of the gene product and antibodies (in some cases, antibodies can also be used to activate a gene product). A few studies have been published using small interfering RNAs (siRNA) in *Xenopus*, but it is not yet clear if this technique will prove widely effective. In each case, specific perturbation can be delivered to the whole embryo or targeted to specific regions via fate map-based blastomere injection. Temporal control of the treatment can also be achieved by controlling the expression of dominant negative constructs through the use of hormone-regulatable domains or specific promoters in transgenic studies.

Mutational Screens

Several efforts are currently underway to perform mutagenic screens in *X. tropicalis* using chemical treatments, irradiation and transposon-insertion (http://tropicalis.berkeley.edu/home/molecular_resources/mutants/mutants.html). These screens should prove to be

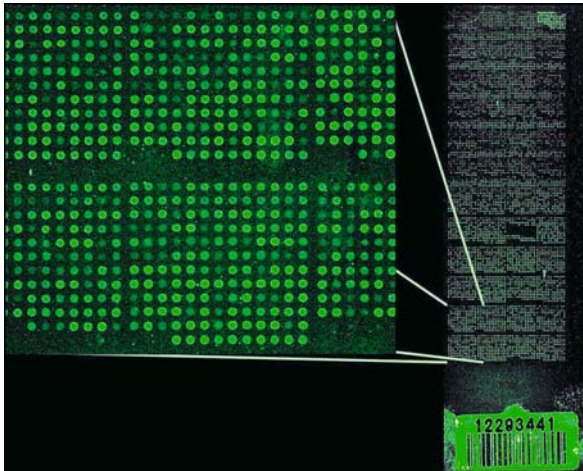
valuable additions to those performed in other genetically amenable organisms.

Enhancer Trapping Screens

Using insertional transgenic techniques, enhancer (or gene) trapping screens are currently underway, in both *X. laevis* and *tropicalis*. The most common approach is to use a fluorescent reporter gene in a construct that allows for the cloning and sequencing of flanking genomic sequences.

Microarrays

The most common tool today for high-throughput functional studies is the DNA [microarray](#), which enables investigators to monitor transcriptional activity for thousands of genes simultaneously, and in some cases across entire genomes. Several *X. laevis* spotted cDNA microarrays have been fabricated and are in use in gene expression studies (Fig. 4). Additionally, a consortium of researchers is in the process of designing oligonucleotide probes for 15,600 *X. laevis* genes for the creation of a spotted oligo microarray. Additionally,



***Xenopus* as a Model Organism for Functional Genomics: Rich History, Promising Future.**

Figure 4 A spotted microarray fashioned from a *Xenopus laevis* embryonic heart cDNA library. This array has over 7,000 cDNA probes spotted onto the surface of a 1 inch by 3 inch glass slide (photo courtesy of Grow lab).

Affymetrix, Inc. (<http://www.affymetrix.com>) markets a commercial “Genechip™” microarray with 15,503 oligo probe sets for the detection of approximately 14,400 transcripts.

***Xenopus* Gene Expression Atlas**

Microarray-based experiments are certain to lead to the identification of thousands of genes for which follow-up characterization is desired. In order to avoid wasting valuable resources on redundant studies, a centralized database of *Xenopus* gene expression is clearly essential. This database should not only serve as a repository for published *Xenopus* microarray expression profiles, but should also curate the temporal/spatial expression data for characterized genes (recording tissue-specific expression across multiple embryonic stages, for instance). One such early effort is the Axelldb database (see link below). As *in situ* hybridization techniques improve and move to high-throughput platforms, the large numbers of nearly identical stage-matched embryos generated by *Xenopus* females will greatly facilitate data collection and analysis. Recently, a large-scale microarray profiling effort has been made in an attempt to characterize ~21,000 *Xenopus* genes; the results of this effort have been placed into a searchable “iChip” database (16) (see link below).

Online *Xenopus* Resources

The following are web addresses for several sites of interest. This is not meant to be a comprehensive list, but rather a starting point for additional research on *Xenopus*:

- ▶ <http://www.xenbase.org> – Xenbase: A *Xenopus* web site with information on many aspects of *Xenopus* research.
- ▶ <http://www.nih.gov/science/models/Xenopus/index.html> – The Trans-NIH *Xenopus* Initiative.
- ▶ http://www.tigr.org/tigr-scripts/tgi/T_index.cgi?species=Xenopus – The TIGR *Xenopus laevis* Gene Index.
- ▶ <http://genome.jgi-psf.org/Xentr3/Xentr3.home.html> – *Xenopus tropicalis* Genome Sequencing at the Joint Genome Institute
- ▶ http://www.sanger.ac.uk/Projects/X_tropicalis/ – *Xenopus tropicalis* EST Sequencing at the Sanger Institute.
- ▶ http://www.dkfz-heidelberg.de/molecular_embryology/axelldb.htm – Axelldb, a *Xenopus laevis* gene expression database.
- ▶ http://www.dkfz-heidelberg.de/molecular_embryology/microarraydb.html – iChip database, databases of materials and methods for *Xenopus laevis* and such (extremely incomplete)
- ▶ <http://web.wi.mit.edu/sive/pub/frogfish.html> – Hazel Sive’s Frog and Fish Resource site
- ▶ <http://faculty.washington.edu/rtmoon/xenopus.express.php> – Randy Moon’s plasmid page
- ▶ <http://spot.colorado.edu/~klym/> – The Klymkowsky Lab; *Xenopus* techniques.
- ▶ <http://cmg.iupui.edu/grow/xenopus.html> – The Grow Lab: *Xenopus* techniques and microarray analysis
- ▶ <http://faculty.virginia.edu/xtropicalis/> – The Grainger Lab *Xenopus tropicalis* web site.
- ▶ <http://tropicalis.berkeley.edu/home/> – The Harland Lab *Xenopus tropicalis* web site.
- ▶ <http://froglab.biology.utah.edu/next.html-ssi> – Dave Gard’s introduction to the *Xenopus* oocytes/early embryo and its cytoskeleton.
- ▶ <http://www.morpholino.com> – The web site for information on Morpholino antisense oligonucleotides from Gene Tools, LLC.
- ▶ <http://labs.systemsbio.org/bolouri/software/Bio-Tapestry/> – a useful Java tool for assembling gene networks.
- ▶ <http://www.nottingham.ac.uk/biology/Genetics/staff/rogerpatient/networks/mesendoderm/results/late.htm> – a Flash-based interactive network approach to gene regulatory networks in *Xenopus*.

A Future for *Xenopus* in Functional Genomics

It is clear that no single model organism can fill all of the requirements for future functional studies, and that

the greatest knowledge is often gained through comparative genomics. There is no doubt, however, that the many and unique advantages of *Xenopus* as an experimental system, as well as the growing genomic resources and critical mass in terms of on-going intellectual effort, will ensure its position as a fundamental vertebrate model. It is likely that future contributions from *Xenopus* research will lead to rapid progress in our understanding of the interactions between genes and their regulatory interactions.

Clinical Relevance

Perhaps the most surprising biological fact to emerge during the modern era is the high degree of conservation in the molecular and cellular mechanisms that underlie metazoan development. It is this conservation of mechanism that makes it possible to draw meaningful conclusions about human development and disease from studies of fruit flies and roundworms. As a vertebrate, rather closer in the evolutionary lineage leading to humans than many other experimental organisms, *Xenopus* offers experimental features, and low costs compared to mouse and primate systems, while preserving the promise of practical relevance to human biology.

References

- Kumano G, Smith WC (2002) Revisions to the *Xenopus* gastrula fate map: implications for mesoderm induction and patterning. *Dev Dyn* 225:409–21
- Dale L, Slack JM (1987) Fate map for the 32-cell stage of *Xenopus laevis*. *Development* 99:527–51
- Kroll KL, Amaya E (1996) Transgenic *Xenopus* embryos from sperm nuclear transplantations reveal FGF signaling requirements during gastrulation. *Development* 122:3173–83
- Hirsch N, Zimmerman LB, Grainger RM (2002) *Xenopus*, the next generation: *X. tropicalis* genetics and genomics. *Dev Dyn* 225:422–33
- Khokha MK, Chung C, Bustamante EL et al (2002) Techniques and probes for the study of *Xenopus tropicalis* development. *Dev Dyn* 225:499–510
- D'Souza A, Lee M, Taverner N et al (2003) Molecular components of the endoderm specification pathway in *Xenopus tropicalis*. *Dev Dyn* 226:118–27
- Green J (1999) The animal cap assay. *Methods Mol Biol* 127:1–13
- Ariizumi T, Asashima M (2001) In vitro induction systems for analyses of amphibian organogenesis and body patterning. *Int J Dev Biol* 45:173–179
- Zhang C, Basta T, Klymkowsky MW (2005). SOX7 and SOX18 are essential for the induction of cardiogenesis in *Xenopus*. *Dev Dyn*, in press
- Hausen P, Riebesell M (1991) The early development of *Xenopus laevis*: an atlas of histology. Verlag der Zeitschrift für Naturforschung, Tübingen
- Sive HL, Grainger RM, Harland RM (2000) Early development of *Xenopus laevis*: a laboratory manual. Cold Spring Harbor Press
- Grammer TC, Liu KJ, Mariani FV, Harland RM (2000) Use of large-scale expression cloning screens in the *Xenopus laevis* tadpole to identify gene function. *Dev Biol* 228:197–210
- Kolker SJ, Tajchman U, Weeks DL (2000) Confocal imaging of early heart development in *Xenopus laevis*. *Dev Biol* 218:64–73
- Hartley KO, Nutt SL, Amaya E (2002) Targeted gene expression in transgenic *Xenopus* using the binary Gal4-UAS system. *Proc Natl Acad Sci USA* 99:1377–82
- Heasman J (2002) Morpholino oligos: making sense of antisense? *Dev Biol*. 243:209–14.
- Baldessari D, Shin Y, Krebs O et al (2005) Global gene expression profiling and cluster analysis in *Xenopus laevis*. *Mech Dev* 122:441–75

Xenopus Laevis

Definition

Xenopus laevis is the name for the African Clawed Frog which is frequently used in developmental biology and biomedical research. It is considered to be allotetraploid, with mature females growing to an average of 10–12 cm in length. The oocytes of *Xenopus* are commonly used cells for electrophysiological recordings. They are a popular model because of their large size (~1.0 mm) which makes their handling and manipulation easy.

► [Axis Formation – Formation and Function of the Dorsal Organizer](#)

► [Xenopus as a Model Organism for Functional Genomics](#)

Xenopus Tropicalis

Definition

Xenopus tropicalis is the name for the Western Clawed Frog, which is gaining acceptance as a laboratory model organism for genetic studies. Also named *Xenopus (Silurana) tropicalis* or *Silurana tropicalis*, this frog has a diploid genome, and grows to approximately half the size of *Xenopus laevis*.

► [Xenopus as a Model Organism for Functional Genomics](#)

Xeroderma Pigmentosum

Definition

Xeroderma pigmentosum (XP) is a very rare skin disorder, which is caused by cellular hypersensitivity to

ultraviolet (UV) light as a result of a defect in the DNA repair system.

► [DNA-Repair Mechanisms](#)

X-Linked Adrenoleukodystrophy

Definition

X-linked adrenoleukodystrophy (XALD) is a X-chromosomal recessively transmitted disorder with two main phenotypes including childhood-cerebral ► [ALD](#) (► [CCALD](#)) and adrenomyeloneuropathy (► [AMN](#)).

► [Peroxisomal Disorders](#)

X-Linked Inheritance

Definition

X-linked designates a gene carried on the X chromosome. When a gene is situated on the X-chromosome, males will more commonly be affected by a mutation because males have only one Chromosome.

► [Hereditary Hemostatic Defects and Recombinant Proteins for Treatment](#)

► [Mitochondrial Myopathies](#)

XML

Definition

XML refers to 'eXtensible Markup Language', a data representation scheme. This enables mark up of data in flat files, in a standard and uniform way. XML has flexibility and is suitable as a data description language. Therefore, one can write flat files of structural data in a standard format. XML was suggested by the W3C (World Wide Web Consortium) in 1996.

► [Protein Databases](#)

XP

► [Xeroderma Pigmentosum](#)

X-Ray Radiation

Definition

X-ray radiation is an electromagnetic radiation (light) with wavelength(s) from about 10^{-7} to about 10^{-11} meters (100-0.01 nm). X-ray crystallography uses wavelengths typically around 0,1 nm (atomic dimensions) in order to achieve atomic resolution.

► [X-ray Crystallography, Basic Principles](#)

X-Ray Crystallography

Definition

X-ray crystallography is a technique of determining the three-dimensional molecular structure of e.g. biomacromolecules, by analyzing x-ray diffraction patterns of crystals of the molecule in question.

► [Protein Databases](#)

► [Protein-Protein Interaction](#)

► [Two Hybrid System](#)

► [X-ray Crystallography, Basic Principles](#)

X-Ray Crystallography, Basic Principles

RICHARD A. ENGH

Max Planck Institute of Biochemistry, Martinsried,
Germany
engh@biochem.mpg.de

Definition

X-ray crystallography reveals the spatial structure of molecules by measuring how they scatter ► [X-ray radiation](#) when arranged in a ► [crystal lattice](#). Two broad fields may be distinguished: small molecule crystallography deals with a small number of atom positions and typically well-ordered crystals, while macromolecular (usually protein) crystallography determines a much larger number of atomic positions, usually despite considerable ► [crystalline disorder](#).

Their common goal is to calculate the ► [electron density](#) distribution in the crystal from measured X-ray ► [diffraction intensities](#). The electron density and diffraction intensities are represented by mathematical functions, which may be interconverted variously using the ► [Fourier transform](#), ► [convolution](#) and ► [complex](#)

product operations if the phases of the diffracted X-rays are known. Because the phases are not directly measurable, the dilemma known as the “phase problem” arises. Several methods exist that enable phase estimation and thereby the solution of the spatial structure. The final structure is typically modeled as a set of unique atomic positions with approximate corrections for atomic motions. The accuracy of the structure depends on the quality of the diffraction data, but also on the validity of assumptions regarding atomic positions, their relationships to the electron density distribution and chemical geometries.

Description

X-Ray Radiation, Scattering

X-ray radiation is electromagnetic radiation with wavelength(s) from about 10^{-7} to about 10^{-11} meters (100–0.01 nm), appropriate for the resolution of atomic details (*ca.* 0.1 nm). These correspond approximately to frequencies ranging from 10^{15} – 10^{19} Hz, or photon energies ranging from 10–100 keV. A classical (non-quantum mechanical) description of X-ray – matter interactions provides an intuitive explanation of much of protein crystallography. As a light wave, the incoming X-ray radiation is an oscillating electrical field that induces an oscillation of charges found within the field. This secondary oscillation, which in atoms is significant only for electrons (due to their low mass relative to protons), is itself a source of radiation and scatters thereby some of the incident radiation. The electrons cannot be considered point charges, but rather as a cloud of (probability) density. The amplitude of the scattered wave is proportional to the number of electrons in the cloud and the angular dependence of the scattering depends on its shape. To a good approximation, the electron cloud of an atom is spherically symmetric and dependent on atom type.

The total scattered intensity from a molecule is the sum of all of the waves scattered by the individual atoms, wherein constructive and destructive ►interference between the waves amplify and diminish the resultant wave respectively, depending on the scattering angle. In theory, measurement of the angular (two dimensions) dependence of the waves (amplitudes and phases) scattered by a molecule over all possible orientations about a rotation axis (the third dimension) would enable the calculation of the (three dimensional) spatial molecular electron density distribution by means of the Fourier transform. In practice, however, the waves scattered by a single molecule would be too weak to measure.

Crystal Lattice

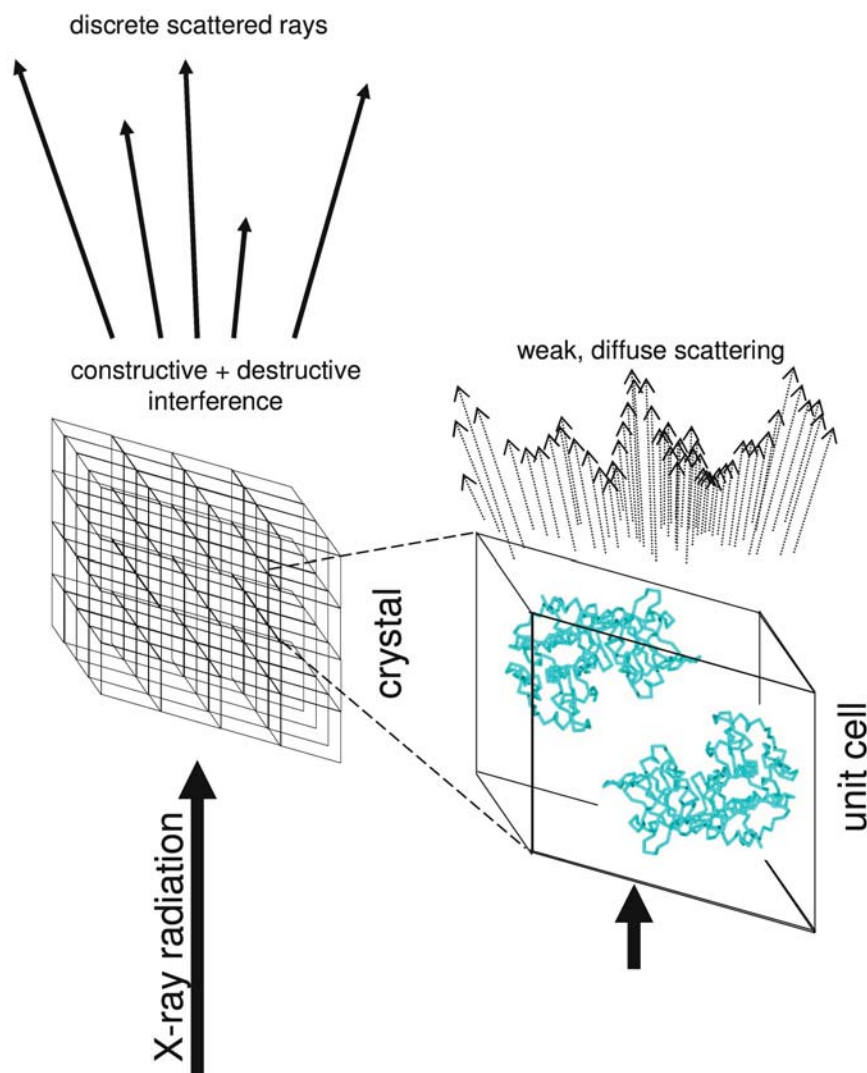
The discussion above considers scattering from atoms and molecules without reference to crystals. The periodicity of molecules packed in a crystal lattice

transforms the diffuse scattering from an individual molecule *via* constructive and destructive interference into a set of discrete diffracted spots. The ideal crystal consists of an entity called the ►unit cell, which is repeated identically, periodically and infinitely in three dimensions. The unit cell is defined by three axes and dimensions that define the edges and faces of the unit cell. Neighboring identical unit cells pack against one another along shared edges and faces such that displacement (translation) of one (or any integer number) cell dimension(s) along the corresponding cell moves an observer to an indistinguishable position in a neighboring unit cell.

The unit cell contains the molecules in a ►crystal packing arrangement that may include multiple copies of the molecule(s); these in turn may or may not be identical to each other. The transformations of the unit cell that generate a unit cell indistinguishable from the first are the ►symmetry operations that determine the cell symmetry. These operations in general comprise rotations, screw rotations, mirror planes, inversions and inversion axes; the inherent chirality of natural proteins eliminates inversion and mirror symmetries from protein unit cell symmetries. The symmetry operations map the molecules of the unit cells onto identical molecules. A minimal set of molecules that cannot be mapped onto each other using the cell symmetry operations defines the ►asymmetric unit of the crystal cell. The molecules of the asymmetric unit need not be identical and are always distinguishable at least by differing crystal environments.

The periodicity of the crystal lattice results in total destructive interference of the scattered X-ray waves except at specific scattering angles, for which constructive interference maximizes their intensities. A simple comparison with a one-dimensional diffraction grating provides an intuitive explanation of this phenomenon. Visible light that is scattered by a single groove engraved on a plate is weak and diffuse. If however there is a row of identical grooves to form a diffraction grating, the separately scattered waves recombine at some distance from the grating into alternating light and dark bands, corresponding to regions of constructive and destructive interference, respectively. ►Bragg’s law is the mathematical formula that relates the spacing of the grooves to the wavelength of light and the diffraction angles of the series of bands of constructive interference. For a three dimensional crystal, the spacing of the grooves corresponds to the lengths of the unit cell; higher order reflections can be rationalized by the description of a three dimensional array of stacked planes (►Miller planes) that are associated *via* the Miller indices (h, k, l) with each diffracted spot.

A crystal with a lone X-ray scattering atom in the asymmetric unit would produce a diffraction pattern characteristic of the unit cell size and symmetry. The



X-Ray Crystallography, Basic Principles. Figure 1

intensities of the diffracted spots would be uniform for a given scattering angle and monotonically decreasing for increasing scattering angles (depending on the atom type). More complicated molecules in the same unit cell (or asymmetric unit) would produce the same diffraction pattern but with more highly varied intensities. One way to visualize the origin of this phenomenon is to consider the diffuse diffraction pattern of the unit cell. The intensity of a particular diffraction angle is a function of the diffuse scattering and is identical for each unit cell. If a particular scattering angle does not correspond to the diffraction angle for a crystal lattice reflection, the waves from all the unit cells will interfere destructively resulting in zero intensity. If the angle does correspond to a crystal lattice reflection, there will be constructive interference to produce an amplified intensity proportional to the

intensity of the diffuse diffraction pattern. The result is that the discrete diffraction pattern of a crystal is effectively the diffuse diffraction of the unit cell sampled at positions determined by the crystal lattice. Crystalline disorder weakens and broadens the diffraction spots and limits the resolution of the atomic positions. The perfect crystal lattice does not exist – it cannot be infinite, for example – and damage and a variety of growth defects lower the quality of the crystal and the diffraction data. Several measures of crystalline disorder have been defined, including “mosaicity” – the extent to which a crystal should be described as a mosaic arrangement of smaller perfect crystals with small statistical variations in orientation and “temperature factor” – the broadening of reflections due to overall statistical disorder of the unit cell contents.

Crystal Growth

Crystallography begins with crystal nucleation and growth. Nucleation occurs after solution properties are changed so that the solution becomes supersaturated with the molecule to be crystallized. In this state, if attractive forces dominate during spontaneous collisions of the molecules, dimers and higher multimers are formed. These may continue to grow into crystals if the attractive forces between molecules in the initial aggregate are strong and specific enough to orient the molecules uniformly and if the specific orientations are compatible with a periodic lattice. Crystal needles and plates are formed when these conditions hold only in one or two dimensions respectively; growth in the other dimensions is much slower and often disordered. Precipitate is formed if the aggregates are not compatible with an ordered crystal lattice. Crystal growth continues until the solution is no longer supersaturated. The crystals can also be modified after they are grown.

Theoretical Principles

As described above, the diffraction pattern is a combination of the diffuse scattering from a unit cell and the constructive/destructive interference caused by the regularity of the crystal lattice. Each diffraction spot can be indexed (h,k,l) to the Miller plane associated with the angle of reflection. The measured intensities I are thus a function of the index and can be expressed as $I(h,k,l)$, and represent the squared amplitudes of the diffracted waves. The waves are characterized additionally by phases and are described by $F(h,k,l)$ or $F(\mathbf{h})$ for short, where the F values are **complex numbers** that account for both amplitudes and phases. This function is directly related by the Fourier transform to the function that describes the electron density distribution in the unit cell. The electron density function is commonly denoted $\rho(x,y,z)$ or simply $\rho(\mathbf{x})$. Because the intensities $I(h,k,l)$ of the diffracted waves are measured in crystallography and not the phases, calculation of the electron density requires supplemental phase information in order to derive the $F(\mathbf{h})$ function (and therefore the electron density distribution) from the intensities $I(h,k,l)$. This is the so-called phase problem of crystallography.

Details of the methods to estimate phases go beyond the scope of this article. Direct methods rely on necessary characteristics of the electron density (e.g. always positive) and on resulting phase relationships that exist among sets of reflections, leading to a finite set of possible phase combinations that can be systematically tested by computer. Because of the relatively small number of reflections, small molecule crystal structures are typically solved using direct method phase estimation. For protein crystallography, phases may be estimated using closely related protein structures

(molecular replacement or MR), measurements of multiple crystals that differ only by the presence of one or few fixed and strongly scattering atoms (multiple isomorphous replacement or MIR) or measurement of diffraction data at multiple wavelengths after incorporation of anomalous scattering atoms (multiple wavelength anomalous diffraction or MAD).

Successful phase estimation allows the calculation of an interpretable electron density map using the Fourier transform. This density allows the generation of an initial model in the case of MIR or MAD or correction of the initial model used in MR. The improved models improve the phase estimation for a second cycle of electron density calculation and model building. Depending on the resolution of the data and number of measured reflections, additional information, such as bond lengths and angles, may be used to maintain a chemically reasonable model. Protein structure refinement proceeds until there is no clear improvement of the structure.

References

1. Drenth J (1994) Principles of Protein X-ray Crystallography. Springer-Verlag, New York
2. Rhodes G (2000) Crystallography Made Crystal Clear, 2nd edn. Academic Press

X-Ray Data Collection from Macromolecular Crystals

ANDREA SCHMIDT, VICTOR S. LAMZIN
European Molecular Biology Laboratory (EMBL),
Hamburg c/o DESY, Hamburg, Germany
andrea@embl-hamburg.de
victor@embl-hamburg.de

Definition

The actual crystallographic experiment is data collection, i.e. the measurement of X-ray diffraction intensities from a crystal. Both the location (h , k and l coordinates in reciprocal space) and the value of the diffracted intensity are measured.

Description

X-ray data collection is central to crystallographic structure determination and represents the final experimental step. The manner in which the data collection is carried out will determine the data quality, which in turn will define the success of subsequent steps in structure determination and the quality of the macromolecular model. A number of considerations need to

be taken into account. These are briefly outlined in this contribution.

Protein Crystal Properties

Protein molecules are comparatively large and flexible. Once protein crystals are formed, the molecules pack loosely into the crystalline lattice and are surrounded by layers of solvent molecules from the crystallisation cocktail. The crystal is formed of identical building blocks, known as “unit cells” and may then be mathematically defined as a three-dimensional array of such unit cells. The size of the protein molecule defines their dimensions, so that larger proteins tend to produce crystals with larger unit cells. The number of reflections increases with the unit cell size and the diffracted intensities are inversely proportional to its volume. Thus, a larger number of reflections results in weaker diffraction intensities. As protein crystals contain a high amount of solvent (typically 20–80% (v/v)), there is a relatively high degree of mobility (atomic displacement) of some of the individual protein loops and side chains.

Data Collection Setup

The experimental setup for crystallographic data collection involves the X-ray source, the mounting system and the detector.

X-Ray Sources

In order to account for the diffraction properties of protein crystals, the X-rays used for data collection need to fulfil a number of requirements. Due to the generally weak diffraction of protein crystals, high intensity of the incoming radiation is a prerequisite for successful data collection. The experiments require the measurement of a potentially large number of reflections, which are densely arranged in reciprocal space. Such “crowding” can lead to an overlap of diffraction spots, resulting in loss of data for those reflections whose intensities are not uniquely defined. In order to minimise this effect the primary beam should have low divergence.

The wavelength of the X-rays determines the maximum resolution measurable with the experiment. Given the same detector, a shorter wavelength (higher energy) allows measurement of data to higher resolution. Crystals also deteriorate in quality due to the absorption of radiation in an energy-dependent manner. The use of shorter wavelengths is advantageous since it reduces the radiation absorption in the crystal.

While a rotating anode can be suitable for well-diffracting, stable crystals, the preferred choice of a radiation source for protein crystals is a synchrotron. This is due to the superior beam properties of synchrotrons, such as high intensity, low divergence and small focus size.

Mounting Techniques

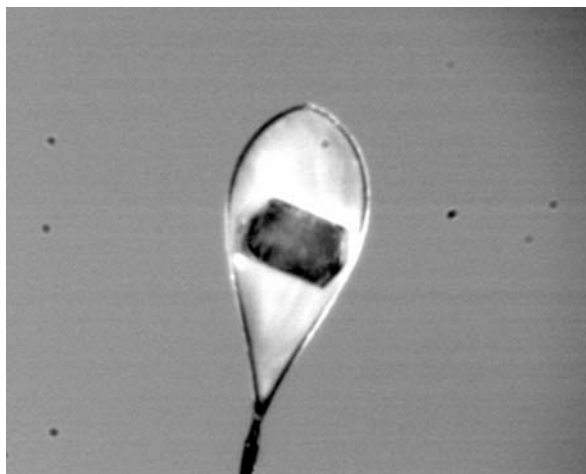
Prior to the data collection the crystalline sample should be mounted onto a beamline goniostat. There exist a variety of mounting techniques, which will not be described here. However, the important parameters are the data collection temperature and the type of goniostat used in the experiment. Typically, a protein diffraction experiment is carried out at cryogenic temperature (100 K), where the crystal is flash-cooled in a nitrogen stream. The use of cryogenic temperature reduces the effects of radiation damage and the crystal can be kept in the beam for a longer exposure. Cryogenic freezing is an almost absolute requirement for data collection at a synchrotron, due to the high radiation dose rates incident upon the crystals. An important issue for cryogenic data collection is the prevention of ice formation both in the solvent channels inside the crystal and on the surface of the sample. This is achieved by soaking the crystal in a “cryo-protectant” solution prior to mounting it on the goniostat ((1) and references therein). The crystal is then mounted into the beam in a small nylon fibre loop which holds the crystal encased in a thin film of cryo-protectant in a glass state (Fig. 1).

Owing to the large number of observations to be recorded for a protein crystal, it is not feasible to measure each reflection separately. The way in which data collection is typically carried out is to mount the crystal on a rotation axis (phi-axis), rotate it by a certain increment and record the diffraction pattern for each such slice on an area detector – a method commonly known as the “rotation method”.

Detector Types

Various detectors exist with different construction types and properties, which determine the suitability of a certain detector for any given experiment. Detectors should ideally have high detective quantum efficiency (DQE), a large dynamic range, linear response, high spatial resolution, large sized active area and uniform response and should accept high count rates. Examples of detectors commonly used in protein crystallography are image plates (IP) and charge-coupled device (CCD) cameras. The image plate consists of a support plate covered with a photosensitive layer and a protective cover (2). During the exposure, the phosphor layer stores the absorbed X-ray energy and emits upon read-out with a visible-light laser a phospholuminescence. The emitted intensity is proportional to the absorbed X-ray intensity at each point on the image plate (typical pixel sizes are 75–150 microns). The IP detectors are large and have a large dynamic range and low background noise, giving rise to a good DQE. However, the read-out and erasure after each image are often slow compared to the actual exposure time. In CCDs, the incoming X-rays cause visible light phospholuminescence. This creates “electron holes” in a

silicon layer *via* the photoelectric effect (3). The active surface consists of an arrangement of pixels (typically of a size of 50 microns) that accumulate the generated charge (electrons) in an electrostatic well. The charge is then shifted on a readout circuit and read as the signal. Although there is a certain amount of noise associated with readout and background generation of arbitrary photoelectrons, the speed and good spatial resolution of CCDs make them highly suitable for data collection. This is especially true at synchrotron sources where shorter exposure times are required.



X-Ray Data Collection from Macromolecular Crystals. Figure 1 Flash-cooled crystal mounted in a fibre loop for data collection.

Properties of the Diffraction Pattern

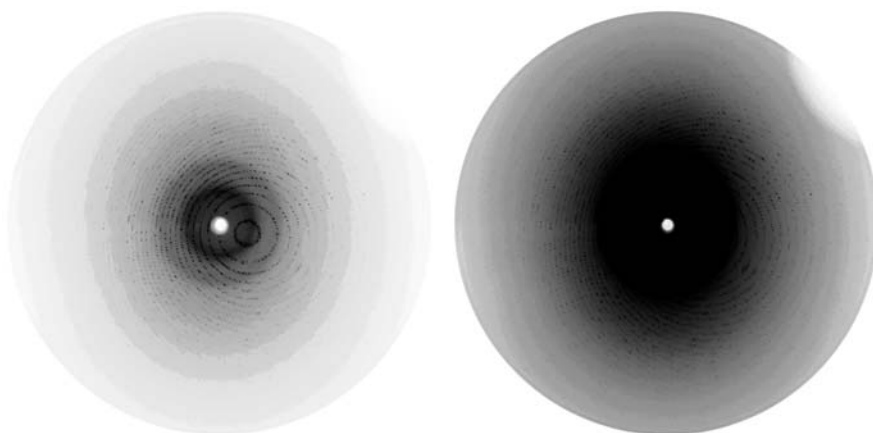
Diffraction images from protein crystals show a number of properties that have direct impact on the choice of data collection strategy. Due to the high mobility and the presence of disordered solvent, there is a steep intensity falloff with resolution (Fig. 2). The reflections are close to each other due to large cell parameters and their spots are broadened by the high mosaicity of protein crystals. Reflections arising from a set of lattice planes appear in circles on the detector plane (Fig. 3). The width of those circles or lunes is determined by the mosaicity of the crystal and the oscillation increment. A drawback of the rotation method is the presence of “blind regions” (Fig. 4), within which the reflections cannot be recorded. An additional feature is the presence of a region with higher background arising from the presence of solvent water.

Strategy

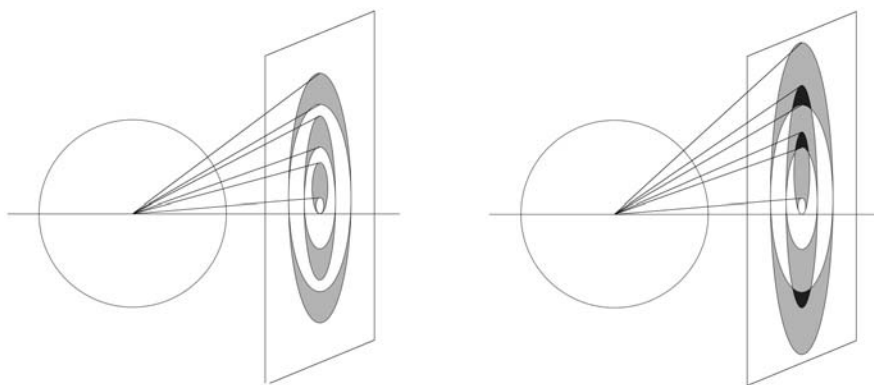
The aim of a good data collection strategy is to obtain a complete dataset while:

1. keeping the data collection time to a minimum
2. keeping the total exposure to a minimum in order to reduce the effects of the radiation damage
3. bringing the signal-to-noise ratio to a maximum in order to achieve the highest possible resolution and accurate data.

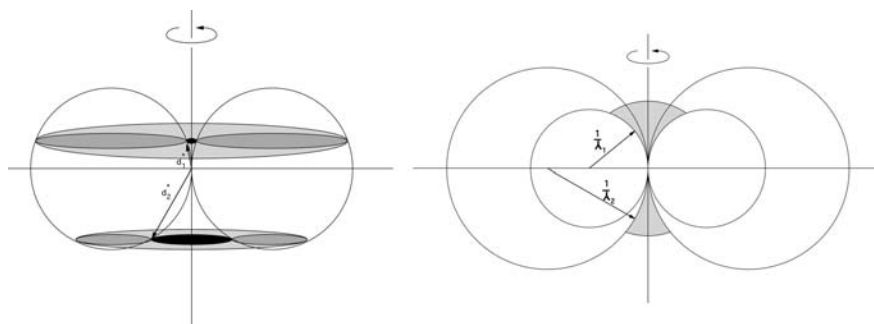
As a result, the special properties of protein crystal diffraction have to be taken into account. Additionally the available equipment (e.g. detector type), time constraints (at a synchrotron) and, naturally, the quality of the crystals have to be considered. The main parameters to adjust for a particular case are the



X-Ray Data Collection from Macromolecular Crystals. Figure 2 Diffraction pattern from a protein crystal showing the density of the diffraction spots and the steep intensity falloff. The picture on the left is displayed with a higher background threshold, the one on the right with a lower threshold.



X-Ray Data Collection from Macromolecular Crystals. Figure 3 Lunes in the rotation method of data collection. Left: correct setup with small oscillation increment; right: increment too large causing the lunes to overlap.



X-Ray Data Collection from Macromolecular Crystals. Figure 4 Schematic representation of the diffraction geometry indicating the blind regions dependent on resolution and wavelength. Left: black circles indicate blind regions for different resolution cutoff; right: blind regions at different wavelengths.

crystal-to-detector distance (and the crystal-to-beam-stop distance), the oscillation increment per frame, the exposure time per frame and the total oscillation range. For a high-quality structure, one should collect as high resolution data as possible. However, time and the risk of radiation damage have to be taken into account. The presence of overlapping reflections, overloads or an incorrect choice for the high resolution limit (detector being positioned too close or too far) will have a negative effect and often a compromise has to be found between the diffraction properties of the crystal and the capabilities of the hardware set-up. An excellent overview of the tools for the optimal choice of parameters can be found elsewhere (4).

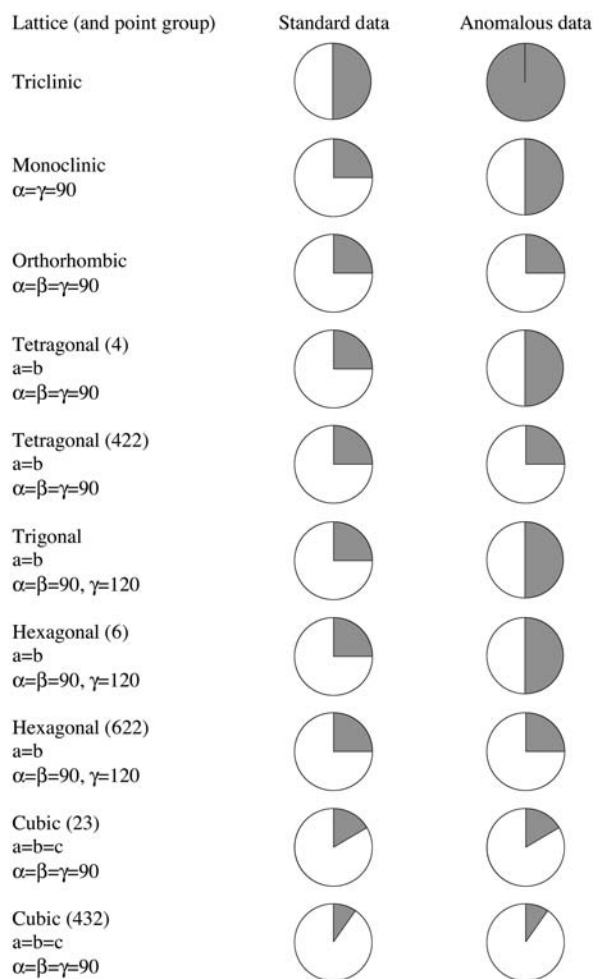
Asymmetric Unit

It is highly advisable to make use of the crystal orientation and crystal symmetry in order to minimise the exposure time and the total oscillation range. The required minimum regions in reciprocal space are given in Fig. 5. This is easy if the crystal is well oriented with

its cell axes corresponding to the “laboratory system” determined by: the spindle (x), the beam (z) and the vertical axis (y) perpendicular to x and z . Usually, the orientation of the crystal lattice is determined by indexing the first image with respect to the xyz coordinate system. For certain orientations of the crystal (e.g. c^* being parallel to the rotation axis), the required rotation range may be smaller than in Fig. 5. Because of the presence of blind regions (Fig. 4), the orientation of the crystal will also determine the completeness of the final dataset. In cases of low symmetry and perfect alignment, re-orientation is necessary to achieve full completeness.

Resolution and Detector Setup

In order to obtain the best possible signal-to-noise ratio and spot separation the detector should be positioned as far away as possible from the crystal while still collecting the highest resolution reflections. The diffraction limit is sometimes hard to determine by



X-Ray Data Collection from Macromolecular Crystals. Figure 5 Minimum required oscillation range for different crystalline lattices and point groups.

eye. Nowadays there are software strategy programs available which can greatly assist in the estimation of the resolution limit. The exposure time should be chosen so that the required resolution is achieved without (significant) radiation damage to the crystal.

Oscillation Increment and Mosaicity

Due to the high spot density in reciprocal space there is a risk of having overlapping spots from neighbouring lunes when the oscillation increment is too large (Fig. 3). Another criterion determining the oscillation width is the length of the unit cell axis normal to the spindle. The estimation for the maximum increment (in degrees) can be obtained using the expression: $\Delta\varphi = \frac{180d \cos \theta}{\pi a}$ where d is the highest resolution, a is the length of unit cell perpendicular to the spindle and θ is the maximum diffraction angle. Note that this formula does not account for the mosaicity and beam

divergence, both of which make the lunes wider. The choice of the oscillation width will also determine how many reflections will be fully recorded and how many will be partially recorded (i.e. on two or more consecutive images). A larger fraction of fully recorded reflection is advantageous for the scaling of images. However, a large amount of partially recorded reflections is advantageous for the post-refinement. Here the reader is referred to a specialised literature (5).

Exposure Times and Multiple Passes

The exposure time per image is the primary factor, which ultimately determines the signal-to-noise ratio of the data. Collection of higher resolution data requires longer exposure times. Due to the steep intensity falloff and the inherent limits of the dynamic range of the detector, the strong reflections at low resolution may often be overloaded and show truncated profiles. However, it cannot be overstated that these reflections have the highest information content and need to be recorded accurately. A solution is to split the data collection in two (or more, if required) passes, a high-resolution pass, which will extend to the maximum desired resolution and a low-resolution pass to record otherwise overloaded reflections. The low-resolution pass should be set up so that it covers the same range in reciprocal space and overlaps sufficiently with the high-resolution pass and the exposure time should be appropriately shortened. A rule of thumb is to move beamstop and detector together; in the high-resolution pass the beamstop should be close to the crystal in order to reduce the background. In the low-resolution pass the beamstop should be placed further from the crystal, so that it does not obscure the low-resolution reflections.

Data Quality and Statistics

The quality of the collected data should be inspected immediately, ideally while the crystal is still on the goniostat. There are several parameters that can be used to assess the quality of the data set and to determine the high-resolution limit. The most traditional is the R_{sym} or R_{merge} , although its value depends on the data redundancy (the number of times each unique reflection has been measured). More elaborate indicators use redundancy-corrected R-factors ($R_{\text{r.i.m.}}$, $R_{\text{p.i.m.}}$) (6) or the $I/\sigma(I)$ ratio, under the assumption that the standard uncertainties are estimated correctly.

Automatic Data Collection

Over recent years there has been an increasing tendency towards higher throughput, faster data collection, rapid screening and on-line assessment of crystal quality. Many of the decision-making steps during the data collection can be done by the appropriate software. One example is the package

BEST (7) which analyses an initial image (or a set of images), derives the crystal orientation and estimates the intensity falloff and the diffraction limit of the crystal. The package then proposes a data collection strategy based on the diffraction properties, resolution requirements and further data quality indicators.

Special Cases

Multi-Wavelength Anomalous Dispersion

Multi-wavelength anomalous dispersion has become a very popular method of gathering experimental phase information. It involves data collection on a structure containing heavy atoms (artificially introduced or naturally present in a macromolecule) at different wavelengths. These heavy atoms scatter “anomalously”, resulting in a breakdown of Friedel’s law. The phase information is extracted from the small differences in the intensities of reflections that form a Friedel pair (also called Bijvoet pair). For the data collection setup, this means that:

1. an additional experiment (a fluorescence scan) is needed to determine the X-ray energy at which the data collection should be carried out,
2. the Friedel pairs should be recorded for each reflection,
3. the measurement of the intensities needs to be sufficiently accurate – which can be achieved by high data redundancy and
4. the crystal should ideally be stable throughout the data collection at different wavelengths.

In order to obtain a complete anomalous dataset, one may need to increase the total rotation range (Fig. 5). The advantage of MAD is that all datasets come from one crystal, thereby bypassing the problem of non-isomorphism and of scaling various datasets together. Naturally, as tuneability of the X-ray source is required, such data collections are carried out at synchrotron beamlines. The use of one wavelength only (SAD) does not require tuneability and in many cases provides phase information of a quality comparable to that from MAD.

Conclusions

With the continuously increasing usage of X-ray crystallography, the number of macromolecular structures deposited in the PDB, their size and the complexity are rapidly increasing. It is important, that

the data for these structures are obtained using the state of the art technologies.

References

1. Garman E (1999) Cool data: quantity AND quality. *Acta Cryst* 55:1641–1653
2. Amemiya Y (1997) X-ray storage-phosphor imaging-plate detectors: high-sensitivity X-ray area detector. In: Carter CWJ, Sweet RM (eds) *Methods in Enzymology*. Academic Press, pp 233–243
3. Westbrook EM, Naday I (1997) Charge-coupled device-based area detectors. In: Carter CWJ, Sweet RM (eds) *Methods in Enzymology*. Academic Press, pp 244–268
4. Dauter Z (1997) Data Collection Strategy. In: Carter CWJ, Sweet RM (eds) *Methods in Enzymology*. Academic Press, pp 326–358
5. Evans P (1993) Data reduction. *Proceedings of the CCP4 study weekend*, pp 114–122
6. Weiss MS (2001) Global indicators of data quality. *J Appl Cryst* 34:130–135
7. Popov AN, Bourenkov GP (2003) Choice of data-collection parameters based on statistic modelling. *Acta Cryst* 59:1145–1153

XRN1 and XRN2

Definition

XRN1 and XRN2 designate exonucleases that degrade uncapped RNA substrates in a 5′ – 3′ direction.

XX Male Syndrome

►SRY – Sex Reversal

XY Gonadal Dysgenesis

►SRY – Sex Reversal

YAC and PAC Maps

LEONARD C SCHALKWYK

Social, Genetic and Developmental Psychiatry,
Institute of Psychiatry, London, UK
spjglcs@iop.kcl.ac.uk

Synonyms

Contig maps

Definition

YAC is an acronym for yeast artificial chromosome, a linear DNA construct of potentially unlimited size containing autonomously replicating (ARS) sequence(s), a centromere (CEN), telomere (TEL) sequences, one or more ►selectable markers capable of functioning in yeast (*Saccharomyces cerevisiae*) as a ►vector, and an arbitrary DNA insert.

PAC stands for bacteriophage P1 artificial chromosome, named by analogy with YACs. These are circular constructs similar to bacterial artificial chromosomes (BACs), but using vectors derived from bacteriophage P1. As with YACs, libraries are made using partial restriction digests of genomic DNA and ►pulsed-field gel electrophoresis, but are then introduced into *Escherichia coli* by ►electroporation.

YAC and PAC maps are representations of a genome, chromosome or chromosomal region based on overlapping cloned DNA fragments.

A contig is a group of ►genomic clones (or more recently DNA sequence fragments) joined by a path of overlaps, which represent a contiguous stretch of genomic DNA.

FISH is an acronym for fluorescent *in-situ* hybridization.

Characteristics

History

The first genomic-clone-based maps were generated by ►genome walking, in which ►colony or plaque lift filters of a ►cosmid or ►bacteriophage lambda library are repeatedly screened with probes isolated from ►clones isolated in previous steps. This approach was

crucial in some early triumphs of genomics, such as the positional cloning of the cystic fibrosis gene, but is by its nature slow and arduous. In the mid-1980s several groups had the insight that entire clone libraries could be picked into ►microtiter dishes and maintained as a series of frozen replicas, not only revolutionizing ►library screening, but also allowing the development of multiple-entry-point, parallel mapping approaches. The field was revolutionized again by the advent of PCR, which opened up many more ways of determining overlaps between clones, in particular efficient screening of multidimensionally pooled representations of libraries.

An analysis of mapping methods (5) made clear that the size of the clones and the sensitivity for detection of clone overlaps were the key criteria governing the progress of mapping projects.

Clone and Library Characteristics

In the original (1) and most subsequent YAC vectors, wild type URA and TRP genes are used as selectable markers (in an appropriately auxotrophic yeast background). ARS1 and CEN4 are adjacent to TRP1 and so these are used as a cassette. Sequences with ARS function are also common in mammalian DNA. These vectors also contain bacterial ►replication origins and antibiotic markers so that they can be propagated (in the empty state) in *E. coli*. The bacterial fragments are also used in methods for cloning YAC insert ends. To produce a library of YAC clones, partial restriction digests (most commonly EcoRI) of very high molecular weight genomic DNA are typically size fractionated by pulsed-field gel electrophoresis, ligated with arms prepared from the (*E. coli*-propagated) vector, further purified and transformed into yeast by a polyethylene-glycol-sphaeroplast method. Colonies are picked into microtiter plates and maintained frozen. Efficiency considerations such as those highlighted by Lander and Waterman focused attention on YACs, which are the cloning system with the largest insert size by far. Numerous YAC libraries with insert size averages around 1Mb have been produced for man and other species. These libraries were key reagents in positional cloning and in larger scale genome mapping. These large scale mapping projects had to take account of the characteristics of YAC clones, which present

several challenges. YAC libraries are particularly difficult to produce and only a few laboratories have succeeded in producing megabase-sized YAC libraries. For human, mouse and rat genome studies, this had the benefit of encouraging the use of shared resources, which facilitated the integration of data between laboratories. This custom has continued with PAC and BAC libraries and these are available from both commercial sources and academic resource centers such as the German RZPD (<http://www.rzpd.de/>), the British HGMP (<http://www.hgmp.mrc.ac.uk/geneservice/>) and CHORI (<http://bacpac.chori.org/>) in the USA.

The main limitation on producing YAC clones is the inefficiency of ►transformation of yeast with large DNA fragments. The only way to produce usable numbers of clones is to transform with high concentrations of a ligation mixture, which necessarily contains quantities of unligated or broken genomic DNA. This material activates recombination in yeast, resulting in the assembly of numerous ►chimeric clones. Chimerism rates of the order of 50% were found for nearly all libraries. There were partially successful attempts to use recombination-deficient yeast strains, but none produced megabase range insert sizes. A related problem is that unlike in bacterial cloning systems, there is no ►exclusion mechanism for yeast chromosomes, so multiple cloned fragments of DNA can coexist in a cell.

The lack of host strains with an effectively crippled recombination system, combined with the high frequency of repeated sequence in mammalian DNA results in further problems. Internal deletions are more difficult to detect than chimerism, but are likely to be quite common in YAC libraries. As a host, yeast is relatively tractable and easier to keep alive than *E. coli*, persisting for long periods at 4 degrees or frozen at -20, but it does not offer an easy way of isolating the cloned DNA from the host background. This can be done only by pulsed-field gel electrophoresis and generally requires the clones to be sized by ►Southern blot beforehand.

PACs, on the other hand, take advantage of the *E. coli* hosts much more familiar to molecular biologists. The clones (ten-fold smaller than YACs) can readily be separated from the *E. coli* background by refined alkaline lysis methods and can be used for ►hybridization probe preparation, subcloning or direct sequencing. They are part of a long series of cloning systems striving to produce ever larger clone inserts in *E. coli*, which may have reached its limit with BAC and PAC clones. These systems have many similarities, and high quality libraries of both types have insert size averages of 100–150kb, sometimes approaching 200kb. This to a large extent reflects the properties of the electroporation method used to introduce the ligated DNA into the

host. It is important to remember that with YAC, PAC and BAC cloning systems there is no upper limit on the clone size, simply a decreasing probability with size. Individual clones from 0.25–4 times the average size are commonly found.

The earliest libraries in *E. coli* hosts were produced by empirical methods of enhancing transformation. These methods can be quite efficient for small constructs, but libraries with insert size averages above 20kb cannot practically be produced. To get around the limits of transformation, several cloning systems were devised that used the DNA packaging machinery of bacteriophage lambda *in vitro*, so that the DNA could be delivered by ►transfection with very high efficiency. The lambda phage head naturally holds 48.5kb of DNA. The earliest vectors supplemented the phage DNA with a few kb of insert or replaced the entire central section of the phage DNA (functions concerned with regulation of virulence and lysogeny). These were propagated as bacteriophage, but it was realized that only the bacteriophage DNA ends are required for packaging and the system could be used to deliver practically the full DNA capacity of the bacteriophage head into the cell, to be maintained as a plasmid. These cosmid vectors, with a capacity of 40kb, have been workhorses of genome mapping. Many lessons about how to maintain mammalian DNA stably came from problems with early cosmid libraries. The key points, in addition to avoiding ►library amplification in liquid culture, proved to be low copy number, selection with a stable antibiotic such as kanamycin and recombination-deficient hosts. The most successful cosmid vectors, the lambda-origin loris/lawrist series, served as the model for the bacteriophage P1 cloning system, which takes advantage of bacteriophage P1's larger cloning capacity. The advent of electroporation, at least tenfold more efficient than chemical transformation methods, allowed DNA to be introduced by a straightforward physical method. PAC cloning (4) essentially uses the P1 vector, but abandons the P1 packaging method in favor of introducing the cloned DNA into cells by electroporation, exactly as is done in BAC cloning.

Mapping Methods

Production of contig maps, whatever the cloning system used, is expensive and must be done with carefully considered strategy. Three major strategic choices need to be considered; whether the map is to be global or regional, the type and number of clones to be analyzed and what method of detecting overlaps is to be used. A further consideration is how the map will be linked to (or based upon) other information such as genetic or ►radiation hybrid maps.

In making a regional map, a chief task is identifying those clones from a whole-genome library that

represent the region of interest. Although in principle chromosomes or large restriction fragments can be isolated by flow cytometry or pulsed field electrophoresis, in practice successful large-insert libraries, especially in the YAC and PAC size range, have generally been produced only from whole-genome DNA. Countless regional maps have been produced as part of the process of positional cloning. Whole genome maps have been produced for human (2, 3), mouse (6, 7) and other species, largely with the objective of making positional cloning easier, but also as a starting point for genome sequencing. Whole genome maps are in principle more efficient to produce since they can be organized on more highly parallel, industrial principles and because the initial selection of region-specific clones can be avoided. The undertaking of producing such a map on a whole genome scale for a mammalian sized genome is a very large one and progress towards a complete, error-free map is inevitably asymptotic. As a consequence, whole genome maps inevitably contain numerous, often small, gaps and discrepancies and can be viewed as a draft, which can be refined as necessary when particular regions are of interest.

Genomic libraries are thought of as having a notional 'coverage' which is the average insert length, multiplied by the number of clones available, expressed as a multiple of genome size ('fold coverage'). The probability of a given sequence being present at least once in such a library can be calculated using the Clarke-Carbon equation $1-(1/e)^n$, where n is the fold coverage. This gives 95% of sequences available from 3-fold coverage, and over 99% from 5-fold. Based on these numbers one might arrive at a serious underestimate of what is required to produce a long-range contig map. Much higher levels of redundancy are required to make the mapping process robust to various departures from ideality. One of the most obvious and troublesome of these is false joins, which result in apparent branching of contigs. Every mapping method will produce some proportion of false positive linkages between clones, and chimeric clones give rise to experimentally correct results that nonetheless embody further (with many libraries quite numerous) false joins. In 'deep' (i.e. highly redundant) contigs, however, these false connections (each of which is unique) will be outnumbered by correct sequences. High coverage also reduces the number of genuine overlaps that are missed and the number of poorly represented segments, which result in gaps in the map. A sensible rule of thumb would probably be 20-fold coverage. This would of necessity be composed of several libraries, perhaps of different types, taken together. Although this adds to the logistical difficulty of producing a map, it is likely to further improve the

robustness of the mapping process, as different libraries will have complementary strengths.

Methods of detecting overlaps between clones fall into two broad groups, landmark and fingerprint. In landmark methods, clones are assayed for a short sequence (which can be thought of as a single point). This could be a PCR amplicon or a short (<5 kbp) hybridization probe. One way of generating such probes (and reduced-complexity targets as well) is inter-repeat-sequence (IRS) PCR, which uses PCR primers complementary to the ends of common repeated sequences to amplify sequences flanked by such repeats. Landmark assays tend to be simple and scalable and distinctive enough so that if two clones have a single landmark in common their overlap is highly probable. In fingerprint methods a more diffuse form of information (typically restriction fragment sizes) is obtained and the certainty of identifying an overlap depends on its size. Restriction fingerprint methods are only practical for cosmid, PAC and BAC clones, although fingerprint methods based on inter-repeat-sequence PCR can work with YACs. Both types of approaches have been used with success and questions of operational efficiency in data collection and analysis are at least as important as the efficiency of the experimental method.

Clinical Relevance

The human DNA sequence has become the canonical genome map and primary reference for identifying genes relevant to disease and disease susceptibility, superseding YAC, PAC and other contig maps upon which the sequence was largely built. YAC and PAC maps still have relevance for resolving problems in sequence assembly. YACs and PACs from such maps can be used as hybridization probes for FISH, which can be used to diagnose chromosomal abnormalities. YACs and PACs are important in constructing transgenic mice. Many genes span PAC, YAC or larger sized regions in the genome and some amount of flanking sequence is generally necessary for normal expression. A YAC, PAC (or BAC) from a well-characterized map is thus an ideal reagent for making transgenics.

References

- 1) Burke DT, Carle GF, Olson MV (1987) Cloning of large segments of exogenous DNA into yeast by means of artificial chromosome vectors. *Science* 236:806–812
- 2) Cohen D, Chumakov I, Weissenbach J (1993) A first-generation physical map of the human genome. *Nature* 366:698–701
- 3) Hudson TJ, Stein LD, Gerety SS et al (1995) An STS-based map of the human genome. *Science* 270:1945–1954

- 4) Ioannou PA, Amemiya CT, Barnes J et al (1994) A new bacteriophage P1-derived vector for the propagation of large human DNA fragments. *Nat Genet* 6:84–89
- 5) Lander ES, Waterman MS (1998) Genomic mapping by fingerprinting random clones: a mathematical analysis. *Genomics* 2:231–239
- 6) Nusbaum C, Slonim DK, Harris KL et al (1999) A YAC-based physical map of the mouse genome. *Nat Genet* 22:388–393
- 7) Schalkwyk LC, Cusack B, Dunkel I et al (2001) Advanced integrated mouse YAC map including BAC framework. *Genome Res* 11:2142–50

YAC Transgene

► YAC/PAC/BAC Transgene

YAC/PAC/BAC Transgene

Definition

YAC/PAC/BAC transgene designate large DNA fragments derived from Yeast (YACs), P1-derived (PACs) or Bacterial artificial chromosomes (BACs). YAC/BAC/PAC transgenes improve stability of transgene integration and predictability of their expression.

► Transgenic and Knockout Animals

YACs

► Yeast Artificial Chromosomes

YAG

Definition

Yttrium-aluminium-garnet (YAG) is a solid-state crystal that is composed of yttrium and aluminium oxides. Doped by rare-earth ions such as neodymium (Nd:YAG) or Erbium (Er:YAG) it is used as the active medium in YAG lasers.

► Mass Spectrometry: MALDI

Y-Box

Definition

Y-box describes a promoter element that generally contains a central attgg sequence and interacts with a family of transcription factors, called y-box binding proteins.

► Tight Junctions

Yeast Artificial Chromosomes

Definition

Yeast artificial chromosomes (YACs) are vectors that contain an origin of replication, and a centromere from yeast that behaves like a chromosome. The size of the inserts incorporated (usually 100–200 kb) in this vector cloning system results in a good coverage of long genomic regions. The major disadvantage of this system is that they commonly undergo deletions, rearrangements and chimerism.

► *C. elegans* Genome, Comparative Sequencing

Yeast Two-Hybrid Analyses

Definition

Yeast two-hybrid analysis (yeast two-hybrid system) refers to a yeast genetic assay that uses a reporter gene to detect the physical interaction of a pair of proteins in a yeast cell nucleus. The principle of this method lies in the fact that when the target protein binds to its interacting partner, the interaction combines the two halves of a transcriptional activator, which then switches on the expression of the reporter gene. The assay detects the binding of two hybrid proteins: a bait (fusion between a DNA binding domain of a transcription factor and the protein of interest) and a prey (fusion between an activation domain of a transcription factor and the protein to be tested).

► *C. Elegans* as a Model Organism for Functional Genomics

► *C. elegans* Genome, Comparative Sequencing

► Peptide Aptamers

► Proteomics in Human-Pathogen Interactions

► Protein Microarrays as Tool for Protein Profiling and Identification of Novel Protein-Protein Interactions

Z-Disc

Definition

Z-disc refers to a multiprotein lattice that separates the sarcomeres in a myofibril.

- [Limb Girdle Muscular Dystrophies](#)

Zebrafish

Definition

Zebrafish is a tropical fish species, which as a model animal is used for studies of genetics and developmental biology.

- [Axis Formation – Formation and Function of the Dorsal Organizer](#)
- [Mutagenesis Approaches in the Zebrafish](#)

Zellweger Syndrome

Definition

Zellweger syndrome is an autosomal recessively inherited syndrome with anomalies of peroxisomes in liver and kidneys. It is characterized by muscular hypotrophy seizures, severe psychomotoric retardation and facial anomalies.

- [Cerebro-Hepato-Renal Syndrome](#)
- [Peroxisomal Disorders](#)

Zero Frame

Definition

Zero frame refers to the open reading frame extending downstream from any translation initiation site.

- [Translational Frameshifting, Non-standard Reading of the Genetic Code](#)

Zero-Length Cross-Linkers

Definition

Zero-length cross-linkers are chemical reagents that chemically activate a molecule, enabling it to couple with another molecule directly through covalent bonding.

- [Protein Interaction Analysis: Chemical Cross-Linking](#)

Zinc Finger Domain

Definition

The zinc finger domain is a DNA-, RNA-or protein-binding protein motif which creates a finger-like loop. The classical C2H2 Zinc finger binds to zinc via two cysteine(C) residues located one in each beta-sheets, and two histidine (H) residues located in an alpha-helix. Two or more of these motifs facilitate sequence specificity in DNA binding.

- [Bone and Cartilage](#)
- [Chromatin Acetylation](#)
- [DNA Ligases](#)
- [NFκB Pathway](#)
- [Nuclear Hormone Receptors: Regulators of Gene Transcription and Cellular Signaling](#)
- [Protein/DNA Interaction](#)

ZIP

- [Leucine Zipper](#)

ZNF202

Definition

The zinc finger protein ZNF202 is a transcriptional repressor that binds to promoter elements predomi-

nantly found in genes involved in lipid metabolism, thus playing a critical role in high-density lipoprotein homeostasis.

- ▶ Hereditary Neuropathies, Motor and/or Sensory
- ▶ Tangier Disease
- ▶ Transcription Factors and Regulation of Gene Expression

ZO

- ▶ Zona Occludens

Zona Occludens

Definition

- ▶ PDZ Domain

Zona Pellucida

Definition

Zona pellucida designates a relatively thick extracellular coat composed of glycoproteins that surrounds all mammalian eggs and performs important functions during oogenesis, fertilization, and preimplantation development.

- ▶ Mammalian Fertilization

Zona Reaction

Definition

Zona reaction refers to a modification of the ▶ [zona pellucida](#) in response to fertilization, which alters the solubility and binding properties of the zona pellucida, and leads to a slow block to polyspermy.

- ▶ Mammalian Fertilization

Zone of Polarizing Activity

Definition

Zone of Polarizing Activity (ZPA) comprises of a group of mesenchymal cells that are located in the posterior

part of the limb bud. The ZPA organises structures within the limb along the anterior-posterior axis (thumb to little finger).

- ▶ Limb Development

Zonula Adherens

Definition

Zonula adherens denotes an adhesive junction that is located at the apico-lateral border of the epithelial layer, located basal to the tight junctions.

- ▶ Adherens Junctions
- ▶ Adhesion Molecules

ZPA

- ▶ Zone of Polarizing Activity

Zygote

Definition

The zygote is a one-cell embryo resulting from fertilization of an egg by the sperm, that is, fusion of the female germ cell with the male germ cell.

- ▶ Cre/Lox P Strategies

Zygotene

Definition

Zygotene describes a meiotic division stage that is characterized by the beginning of chromosome pairing, and becomes twisted around one another (relational coiling).

- ▶ Meiosis and Meiotic Recombination

Zygotic Transcript

Definition

Zygotic transcript refers to an mRNA that is expressed by the transcriptional machinery of the embryo or adult.

- ▶ Morpholino Oligonucleotides, Functional Genomics by Gene ‘Knock-down’

List of Entries

Essays are shown in bold

- ▶ AAA ATPase
- ▶ AAV
- ▶ Abasic Site
- ▶ ABCB1
- ▶ ABC Transporters
- ▶ Absorption Coefficient
- ▶ Absorption Edge
- ▶ Absorptive Intestinal Cells
- ▶ ACAMP
- ▶ Acamprosate
- ▶ Acantholysis
- ▶ Acanthosis
- ▶ Acceptor
- ▶ Accessible Surface Area
- ▶ ACE
- ▶ ACE Inhibitor
- ▶ Acetabular Protrusion
- ▶ Acetal Linkage
- ▶ Acetyl Choline Esterase
- ▶ AchE
- ▶ Achondrogenesis
- ▶ Achondroplasia
- ▶ Acousto-Optical Filters
- ▶ Acousto-Optical Modulator
- ▶ Acousto-Optical Tunable Filter
- ▶ Acquired Immunity
- ▶ Acquisition Time
- ▶ Acrocentric Chromosome
- ▶ Acrosome
- ▶ Acrosome Reaction
- ▶ ACTH
- ▶ Actin
- ▶ **Actin Cytoskeleton**
RAJESH ARASADA, MICHAEL SCHLEICHER
- ▶ Actin Filaments
- ▶ Actin Polymerisation
- ▶ Actin Related Proteins
- ▶ Action Potential
- ▶ Action Potential Backpropagation
- ▶ Activation (of a Molecule)
- ▶ Activators of G Protein Signaling
- ▶ Active Site
- ▶ Activin
- ▶ Activin Receptor-Like Kinase
- ▶ Actomyosin
- ▶ ActR-IB/-IC
- ▶ ActR-IIA/-IIB
- ▶ **Acute Intermittent Porphyria**
ULRICH GROSS
- ▶ Acute Lymphoblastic Leukaemia
- ▶ Acute Myeloid Leukaemia
- ▶ Acute Phase Proteins
- ▶ Acute Phase Reaction
- ▶ Acute Promyelocytic Leukaemia
- ▶ Acyltransferase
- ▶ AD
- ▶ ADAM
- ▶ Adaptive Immunity
- ▶ Adaptor Complexes
- ▶ Adaptor Protein Complexes
- ▶ Adaptor Proteins
- ▶ **Addiction, Molecular Biology**
WALTER ZIEGLGÄNSBERGER
- ▶ Addison's Disease
- ▶ Adeno-Associated Virus
- ▶ Adenoma
- ▶ Adenomas
- ▶ Adenomatous Polyposis Coli
- ▶ Adenylyl Cyclase
- ▶ ADH

- ▶ ADHD
- ▶ **Adherens Junction**
YASUYUKI FUJITA, CATHERINE HOGAN
- ▶ **Adhesion Molecules**
KERSTIN DANKER
- ▶ Adipogenesis
- ▶ Adiponectin
- ▶ Adjuvant
- ▶ ADMET or ADME / Tox
- ▶ Adoptee's Family Studies
- ▶ ADPKD (Autosomal Dominant Polycystic Kidney Disease)
- ▶ Adrenal Insufficiency
- ▶ Adrenocorticotropin
- ▶ Adrenoleucodystrophy
- ▶ Adrenomyeloneuropathy
- ▶ Adult Neurogenesis
- ▶ Adult Onset Polycystic Kidney Disease
- ▶ Adult Stem Cells
- ▶ AER
- ▶ Affinity
- ▶ Affinity Chromatography
- ▶ **Affinity Chromatography and *In Vitro* Binding (Beads)**
KONRAD BÜSSOW
- ▶ Affinity Constant
- ▶ Affinity Separation
- ▶ Affinity Tag
- ▶ Agarose
- ▶ Aggrecan
- ▶ Aggregation (Chimera)
- ▶ Aggregophore
- ▶ Aging
- ▶ Agonist
- ▶ Agouti
- ▶ AGS
- ▶ AGT
- ▶ AHO
- ▶ AIDS
- ▶ AJ
- ▶ Akt
- ▶ Alagille Syndrome
- ▶ Albright's Hereditary Osteodystrophy
- ▶ Alcohol Abuse, Consequences
- ▶ ALD
- ▶ Aldosterone
- ▶ ALFPs
- ▶ ALK
- ▶ Alkaline Phosphatase
- ▶ Alkanethiol
- ▶ **Alkyltransferases**
BERND KAINA, MARKUS CHRISTMANN
- ▶ ALL
- ▶ Allele
- ▶ Allele-Specific
- ▶ Allelic Association
- ▶ Allelic Series
- ▶ Allergic Disorders, Genetics
- ▶ Allergic Drug Reactions
- ▶ Allergy
- ▶ Allostery
- ▶ Allotetraploid
- ▶ Allozygous
- ▶ Alpha-Antitrypsin
- ▶ Alpha-Glucosidase/Alpha-Mannosidase
- ▶ Alpha-Helix (α -Helix)
- ▶ Alpha-Mannosidase
- ▶ Alpha-Oxidation
- ▶ Alpharetroviruses
- ▶ Alpharetroviruses CA (Capsid)
- ▶ Alpha-Tubulin / Beta-Tubulin
- ▶ Alphoid DNA
- ▶ **Alternative Splicing**
STEFAN STAMM
- ▶ Alu Repetitive Elements
- ▶ Aly/REF
- ▶ Alzheimer Dementia
- ▶ **Alzheimer's Disease**
ULRICH FINCKH
- ▶ AMACR Deficiency
- ▶ Amadori Rearrangement
- ▶ AME
- ▶ AMH / MIS
- ▶ AMHR-II
- ▶ Amide Bonds
- ▶ Amide-Linkage
- ▶ **Amino Acids: Physicochemical Properties**
EPHRAIM KATCHALSKI-KATZIR,
RONI KASHER, MATI FRIDKIN
- ▶ AML
- ▶ AMN
- ▶ AMP

- ▶ Amphetamine
- ▶ Amphipathic
- ▶ Amplification
- ▶ Amplified Fragment Length Polymorphic Markers
- ▶ Amygdala
- ▶ Amyloid
- ▶ Amyloidosis
- ▶ Amyotrophy
- ▶ **Analytical Ultracentrifugation**
PETER SCHUCK
- ▶ Anaphase
- ▶ Anaphase Movement
- ▶ Anaphase Promoting Complex (APC)/Cyclosome
- ▶ ANAs
- ▶ Anchorage Dependence/Independence
- ▶ Androgen Insensitivity
- ▶ Androgen Receptor (Gene)
- ▶ Aneuploidy
- ▶ Angelman Syndrome
- ▶ Angiofibroma
- ▶ **Angiogenesis**
KARL H. PLATE
- ▶ Angiogenesis Factors
- ▶ Angiogenesis-Dependent Diseases
- ▶ Angiomyolipoma
- ▶ Angiotensin Converting Enzyme
- ▶ Ångström (Å)
- ▶ Animal Cap
- ▶ Animal Model
- ▶ Anionic C-terminal Carboxylate
- ▶ Ankyrin Repeat
- ▶ Anlage
- ▶ Annotation
- ▶ Anomalous Diffraction, Anomalous Scattering
- ▶ Anomalous Scatterer
- ▶ Anomalous Scattering
- ▶ Anomer
- ▶ ANT1
- ▶ Antagonist
- ▶ ANT-C
- ▶ Antennapedia Complex
- ▶ Anterior Horn Cells
- ▶ Anthracyclines
- ▶ Antibody
- ▶ Antibody Array
- ▶ Antibody Chip
- ▶ Antibody Cross Reactivity
- ▶ Anticipation
- ▶ Anticodon
- ▶ Anti-Craving Substances
- ▶ Antigen
- ▶ Antigen Expression
- ▶ Antigen Receptors
- ▶ Antimüllerian Hormone
- ▶ Anti-Nuclear Autoantibodies
- ▶ Antiparallel
- ▶ Antiretroviral Agents
- ▶ Antisense (Oligonucleotides)
- ▶ Antisense RNA
- ▶ Antisera
- ▶ AOM
- ▶ Aorta
- ▶ AOTF
- ▶ AP Endonuclease
- ▶ AP-1
- ▶ APC
- ▶ APC
- ▶ APC/C
- ▶ APCI
- ▶ APH-1
- ▶ Apheresis
- ▶ Apical Ectodermal Ridge
- ▶ Apical Junctional Complex
- ▶ Apical Membrane
- ▶ APL
- ▶ APOE
- ▶ APO-2 ligand
- ▶ apoE Null (Mutant, Knockout) Mice
- ▶ Apolar
- ▶ Apolipoprotein
- ▶ Apoptosis
- ▶ **Apoptosis, Regulation and Clinical Implications**
HANS-UWE SIMON,
UWE ZANGEMEISTER-WITTKÉ
- ▶ Apoptosome
- ▶ APP/*App*
- ▶ Apparent Mineralocorticoid Excess
- ▶ Appositional Growth
- ▶ APR

- ▶ Aptamer
- ▶ Apurinic or Apyrimidinic (AP) Sites
- ▶ AQP2
- ▶ Aquaporin
- ▶ Aquaporin–2
- ▶ Arabinogalactan
- ▶ Arachnodactyly
- ▶ Archimedes Principle
- ▶ ARE
- ▶ ARF
- ▶ Armadillo
- ▶ Arp2/3 Complex
- ▶ Array
- ▶ ARS
- ▶ ARVs
- ▶ Arylsulfatase A
- ▶ Ashkenazi
- ▶ Asn-Linked Glycosylation
- ▶ Assembly
- ▶ Assisted Reproduction
- ▶ Association
- ▶ Association Studies
- ▶ Association-Dissociation Equilibrium
- ▶ Aster
- ▶ Asthma
- ▶ Astral Microtubule
- ▶ Astrocyte
- ▶ Asymmetric Carbon Atom
- ▶ Asymmetric Cell Division
- ▶ Asymmetric Unit
- ▶ Asynchronous Replication
- ▶ Ataxia
- ▶ Ataxia Telangiectasia (AT)
- ▶ Ataxia Telangiectasia-Like Disease
- ▶ Ataxia Teleangiectasia Mutated Gene
- ▶ Atelosteogenesis
- ▶ AT-LD
- ▶ ATM
- ▶ **Atomic Force Microscopy**
FILIPP OESTERHELT
- ▶ **Atopy, Genetics**
SABINE HOFFJAN
- ▶ ATP
- ▶ ATPase
- ▶ ATP-Binding Cassette (ABC) Transporters
- ▶ ATP-Dependent Nucleosome Remodelling

- ▶ Atrophia Gyrata
- ▶ Attention Deficit Hyperactivity Disorder
- ▶ Attenuation
- ▶ AUG Translational Start Site
- ▶ AU-Rich Elements
- ▶ Aurora Kinases
- ▶ Autoantibody
- ▶ Autoantigen
- ▶ Autocorrelation
- ▶ Autocrine
- ▶ Autogenous Translational Control
- ▶ **Autoimmune Diseases**
WESTLEY H. REEVES
- ▶ Autoimmune Thyroid Disease
- ▶ Autoimmunity
- ▶ Autoinflammatory Disease
- ▶ **Automated High Throughput Functional Characterization of Human Proteins**
DIETER LINK
- ▶ Autonomic Nervous System
- ▶ Autonomous Transposable Element
- ▶ Autonomously Replicating Sequence
- ▶ Autophosphorylation
- ▶ Autosomal
- ▶ Autosomal Dominant (Inherited Disorder)
- ▶ Autosomal Dominant Hypercholesterolemia
- ▶ Autosomal Dominant Polycystic Kidney Disease
- ▶ Autosomal Recessive (Disorder)
- ▶ Autosomal Recessive Spinal Muscular Atrophy
- ▶ Autosome
- ▶ Auxotrophic
- ▶ Avascular
- ▶ Avidin
- ▶ Avidity/Affinity
- ▶ Axial Element
- ▶ Axin/Conductin
- ▶ **Axis Formation – Formation and Function of the Dorsal Organizer**
MASAHIKO HIBI
- ▶ Axon
- ▶ Axon Guidance
- ▶ Axon Initial Segment
- ▶ Axonal Outgrowth
- ▶ B Lymphocyte

- ▶ β -TrCP
- ▶ BAC
- ▶ BAC Transgene
- ▶ Back-Projection
- ▶ Bacterial Artificial Chromosome
- ▶ Bacteriophage
- ▶ Bacteriophage Display
- ▶ Bacteriophage Lambda
- ▶ Bacteriophage Surface Display
- ▶ Bait
- ▶ Balancer Chromosome
- ▶ Balbiani Ring mRNA
- ▶ Band Keratopathy
- ▶ Baroreceptor Reflex/Baroreflex
- ▶ Barr Body
- ▶ Basal Cell Carcinoma
- ▶ Basal Ganglia
- ▶ Basal Promoter
- ▶ **Base Excision Repair**
MARIA KARTALOU, LEONA D. SAMSON
- ▶ Basement Membrane (BM)
- ▶ Base-Pairs (bp)
- ▶ Basic Helix-Loop-Helix
- ▶ Basic Helix-Loop-Helix Transcription Factors
- ▶ Basic-Region-Leucine Zipper
- ▶ Basolateral Domain
- ▶ Batch Entrez
- ▶ BC1/BC100
- ▶ BCC
- ▶ B-Cell Epitope
- ▶ Bcl-2/Bcl-xL
- ▶ B-CLL
- ▶ BCR-ABL
- ▶ BCRP
- ▶ BCYRN1
- ▶ BDNF
- ▶ Bead Models
- ▶ Beads
- ▶ Becker Muscular Dystrophy
- ▶ Beckwith-Wiedemann Syndrome
- ▶ BER
- ▶ Bernard-Soulier Syndrome
- ▶ Beta Adrenergic Blocker
- ▶ Beta2-Syntrophin/Utrophin Complex
- ▶ Beta-Amyloid (β -Amyloid)
- ▶ Beta-Blocker (β -Blocker)
- ▶ Beta-Catenin (β -Catenin)
- ▶ Beta-Oxidation
- ▶ Beta-Sheet Oligomers (β -Sheet Oligomer)
- ▶ Beta-Sheet Structure (β -Sheet Structure)
- ▶ Beta-Tubulin
- ▶ bHLH
- ▶ bHLH Transcription Factors
- ▶ Biallelic
- ▶ Bicistronic
- ▶ Bijvoet Difference
- ▶ Bile Acids
- ▶ Bile Ducts (Choledochus)
- ▶ Biliary Atresia
- ▶ Bimolecular Process
- ▶ Binary GAL4-Expression System
- ▶ Binding Affinity
- ▶ Binding Isotherm
- ▶ Bio-Beads
- ▶ **Biochemical Engineering of Glycoproteins**
RÜDIGER HORSTKORTE
- ▶ Biochip
- ▶ Biochip Technology
- ▶ Bioconjugation
- ▶ Bioethics
- ▶ Biogerontology
- ▶ Bioinformatics
- ▶ **Biological Membranes**
THOMAS POMORSKI, ANDREAS HERRMANN
- ▶ Biologicals
- ▶ Biologics
- ▶ Biomarker
- ▶ Biomimetic Ligands
- ▶ Biopanning
- ▶ Biopharmaceuticals
- ▶ Biosensors
- ▶ Bipartite NLS
- ▶ **Bipolar Affective Disorder**
MARKUS M. NÖTHEN, SVEN CICHON,
CHRISTINE SCHMÄL, MARCELLA RIETSCHEL
- ▶ Bispecific Antibodies
- ▶ Bithorax Complex (BX-C)
- ▶ BLAST
- ▶ BLAST Link
- ▶ Blastocyst
- ▶ Blastocyst Injection
- ▶ Blastomere

- ▶ BLink
- ▶ Blocked 5' Structure
- ▶ Blocks
- ▶ **Bloom Syndrome**
EBERHARD PASSARGE
- ▶ Blotting
- ▶ BMP
- ▶ BMPR-IA/-IB
- ▶ BMPR-II
- ▶ Boc Chemistry
- ▶ **Bone and Cartilage**
CHRISTINE HARTMANN
- ▶ **Bone Disease and Skeletal Disorders, Genetics**
STEFAN MUNDLOS
- ▶ Bone Lining Cell
- ▶ Bone Marrow
- ▶ Bone Morphogenetic Proteins
- ▶ Bootstrapping
- ▶ Boundary Conditions
- ▶ Bovine Spongiform Encephalopathy
- ▶ Brachydactyly
- ▶ Bragg's Law
- ▶ **Brain**
INGO BECHMANN
- ▶ Brain-Derived Neurotrophic Factor
- ▶ Branchial Arches
- ▶ BRCA
- ▶ BRCA1
- ▶ BRCA2
- ▶ BRCT Domain
- ▶ BrdU
- ▶ BRE
- ▶ **Breast Cancer**
ROLAND MOLL
- ▶ Breast Cancer Resistance Protein
- ▶ Breast Cancer Susceptibility Gene 1
- ▶ Breast Cancer Susceptibility Gene 2
- ▶ Bromodomain
- ▶ Bronchial Asthma
- ▶ Bronchospasm
- ▶ Bronzed Diabetes
- ▶ Brown Tumor
- ▶ Brush Border
- ▶ BSE
- ▶ BTB/Kelch
- ▶ Bud Neck
- ▶ Bullous (Blistering)
- ▶ Bullous Impetigo
- ▶ BX-C
- ▶ bZip
- ▶ bZIP Proteins
- ▶ C Type Lectin Domain
- ▶ *C. Elegans*
- ▶ ***C. elegans* Genome, Comparative Sequencing**
JAGAN SRINIVASAN, RALF J. SOMMER
- ▶ C2H2 Zinc Finger Gene
- ▶ C57BL/6J
- ▶ CAAX-Box
- ▶ CADASIL
- ▶ **Cadherins**
AKIRA NAGAFUCHI, MASATOSHI TAKEICHI
- ▶ *caenorhabditis elegans*
- ▶ ***Caenorhabditis elegans* as a Model Organism for Functional Genomics**
ADAM ANTEBI
- ▶ Caged Compound
- ▶ Caging Group
- ▶ Calcified
- ▶ Calcium Release Channel
- ▶ Calnexin
- ▶ Calpain 10
- ▶ Calreticulin
- ▶ **Camel as a Model for Functional Genomics**
DIRK SAERENS, MIREILLE PELLIS,
CÉCILE VINCKE, ELS PARDON,
KATJA CONRATH, SERGE MUYLDERMANS
- ▶ Camelid
- ▶ cAMP
- ▶ Cancer
- ▶ **Cancer Invasion and Metastasis: Cellular, Molecular and Clinical Aspects**
MARIA JOSÉ OLIVEIRA, MARC MAREEL,
ANCY LEROY
- ▶ Candidate Gene
- ▶ Candidate Gene Approach
- ▶ 5' Cap
- ▶ Cap/Cap Structure
- ▶ Capacitation
- ▶ Cap-Dependent/Cap-Independent Translation
- ▶ Capillary Electrophoresis

► **Cap-Independent Translational Control**

MARTIN HOLČÍK

- Capping Site
- Cap-Snatching
- CAR
- Carbohydrate Sulfotransferase
- Carboxymethylated Dextran
- Carboxy-Terminal Transactivation Domain
- Carcinoma
- CARD15
- Cardiac Connexins
- **Cardiac Signaling: Cellular, Molecular and Clinical Aspects**

MICHAEL GOTTHARDT, KATY RADDATZ

► **Cardiochip**

CHOONG-CHIN LIEW

- Cardiogenesis
- Cardiomyocyte
- Cardiomyopathy
- Cardiovascular Disease
- Cardiovascular DNA Chip
- Cardiovascular Microarray
- Caretaker Gene
- Cargo
- Carpopedal Spasm
- Carrier Ethics
- Carrier Frequency
- Carrier(s)
- Cartilage
- Cas
- Casein Kinase I
- Casein Kinase II (CKII)
- Caspase
- CAT
- Catalysis
- Catalytic Antibodies
- **Catalytic RNAs**

MASAYUKI SANO, KAZUNARI TAIRA

- Cataplexy
- Catecholamines
- Catenin
- Caveolin-1
- CBF
- CBP
- CBP (CREB-Binding Protein)/p300
- CCALD

► CCD

- CCFDN
 - CD
 - CD8
 - CD147
 - CDC42
 - CDGs
 - CDK
 - cDNA
 - cDNA (Expression) Library
 - cDNA Collection (Full Length)
 - cDNA Microarray
 - Cdx Genes
 - ced
 - Cell Adhesion
 - Cell Based Assay
 - Cell Cycle Checkpoints
 - **Cell Cycle – Overview**
- LUDGER HENGST, ERICH A. NIGG
- Cell Death
 - Cell Differentiation
 - **Cell Division**
- CAROLINE BODEN, ULRIKE ZIEBOLD
- Cell Division Cycle 42 Protein
 - Cell Lineage Specification
 - Cell Membrane
 - Cell Migration
 - Cell Motility
 - **Cell Polarity**
- ELENA CIBRIAN-UHALTE,
SALIM ABDELILAH-SEYFRIED
- Cell Transformation
 - Cellular Disease Model
 - Cellular Immune Response
 - Cellular Localization
 - Cellular Proteomics
 - **Cellular Senescence**
- IGNACIO PALMERO
- CENPs
 - Central Element
 - **Centromere**
- WILLIAM R. BRINKLEY, SCOTT SLATTERY
- Centrosome
 - Centrosome Cycle
 - Cephalochordates
 - Cer

- ▶ Cerberus
- ▶ Cerebellum
- ▶ Cerebro-Hepato-Renal Syndrome
- ▶ CF
- ▶ CF I/CF II
- ▶ CFTR
- ▶ CGH
- ▶ Channel Assembly
- ▶ Channelopathy
- ▶ Chaotrope
- ▶ Chaperone (Proteins)
- ▶ Charcot-Marie-Tooth Disease
- ▶ Charge Coupled Device cDNA
- ▶ Charged Coupled Device (CCD) Camera
- ▶ Charge-Tagging
- ▶ CHD (Familial Combined Hyperlipidemia)
- ▶ Checkpoint
- ▶ Checkpoint Kinase 2
- ▶ Chemical Cross-Linking
- ▶ Chemical Exchange
- ▶ Chemical Shift
- ▶ Chemiluminescence
- ▶ **Chemokine Receptors**
RICHARD HORUK
- ▶ Chemokines
- ▶ Chemotaxis
- ▶ Chemotherapy
- ▶ Chenodeoxycholic Acid
- ▶ Chiasma(ta)
- ▶ Childhood Cerebral ALD
- ▶ Chimeras (Chimeric Clones)
- ▶ Chimeric Antibodies
- ▶ Chimeric Clones
- ▶ Chimeric Proteins
- ▶ ChIP on Chip
- ▶ **Chip Technologies, Basic Principles**
STEFAN WÖLFL
- ▶ Chirality
- ▶ CHK2
- ▶ CHN
- ▶ 5–Chlorodeoxyuridine (CldU) &
5–Bromodeoxyuridine (BrdU)
- ▶ CHN
- ▶ Cholera Toxin
- ▶ Cholesterol Efflux
- ▶ Cholesterol Modification

- ▶ Cholic acid
- ▶ Chondroblast
- ▶ Chondrocyte
- ▶ Chondrodysplasia
- ▶ Chorea Huntington
- ▶ Chorea Major
- ▶ Chorionic Villus Sampling
- ▶ Choroideremia
- ▶ Chromatid
- ▶ Chromatin
- ▶ **Chromatin Acetylation**
ANN EHRENHOFER-MURRAY
- ▶ Chromatin Assembly
- ▶ Chromatin Modifications
- ▶ Chromatin Remodellers
- ▶ **Chromatin Remodelling**
LEONIE RINGROSE
- ▶ Chromatin Remodelling Complexes
- ▶ Chromatin Remodelling Enzymes
- ▶ Chromatin Remodelling Factors
- ▶ Chromatographic Procedures
- ▶ Chromodomain
- ▶ Chromophore
- ▶ Chromoprotein
- ▶ Chromosomal Aberration
- ▶ **Chromosomal Instability Syndromes**
MARTIN DIGWEED
- ▶ Chromosomal Microdeletion Syndromes
- ▶ Chromosomal Non-Histone proteins
- ▶ Chromosomal Translocation
- ▶ Chromosome
- ▶ **Chromosome 21, Disorders**
SAMUEL DEUTSCH,
STYLIANOS E. ANTONARAKIS,
ALEXANDRE REYMOND
- ▶ Chromosome Bands
- ▶ Chromosome Breakage Syndromes
- ▶ **Chromosome Condensation**
ELLADA SAVVIDOU, MARGARETE M. S. HECK
- ▶ Chromosome Duplication
- ▶ Chromosome Engineering
- ▶ Chromosome Instability
- ▶ Chromosome Passengers
- ▶ Chronic Airflow Limitation
- ▶ Chronic B-Cell Lymphocytic Leukaemia
- ▶ Chronic Myelogenous Leukaemia

- ▶ Chronic Obstructive Airway Disease
- ▶ Chronic Obstructive Lung Disease
- ▶ Chronic Obstructive Pulmonary Disease
- ▶ Chronic Pancreatitis
- ▶ Chylomicron
- ▶ Cilia
- ▶ CIN
- ▶ Circadian
- ▶ **Circadian Clocks**
ACHIM KRAMER
- ▶ Circumoral Numbness
- ▶ Cis-Acting Elements
- ▶ Cis-Element
- ▶ Cistron
- ▶ CJD
- ▶ CK2
- ▶ CKI
- ▶ CKI
- ▶ CKII
- ▶ c-KIT
- ▶ Classification
- ▶ **Classification of Active Centers**
CORNELIUS FRÖMMEL
- ▶ Classification of Expression Profiles
- ▶ Clathrin
- ▶ Cl_{Ca}
- ▶ CldU
- ▶ Cleavable Cross-Linkers
- ▶ Cleavage and Polyadenylation Specificity Factor
- ▶ Cleavage Factor I (CF I)/Cleavage Factor II (CF II)
- ▶ Cleavage Stimulation Factor
- ▶ **Cleft Lip Palate**
MARY L. MARAZITA
- ▶ Cleidocranial Dysplasia
- ▶ Clinical Bioinformatics
- ▶ **Clinical Gene Transfer**
KLAUS CICHUTEK, MATTHIAS SCHWEIZER,
CHRISTIAN BUCHHOLZ, EGBERT FLORY
- ▶ Clinical Marker Gene
- ▶ Clonal Analysis
- ▶ Clones
- ▶ Cloning
- ▶ Cloning Vector
- ▶ CLP
- ▶ CLSM
- ▶ Cluster Analysis
- ▶ Cluster of Differentiation
- ▶ Clustering
- ▶ Clustering of Expression Profiles
- ▶ CMAP
- ▶ CML
- ▶ CMS
- ▶ CMT
- ▶ c-Myc
- ▶ Coactivator
- ▶ Coactivator Complexes
- ▶ Coat Proteins
- ▶ Cocaine
- ▶ Cockayne Syndrome (CS)
- ▶ Coding Regions (of the Human Genome)
- ▶ Codon
- ▶ Codon Redefinition
- ▶ Coefficient of Variance in Percent (%CV)
- ▶ Coenzyme Q
- ▶ Cohesin
- ▶ Coiled-Coil Domain
- ▶ Co-Immunoprecipitation
- ▶ Colchicine
- ▶ Colinearity Rule
- ▶ Collagen
- ▶ Collecting Duct
- ▶ Colonies
- ▶ Colony Stimulating Factors
- ▶ **Colorectal Cancer**
JÜRGEN BEHRENS, BARBARA LUSTIG
- ▶ Combined Pituitary Hormone Deficiencies
- ▶ Commitment
- ▶ Common (Multifactorial) Diseases
- ▶ **Common Diseases, Genetics**
M. JULIA BROSNAN, ANNA F. DOMINICZAK
- ▶ Common Lymphoid Progenitor
- ▶ Comparative Genomic Hybridization
- ▶ Comparative Modeling
- ▶ Compartment Modeling
- ▶ Complement System
- ▶ Complementary Base
- ▶ Complementation
- ▶ Complementation Studies
- ▶ Complex Diseases/Complex (Multigenic) Traits

- ▶ Complex Numbers
- ▶ Complex Product
- ▶ Complex-Type
- ▶ Compound Heterozygote
- ▶ Compulsive Drug Seeking
- ▶ **Computational Diagnostics**
RAINER SPANG, FLORIAN MARKOWETZ
- ▶ Concordance (Rate)
- ▶ Condensin
- ▶ Conditional Gene Knockout
- ▶ Conditional Gene Targeting
- ▶ Conditional Knockout (KO)
- ▶ Conduction-System Disease
- ▶ Cone-Rod Dystrophies
- ▶ Cones
- ▶ Configuration
- ▶ Configuration Space
- ▶ Confluent Monolayer/Confluence
- ▶ Confocal
- ▶ Confocal Laser Scanning Microscope
- ▶ Conformation
- ▶ Conformation of Double-Stranded DNA
- ▶ Conformational Analysis
- ▶ Conformational Order
- ▶ Congenic Dissection
- ▶ Congenic Strain
- ▶ Congenital
- ▶ Congenital Disorders in Glycosylation
- ▶ Congenital Dyserythropoietic Anemia
- ▶ Congenital Heart Disease
- ▶ Congenital Hypomyelinating Neuropathy
- ▶ Congenital Myasthenic Syndromes
- ▶ Congestive Heart Failure
- ▶ Congressional Movement
- ▶ Conjugal Pair
- ▶ Connective Tissue
- ▶ Connexin Channelopathies
- ▶ Connexins
- ▶ Consanguinity
- ▶ Consensus Map
- ▶ Consensus Sequence (Sequon)
- ▶ Conservation
- ▶ Conserved Synteny
- ▶ Constitutive Expression
- ▶ Constitutive Mutation

- ▶ Constitutive Splicing
- ▶ Contact Inhibition
- ▶ Contact Sites
- ▶ Contig Maps
- ▶ Contiguous Gene Syndromes
- ▶ Contractures
- ▶ Convergence
- ▶ Conversion Technology
- ▶ Convolution
- ▶ Cooperativity
- ▶ COPD
- ▶ **COPD and Asthma, Genetics**
JIAN-QING HE, ANDREW SANDFORD
- ▶ COPI Vesicles
- ▶ COPII Complex/COPII Vesicles
- ▶ Core Binding Factor
- ▶ **Core Promoters**
ANANDA L. ROY
- ▶ Coreceptor
- ▶ Corepressor
- ▶ Corneal Arcus
- ▶ Corneal Opacity
- ▶ Coronary Artery
- ▶ Corpus Callosum
- ▶ Correlation
- ▶ Correlation Analysis
- ▶ Correlative Microscopy
- ▶ Cortical Column
- ▶ Cortical Granules
- ▶ Cortical Reaction
- ▶ Cortical Tuber
- ▶ Corticotropin-Releasing Hormone
- ▶ Cosmid
- ▶ COSY
- ▶ Co-Transcriptional Capping
- ▶ Co-Translational Modification
- ▶ Coulomb Electrostatics
- ▶ Covalent Coupling
- ▶ COX
- ▶ CpG
- ▶ **CpG Islands**
CHRISTOPH PLASS, LAURA J. RUSH
- ▶ CpG Methylation
- ▶ CPSF
- ▶ Craving
- ▶ CRC

- ▶ CRE
- ▶ Cre
- ▶ **Cre/loxP Strategies**
BRIAN SAUER
- ▶ C-Reactive Protein
- ▶ Creatine Kinase
- ▶ CREB
- ▶ Creb Binding Protein
- ▶ Cre-ERT
- ▶ Creutzfeldt-Jakob Disease
- ▶ CRH
- ▶ Cri-Du-Chat Syndrome
- ▶ Crisis
- ▶ Cristae
- ▶ Cristae Junctions
- ▶ Critical Micellar Concentration
- ▶ Crk
- ▶ cRNA
- ▶ **Crohn Disease**
STEFAN SCHREIBER
- ▶ Cross-Bridge
- ▶ Cross-Fostering Studies
- ▶ Crossing-Over
- ▶ Cross-Linking Patterns
- ▶ Cross-Linking Reagents
- ▶ Cross-Talk
- ▶ Crossvalidation
- ▶ CRP
- ▶ Cryo Electron Crystallography
- ▶ **Cryo-Electron Microscopy: Single-Particle Reconstruction**
ROLAND BECKMANN
- ▶ Cryo-EM
- ▶ Cryptic Splice Signals
- ▶ Cryptochrome
- ▶ Crystal Lattice
- ▶ Crystal Packing
- ▶ Crystalline Disorder
- ▶ CsA
- ▶ CSF
- ▶ CstF
- ▶ C-TAD
- ▶ CTCF
- ▶ CTD
- ▶ CTL
- ▶ Current-Clamp Recording
- ▶ CVS
- ▶ Cyclic AMP Receptor
- ▶ Cyclin
- ▶ Cyclin D1
- ▶ Cyclin-Dependent Kinase
- ▶ Cycling Genes
- ▶ Cyclooxygenases
- ▶ Cyclopamine
- ▶ Cyclophilins
- ▶ Cyclosporin A
- ▶ Cysteine Proteases
- ▶ Cystic Fibrosis (CF)
- ▶ Cystic Fibrosis Transmembrane Conductance Regulator
- ▶ Cystine Knot
- ▶ Cytochrome P450
- ▶ Cytochrome C
- ▶ Cytofluorometry
- ▶ Cytogenetic
- ▶ Cytokine Cluster Region
- ▶ Cytokine Receptors
- ▶ **Cytokines**
MICHAEL KRACHT, HELMUT HOLTSMANN
- ▶ Cytokinesis
- ▶ Cytolysis
- ▶ Cytolytic T Lymphocytes
- ▶ Cytoplasm
- ▶ Cytoskeletal Linkers
- ▶ Cytoskeleton
- ▶ **Cytoskeleton: Microtubules and Intermediate Filaments**
DIETER K. MEYER
- ▶ Cytosol
- ▶ DA
- ▶ daf
- ▶ Dalton
- ▶ Dapper (Dpr)/Frodo
- ▶ Data Availability
- ▶ Data Normalisation
- ▶ **Data-Mining in Biology, “How to Find a Needle in a Haystack?”**
KARYN MÉGY
- ▶ DCM
- ▶ DCp
- ▶ DCP1/DCP2
- ▶ DcpS

- ▶ ddNTP
- ▶ De Novo (Mutation)
- ▶ De Novo Pathway
- ▶ Deacylation
- ▶ Deadenylation
- ▶ Deamination
- ▶ Death Domain Fold Protein Superfamily
- ▶ Death Inducing Signaling Complex
- ▶ Decision Tree

▶ **Defective Protein Folding Disorders**

CLAUDIO SOTO

- ▶ Degradation signals
- ▶ Dejerine-Sottas Neuropathy
- ▶ Deletion
- ▶ Deletion Loop
- ▶ Delta 1-Like (DII) 1,3,4
- ▶ Dementia Praecox
- ▶ Demyelination
- ▶ Denaturing High-Performance Liquid Chromatography
- ▶ Dendrites
- ▶ Dendritic Extensions
- ▶ Dendritic Protein Synthesis
- ▶ Dentatorubral Pallidoluysian Atrophy
- ▶ Deoxyhexose
- ▶ Deoxyribophosphate Lyase
- ▶ Dependence
- ▶ Depolarization
- ▶ Depurination
- ▶ Dermal Papilla
- ▶ Dermis
- ▶ Dermomyotome
- ▶ Des Species
- ▶ Desferrioxamine
- ▶ Desmal Ossification
- ▶ **Desmosomes**

AMANDA E. BASS, KATHLEEN J. GREEN

- ▶ Desorption
- ▶ Destruction Complex
- ▶ Detergent
- ▶ Developmental Timing
- ▶ Devic's Disease
- ▶ DG
- ▶ DG/VCFS
- ▶ DH
- ▶ DHF

- ▶ DHPLC
- ▶ Di- and Trihydroxycholestanoic Acid
- ▶ Diabetes
- ▶ Diabetes Insipidus
- ▶ **Diabetes Insipidus, a Water Homeostasis Disease**

MARIEL BIRNBAUMER

- ▶ Diabetes Mellitus
- ▶ **Diabetes Mellitus, Genetics**

DONALD W. BOWDEN

- ▶ Diabetes Polyuria
 - ▶ Diagnostic Gene Testing
 - ▶ Diaphragm
 - ▶ Diaphysis
 - ▶ Diastole
 - ▶ Diastrophic Dysplasia
 - ▶ Dicentric Chromosomes
 - ▶ Dichromatic Beamsplitter
 - ▶ Dickkopf
 - ▶ DI-CMT
 - ▶ *Dictyostelium Discoideum*
 - ▶ Differential Display
 - ▶ Differential in-Gel Electrophoresis
 - ▶ **Differential Scanning Calorimetry**
- T. REINISCH, J. RÖSGEN, H.-J. HINZ
- ▶ Differentially Methylated Region
 - ▶ Differentiation
 - ▶ Diffraction Intensities
 - ▶ Diffusion
 - ▶ DIGE
 - ▶ DiGeorge / Velocardiofacial Syndrome
 - ▶ Dihydrofolate Reductase
 - ▶ Dilated Cardiomyopathy
 - ▶ Dimer
 - ▶ Dioxygenases, Monooxygenases and Oxidases
 - ▶ Diploid
 - ▶ Diploid Genome
 - ▶ DISC
 - ▶ Discordance / Discordant
 - ▶ Discovery Approach
 - ▶ Disease Models
 - ▶ Disease-Modifying Anti-Rheumatic Drugs
 - ▶ Dishevelled
 - ▶ Dispersive Difference
 - ▶ Dissolution Point
 - ▶ Distributive and Processive DNA Synthesis

- ▶ Disulfide Regeneration
- ▶ Disulfide Reshuffling
- ▶ Disulfide Species
- ▶ Disulfide-Insecure Species
- ▶ Disulfide-Protected Species
- ▶ Disulfide-Secure Species
- ▶ Diuretics
- ▶ Dizygotic
- ▶ Dkk
- ▶ DM
- ▶ DM Domain
- ▶ DM1
- ▶ DM2
- ▶ DMPK
- ▶ DMR
- ▶ DMs
- ▶ DNA
- ▶ **DNA Amplification**
ERHARD WINTERSBERGER
- ▶ DNA Arrays
- ▶ DNA Binding Domains
- ▶ DNA Catenanes
- ▶ DNA Chips
- ▶ DNA Chip Technology
- ▶ DNA Conformations
- ▶ **DNA Damaging Agents**
BERND KAINA, GERHARD FRITZ
- ▶ DNA Damaging Agents
- ▶ DNA Double-Strand Break Repair
- ▶ DNA Gyrase
- ▶ **DNA Helicases**
NAYEF A. MAZLOUM, NATHAN A. ELLIS
- ▶ DNA Lesion
- ▶ **DNA Ligases**
ALESSANDRA MONTECUCCO, ROSSELLA ROSSI
- ▶ DNA Marker
- ▶ DNA Methylation
- ▶ DNA Microarrays / DNA Arrays
- ▶ DNA Polymerase Switch
- ▶ DNA Polymerase δ Holoenzyme
- ▶ **DNA Polymerases**
HEINZ-PETER NASHEUER, HELMUT POSPIECH,
JUHANI SYVÄOJA
- ▶ DNA Primase
- ▶ **DNA Recombination**
JEAN-MARIE BUERSTEDDE
- ▶ DNA Relaxation
- ▶ DNA Repair
- ▶ **DNA Repair Mechanisms**
WOLFGANG GOEDECKE
- ▶ DNA Replication
- ▶ **DNA Replication Initiation**
MASATOSHI FUJITA
- ▶ DNA Structure
- ▶ **DNA Topoisomerases**
JOHN L. NITISS,
ALEXANDER STEPANOV
- ▶ **DNA-based Vaccination**
REINHOLD SCHIRMBECK, MARTIN SCHLEEF,
JÖRG REIMANN
- ▶ DNA-Dependent RNA Polymerase I
- ▶ DnaJ
- ▶ dNTPs
- ▶ DOCK-180
- ▶ Docking
- ▶ Dolichol
- ▶ Domain
- ▶ Domain Fold
- ▶ Dominant
- ▶ Dominant Allele
- ▶ Dominant Inhibitory Mutant / Dominant
Negative Mutant
- ▶ Dominant Mutation
- ▶ Dominant Negative Mutation
- ▶ Donor
- ▶ Dopamine
- ▶ Dopamine Replacement Therapy
- ▶ Dopamine Responsive Parkinsonism
- ▶ Dormancy
- ▶ Dosage Compensation
- ▶ Doss Porphyria
- ▶ Double Mutant
- ▶ **Double-Strand Break Repair**
ALBERT PASTINK,
MALGORZATA Z. ZDZIENICKA
- ▶ Double-Stranded RNA-Dependent Protein
Kinase
- ▶ Double-Stranded RNA-Mediated Gene
Silencing
- ▶ Down Syndrome
- ▶ Downstream
- ▶ DPE

► ***Drosophila* as a Model Organism for Functional Genomics**

MARC HILD, RENATO PARO

► ***Drosophila* as a Model to Study Cardiac Disease Genes**

ROLF BODMER, ETHAN BIER

- *Drosophila Melanogaster*
- DRPLA
- Drug Abuse
- Drug Substitution and Anti-Craving Medication
- Drug Target
- Drug Tolerance
- Drug-Seeking Behavior
- Drusen
- DS
- DSC
- DSL-Family
- DSN
- **Duchenne Muscular Dystrophy**
ERYNN S. GORDON, ERIC P. HOFFMAN
- Duplication
- Dural Ectasia
- Duty Cycle
- Dvl
- Dynamic Instability
- Dynamic Mutation
- Dynamic Range
- Dysautonomia
- Dyshormogenesis
- Dyslipidemia
- Dysmyelination
- Dysostoses
- Dystroglycan
- Dystrophia Myotonica
- Dystrophin
- Dystrophin Glycoprotein Complex
- E2F
- E3–Ubiquitin Ligase
- Early Pressure
- EBD
- EBI
- E-Box
- EC
- ECD
- Ecdysone

- ECG
- ECM
- Ectodomain Shedding
- Ectopia Lentis
- EEG
- EF Hand
- Effector
- EGF
- EGFP
- EGFR
- Egg
- eIF–4F
- EJC
- Electrical Pacing
- Electrocardiogram (ECG)/Electrocardiography
- Electromechanical Coupling
- Electron Beam Deposition
- Electron Crystallography
- Electron Density
- Electron Microscopy
- Electron Multiplying CCD Camera
- **Electron Tomography**
MANFRED AUER
- Electronic Transition Energy
- Electrophoresis
- Electroporation
- Electrospray Ionization Mass Spectrometry
- ELISA
- EM
- EMBL
- Embryo Transfer
- Embryoid Body
- Embryonic Lethality
- Embryonic Stem Cell
- EMCCD
- Emission Spectrum
- EMT
- 5' End/3' End
- 5' End Processing
- Endocarditis
- Endochondral
- Endocrine
- Endocrine Cells
- Endocytic Pathway
- Endocytosis
- Endoderm

- ▶ Endogenous Proviruses
- ▶ Endonuclease
- ▶ Endoplasmic Reticulum
- ▶ Endosomal Compartment
- ▶ Endosome
- ▶ Endosteum
- ▶ Endothelial Cells
- ▶ Endothelium
- ▶ Enhanced Green Fluorescent Protein
- ▶ Enhanceosome
- ▶ **Enhancer**
WALTER SCHAFFNER
- ▶ Enhancer Trap
- ▶ Enhancer/Suppressor Screens
- ▶ Enophthalmos
- ▶ Ensemble Average
- ▶ Enteric Nervous System
- ▶ Enterobacteriaceae
- ▶ Enthalpy/Enthalpy Change (DH)
- ▶ Entrainment
- ▶ Entrez
- ▶ Entropy Change (DS)
- ▶ ENU
- ▶ *env*
- ▶ Envelope
- ▶ **Enzyme Catalyzed Post-Translational Hydroxylation of Proteins**
CHRISTOPHER SCHOFIELD,
THOMAS MURRAY-RUST
- ▶ Enzyme Linked Immuno-Sorbent Assay
- ▶ Enzymes
- ▶ Epidermal Growth Factor Receptor
- ▶ Epidermis
- ▶ Epidermolysis Bullosa (Simplex)
- ▶ Epidermolytic Hyperkeratosis
- ▶ Epidermolytic Palmoplantar Keratoderma
- ▶ Epigenetic (Epigenetics)
- ▶ Epigenetic Determinants
- ▶ Epigenetic Modification
- ▶ Epiglottis
- ▶ Epiphysis
- ▶ Epistasis
- ▶ Epistasis Group
- ▶ **Epistasis in Cystic Fibrosis**
FRANCESCO SALVATORE, GIUSEPPE CASTALDO
- ▶ Epithelia
- ▶ **Epithelial Cells**
APARNA LAKKARAJU,
ENRIQUE RODRIGUEZ-BOULAN
- ▶ Epithelial-to-Mesenchymal Transition
- ▶ Epitope
- ▶ Epitope Tag
- ▶ EPR
- ▶ EPSP/IPSP
- ▶ Epstein-Barr Virus
- ▶ Equilibrium
- ▶ ER
- ▶ ERK
- ▶ ERM Protein
- ▶ Erythropoietic Porphyrins
- ▶ Erythropoietic Protoporphyria
- ▶ Erythropoietin
- ▶ **ES Cell Differentiation as a Model System for Functional Genomics**
KENNETH R. BOHLER
- ▶ ES Cells
- ▶ ESCRT
- ▶ ESE
- ▶ ESI
- ▶ ESI-MS
- ▶ Essential Hypertension
- ▶ EST
- ▶ Estrogen Receptor
- ▶ Etherphospholipids
- ▶ **Ethical Issues in Medical Genetics**
HANS MARTIN SASS
- ▶ Etiology
- ▶ Euchromatin
- ▶ Evanescent Electric Field
- ▶ Evolution Time
- ▶ Evolutionary Conservation
- ▶ Ex Vivo
- ▶ Excitable membranes
- ▶ Excitation Spectrum
- ▶ Exclusion
- ▶ Executive Function
- ▶ Exocrine Cells
- ▶ Exocyst
- ▶ **Exocytic Pathway**
ROMAN S. POLISHCHUK,
ALEXANDER A. MIRONOV, ALBERTO LUINI
- ▶ Exocytosis

- ▶ Exon
- ▶ Exon Junction Complex
- ▶ Exon / Intron Junction
- ▶ Exonic Sequence Element
- ▶ Exonic Splicing Enhancer/Silencer
- ▶ Exonuclease
- ▶ Exosome
- ▶ Export Adapters
- ▶ Export Receptors
- ▶ Exportin
- ▶ Expressed Sequence Tag
- ▶ Expression
- ▶ Expression Difference Mapping
- ▶ Expression Landscape / Mountains
- ▶ Expression Profile
- ▶ Expression Screening
- ▶ Expression System
- ▶ Expression Vector/Expression Construct
- ▶ Expressivity
- ▶ Extinction Coefficient (ϵ)
- ▶ **Extracellular Matrix**
JAN F. TALTS, PETER EKBLOM
- ▶ Extrachromosomal Elements
- ▶ Ezrin
- ▶ Fabrication
- ▶ Fabry Disease
- ▶ FACS
- ▶ F-Actin
- ▶ Factor Inhibiting HIF-1
- ▶ FAD
- ▶ FADD
- ▶ FAK
- ▶ False Positives/False Negatives
- ▶ Familial Adenomatous Polyposis
- ▶ Familial Aggregation
- ▶ Familial Alzheimer Disease
- ▶ Familial Amyotrophic Lateral Sclerosis
- ▶ Familial Combined Hyperlipidemia
- ▶ Familial DCM
- ▶ **Familial Dilated Cardiomyopathy**
DIANE FATKIN
- ▶ Familial Endocrinopathy
- ▶ Familial Exudative Vitreoretinopathy
- ▶ **Familial Hypercholesterolemia**
HENRICK K. JENSEN, TROND P. LEREN
- ▶ Familial Hypertrophic Cardiomyopathy

- ▶ Familial Spastic Paraplegia
- ▶ Fanconi Anaemia
- ▶ FAP
- ▶ Farnesyl
- ▶ Farnesylation
- ▶ FAS
- ▶ Fas Signaling
- ▶ Fas-Associated Death Domain
- ▶ Fast Exchange Limit
- ▶ FASTA
- ▶ Fat Metabolism
- ▶ **Fatty Acid Acylation of Proteins**
MICHAEL VEIT, MICHAEL F.G. SCHMIDT
- ▶ FCMD
- ▶ Feedback
- ▶ FEP
- ▶ Fermentation
- ▶ Ferritin
- ▶ Fertilization
- ▶ FEVR
- ▶ FG Repeats
- ▶ FGF
- ▶ Fibrinolysis
- ▶ **Fibroblast**
GIULIO GABBANI
- ▶ Fibronectin
- ▶ FID
- ▶ FIH
- ▶ Filaggrin
- ▶ Filamentous Actin
- ▶ Filopodia
- ▶ Filtering
- ▶ Fingerprint/Fingerprinting
- ▶ FISH
- ▶ Fission Yeast
- ▶ Fitness
- ▶ Fixation
- ▶ FK506 (Tacrolimus)
- ▶ FKBP12
- ▶ FLAG-Tag
- ▶ FIAsh/ReAsH System
- ▶ Flexible Docking
- ▶ FLIM
- ▶ Flippase
- ▶ Flow Cytometry
- ▶ Floxed

► FLP/FRT System

► FLT3

► Fluorescence

► Fluorescence Anisotropy

► **Fluorescence Correlation Spectroscopy**

PETRA SCHWILLE

► Fluorescence *In Situ* Hybridization

► Fluorescence Labelling

► Fluorescence Lifetime Imaging

► Fluorescence Microscopy

► **Fluorescence Microscopy: Single Particle Tracking**

ULRICH KUBITSCHKE

► **Fluorescence Microscopy: Spectral Imaging vs. Filter-based Imaging**

M. CRISTINA CARDOSO

► Fluorescence Quantum Yield

► Fluorescence Recovery after Photobleaching

► Fluorescence Resonance Energy Transfer

► Fluorescence-Activated Cell Sorter

► Fluorescent Proteins

► Fluorochrome

► 5-Fluoroorotic Acid

► 5-FOA

► Fluorophore

► FlyBase

► Fmoc Chemistry

► FMR Protein

► FMR1 Knockout Mouse

► *fmr1*/FMR1

► fMRI

► FMRP

► FMS-Like Tyrosine Kinase 3 Gene

► **Focal Adhesions**

BERHARD WEHRLE-HALLER,

BEAT A. IMHOF

► Focal Complexes/Focal Contacts

► Folding Funnel

► Follicle

► Follicle Stimulating Hormone

► Fölling Disease

► Force Field

► Force Generation

► Force Transmission

► Forkhead Genes or Proteins

► **Förster Resonance Energy Transfer**

MARTIN BEUTLER,

RAINER HEINTZMANN

► Forward Genetics

► Fos

► Founder (Gene) Effect

► Founder Mutation

► Fourier Shell Correlation

► Fourier Transform

► Fourier Transform Ion Cyclotron Resonance Mass Spectrometer

► Fovea

► Fox Genes or Proteins

► Fps

► Fragile Site

► **Fragile X Mental Retardation Syndrome**

PETER STEINBACH

► Fragile X Syndrome

► Fragile X(A) Syndrome

► Fragment-Based Ligand Design

► Frame (+1 Frame/-1 Frame)

► Frameshift (Mutation)

► Frameshift Stimulator Sequences

► Frameshift Suppression/Suppressor

► FRAP

► **FRAP and Other Photobleaching Methods**

ERIK LEE SNAPP

► Frasier Syndrome

► FRAXA

► Free Energy

► Free Energy Change (dG)

► Free Energy Perturbation

► Free Induction Decay

► **Free Radicals**

LEOPOLD FLOHÉ, HEIKE BUDDE

► FRET

► Frictional Ratio

► Friedreich Ataxia

► Frizzled

► FSC

► FSH

► FTI

► FT-ICR

► Fukutin

- ▶ Fukuyama Myopathy
- ▶ Full Mutation
- ▶ Full-Length cDNA
- ▶ **Full-Length cDNA Sequencing**
STEFAN WIEMANN, REGINA ALBERT,
PETRA MOOSMAYER, INGO SCHUPP
- ▶ Full-Size cDNA
- ▶ **Functional Assays**
JEREMY C. SIMPSON
- ▶ Functional Group
- ▶ **Functional Imaging**
BURKHARD WIESNER
- ▶ Functional Magnetic Resonance Imaging (fMRI)
- ▶ Functional Proteomics
- ▶ Functional Studies
- ▶ Furin
- ▶ Fusion
- ▶ Fusion Domain
- ▶ Fusion Proteins
- ▶ Fusion Tag
- ▶ Fyn
- ▶ FzB
- ▶ G4 DNA
- ▶ GABA
- ▶ Gag
- ▶ Gain-of-Function Mutations
- ▶ Gain-of-Function Screens
- ▶ GAL4–Expression System
- ▶ Gamates
- ▶ Gamma Rays
- ▶ Gammaretroviruses
- ▶ Gammaretroviruses IN (Integrase)
- ▶ Gamma-Secretase (Complex)
- ▶ Gap
- ▶ **Gap Junctions**
W. HOWARD EVANS
- ▶ GAPs
- ▶ GAR Motif
- ▶ Gas Chromatography (GC)
- ▶ Gastrulation
- ▶ Gatekeeper Genes
- ▶ Gating
- ▶ GBP/FRAT
- ▶ GDIs
- ▶ GEF
- ▶ 2D-Gel Electrophoresis
- ▶ Gelsolin
- ▶ Gemins
- ▶ GenBank
- ▶ Gene
- ▶ **Gene Annotation in Plants**
S. ROMBAUTS, Y. VAN DE PEER,
PIERRE ROUZÉ
- ▶ Gene Chip Technology
- ▶ **Gene Chip Technology and Its Application to Molecular Medicine**
HEIKE ZIMDAHL, NORBERT HÜBNER
- ▶ Gene Cluster
- ▶ Gene Conversion
- ▶ Gene Dosage Analysis
- ▶ **Gene Duplications**
SEBASTIAN M. SHIMELD
- ▶ Gene Expression
- ▶ Gene Expression Data Analysis: Classification
- ▶ Gene Expression Data Analysis: Supervised Analysis
- ▶ Gene Expression Data Analysis: Unsupervised Analysis
- ▶ Gene Expression Data Matrix
- ▶ Gene Expression Profile
- ▶ Gene Gun
- ▶ Gene Mapping
- ▶ Gene Ontology
- ▶ Gene Silencing
- ▶ Gene Silencing by Double-Stranded RNA
- ▶ Gene Targeting
- ▶ Gene Therapy
- ▶ Gene Trapping
- ▶ Gene-Based Therapies
- ▶ Gene-Environment Interaction
- ▶ Gene-Gene Interaction
- ▶ General Transcription Factors
- ▶ Genetic Algorithm
- ▶ **Genetic Background**
ROBERT GERLAI
- ▶ **Genetic Code**
JEAN LEHMANN
- ▶ Genetic Counselling

- ▶ Genetic Distance
- ▶ **Genetic Epidemiology**
MICHAEL KRAWCZAK
- ▶ Genetic Hearing Disorder
- ▶ Genetic Heterogeneity
- ▶ Genetic Immunization
- ▶ Genetic Interactions
- ▶ Genetic Map
- ▶ Genetic Modification
- ▶ Genetic Polymorphism
- ▶ **Genetic Predisposition to Multiple Sclerosis**
ALASTAIR COMPSTON
- ▶ Genetic Redundancy
- ▶ Genetic Screen
- ▶ **Genetic Screening in Populations**
VILMUNDUR GUDNASON
- ▶ Genetic X-Linked Disease
- ▶ Genetically Engineered Animals
- ▶ Genome
- ▶ Genome Analysis in Plants
- ▶ Genome Engineering
- ▶ Genome Functionalization by Arrayed cDNA Transduction
- ▶ Genome Instability
- ▶ Genome Scan
- ▶ Genome Screen
- ▶ Genome Walking
- ▶ Genome-Wide Analysis
- ▶ **Genomic Analysis of Single Disseminated Cancer Cells**
CHRISTOPH A. KLEIN
- ▶ Genomic Clone
- ▶ Genomic Control
- ▶ **Genomic Imprinting**
TAKUYA IMAMURA, ANDRAS PALDI
- ▶ **Genomic Information and Cancer**
TRAVIS DUNCKLEY, KEITH D. COON,
DIETRICH A. STEPHAN
- ▶ Genomic Instability
- ▶ Genomics
- ▶ Genomics
- ▶ Genotoxin
- ▶ Genotype
- ▶ Genotype-Driven Approach
- ▶ Genotype-Phenotype Correlations
- ▶ Genotyping
- ▶ Geranylgeranyl Pyrophosphate
- ▶ Germ Cells
- ▶ Germinal Vesicle
- ▶ Germline (Gonadal) Mosaicism
- ▶ Germline Mutation
- ▶ Germline Transmission
- ▶ GFACT Expression screening
- ▶ GFAP
- ▶ GFFKR Domain
- ▶ GFP
- ▶ GGA Proteins
- ▶ Giga-Seal
- ▶ Glanzmann's Thrombasthenia
- ▶ GLI
- ▶ Glial Cells
- ▶ **Glial Cells and Myelination**
HAUKE WERNER, KLAUS-ARMIN NAVE
- ▶ Global Genome Repair
- ▶ Glomerular Filtrate
- ▶ Glomerulonephritis
- ▶ Glomerulus
- ▶ Glucocorticoid/Mineralocorticoid Receptors
- ▶ Glucocorticoid/Mineralocorticoid Resistance
- ▶ Glucocorticoids
- ▶ Glutamate
- ▶ Glutathione Peroxidase
- ▶ Glycan
- ▶ Glycated Protein
- ▶ Glycine
- ▶ Glycoconjugate
- ▶ Glycoform
- ▶ Glycogen Synthase Kinase-3
- ▶ Glycohemoglobin
- ▶ Glycolysis
- ▶ Glycoprotein
- ▶ Glycoproteomics
- ▶ Glycosaminoglycan
- ▶ Glycosidic Linkage
- ▶ Glycosylase
- ▶ **Glycosylation of Proteins**
RICHARD D. CUMMINGS
- ▶ Glycosylphosphatidylinositol Anchors
- ▶ Glycosyltransferase
- ▶ Glyoxylate

- ▶ Glypidation
- ▶ Golgi Apparatus (Golgi Complex)
- ▶ Gomori Trichrome
- ▶ Gonadal Mosaicism
- ▶ Gonadotropin Deficiency
- ▶ Gonadotropins
- ▶ Gordon's Syndrome
- ▶ Gorlin's Syndrome
- ▶ GPCRs
- ▶ G-Phase
- ▶ GPI (-Anchored) Protein
- ▶ GPIIb/IIIa Complex
- ▶ G-Protein Coupled Proteolytic Site
- ▶ G-Protein Coupled Receptors
- ▶ G-Proteins
- ▶ **G-Proteins and G-Protein Mutations in Human Diseases**
THOMAS GUDERMANN
- ▶ GPx
- ▶ G-Quartet DNA
- ▶ Graafian Follicles
- ▶ Grade of Malignancy
- ▶ Granulation Tissue
- ▶ Granuloma
- ▶ Granulomatosis
- ▶ GRAS
- ▶ Grb2
- ▶ Green Fluorescent Protein
- ▶ Greig's Cephalopolysyndactyly
- ▶ Growth Factor Receptors
- ▶ **Growth Factors**
ULF HEDIN, JOY ROY
- ▶ Growth Hormone Deficiency
- ▶ Growth Plate
- ▶ Gamma-Secretase (Complex)
- ▶ GSK3
- ▶ GST
- ▶ GST Pull-Down Experiment
- ▶ GTP
- ▶ GTPase-Activating Proteins
- ▶ GTPases
- ▶ Guanine Nucleotide Dissociation Inhibitors
- ▶ Guanine Nucleotide Exchange Factors
- ▶ **Gut Epithelium**
MICHÈLE KEDINGER, JEAN-NOËL FREUND
- ▶ Hair

- ▶ Hair Bulge
- ▶ Hair Matrix
- ▶ Hairpin End
- ▶ Hairpin Ribozyme
- ▶ Half-Life
- ▶ Hamartoma
- ▶ Hammerhead Ribozyme
- ▶ Haploid
- ▶ Haploid Analysis
- ▶ Haploinsufficiency
- ▶ Haplotype
- ▶ Haplotype Block
- ▶ Hapten
- ▶ Harker Section
- ▶ Hayflick Limit
- ▶ HCAb
- ▶ HCC
- ▶ HCM
- ▶ HCS
- ▶ HD
- ▶ HDAC
- ▶ HDAC Inhibitor(s)
- ▶ HDL
- ▶ Health Literacy
- ▶ **Heart**
INGO MORANO
- ▶ Heat Capacity Change
- ▶ Heat Shock
- ▶ Heat Shock Protein
- ▶ Heat Shock Response
- ▶ Heavy-Chain Antibody
- ▶ HECT-Domain
- ▶ Hedgehog
- ▶ **Hedgehog Signalling**
CAROL WICKING, TIMOTHY M. EVANS
- ▶ Helicases
- ▶ Helix
- ▶ Helix Boundary Motif
- ▶ Helix Initiation
- ▶ Helix Initiation Peptide Helix Termination Peptide
- ▶ Helix Notation
- ▶ Helix Propagation
- ▶ Helix Termination Peptide
- ▶ Helix-Loop-Helix
- ▶ Helix-Turn-Helix

- ▶ Helix-Turn-Helix Transcription Factors
- ▶ Helper Phage
- ▶ Hematopoiesis
- ▶ Hematopoietic Stem and Progenitor Cells
- ▶ Hemichannels
- ▶ **Hemidesmosomes**
SANDY H.M. LITJENS, ARNOUD SONNENBERG
- ▶ Hemizygous
- ▶ **Hemochromatosis**
ERNEST BEUTLER
- ▶ Hemophilia A
- ▶ Hemophilia B
- ▶ Heparin Sulfate Proteoglycans
- ▶ Hepatitis B Virus
- ▶ Hepatocellular Carcinoma
- ▶ Hepatocyte Nuclear Factor 4a
- ▶ HER1
- ▶ Hereditary Cancer
- ▶ Hereditary Coproporphyrria
- ▶ **Hereditary Disease, Genetic Basis**
JÖRG SCHMIDTKE
- ▶ **Hereditary Hemostatic Defects and Recombinant Proteins for Treatment**
ELISABETH ERHARDTSEN
- ▶ Hereditary Motor Neuropathies
- ▶ **Hereditary Neuropathies, Motor and/or Sensory**
BERND RAUTENSTRAUß, JAMES R. LUPSKI, EVA NELIS
- ▶ Hereditary Neuropathy with Liability to Pressure Palsies
- ▶ **Hereditary Nonpolyposis Colorectal Cancer**
PÄIVI PELTOMÄKI
- ▶ Hereditary Sensory Neuropathies
- ▶ Hereditary Sideroblastic Anemia
- ▶ **Hereditary Spastic Paraplegia**
EVAN REID
- ▶ Heritability
- ▶ **Heritable Skin Disorders**
GABRIELE RICHARD, JOUNI UITTO
- ▶ Heroin
- ▶ Heterochromatin
- ▶ Heterodimer/Heterodimeric protein
- ▶ Heterogeneous, Genetically/Heterogeneity, Genetic
- ▶ Heterologous Inducers
- ▶ Heterologous Promoter Control
- ▶ Heteronuclear NMR Experiments
- ▶ Heteronuclear RNP Proteins
- ▶ Heteronuclear Single Quantum Coherence
- ▶ Heteroplasmy
- ▶ Heterotrimeric Guanine Nucleotide-Binding Protein
- ▶ Heterozygote (Heterozygous)
- ▶ Hexosamine
- ▶ Hexose
- ▶ HFE
- ▶ HGPRT
- ▶ HH
- ▶ HH Receptor
- ▶ HHV-8
- ▶ Hidden Markov Models
- ▶ Hierarchical Clustering
- ▶ HIF
- ▶ HIF-Prolyl-Hydroxylase
- ▶ HIF- α
- ▶ High Content Screening
- ▶ High Content Screening
- ▶ High Mannose-Type
- ▶ High Performance Liquid Chromatography
- ▶ High Throughput Functional Cell-Based Screening
- ▶ High-Density Lipoprotein
- ▶ **High-HDL Syndrome**
AKIHIRO INAZU, HIROSHI MABUCHI
- ▶ High-Mobility-Group Transcription Factor
- ▶ **High-Throughput Analysis of Gene Function in Mammalian Cells Using Transfected Cell Arrays**
DOMINIQUE VANHECKE, MICHAL JANITZ
- ▶ High-Throughput-Screening
- ▶ Hippocampus
- ▶ His-Tag/6xHis
- ▶ Histocompatibility
- ▶ Histone
- ▶ Histone Acetylation
- ▶ Histone Acetyltransferases
- ▶ Histone Code
- ▶ Histone Deacetylase
- ▶ Histone Methylation
- ▶ Histone Methyltransferase

- ▶ Histone Octamer
- ▶ HIV-1
- ▶ HLA
- ▶ **HLA and Disease**
KATSUSHI TOKUNAGA
- ▶ HLH
- ▶ HMG Box
- ▶ HMM
- ▶ HMN
- ▶ HMSN
- ▶ HNA
- ▶ HNF4A Gene
- ▶ HNPCC
- ▶ HNPP
- ▶ hnRNA
- ▶ Holiday Junction
- ▶ Holoprosencephaly
- ▶ HOM-C
- ▶ Homeobox
- ▶ Homeobox Genes
- ▶ Homeobox Transcription Factor
- ▶ **Homeodomains**
MARY TRUSCOTT, ALAIN NEPVEU
- ▶ Homeostasis
- ▶ Homeotic Genes
- ▶ Homeotic Mutations
- ▶ Homogeneous Assay
- ▶ Homologous Recombination
- ▶ Homologous Recombination Repair (HRR)
- ▶ Homology
- ▶ **Homology Modeling**
MICHAEL NILGES
- ▶ Homology-Derived Secondary Structure of Proteins Database
- ▶ Homonuclear NMR Experiments
- ▶ Homophilic Binding
- ▶ Homozygosity Mapping
- ▶ Homozygote/Homozygous
- ▶ Horizontal Transmission
- ▶ Hormone(s)
- ▶ Horseradish Peroxidase
- ▶ Host Genome
- ▶ Hotspot Mutation
- ▶ Housekeeping Genes
- ▶ Hox Cluster
- ▶ Hox Code
- ▶ Hox Genes
- ▶ HpaII Tiny Fragments
- ▶ HPLC
- ▶ HPRT Vector
- ▶ HRR
- ▶ HS(A)N
- ▶ HSCs/HPCs
- ▶ HSF
- ▶ HSN
- ▶ HSP
- ▶ Hsp90
- ▶ HSPGs
- ▶ HSQC
- ▶ HSRs
- ▶ HSSP Database
- ▶ HTF Islands
- ▶ HTH
- ▶ HTLV
- ▶ Htr
- ▶ HTS
- ▶ HuCAL[®]
- ▶ Human Genome (Sequence) Project
- ▶ Human Herpesvirus Type 8
- ▶ Human Immunodeficiency Virus Type 1
- ▶ Human Leukocyte Antigens
- ▶ **Human Repetitive DNA**
ADAM PAVLICEK, VLADIMIR V. KAPITONOV,
JERZY JURKA
- ▶ Human T-Cell Leukemia Virus
- ▶ Humanization
- ▶ Humanized Antibodies
- ▶ Humoral Immune Response
- ▶ Huntington Chorea
- ▶ **Huntington's Disease**
STEFAN WIECZOREK, JÖRG T. EPPLEN
- ▶ HUPO
- ▶ Hybridization Probes
- ▶ Hybridization
- ▶ Hydrogel
- ▶ Hydrogen Bond
- ▶ Hydrolysis
- ▶ Hydropathy
- ▶ Hydrophilic-/Hydrophobic Cross-Linkers
- ▶ Hydrophobic
- ▶ Hydrophobic Interaction Chromatography
- ▶ Hydrophobic Interactions

- ▶ Hydrophobicity
- ▶ Hydroxyl
- ▶ Hydroxylase
- ▶ Hydroxyproline-Rich Glycoprotein
- ▶ **Hyper- and Hypoparathyroidism**
GEOFFREY N. HENDY
- ▶ Hypercalcemia of Malignancy
- ▶ Hyperferritinemia Cataract Syndrome
- ▶ Hyperglycemia
- ▶ Hyperkeratosis
- ▶ Hypermethylation
- ▶ Hyperparathyroidism
- ▶ Hyperphagia
- ▶ Hyperphosphorylation
- ▶ Hyperpolarization
- ▶ Hypersensitivity Drug Reactions
- ▶ Hypertension
- ▶ Hyperthyroidism
- ▶ Hyperthyroxinemia
- ▶ Hypertriglyceridemia
- ▶ Hypertrophic Chondrocyte
- ▶ Hypertrophy
- ▶ Hypnagogic Hallucinations
- ▶ Hypocholesterolemia
- ▶ Hypochondroplasia
- ▶ Hypocretins
- ▶ Hypogammaglobulinemia
- ▶ Hypoglossal Cord
- ▶ Hypogonadism
- ▶ Hypoinsulinemic
- ▶ Hypokalemia
- ▶ Hypomethylation
- ▶ Hypomorph/Hypomorphic Mutation
- ▶ Hypopituitarism
- ▶ Hypospadias
- ▶ **Hypothalamic and Pituitary Diseases, Genetics**
SALLY RADOVICK,
ELIZABETH A. ROCHOWICZ
- ▶ Hypothyroidism
- ▶ Hypoxanthine-Guanine-Phosphoribosyl-Transferase
- ▶ **Hypoxia Inducible Factors**
TILL ACKER
- ▶ Hypoxia/Normoxia
- ▶ Hypoxic Response
- ▶ I/A Domain
- ▶ IAP
- ▶ IAP
- ▶ IBD
- ▶ ICAM
- ▶ ICAT
- ▶ ICAT
- ▶ I-Cell Disease
- ▶ ICG-HNPCC
- ▶ Ichthyosis
- ▶ Ichthyosis Bullosa of Siemens
- ▶ Ichthyosis Hystrix Curth-Macklin
- ▶ Identity-by-Descent
- ▶ Idiopathic
- ▶ **Idiosyncratic Drug Reactions**
JACK UETRECHT
- ▶ IEF
- ▶ Ig Domain
- ▶ IGF 1
- ▶ IgV_H
- ▶ IHC
- ▶ IL
- ▶ IL-6 Family
- ▶ IMAC
- ▶ Image Analysis
- ▶ Imatinib Mesylate
- ▶ Imidazolium
- ▶ Immediate Early Genes
- ▶ Immobilized Metal-Affinity Chromatography
- ▶ Immobilized pH Gradients
- ▶ Immortalisation
- ▶ Immune Cells
- ▶ Immune System
- ▶ Immune Tolerance
- ▶ Immunity
- ▶ Immunization, Active and Passive
- ▶ Immunoblot
- ▶ **Immunochemical Methods, Localization**
LEE BERGMAN, STEPHANIE BECHTEL,
STEFAN WIEMANN
- ▶ Immunocytochemistry
- ▶ Immunofluorescence
- ▶ Immunogenic
- ▶ Immunoglobulin
- ▶ Immunoglobulin-Fold
- ▶ Immunohistochemistry

- ▶ Immunological Dysbalance
- ▶ **Immunological Methods, the Use of Antibodies to Discover the Function of Novel Gene Products**
MARKUS ENZELBERGER, THOMAS VON RÜDEN
- ▶ Immunophilins and Cyclophilins
- ▶ Immunoprecipitation
- ▶ Immunoproteomics
- ▶ Immunostaining
- ▶ Immunosuppressant
- ▶ Impedance
- ▶ Importin
- ▶ Imprinting
- ▶ In Silico (Procedures)
- ▶ *In Situ* Hybridization
- ▶ *In Vitro*
- ▶ In Vitro Binding (Beads)
- ▶ *In Vitro* Fertilization
- ▶ *In Vivo*
- ▶ *In Vivo* **Imaging of Transgenic Mice with Fluorescent Protein Expression**
FRITJOF HELMCHEN, FRANK KIRCHHOFF
- ▶ Inbred Strain/Inbreeding
- ▶ Incidence
- ▶ Inclusion Body
- ▶ Indifferent Embryonic Gonad
- ▶ Indirect Readout
- ▶ Induced Fit
- ▶ Inducible Expression
- ▶ Ineffective Erythropoiesis
- ▶ Infantile Refsum Disease
- ▶ Infantile Spasm
- ▶ Infantile Spinal Muscular Atrophy
- ▶ Inflammation
- ▶ **Inflammatory Response**
AHMED SHERIFF, MARTIN HERRMANN,
REINHARD E. VOLL, UDO S. GAIP, L,
JOACHIM R. KALDEN
- ▶ Informational Property Right
- ▶ Informed Consent
- ▶ Informed Contract
- ▶ In-Frame Deletion
- ▶ **Inherited Mental Retardation Syndromes**
NIELS TOMMERUP
- ▶ Inherited Neurodegenerative Disease
- ▶ Inhibitor of β -Catenin and TCF-4
- ▶ Inhibitor of Apoptosis Proteins
- ▶ Initiation Complex
- ▶ Initiator
- ▶ INK4a/ARF
- ▶ Innate Adjuvants
- ▶ Innate Immunity
- ▶ Inner Boundary Membrane
- ▶ Inner Ear
- ▶ Innexins
- ▶ Inr
- ▶ Insertion
- ▶ Insertion/Deletion Loop
- ▶ Insertional (Oncogen) Activation
- ▶ Insertional Activation/Inactivation
- ▶ Insulator
- ▶ Insulin
- ▶ Insulin Dependent Diabetes
- ▶ Insulin Receptor
- ▶ Insulin Resistance
- ▶ Integrale Membrane Proteins
- ▶ Integrase
- ▶ Integration
- ▶ Integrin Ligand
- ▶ **Integrin Signaling**
LAWRENCE E. GOLDFINGER,
MARK H. GINSBERG
- ▶ Integrin-Associated Protein
- ▶ Integrins
- ▶ Intein
- ▶ Intensified CCD Camera
- ▶ ICCD Camera
- ▶ Interaction Chromatography
- ▶ Interaction Discovery MappingTM
- ▶ Interaction Map
- ▶ Interaction Trap
- ▶ Interactome
- ▶ Intercalated Disc
- ▶ Intercellular Adhesion Molecule
- ▶ Intercellular Junctions
- ▶ Interference
- ▶ Interference Filter
- ▶ Interferon
- ▶ Interferon-Regulatory Factor 1
- ▶ Interleukin-1 Receptor Associated Kinase
- ▶ Interleukins

► **Interleukins and Their Receptors, a Role in Regulating Hematopoietic Stem Cells**

CLAUDIA ORELIO, ELAINE DZIERZAK

► **Intermediate Filaments**

E. BIRGITTE LANE

- Intermediate Mesoderm
- Intermembrane Space
- Internal Ribosomal Entry Sites
- Interneuron
- Internode
- Interphase
- Interpolar Microtubule
- Interspersed Repeats
- Intervertebral Disc
- Intrabodies
- Intron
- Intronic Mutation
- Intronic Sequence Element
- Inversin

► **Ion Channels/Excitable Membranes**

HEINRICH TERLAU, FRANK KIRCHHOFF

- Ion Trap
- Ion Trap Mass Spectrometer
- Ion-Exchange Chromatography
- Ionic Interactions
- Ionisation
- iProClass
- IQ
- IRAK
- IRD
- IRES
- IRF-1
- Iron Storage Disease
- I-Sce I Meganuclease
- Isochromosome
- Isoelectric Focusing
- Isoenzymes/Isozymes
- Isoform
- Isolated Tumour Cells
- Isometric Contraction
- Isomorphic Replacement
- Isopeptide Bond
- Isoprenoid
- Isoprenylation
- **Isothermal Titration Calorimetry**

TIMM REINISCH, HANS-JURGEN HINZ

► Isotonic Contraction

- Isotope Labeling
 - Isozymes
 - ITC
 - ITIM
 - Jag1 and 2
 - Jagged 1 and 2
 - JAK
 - JAK/STAT
 - Janus Kinase
 - Jet Lag
 - JNK
 - Joule Heating
 - **Jun/Fos**
- PETER ANGEL, MARINA SCHORPP-KISTNER
- Juvenile Diabetes
 - Juvenile Polyposis
 - Juvenile Spinal Muscular Atrophy
 - kanMX4
 - Karyopherins
 - Karyotype
 - Karyotypic Abnormalities
 - kb
 - KD
 - Kennedy Disease
 - Keratin
 - Keratin End Domains (Head and Tail)
 - Keratin Homologous Domain
 - Keratin Linker Domains/Regions
 - Keratin Rod Domain
 - Keratinocyte
 - **Kidney**
- SEPPO J. VAINIO
- Kinase
 - Kinetochore
 - Klinefelter Syndrome
 - Knockdown
 - Knock-In (Mutation)
 - Knock-out
 - Knock-Out Animals
 - Knock-Out Mice
 - Kremen
 - Kugelberg-Welander Disease
 - Kupffer Cells
 - Labeling
 - Laboratory Informatics Management System

- ▶ LacZ
- ▶ Lagging Strand
- ▶ Lamellipodium
- ▶ Laminar Flow
- ▶ Laminin
- ▶ Lamm Equation
- ▶ Langer-Giedion Syndrome
- ▶ LAP
- ▶ **Large Scale ENU Mutagenesis In Mice**
MARTIN HRABÉ DE ANGELIS,
JOHANNES BECKERS, SIBYLLE WAGNER,
DIAN SOEWARTO
- ▶ Large Scale Protein Production
- ▶ **Large-Scale Gene Trap Approaches in Mice**
PATRICIA RUIZ
- ▶ **Large-Scale Homologous Recombination Approaches in Mouse Embryonic Stem Cells**
MICHAEL V. WILES, GABRIELE PROETZEL
- ▶ Laryngospasm
- ▶ Laser Capture Microdissection
- ▶ Laser Interferometer
- ▶ Lateral Element
- ▶ Lateral Inhibition
- ▶ LCA
- ▶ LCM
- ▶ LCR
- ▶ LCRs
- ▶ LD
- ▶ LDL
- ▶ LDL Receptor
- ▶ LDL-A Domain
- ▶ LDLR-/-Mice
- ▶ Leader Signal
- ▶ Leading Strand
- ▶ Learning Disabilities
- ▶ Leber's Congenital Amaurosis
- ▶ Leber's Hereditary Optic Neuropathy
- ▶ LEF/TCF Family
- ▶ Legless(Lgs)/BCL19
- ▶ Lens Connexins
- ▶ Lentiviruses
- ▶ Leptin
- ▶ Leptomycin B
- ▶ Leptotene
- ▶ Lethal Mutations
- ▶ Leucine Rich Repeat
- ▶ Leucine Zipper
- ▶ **Leucine Zipper Transcription Factors: bZIP Proteins**
ANDREAS G. BADER, PETER K. VOGT
- ▶ **Leukemia**
RALF BARGOU, WOLF-DIETER LUDWIG
- ▶ Leukocytes
- ▶ Leukodystrophies
- ▶ Lewy Body
- ▶ Leydig Cells
- ▶ LFS
- ▶ Lgs/BCL19
- ▶ LH
- ▶ LH Domain
- ▶ Library
- ▶ Library Amplification
- ▶ Liddle's Syndrome
- ▶ LIE
- ▶ Li-Fraumeni Syndrome
- ▶ Ligand
- ▶ Ligation
- ▶ LIM Gene
- ▶ **Limb Development**
NATALIA SOSHNIKOVA, WALTER BIRCHMEIER
- ▶ **Limb Girdle Muscular Dystrophies**
ISABELLE RICHARD
- ▶ Limbic System
- ▶ LIMS
- ▶ Lin
- ▶ LINE
- ▶ Lineage Determination
- ▶ Linear Unmixing
- ▶ Linkage
- ▶ Linkage Analysis
- ▶ Linkage Disequilibrium
- ▶ Linkage Map
- ▶ Linker DNA
- ▶ Lipid Asymmetry
- ▶ Lipid Peroxidation
- ▶ Lipid Rafts
- ▶ Lipidophilicity
- ▶ Lipopolysaccharide
- ▶ Lipoprotein
- ▶ Liposomes
- ▶ Liquid Chromatography

- ▶ Liquid Phase Photo-Polymerization
- ▶ LOAD
- ▶ Loader
- ▶ Localization Precision
- ▶ Locus
- ▶ Locus Control Region
- ▶ Locus Heterogeneity
- ▶ Lod Score
- ▶ LOH
- ▶ Long (Range) PCR
- ▶ Long Interspersed Repeat
- ▶ Long Term Depression
- ▶ Long Term Potentiation
- ▶ Long Terminal Repeats
- ▶ Loss of Function Mutations
- ▶ Loss of Heterogeneity
- ▶ Loss of Heterozygosity
- ▶ Loss-of-Function Screen
- ▶ Loss-of-Function Studies
- ▶ LOV Domain
- ▶ Lov-1
- ▶ Low Copy Repeats (LCRs)
- ▶ Low Density Lipoprotein
- ▶ Low Density Lipoprotein Receptor Related Proteins 5 and 6 (LRP5/6)
- ▶ Lower Motor Neurons
- ▶ *LoxP* and Cre Recombinase System
- ▶ *LoxP* Site
- ▶ LP3
- ▶ LPS
- ▶ LPS Recognition Complex
- ▶ LRP5/6
- ▶ LR-PCR
- ▶ LTBP
- ▶ LTD
- ▶ LTP
- ▶ LTRs
- ▶ L-Type Calcium Channel
- ▶ **Lung**
SP DE LANGHE, FG SALA, AA MAILLEUX,
S BELLUSCI
- ▶ Lupus
- ▶ Luteinizing Hormone
- ▶ Lymphangiogenesis
- ▶ Lymphangiomyomatosis
- ▶ Lymphocyte
- ▶ Lynch Syndrome
- ▶ Lyonization
- ▶ Lysolecithin
- ▶ Lysosomal Acid Hydrolases
- ▶ Lysosomes
- ▶ M Phase
- ▶ MA
- ▶ Machado-Joseph Disease
- ▶ Machine Learning
- ▶ Macromolecular Crowding
- ▶ Macrophage
- ▶ Macula
- ▶ **MAD Phasing**
JÜRGEN J. MÜLLER, UDO HEINEMANN
- ▶ Madin-Darby Canine Kidney Cells
- ▶ Maguk Proteins
- ▶ Major and Minor Groove
- ▶ Major Histocompatibility Complex
- ▶ Major Locus
- ▶ Maladie de G lineau
- ▶ Malar Hypoplasia
- ▶ MALDI
- ▶ MALDI-MS
- ▶ MALDI-TOF-MS
- ▶ Malignancy
- ▶ Malignant Transformation
- ▶ **Mammalian Fertilization**
EVELINE S. LITSCHER, PAUL M. WASSARMAN
- ▶ Manic Depression
- ▶ ManNAc
- ▶ Mannose 6-Phosphate Receptor
- ▶ Map Distances
- ▶ MAPK
- ▶ MAPK Pathway
- ▶ Mapping Function
- ▶ MAPs
- ▶ **Marfan Syndrome**
PHILIP F. GIAMPIETRO
- ▶ Marker
- ▶ Marker X Syndrome
- ▶ Martin Bell Syndrome
- ▶ MAS
- ▶ Mask
- ▶ Mass Action Law
- ▶ Mass Spectrometer
- ▶ Mass Spectrometric Fragmentation

- ▶ Mass Spectrometry
- ▶ **Mass Spectrometry: ESI**
EVA-CHRISTINA MÜLLER
- ▶ **Mass Spectrometry: MALDI**
JOHAN GOBOM
- ▶ **Mass Spectrometry: MS/MS**
EBERHARD KRAUSE
- ▶ **Mass Spectrometry: Quantitation**
OLIVER FIEHN, WOLFRAM WECKWERTH
- ▶ **Mass Spectrometry: SELDI**
GURU REDDY, RICHARD RUBIN,
SCOT WEINBERGER
- ▶ Maternal Transcript
- ▶ Matrigel™
- ▶ Matrix
- ▶ Matrix Metalloproteinases
- ▶ Matrix-Assisted Laser Desorption/Ionization
Mass Spectrometry
- ▶ Matrix-CGH (Comparative Genomic
Hybridization)
- ▶ Maturation Process
- ▶ Maximal Shortening Velocity
- ▶ Maxizyme
- ▶ MBP
- ▶ MCA/MR Syndrome
- ▶ McCune-Albright Syndrome
- ▶ MCD
- ▶ McKusick Number
- ▶ MCM
- ▶ MDCK Cells
- ▶ MDR
- ▶ MDR Modulators
- ▶ MDR1
- ▶ Meconium Ileus
- ▶ **Medaka as a Model Organism for
Functional Genomics**
FELIX LOOSLI, JOACHIM WITTBRODT
- ▶ Medial/Lateral Ganglionic Eminence
- ▶ Medicinal Product
- ▶ Medulloblastoma
- ▶ Meesmann Corneal Dystrophy
- ▶ MEF
- ▶ Megalin
- ▶ Meiosis
- ▶ **Meiosis and Meiotic Recombination**
CHARLES TEASE
- ▶ Meiotic Mapping
- ▶ Meiotic Maturation
- ▶ Meiotic Mismatch
- ▶ Meiotic Recombination
- ▶ Melanocytes
- ▶ Melanosomes
- ▶ MELAS
- ▶ Melatonin
- ▶ Melting Temperature
- ▶ Membrane
- ▶ Membrane Protein
- ▶ Memory of Addiction
- ▶ MEN1
- ▶ Mendelian (Inheritance)
- ▶ Mendelian Disorder
- ▶ **Mendelian Forms of Human Hypertension
and Mechanisms of Disease**
FRIEDRICH C. LUFT
- ▶ Mendelian Hypertension
- ▶ Mendelian Inheritance in Man
- ▶ Mental Retardation
- ▶ Mesenchymal Stem Cells
- ▶ Mesenchyme
- ▶ Mesoderm
- ▶ Mesonephros
- ▶ Messenger RNA
- ▶ **Messenger RNA Stability**
SHIRLEY R. BRUCE,
MILES F. WILKINSON
- ▶ Meta-Analysis
- ▶ Metabolic Alkalosis
- ▶ Metabolite Fingerprinting
- ▶ Metabolite Phenotype
- ▶ Metabolite Profiling
- ▶ Metabolome Profiling
- ▶ **Metabolomics**
JOACHIM KOPKA
- ▶ Metacentric Chromosome
- ▶ Metachromatic Leukodystrophy
- ▶ Metalloprotease
- ▶ Metanephric Blastema
- ▶ Metaphase
- ▶ Metaphase Plate
- ▶ Metastable Disulfide Species
- ▶ Metastasis
- ▶ Methotrexate

- ▶ 5-Methylcytosine
- ▶ Methyl Ester
- ▶ 7-Methyl-Guanosine
- ▶ Methyl-AcylCoA Racemase Deficiency

▶ **Methylation of Proteins**

MARK T. BEDFORD

- ▶ Mf
- ▶ M-FISH
- ▶ MFS
- ▶ MGMT
- ▶ MHC
- ▶ MIAME

▶ **Microarray Data Analysis**

ALVIS BRAZMA

- ▶ Microarray Technology
- ▶ Microarrays

▶ **Microarrays in Colorectal Cancer**

JUDITH M. BOER

▶ **Microarrays in Pancreatic Cancer**

MALTE BUCHHOLZ, THOMAS M. GRESS

▶ **Microarrays in Plant Genomics**

NIKOLAUS L. SCHLAICH,
ALAN J. SLUSARENKO

▶ **Microarrays in Rheumatoid Diseases**

THOMAS HÄUPL

- ▶ Microcalorimetry
- ▶ Microcephalus
- ▶ Microchannel
- ▶ Microdeletion

▶ **Microdeletion Syndromes**

OLIVER BARTSCH, EVA SEEMANOVÁ

- ▶ Microfilaments
- ▶ Microfluidics
- ▶ Microglia
- ▶ Micrometastasis
- ▶ Micrometastatic Tumour Cells
- ▶ Micromolding

▶ **MicroRNA**

V. NARRY KIM

- ▶ Microsatellite/Microsatellite Marker
- ▶ Microscope Slide
- ▶ Microtiter Dishes
- ▶ Microtubule
- ▶ Microtubule Associated Proteins
- ▶ Microtubule Motor Proteins
- ▶ Microtubule-Organizing Center

▶ **Microvilli**

EVELYNE FRIEDERICH, DANIEL LOUVARD

- ▶ MIDAS
- ▶ MIF
- ▶ Migration
- ▶ Milk-Alkali Syndrome
- ▶ Miller-Dieker Syndrome
- ▶ Miller Planes
- ▶ MIM
- ▶ MIN/MSI
- ▶ Mineralocorticoid Receptors
- ▶ Mineralocorticoid Resistance
- ▶ Mineralocorticoids
- ▶ Minisatellites
- ▶ Minor Groove
- ▶ miRNA
- ▶ miRNP
- ▶ Misexpression
- ▶ Misfolded Proteins
- ▶ Mismatch (Nucleotide)
- ▶ Mismatch Repair
- ▶ Missense

▶ **Mitochondria – Biogenesis and Structural Organization**

ANDREAS S. REICHERT, WALTER NEUPERT

- ▶ Mitochondrial Disease
- ▶ Mitochondrial DNA
- ▶ **Mitochondrial Genome**
- ▶ Mitochondrial Membrane
- ▶ **Mitochondrial Myopathy**

LENE SÖRENSEN, NILS-GÖRAN LARSSON

- ▶ Mitochondrial Permeability Transition Pore
- ▶ Mitogen-Activated Protein Kinases
- ▶ Mitosis
- ▶ Mitotic Figure
- ▶ **Mitotic Recombination**
- ▶ Mitotic Recombination
- ▶ Mitotic Spindle Poles
- ▶ **Mitotic Spindles**
- ▶ RYOKO KURIYAMA
- ▶ Mitral Valve Prolapse
- ▶ Mixed Disulfide
- ▶ Mixed Gonadal Dysgenesis

- ▶ Mixing Time
- ▶ MJD
- ▶ MLC
- ▶ MLH
- ▶ MM-PBSA
- ▶ MMPs
- ▶ MMR
- ▶ MMTV
- ▶ Mobile Fraction
- ▶ Mobile Genetic Elements
- ▶ Mode of Release
- ▶ Modified Antisense Oligonucleotide
- ▶ Modifiers
- ▶ MODY
- ▶ Moesin
- ▶ **Molecular Aging Research**
CHRISTIAN BEHL
- ▶ Molecular Chaperones and Cochaperones
- ▶ Molecular Clock
- ▶ Molecular Complementarity
- ▶ **Molecular Docking**
KARIN SCHLEINKOFER, TING WANG,
REBECCA C. WADE
- ▶ Molecular Dynamics Simulations
- ▶ **Molecular Dynamics Simulations in Drug Design**
CHRISTOPH A. SOTRIFFER
- ▶ Molecular Imaging
- ▶ Molecular Imprints
- ▶ Molecular Mechanics Force Field
- ▶ Molecular Mimicry
- ▶ **Molecular Motors**
MANFRED SCHLIWA
- ▶ Molecular Symptom
- ▶ Molten Globule
- ▶ Monilethrix
- ▶ **Monoclonal Antibodies**
BURKHARD MICHEEL
- ▶ Monoclonality/Monoclonal
- ▶ Monogenic Disorder/Monogenic
- ▶ Monogenic Hypertension
- ▶ Monogenic Inheritance
- ▶ Monolayer
- ▶ Monooxygenases
- ▶ Monosaccharide
- ▶ Monosomy 1p36 Syndrome
- ▶ Mono-Ubiquitination
- ▶ Monozygotic Twins
- ▶ **Morbus Wegener**
JULIE M. WILLIAMS,
CAROLINE O.S. SAVAGE
- ▶ Morgan Unit(s)
- ▶ Morphant
- ▶ Morphogen/Morphogenic Factor
- ▶ Morphogenesis
- ▶ **Morpholino Oligonucleotides, Functional Genomics by Gene ‘Knock-down’**
SALIM ABDELILAH-SEYFRIED
- ▶ Mosaic/Mosaicism
- ▶ Motif
- ▶ Motif Databases
- ▶ Motor Proteins
- ▶ **Mouse Genomics**
LOUISE VAN DER WEYDEN, DAVID J. ADAMS,
ALLAN BRADLEY
- ▶ Moving Platform
- ▶ MPNSTs
- ▶ MPTP
- ▶ **MR Imaging**
MATTHIAS G. FRIEDRICH
- ▶ mRNA
- ▶ **mRNA Capping**
YASUHIRO FURUICHI, AARON J. SHATKIN
- ▶ mRNA Decay
- ▶ mRNA Processing
- ▶ mRNA Splicing
- ▶ mRNA Stability
- ▶ mRNA Translation
- ▶ mRNP
- ▶ MS
- ▶ MS/MS
- ▶ MSA
- ▶ MSH
- ▶ MSI
- ▶ mtDNA
- ▶ Mucin
- ▶ Mucocutaneous Candidiasis
- ▶ Mucoviscidosis
- ▶ MudPIT
- ▶ Multicloning Site
- ▶ Multidimensional Chromatography

► **Multidimensional NMR spectroscopy**

PETER SCHMIEDER

► **Multidimensional Protein Identification Technology**

► **Multidrug Resistance**

ULRIKE STEIN

- Multidrug Resistance Gene 1
- Multidrug Resistance - Associated Protein 1
- Multifactorial or Common Diseases
- Multifactorial Threshold
- Multifunctional Enzymes
- Multiple Alignments
- Multiple Endocrine Neoplasia Type 1
- Multiple Sclerosis
- Multiple Sequence Alignment
- Multiple System Atrophy
- Multiplex Quantitative PCR
- Multiplexed Expression Fluorescence in Situ Hybridization
- Multiplexing
- Multipotent
- Multipotent Stem Cells
- Multipotent/Multipotency
- Multi-Wavelength Anomalous Diffraction
- Munc13
- Munc18
- Muscle Atrophy

► **Muscle Contraction**

INGO MORANO

► **Muscle Development**

HANS-HENNING ARNOLD

- Muscular Dystrophy
- Muscular Hypotonia
- Mutagen
- Mutagenesis

► **Mutagenesis Approaches in Medaka**

HIROYUKI TAKEDA

► **Mutagenesis Approaches in the Zebrafish**

ALEXANDER GRUNDT, WOLFGANG ROTTBAUER

► **Mutagenesis Approaches in Yeast**

MICHAEL BREITENBACH, PETER LAUN

- Mutant (Phenotype)
- Mutation
- Mutation Rate
- Mutator Phenotype
- Myalgia

► Myelin

► Myelination

► Myeloperoxidase

► Myocardial Infarction

► Myocardium

► Myocytes

► Myofibrillogenesis

► Myofibrils

► Myofilament

► Myoglobinuria

► Myosin

► Myotome

► Myotonic Dystrophy Protein Kinase

► **Myotonic Dystrophy Type 1**

DERICK G. WANSINK, BÉ WIERINGA

► Myotonic Dystrophy Type 2

► Myristoylation Signal

► N-Acetyl-D-Mannosamine

► N-Acetylneuraminic Acid

► N-Acylation

► NADPH Oxidase

► NADPH + H⁺/NAD⁺

► Naked Cuticle

► 'Naked' DNA

► NALD

► **Narcolepsy**

THOMAS POLLMÄCHER

► Narcolepsy-Cataplexy

► Native State

► **Natural Antisense Transcripts**

ANDREAS WERNER

► NC

► NCAM

► NCBI

► NCGR

► NCV

► Nearest Neighbour

► Nearest Neighbour Model

► Necrosis

► Neighbour-Joining

► Nemo Like Kinase

► N-End Rule

► Neocentromeres

► Neocortex/Cortical Column

► Neonatal

► Neonatal Adrenoleukodystrophy

- ▶ Neoplasia
- ▶ Nephrocalcinosis
- ▶ Nephrogenic Mesenchyme
- ▶ Nephrolithiasis
- ▶ Nephron
- ▶ Nephropathy
- ▶ NER
- ▶ Nerve Conduction Velocity
- ▶ NES
- ▶ Nestin
- ▶ N-Ethylmaleimide-Sensitive-Factor
- ▶ NeuN
- ▶ **Neural Crest Cells and their Derivatives**
CHAYA KALCHEIM
- ▶ **Neural Development**
THOMAS MÜLLER
- ▶ Neural Networks
- ▶ Neural Plate
- ▶ **Neural Stem Cells**
GERD KEMPERMANN
- ▶ Neural Tube
- ▶ Neuraminic Acid
- ▶ Neuraminidase
- ▶ Neurodegeneration
- ▶ Neuroectoderm
- ▶ Neurofibrillary Tangles
- ▶ Neurofibromas
- ▶ **Neurofibromatosis Type 1 (NF1), Genetics**
MEENA UPADHYAYA, DAVID N. COOPER
- ▶ Neurofibromin
- ▶ Neurogenic Genes
- ▶ Neuromuscular Junction
- ▶ **Neurons**
JOACHIM LÜBKE, DIRK FELDMEYER
- ▶ Neuropathy
- ▶ Neurotransmission
- ▶ Neurotransmitter
- ▶ Neurotransmitter Release
- ▶ Neurotrophic Factor Hypothesis
- ▶ **Neurotrophic Factors**
ANNA MARIA CALELLA,
LILIANA MINICHELLO
- ▶ Neutrophils
- ▶ NEXT
- ▶ NF1 Gene
- ▶ NFT
- ▶ NF- κ B
- ▶ **NF- κ B Pathway**
CLAUS SCHEIDEREIT, DANIEL KRAPPMANN
- ▶ NGF
- ▶ N-Glycan
- ▶ NHEJ
- ▶ Nick-Closing Enzymes
- ▶ Nicotine
- ▶ Nieuwkoop Center
- ▶ Ni-NTA
- ▶ Nintra
- ▶ Nitrosative Stress
- ▶ Nkd
- ▶ N-Linked Oligosaccharide
- ▶ NLK
- ▶ NLS
- ▶ NMD
- ▶ NMDA Receptor
- ▶ NMR
- ▶ NMR Spectroscopy
- ▶ **NMR-based Screening**
WOLFGANG JAHNKE
- ▶ N-Nitroso-N-Ethylurea
- ▶ Nodal
- ▶ Nodes of Ranvier
- ▶ NOE
- ▶ NOESY
- ▶ Non Rapid-Eye-Movement Sleep
- ▶ Noncoding RNA
- ▶ Nondisjunction
- ▶ **Non-Histone Chromatin Proteins**
MONICA FEDELE, ALFREDO FUSCO
- ▶ Non-Homologous End Joining
- ▶ NonREM (Non Rapid-Eye-Movement) Sleep
- ▶ Nonsense Mutation
- ▶ Nonsense-Mediated RNA Decay
- ▶ Non-Silent Substitution
- ▶ Nonsyndromal/Nonsyndromic Deafness
- ▶ Non-Uniform Sampling
- ▶ Normalization
- ▶ Normoxia
- ▶ NORs
- ▶ Northern Blot
- ▶ **Notch Pathway**
THOMAS KLEIN
- ▶ Notochord

- ▶ NPC
- ▶ NPxY/F Motif
- ▶ NSF
- ▶ N-TAD
- ▶ N-Terminal Myristoylation
- ▶ NTPase
- ▶ Nuclear Acidic Proteins
- ▶ **Nuclear Compartments**
ALISTAIR E. T. NEWALL, JAYSON WANG,
DENISE SHEER
- ▶ Nuclear Envelope
- ▶ Nuclear Export Factor 1
- ▶ Nuclear Export Signal
- ▶ **Nuclear Hormone Receptors: Regulators of
Gene Transcription and Cellular Signaling**
ANDREW C.B. CATO, LIUBOV SHATKINA
- ▶ **Nuclear Import and Export**
ANJA PANNEK, KATRIN STADE
- ▶ Nuclear Lamina
- ▶ Nuclear Localization Signal
- ▶ Nuclear Magnetic Resonance
- ▶ Nuclear Matrix
- ▶ Nuclear Medicine
- ▶ Nuclear Poly(A) Binding
Protein 1
- ▶ **Nuclear Pore Complex**
BIRTHE FAHRENKROG, UELI AEBI
- ▶ Nuclear Receptor Co-Activators
- ▶ Nuclear Receptors
- ▶ Nuclease
- ▶ Nucleic Acid
- ▶ Nucleobase
- ▶ Nucleocapsid
- ▶ Nucleolar Organising Regions
- ▶ Nucleolus
- ▶ Nucleophile/Nucleophilic
- ▶ Nucleoporins
- ▶ Nucleoside, Nucleotide
- ▶ Nucleosome Remodelling (Enzymes/Factors)
- ▶ Nucleosome Repositioning
- ▶ **Nucleosomes**
LOREDANA VERDONE, MICAELA CASERTA
- ▶ Nucleotide
- ▶ Nucleotide Binding Fold
- ▶ **Nucleotide Biosynthesis**
RICHARD I. CHRISTOPHERSON
- ▶ **Nucleotide Excision Repair**
JAMES E. CLEAVER
- ▶ Nucleotide Microarrays
- ▶ Nucleus
- ▶ Null Allele/Null Phenotype
- ▶ Null Mice
- ▶ Null Phenotype
- ▶ NXF1
- ▶ O6-Alkylguanine-DNA Alkyltransferases
- ▶ O6-Methylguanine-DNA Methyltransferase
- ▶ Obesity
- ▶ Oculopharyngeal Muscular Dystrophy
- ▶ ODD
- ▶ OFC
- ▶ OFUT-1
- ▶ Okazaki Fragment
- ▶ Oligo Microarray
- ▶ Oligoarticular
- ▶ Oligodendrocyte
- ▶ Oligogenic
- ▶ Oligomerization
- ▶ Oligonucleotides
- ▶ Oligosaccharide
- ▶ Oligosaccharyltransferase
- ▶ OMIM (TM)
- ▶ Oncogen Activation
- ▶ **Oncogene**
MANFRED SCHWAB
- ▶ Oncogenesis
- ▶ Oncoprotein
- ▶ Oncostatin M
- ▶ Onionskin Mechanism
- ▶ Online Mendelian Inheritance in Man
- ▶ Oogenesis
- ▶ Oogonia
- ▶ Open Reading Frame
- ▶ Operons
- ▶ Opiates
- ▶ Opsonization
- ▶ ORC
- ▶ Orexins
- ▶ ORF
- ▶ Organelle
- ▶ Organizer
- ▶ Organogenesis
- ▶ Ori

- ▶ Origin of Bidirectional Synthesis
- ▶ Origin Recognition Complex
- ▶ Orofacial Clefts
- ▶ Ortholog/Orthologous Genes/Orthologous Proteins
- ▶ Orthostatic Hypotension
- ▶ Oscillator
- ▶ OST
- ▶ Osteitis
- ▶ Osteitis Fibrosa Cystica
- ▶ Osteoblastogenesis
- ▶ Osteoblasts
- ▶ Osteochondrodysplasias
- ▶ Osteoclast(s)
- ▶ Osteoclastogenesis
- ▶ Osteocyte
- ▶ Osteogenesis Imperfecta
- ▶ Osteoid
- ▶ Osteomalacia
- ▶ Osteopenia
- ▶ Osteopetrosis
- ▶ Osteoporosis
- ▶ Osteoprogenitor
- ▶ Osteosclerosis
- ▶ Overexpression/Misexpression
- ▶ Overfitting
- ▶ OXA1 Complex
- ▶ Oxidases
- ▶ Oxidation
- ▶ Oxidative Phosphorylation Disease
- ▶ Oxidative Stress
- ▶ 2–Oxoglutarate-Dependent Hydroxylase
- ▶ Oxygen Sensing
- ▶ Oxytocin
- ▶ p21
- ▶ P300/CBP
- ▶ p53
- ▶ PABP
- ▶ PABPN1
- ▶ PAC
- ▶ PAC Transgene
- ▶ Pacemaker
- ▶ Pachyonychia Congenita
- ▶ Paclitaxel
- ▶ PAGE
- ▶ 2D-PAGE Database
- ▶ Paget's Disease
- ▶ PAK
- ▶ Pal-CoA
- ▶ Palindrome
- ▶ Palmitoylation
- ▶ Palmoplantar Keratoderma
- ▶ PAMP
- ▶ Pancreatic Insufficiency
- ▶ Pancreatitis
- ▶ Panhypopituitarism
- ▶ Pannexins
- ▶ Pannus
- ▶ Papilloma Viruses
- ▶ PAR
- ▶ Paracrine
- ▶ Paralog/Paralogous Genes/Paralogous Proteins
- ▶ Paralysis Agitans
- ▶ Parathyroid Hormone
- ▶ Parathyroid Hormone-Related Protein
- ▶ Paratope
- ▶ Paraxial Mesoderm
- ▶ Paresthesia
- ▶ PARG
- ▶ **Parkinson's Disease: Insights from Genetic Causes**
MELISA J BAPTISTA, MARK R COOKSON,
KATRINA GWINN-HARDY
- ▶ PARN
- ▶ PARP1
- ▶ Parsimony
- ▶ Partial Specific Volume
- ▶ Particle-Coated DNA
- ▶ PAS Domain
- ▶ **Patch Clamping**
CLAUDIA EDER
- ▶ Pathomechanism
- ▶ Pathway Particle
- ▶ Pattern Recognition Receptors
- ▶ Patterson Function/Patterson Map
- ▶ PCR
- ▶ PDGF
- ▶ PDZ Domains
- ▶ PE
- ▶ Pectus Carinatum

- ▶ Pectus Excavatum
- ▶ Pedigree
- ▶ P-Element
- ▶ Pemphigus
- ▶ PEN-2
- ▶ Penetrance
- ▶ Peptamers
- ▶ Peptide
- ▶ **Peptide Aptamers**
PIERRE COLAS
- ▶ **Peptide Chips**
RUDOLF VOLKMER-ENGERT
- ▶ Peptide Mass Fingerprinting
- ▶ Peptide Mass Map
- ▶ Peptide Release Factor 2
- ▶ Peptidomimetic
- ▶ **Peptidyl Prolyl *cis/trans* Isomerases**
CORDELIA SCHIENE-FISCHER
- ▶ Peptidylglycine Amidation
- ▶ Peptidyl-Transferase Center
- ▶ Percentage Chimaerism
- ▶ Perforated Patch Recording
- ▶ Perichondrium
- ▶ Perinatal
- ▶ Periosteum
- ▶ Peripheral Membrane Proteins
- ▶ Peripheral Nervous System
- ▶ Periplasm
- ▶ Permeability Barrier
- ▶ Peroxin
- ▶ Peroxiredoxin
- ▶ **Peroxisomal Disorders**
RONALD J.A. WANDERS
- ▶ Peroxisome
- ▶ Peroxisome Biogenesis
- ▶ Peroxisome Proliferator-Activated Receptor
- ▶ Peroxisome Targeting Signal
- ▶ Perturbagens
- ▶ Pertussis Toxin
- ▶ Pes Planus
- ▶ PET
- ▶ Petite Mutant
- ▶ **Peutz-Jeghers Syndrome**
DIETER E. JENNE
- ▶ PEV
- ▶ PEX Gene
- ▶ PEX7 Gene
- ▶ PFAM Database
- ▶ PFGE
- ▶ PGC
- ▶ PGD
- ▶ PHA
- ▶ PHA 1
- ▶ PHA 2
- ▶ Phage Artificial Chromosome
- ▶ Phage Display
- ▶ Phage Surface Display
- ▶ Phagemid
- ▶ Phagocytosis
- ▶ Pharmacology
- ▶ Pharmacogenetics
- ▶ **Pharmacogenomics**
MICHEL EICHELBAUM,
MATTHIAS SCHWAB
- ▶ Pharmacophore
- ▶ Phase
- ▶ Phase Problem
- ▶ Phases of the Cell Cycle
- ▶ Phenocopies
- ▶ **Phenomics**
ROBERT GERLAI
- ▶ Phenotype
- ▶ Phenotypical Characterization
- ▶ Phenotypical Profiling
- ▶ Phenotypical Screening
- ▶ Phenotyping
- ▶ Phenylalanine Hydroxylase Deficiency
- ▶ **Phenylketonuria**
ULRICH LANGENBECK
- ▶ Pheochromocytomas
- ▶ Phlebotomy
- ▶ Phosphatidyl Serine
- ▶ Phosphatidylinositides
- ▶ Phosphatidylinositol 3-Kinase
- ▶ 3'-Phosphoadenylyl-Sulfate
- ▶ Phosphodiester
- ▶ Phosphoinositides
- ▶ Phospholamban
- ▶ Phospholipase A2
- ▶ Phosphor-Imaging
- ▶ Phosphorylation
- ▶ Photobleaching

- ▶ Photoconversion
- ▶ Photolithographic Synthesis
- ▶ Photolithography
- ▶ Photomorphogenesis
- ▶ Photoreceptor Cells
- ▶ **Photoreceptors**
MARTIN ENGELHARD,
KLAUS PETER HOFMANN
- ▶ Photorelease
- ▶ Photoremovable Protecting Group
- ▶ Phototaxis
- ▶ Phototrigger (of Biomolecules)
- ▶ PHP
- ▶ Phylogeny
- ▶ Physical Maps
- ▶ Phytanic Acid
- ▶ Phytanoyl-CoA Hydroxylase
- ▶ PI3K
- ▶ PI3K Pathway
- ▶ PIC
- ▶ Piezo-Electricity
- ▶ Pigment Spot Polyposis
- ▶ Pilomatricomas
- ▶ PIP2
- ▶ Pituitary Diseases
- ▶ PJS
- ▶ PKC
- ▶ PKD Domain
- ▶ PKD1 and PKD2
- ▶ PKR
- ▶ PKU
- ▶ PLA2
- ▶ Plakins
- ▶ Plant Functional Genomics
- ▶ **Plant Genomics**
THOMAS ALTMANN
- ▶ Plasma Membrane
- ▶ Plasmid/Plasmid Vector
- ▶ Pleiotropic/Pleiotropic Gene Effect
- ▶ Plexiform Neurofibromas
- ▶ Pluri-/Multipotent
- ▶ PMF
- ▶ PML Nuclear Bodies
- ▶ PMS
- ▶ PNA
- ▶ **PNA Chips**
ANETTE JACOB, OLE BRANDT,
JÖRG D. HOHEISEL
- ▶ PNET
- ▶ Pneumothorax
- ▶ PNS
- ▶ Podophyllotoxins
- ▶ Point Mutation
- ▶ Point Spread Function
- ▶ Poisson-Boltzmann Electrostatics
- ▶ Pol
- ▶ Pol I
- ▶ Polaris
- ▶ Polarizability
- ▶ POLG
- ▶ Poly(A) Polymerase
- ▶ Poly(A) Ribonuclease
- ▶ Poly(A) Tail
- ▶ Poly(A)-Binding Protein
- ▶ Poly(ADP-Ribose) Glycohydrolase
- ▶ Poly(ADP-Ribose) Polymerase (PARP)–1
- ▶ Poly(ADP-Ribosyl)ation
- ▶ **Poly(ADP-Ribosyl)ation, Pathophysiology**
ALEXANDER BÜRKLE
- ▶ Polyacrylamide Gel Electrophoresis
- ▶ **Polyadenylation**
ELMAR WAHLE, NIELS H. GEHRING
- ▶ Polyarticular
- ▶ Polycistronic Constructs
- ▶ **Polycystic Kidney Disease, Autosomal Dominant**
MICHAEL SUTTERS
- ▶ Polycystin
- ▶ Polygenic/Polygenetic
- ▶ **Polyglutamine Disease, the Emerging Role of Transcription Interference**
J. PAUL TAYLOR, TIM W. RATZLAFF
- ▶ Polymerase
- ▶ Polymerase Chain Reaction
- ▶ Polymorphism/Polymorphic
- ▶ Polynucleotide Kinase
- ▶ Polypeptide
- ▶ Polyploidy
- ▶ Polyprotein
- ▶ Polypos-and-Spots Syndrome

- ▶ Polysaccharide
- ▶ Polytene Chromosomes
- ▶ Population Stratification
- ▶ Porphyria Cutanea Tarda
- ▶ Porphyrias
- ▶ Position Effect
- ▶ Position Effect Variegation
- ▶ Positional Candidate Cloning
- ▶ Positional Cloning
- ▶ Positive Reinforcer
- ▶ Positron
- ▶ **Positron Emission Tomography**
JOERG VAN DEN HOFF
- ▶ Posterior Prevalence
- ▶ Postmeiotic Segregation
- ▶ Post-Mitotic
- ▶ Postnatal
- ▶ Postsynaptic Density
- ▶ Post-Translational Modification
- ▶ POU Transcription Factors
- ▶ PP2A
- ▶ PPAR
- ▶ PPAR
- ▶ PPAR α -Consensus Elements
- ▶ PR
- ▶ PR (Protease)
- ▶ **Prader-Willi and Angelman Syndromes**
BERNHARD HORSTHEMKE
- ▶ Prader-Willi Syndrome
- ▶ Prechondrogenic
- ▶ Precipitant
- ▶ Precursor Cell
- ▶ Predictive Genetic Testing
- ▶ **Predictive Testing and Genetic Counseling**
ANNEMARIE SCHWAN, JÖRG T. EPPLER
- ▶ Preimplantation Genetic Diagnosis
- ▶ Preinitiation Complex
- ▶ Preintegration Complex
- ▶ Premature Termination Codon
- ▶ Premenstrual Manifestation
- ▶ pre-miRNA
- ▶ pre-mRNA
- ▶ Premutation
- ▶ Prenatal Diagnostics
- ▶ Prenyl Group
- ▶ Prenylation
- ▶ Preproorexin
- ▶ Presomitic Mesoderm
- ▶ Presymptomatic Diagnosis
- ▶ Prevalence
- ▶ Prey
- ▶ PRF
- ▶ P-Rib-PP
- ▶ Primary Aldosteronism
- ▶ Primary Body Axis
- ▶ Primary Cells
- ▶ Primary Constriction
- ▶ Primary Hyperparathyroidism
- ▶ Primary Hypertension
- ▶ Primary Structure
- ▶ Primitive Streak
- ▶ Primordial Germinal Cells
- ▶ Principal Cell
- ▶ PRINTS
- ▶ Prion
- ▶ Prion Diseases
- ▶ **Prion Diseases**
HANS A. KRETZSCHMAR
- ▶ Pristanic Acid
- ▶ PRL
- ▶ Pro
- ▶ Proband
- ▶ Processed Pseudogene
- ▶ Processivity
- ▶ ProcheckNMR
- ▶ ProDom
- ▶ Profilin
- ▶ Progenitor Cell
- ▶ Progesterone Receptor
- ▶ Prognosis
- ▶ Programmed +1/−1 Frameshifting
- ▶ Programmed Cell Death
- ▶ Programmed Translational Frameshifting
- ▶ Progressive External Ophthalmoplegia
- ▶ Progressive Supranuclear Palsy
- ▶ Projections
- ▶ Prolactin
- ▶ Proline Isomerase
- ▶ Prolyl Isomerase
- ▶ Prometaphase
- ▶ Promoter
- ▶ Promoter Clearance

- ▶ Proneural Transcription Factors
- ▶ Pronuclear (Micro)Injection
- ▶ Proof Reading Mechanism
- ▶ Proofreading Exonuclease
- ▶ Prophase
- ▶ Prophylactic Immunization
- ▶ Prophylaxis
- ▶ PROSITE
- ▶ Prostaglandins
- ▶ Prosthetic groups
- ▶ Protease
- ▶ Protease Inhibitors
- ▶ **Proteases and Inhibitors**
DORA CAVALLO-MEDVED,
BONNIE F. SLOANE
- ▶ Proteasome
- ▶ Protein
- ▶ Protein A
- ▶ **Protein and Membrane Transport in Eukaryotic Cells**
MICHAEL KRAUSS, VOLKER HAUCKE
- ▶ Protein Array
- ▶ Protein Biosynthesis
- ▶ Protein Chemical Modification
- ▶ Protein Chips
- ▶ Protein Complex
- ▶ Protein Conformational Diseases
- ▶ Protein Conjugation
- ▶ **Protein Crystallization for X-Ray**
JOCHEN MUELLER-DIECKMANN,
MANFRED S. WEISS, MATTHIAS WILMANN
- ▶ Protein Crystallography
- ▶ Protein Data Bank
- ▶ **Protein Databases**
MIN-SEOK KWON, SANG YUN CHO,
YOUNG-KI PAK
- ▶ Protein Degradation
- ▶ Protein Disulfide
- ▶ **Protein Disulfide Bonds**
ERVIN WELKER, MAHESH NARAYAN,
HAROLD A SCHERAGA
- ▶ **Protein Domains**
JÖRG SCHULTZ
- ▶ Protein Expression Profiling
- ▶ Protein Family
- ▶ Protein Fold
- ▶ **Protein Folding**
MATTHIAS P. MAYER
- ▶ Protein G
- ▶ Protein Homology
- ▶ **Protein Interaction Analysis, Variations of the Yeast Two-Hybrid System**
IGOR STAGLJAR
- ▶ **Protein Interaction Analysis: Chemical Cross-Linking**
OWEN W. NADEAU
- ▶ **Protein Interaction Analysis: Phage Display**
ZOLTÁN KONTHUR, JÖRN GLÖKLER,
KARL SKRINER
- ▶ Protein Interactions
- ▶ Protein Kinase
- ▶ **Protein Ligand Interactions Studied by X-ray**
MILTON T. STUBBSII
- ▶ Protein Localization
- ▶ **Protein Microarrays as a Tool for Protein Profiling and Functional Proteomics**
ULRIKE KORF
- ▶ Protein Misfolding Disorders
- ▶ Protein Phosphatase–2A
- ▶ **Protein Prenylation**
MARK D. DISTEFANO, LEILA N. ALBERS,
JU-HUA XU
- ▶ Protein Research Foundation
- ▶ Protein Self-Association
- ▶ Protein Species
- ▶ **Protein Structure Prediction**
VOLKHARD HELMS
- ▶ Protein Synthesis Control
- ▶ **Protein Tags**
KAY TERPE
- ▶ Protein Transduction Domains
- ▶ **Protein/DNA Interaction**
PATRICE MORAND, CHRISTOPH W. MÜLLER
- ▶ **Protein-Protein Analysis, Suppressor Hunting**
RANDY STRICH
- ▶ Proteinase 3
- ▶ ProteinChip Array

- ▶ ProteinChip Reader
- ▶ ProteinChip Technology
- ▶ **Protein-Ligand Interaction Studied by NMR**
CHRISTIAN FREUND
- ▶ Proteinogenic Amino Acids
- ▶ **Protein-Protein Interactions**
PETER UETZ, CAROLINA S. VOLLERT
- ▶ Protein-Tyrosine Sulfotransferase
- ▶ Proteoglycan
- ▶ Proteolytic Degradation
- ▶ Proteome
- ▶ Proteome Profiling
- ▶ Proteomic Signatures
- ▶ Proteomics
- ▶ **Proteomics in Ageing**
OLIVIER TOUSSAINT, FLORENCE CHAINIAUX,
ALINE CHRÉTIEN, VÉRONIQUE FLAMENT,
EDOUARD DELAIVE, JEAN-FRANÇOIS DIERICK
- ▶ **Proteomics in Cancer**
FRANÇOIS LE NAOUR
- ▶ **Proteomics in Cardiovascular Disease**
RODNEY LUI, SEAN LAL,
CRISTOBAL G. DOS REMEDIOS
- ▶ **Proteomics in Human-Pathogen Interactions**
HENGLI TANG
- ▶ **Proteomics in Microfluidic Systems**
JAISREE MOORTHY, DAVID J. BEEBE
- ▶ Proteopathies
- ▶ Prothrombin
- ▶ Protocol Mutagenesis
- ▶ Protonation States
- ▶ Proto-Oncogene
- ▶ Protoplast Transformation
- ▶ Protrusion
- ▶ Provirus
- ▶ Proximal Spinal Muscular Atrophy
- ▶ PRP
- ▶ PrP^C
- ▶ PrP^{Sc}
- ▶ Prx
- ▶ PS
- ▶ PS1/PS2
- ▶ PSD95
- ▶ PSEN1/PSEN2
- ▶ Pseudoachondroplasia
- ▶ Pseudoautosomal Region
- ▶ Pseudogene
- ▶ Pseudohypoaldosteronisms
- ▶ Pseudohypoparathyroidism
- ▶ Pseudoknot
- ▶ Pseudomonomolecular Process
- ▶ Pseudopregnant Mouse
- ▶ Pseudotyped Virus
- ▶ PSF
- ▶ PSI
- ▶ PSI
- ▶ PSI-BLAST
- ▶ **Psoriasis, Molecular Basis**
KHUSRU ASADULLAH, WOLFRAM STERRY
- ▶ PSP
- ▶ PTC
- ▶ PTDs
- ▶ PTH
- ▶ PTHrP
- ▶ PTM
- ▶ Ptoxis
- ▶ PTS
- ▶ P-Type ATPase
- ▶ Pulse Sequence
- ▶ Pulsed-Field Gel Electrophoresis
- ▶ Purifying Selection
- ▶ Purine Nucleotide Biosynthesis
- ▶ Purines
- ▶ pVHL
- ▶ PWS
- ▶ Pyg
- ▶ Pygopus
- ▶ PYP
- ▶ Pyrimidine Nucleotide Biosynthesis
- ▶ Pyrimidines
- ▶ Pyruvate Kinase
- ▶ Q10
- ▶ QC
- ▶ QSAR
- ▶ Q-T Interval
- ▶ Q-TOF
- ▶ Quadrupole
- ▶ Quadrupole Mass Analyzer
- ▶ Quadrupole Time of Flight Mass Spectrometer

- ▶ Quality Control
- ▶ **Quantitative Structure-Activity Relationship**
MICHAEL WIESE
- ▶ Quantum Yield
- ▶ Quasispecies
- ▶ Quaternary Structure
- ▶ Quiescence
- ▶ RA
- ▶ Rac
- ▶ RACE
- ▶ RAD51
- ▶ Radiation Hybrid Map
- ▶ Radiation Sensitivity Gene 51
- ▶ Radioactive Isotope
- ▶ Radioactivity/Radioactive Decay
- ▶ Radixin
- ▶ Rafts
- ▶ Ragged Red Fibres
- ▶ Ran
- ▶ Ran GTPase Activating Protein
- ▶ Ran Guanine Nucleotide Exchange Factor
- ▶ Random Amplification of Polymorphic DNA Markers
- ▶ RanGAP
- ▶ RanGEF
- ▶ Ran-GTP
- ▶ RANKL
- ▶ Rapamycin (Sirolimus)
- ▶ Rapid-Eye-Movement Sleep
- ▶ RAPDs
- ▶ Rare Mutation
- ▶ Ras
- ▶ Ras Guanyl Nucleotide Exchange Factor
- ▶ Ras Proteins
- ▶ **Ras Signalling**
REINHOLD SCHÄFER
- ▶ Ras Superfamily of Small GTP Binding Proteins
- ▶ Ras-Like GTPases
- ▶ RasMol
- ▶ Rb
- ▶ RB1
- ▶ RBS
- ▶ RCA
- ▶ RCDP
- ▶ rDNA
- ▶ Reactive Oxygen Species
- ▶ Reading Frame
- ▶ Readlength
- ▶ Readthrough Suppression
- ▶ Real-time PCR
- ▶ Real-Time Polymerase Chain Reaction
- ▶ Rearrangement of Immunoglobulin Genes
- ▶ RecA
- ▶ Receptor for Egg Jelly Protein
- ▶ **Receptor Serine/Threonine Kinases**
ARISTIDIS MOUSTAKAS,
SERHIY SOUCHELNYTSKYI,
CARL-HENRIK HELDIN
- ▶ Receptor Tyrosine Kinases
- ▶ Receptors
- ▶ Recessive Allele
- ▶ Recessive Mutation
- ▶ Recklingshausen Disease
- ▶ Recoding
- ▶ Recombinant Protein
- ▶ **Recombinant Protein Expression in Bacteria**
CLAUDIA LANGLAIS, BERNHARD KORN
- ▶ **Recombinant Protein Production in Mammalian Cell Culture**
ANJA DRÖGE
- ▶ **Recombinant Protein Production in Yeast**
ROLAND WEIS, FRANZ HARTNER,
ANTON GLIEDER
- ▶ Recombination
- ▶ RecQ Helicase Family
- ▶ Recurrence Risk
- ▶ Redefinition of Codons
- ▶ Reductive Unfolding
- ▶ Redundancy
- ▶ REE, RxRE
- ▶ Reference State
- ▶ RefSeq
- ▶ Refsum Syndrome
- ▶ Region of Interest
- ▶ Regulators of G Protein Signaling
- ▶ REJ
- ▶ Relaxation Techniques/Relaxation
- ▶ Release Factors

- ▶ Release Probability
- ▶ REM (Rapid-Eye-Movement) Sleep
- ▶ REM Sleep
- ▶ Rendering
- ▶ Repeats
- ▶ Repeat Expansion Diseases
- ▶ **Repeat Expansion Diseases, Dynamic Mutations Cause (Neurological) Model Disorders**

JÖRG T. EPPLER

- ▶ Repetitive DNA
- ▶ Repetitive Elements (Repetitive DNA, Repeats)
- ▶ Replication
- ▶ Replication Bubble
- ▶ Replication Foci
- ▶ **Replication Fork**

ULRICH HÜBSCHER, IGOR SHEVELEV

- ▶ **Replication Origins**

ARTURO FALASCHI, GIUSEPPE BIAMONTI,
SILVANO RIVA

- ▶ Replicative Senescence
- ▶ Replicator
- ▶ Replicon
- ▶ Replisome
- ▶ Repolarization
- ▶ Reporter Genes
- ▶ Repressor
- ▶ Repressor Element–1 Transcription Factor
- ▶ Residue Interface Propensity
- ▶ Resolution
- ▶ Respiratory Burst
- ▶ Respiratory Chain
- ▶ Response Elements
- ▶ REST
- ▶ Restriction Enzymes
- ▶ Restriction Fragment Length Polymorphism
- ▶ Restriction Mapping
- ▶ Restriction Point
- ▶ Retinal Pigment Epithelium
- ▶ Retinal Protein
- ▶ **Retinitis Pigmentosa**

EBERHART ZRENNER

- ▶ Retinoblastoma
- ▶ Retinoblastoma (Rb) Protein

- ▶ Retinoblastoma Gene
- ▶ Retinoic Acid
- ▶ Retinoid Cycle
- ▶ Retinopathy
- ▶ Retroposon
- ▶ Retroseudogene
- ▶ Retrotransposition
- ▶ Retrotransposon
- ▶ **Retroviruses**

JOHN M. COFFIN

- ▶ **Rett Syndrome**

HAYLEY ARCHER, ANGUS J. CLARKE

- ▶ Rev, Rex
- ▶ Rev-Erb
- ▶ S Phase
- ▶ S-Acylation
- ▶ SAD
- ▶ SAGA
- ▶ SAGE
- ▶ Saltatory Nerve Conduction
- ▶ Salvage Pathway
- ▶ SAMP Repeats
- ▶ Sampling
- ▶ Sanger Sequencing Method
- ▶ Sanglifehrin
- ▶ SAR1
- ▶ SARA
- ▶ SAR-by NMR
- ▶ Sarcolemma
- ▶ Sarcomas
- ▶ Sarcomere
- ▶ Sarcoplasmic Reticulum
- ▶ Satellite
- ▶ Satellite Cell
- ▶ Saturation Point
- ▶ SBMA
- ▶ SCA
- ▶ Scaffolding Protein
- ▶ Scanning Tunneling Microscope
- ▶ Scar
- ▶ Scar/Wave Proteins
- ▶ Scavenger Receptor(s)
- ▶ SCC
- ▶ SCE
- ▶ SCF Ubiquitin Ligase

- ▶ Schiff Base
- ▶ **Schizophrenia, Genetics**
NADINE NORTON, MICHAEL J OWEN
- ▶ Schwachman-Diamond Syndrome
- ▶ Schwann Cells
- ▶ Sclerotome
- ▶ Scoring Function
- ▶ Scrambled Disulfide Species
- ▶ Scrapie
- ▶ SCTAT
- ▶ SDAT
- ▶ SDH
- ▶ SDS
- ▶ SDS-PAGE
- ▶ Seborrheic Keratosis
- ▶ Second Site Suppression
- ▶ Secondary Hyperparathyroidism
- ▶ Secondary Hypertension
- ▶ Secondary Structure
- ▶ Second-Generation MDR Modulators
- ▶ Secreted Frizzled Related Proteins
- ▶ Secretion
- ▶ Secretory Pathway
- ▶ Secretory Vesicle
- ▶ Sedimentation Coefficient
- ▶ Sedimentation Equilibrium
- ▶ Sedimentation Velocity
- ▶ Segment Duplication
- ▶ Segmentation
- ▶ Segregation Analysis
- ▶ SELDI
- ▶ SELDI-ProteinChip Technology
- ▶ SELDI-Tof
- ▶ Selection, Selective Advantage
- ▶ Selection/Selectable Marker
- ▶ Selectivity
- ▶ Selenocysteine
- ▶ Selenoenzymes
- ▶ **Selenoproteins**
DIETRICH BEHNE, ANTONIOS KYRIAKOPOULOS
- ▶ Self-Assembled Monolayer
- ▶ Self-Renewal
- ▶ Semidominant Allele
- ▶ Senescence
- ▶ Senile Dementia/Alzheimer's Type
- ▶ Senile Plaques
- ▶ Sensitivity
- ▶ Sensory Ganglia
- ▶ Septum Pellucidum
- ▶ Sequence Alignment
- ▶ **Sequence Annotation in Evolution**
STEFFEN HENNIG
- ▶ **Sequence Comparison**
MARTIN VINGRON
- ▶ Sequence Databases
- ▶ Sequence Polymorphism
- ▶ Sequence-Specific Post-Transcriptional Gene Silencing by Double-Stranded RNA
- ▶ Sequencing
- ▶ Ser
- ▶ Serial Analysis of Gene Expression
- ▶ Seropositive
- ▶ Sertoli Cells
- ▶ SET
- ▶ SET Domain
- ▶ Seven-Transmembrane Receptors
- ▶ Sex Chromosome Silencing
- ▶ Sex Reversal
- ▶ Sex-Determining Gene
- ▶ Sex-Determining Region Y
- ▶ SH2/SH3 Domains
- ▶ Shaker/Usher Syndrome
- ▶ Shc
- ▶ SHH
- ▶ Shine-Dalgarno Site
- ▶ Shmoo
- ▶ Short Interfering RNA
- ▶ Short Interspersed Repeat
- ▶ **Shotgun Libraries**
MATTHEW C. JONES, KIRSTEN E. MCLAY
- ▶ Sialic Acid
- ▶ Sialidase
- ▶ Signal Recognition Particle
- ▶ Signal Transduction (Pathway)
- ▶ **Signal Transduction: Integrin-Mediated Pathways**
KERSTIN DANKER
- ▶ Signalling
- ▶ Signalling State
- ▶ Signature
- ▶ Silane Reagent
- ▶ Silencer

- ▶ Silent Substitution
- ▶ Silver-Russell Syndrome
- ▶ Simulated Annealing
- ▶ Simultaneous Slippage
- ▶ SINE
- ▶ Single Gene Disorder
- ▶ Single Nucleotide Mismatch
- ▶ Single Nucleotide Polymorphisms
- ▶ Single Particle Tracking
- ▶ Single Particles
- ▶ Single Photon Emission Computer Tomography
- ▶ **Single-Cell Gene Expression Profiling: Cell-Level Biology by Multiplexed Expression Fluorescence *in Situ* Hybridization**
JEFFREY M. LEVSKY,
ROBERT H. SINGER
- ▶ Single-Particle Reconstruction
- ▶ Single-Stranded Conformational Polymorphism
- ▶ Single-Stranded DNA Binding Protein
- ▶ Singlet State
- ▶ Single-Wavelength Anomalous Diffraction
- ▶ siRNA
- ▶ Short Interfering RNA
- ▶ Sister Chromatid
- ▶ Sister Chromatid Cohesion
- ▶ Sister Chromatid Exchange
- ▶ Sister Kinetochores
- ▶ Site-Specific Protease
- ▶ Site-Specific Recombination
- ▶ SIX5
- ▶ Size Exclusion Chromatography
- ▶ Skeletal Disorders
- ▶ Skeletonization
- ▶ Skewed X-Inactivation
- ▶ Skin Barrier
- ▶ Skin (Permeability) Barrier
- ▶ **Skin and Hair**
JOERG HUELSKEN
- ▶ SL1
- ▶ Slavotinek Syndrome
- ▶ **SLE Pathogenesis, Genetic Dissection**
CHUN XIE, CHANDRA MOHAN
- ▶ Sleep Apnea

- ▶ Sleep Paralysis
- ▶ Slippery Heptamer
- ▶ SLOS
- ▶ Slow and Fast Muscle Fibres
- ▶ Slow Exchange Limit
- ▶ Slow Muscle Fibres
- ▶ SLT
- ▶ SMA
- ▶ Smac
- ▶ Smad
- ▶ Smad Anchor for Receptor Activation
- ▶ Smad Ubiquitin Regulatory Factor
- ▶ Small Dendritic RNA
- ▶ Small GTPases
- ▶ Small Interfering RNA
- ▶ Small Interfering RNA-Mediated Gene Silencing
- ▶ Small Intragenic Mutations
- ▶ Small Molecule Inhibitors
- ▶ Small Vessel Vasculitis
- ▶ SMART Database
- ▶ SMase
- ▶ SMC Proteins
- ▶ Smith-Lemli-Opitz Syndrome
- ▶ Smith-Magenis Syndrome
- ▶ SMN-Complex
- ▶ Smoothened
- ▶ SMS
- ▶ Smurf
- ▶ SNARE Proteins
- ▶ SNP
- ▶ **SNP Detection and Mass Spectrometry**
SASCHA SAUER
- ▶ SOC
- ▶ SOD
- ▶ Solid- or Split-Pin
- ▶ Solubilization
- ▶ Solvation Potential
- ▶ Soma
- ▶ Somatic Hypermutation
- ▶ Somatic Mosaicism
- ▶ Somatic Mutation
- ▶ Somite
- ▶ **Somitogenesis**
PHILIPPE DAUBAS,
MARGARET E. BUCKINGHAM

- ▶ SOP
- ▶ Sorting Signals
- ▶ SOS
- ▶ Southern Blot
- ▶ Spastic Paralysis
- ▶ Specific Locus Test
- ▶ Specific Rotation $[\alpha]^{25^\circ}\text{D}$
- ▶ Specificity
- ▶ SPECT
- ▶ Spectral Imaging
- ▶ Sperm
- ▶ Spermatogenesis
- ▶ Spermatogonia
- ▶ Sphingomyelinase
- ▶ **Spinal Muscular Atrophy**
SHUJI OGINO
- ▶ Spindle Check-Point
- ▶ Spindle Microtubules
- ▶ Spines
- ▶ Spinobulbar Muscular Atrophy
- ▶ Spinocerebellar Ataxia
- ▶ Spinocerebellar Degeneration
- ▶ Splanchnic (Visceral)
- ▶ Splice (Junction) Mutation
- ▶ Splice Acceptor
- ▶ Splice Enhancer
- ▶ 3' Splice Site
- ▶ 5' Splice Site
- ▶ Splice Mutation
- ▶ Splice Site Mutation
- ▶ Spliceosome
- ▶ **Splicing**
CHIN LI, WOAN-YUH TARN
- ▶ Split-Hand-Foot
- ▶ Spo11
- ▶ Sporadic (Cases of Disease)
- ▶ Sporadic Mutation
- ▶ Spot Quantitation Matrix
- ▶ SPR
- ▶ Spreading
- ▶ SP-RING
- ▶ Squamous Cell Carcinoma
- ▶ SR Proteins
- ▶ Src
- ▶ SRP
- ▶ Sry

- ▶ **SRY – Sex Reversal**
STEPHEN S. WACHTEL
- ▶ SSCP
- ▶ SSH
- ▶ Stable Isotope
- ▶ STAGA
- ▶ Standard Operating Procedure
- ▶ Standard Proteomics Technologies
- ▶ Staphylococcal Scalded Skin Syndrome
- ▶ State Function
- ▶ Statistical Thermodynamics
- ▶ STATs
- ▶ Steatocystoma Multiplex
- ▶ Steinert's Disease
- ▶ Stem Cell Plasticity or Stem Cell Transdifferentiation
- ▶ Stem Cell Therapy
- ▶ Stem Cell Transdifferentiation
- ▶ **Stem Cells: an Overview**
JÜRGEN HESCHELER, BERND K. FLEISCHMANN
- ▶ Stem Loop
- ▶ Stereocilia
- ▶ Sterile Mutations
- ▶ **Steroid Hormone Receptor Defects, Molecular Basis**
TOMOSHIGE KINO, GEORGE P. CHROUSOS
- ▶ Steroid Hormones
- ▶ Steroids
- ▶ Stickler Syndrome
- ▶ STM
- ▶ Stochastic
- ▶ Stokes Radius
- ▶ Stokes Shift
- ▶ Stop Codon
- ▶ Storage Polysaccharide
- ▶ Store-Operated Calcium Entry
- ▶ STR
- ▶ Strand Transfer
- ▶ Stratified Epithelium
- ▶ Stratum Basale
- ▶ Stratum Corneum
- ▶ Stratum Granulosum
- ▶ Stratum Spinosum
- ▶ Streptavidin
- ▶ Stress Fibres
- ▶ Stress Response

- ▶ Stress-Induced Premature Senescence
- ▶ Striae Atrophicae
- ▶ Stroma
- ▶ Structural Databases
- ▶ Structural Genomics
- ▶ **Structural Genomics: Structure-to-Function Approaches**
UDO HEINEMANN
- ▶ Structural Proteomics
- ▶ **Structure-based Drug Design**
GERHARD KLEBE
- ▶ **3D Structure Determination by NMR**
ZHEN-YU J. SUN, GERHARD WAGNER
- ▶ **3D Structure Determination by X-Ray**
RICHARD A. ENGH
- ▶ Strümpell Disease
- ▶ SU (Surface)
- ▶ SU (Surface) TAR
- ▶ Subcellular Compartment (or Organelle)
- ▶ Subcellular Localization
- ▶ Subclone
- ▶ Subependymal Giant Cell Astrocytoma
- ▶ Subependymal Nodule
- ▶ Subjective Day/Subjective Night
- ▶ Sub-Proteome
- ▶ Substantia Nigra
- ▶ Substrate
- ▶ Subtelomeric Repeats
- ▶ Subtractive Analysis
- ▶ Succinate Dehydrogenase
- ▶ Sumo
- ▶ **Sumoylation**
GUILLAUME BOSSIS, FRAUKE MELCHIOR
- ▶ Superantigens
- ▶ Supercoiled DNA
- ▶ Superoxide Dismutase
- ▶ Superresolution
- ▶ Supervised Analysis
- ▶ Supervised Learning
- ▶ Suppressive Subtractive Hybridization
- ▶ Suppressor Hunting
- ▶ **Surface Plasmon Resonance**
KARL-WILHELM KOCH
- ▶ Surfactant Proteins
- ▶ Susceptibility Gene
- ▶ Susceptibility Locus/Susceptibility Gene
- ▶ SV
- ▶ Swiss Institute of Bioinformatics
- ▶ Swiss-PdbViewer
- ▶ Symmetry Operations
- ▶ Synapse
- ▶ Synaptic Efficacy
- ▶ Synaptic Vesicle
- ▶ Synaptogenesis
- ▶ Synaptonemal Complex
- ▶ Synaptotagmin I
- ▶ Synchrotron
- ▶ Syndecans
- ▶ Syndetome
- ▶ Syndrome
- ▶ Synergism
- ▶ Synergistic Interaction
- ▶ Synonymous Codon Changes
- ▶ Synpolydactyly
- ▶ Synteny/Syntenic
- ▶ Syringomyelia
- ▶ Systemic Lupus Erythematosus
- ▶ Systems Biology
- ▶ Systole
- ▶ T Cell
- ▶ T Lymphocyte
- ▶ Tachycardia
- ▶ TAFs
- ▶ T-Ag
- ▶ Tagging / Tag
- ▶ TAK1
- ▶ Talin
- ▶ TAMs
- ▶ Tandem Mass Spectrometry
- ▶ Tandem Repeats
- ▶ **Tangier disease**
GERD SCHMITZ, WOLFGANG E. KAMINSKI
- ▶ TAP
- ▶ TAP Tag
- ▶ TAP/NXF1
- ▶ Target Gene
- ▶ Target Registration DataBase
- ▶ TargetDB
- ▶ Targeted Gene Disruption
- ▶ Targeting Vector
- ▶ Tat

- ▶ TATA Binding Protein
- ▶ TATA Box
- ▶ Tauopathies
- ▶ Tau-Protein
- ▶ Tax
- ▶ Taxanes
- ▶ Taxon (pl. Taxa)
- ▶ *TBCE* Gene
- ▶ TbetaR-I/TbetaR-II (TβR-I/TβR-II)
- ▶ TBP
- ▶ TBP-Associated Factors
- ▶ T-Cell Epitope
- ▶ T-Cell Factors
- ▶ TCF
- ▶ TCFs
- ▶ Telocentric Chromosome
- ▶ **Telomerase**
KENKICHI MASUTOMI, WILLIAM C. HAHN
- ▶ Telomerase Reverse Transcriptase
- ▶ Telomere
- ▶ Telophase
- ▶ Temperature Compensation
- ▶ Temperature Factor
- ▶ Temperature-Sensitive Mutant
- ▶ Template Matching
- ▶ TER
- ▶ Teratogens
- ▶ Terminal Ductal-Lobular Unit
- ▶ Terminator
- ▶ TERT
- ▶ Tertiary Structure
- ▶ Testis-Determining Gene
- ▶ Testosterone
- ▶ Tetany
- ▶ Tetraspanins
- ▶ Tetratricopeptide Repeat Motif
- ▶ TFIID
- ▶ TFIIH
- ▶ TGF
- ▶ TGFβ Activated Kinase–1
- ▶ Th–1 Biased T Cell Response
- ▶ TH1/Th2 Cells
- ▶ Thalassaemia
- ▶ Thanatophoric Dysplasia
- ▶ Therapeutic Immunization
- ▶ Thermodynamic Integration

- ▶ **Thermodynamic Properties of DNA**
G. ERIC PLUM, JENS VÖLKER,
KENNETH J. BRESLAUER
- ▶ Thermodynamic State
- ▶ Thermogram
- ▶ THF
- ▶ Thioester-Linkage
- ▶ Thioether
- ▶ Thioflavin T
- ▶ Thiolate
- ▶ **Thiol-dependent Peroxidases**
LEOPOLD FLOHÉ, TIMO JAEGER
- ▶ Thioredoxin
- ▶ Third-Generation MDR Modulators
- ▶ Threshold Liability Model of Multifactorial Inheritance
- ▶ Thromboembolism
- ▶ Thrombosis
- ▶ Thymidine Analog Mutations
- ▶ Thymidine Kinase 2
- ▶ Thymidine Phosphorylase
- ▶ Thyroid Cancer
- ▶ **Thyroid Disorders, Genetic Basis**
PETER KOPP
- ▶ Thyroid Dysgenesis
- ▶ Thyroid Stimulating Hormone
- ▶ TI
- ▶ **Tight Junctions**
SAIMA AIJAZ, STEVEN GOODRICK,
KARL MATTER, MARIA S. BALDA
- ▶ TIGRFAMs
- ▶ TIM22/23 Complexes
- ▶ Time-of-Flight Mass Spectrometer
- ▶ TIR Domain
- ▶ Tissue Micro Array
- ▶ Tissue-Specific Genes
- ▶ TK2
- ▶ TLR
- ▶ TM (Transmembrane)
- ▶ TM (Transmembrane)
Transactivating Proteins
- ▶ TMA
- ▶ TNF
- ▶ **TNF Receptor/ Fas Signaling Pathways**
HARALD WAJANT

► TNF-Related Apoptosis Inducing
Ligand

► TNF- α

► TOB Complex

► TOF

► Tolerance

► Toll/IL-1 Receptor Domain

► Toll-Like Receptor

► TOM Complex

► Tomography

► Tonofibrils

► Topoisomerase II

► TP

► TPR Motif

► Tracer

► Traditional Silicon-Based Protocols

► TRAF

► TRAIL

► Trait

► Trajectory/Trajectorial

► Trans-Acting Factor

► Transactivation Function

► Transacylation

► Transcript

► Transcript Elongation

► Transcription

► Transcription Coupled Repair

► **Transcription Elongation**

YUKI YAMAGUCHI, TADASHI WADA,
HIROSHI HANDA

► Transcription Elongation Complex

► Transcription Elongation Factor

► Transcription Enhancer

► Transcription Factor

► **Transcription Factors and Regulation of
Gene Expression**

RAINER RENKAWITZ

► Transcription Initiation

► Transcription Unit

► Transcriptional Activator

► Transcriptional Elongation

► Transcriptional Enhancer

► Transcriptional Machinery

► Transcriptional Profiling

► Transcriptional Regulation

► **Transcriptional Repression**

GERALD THIEL, MICHAEL LIETZ,
MATTHIAS HOHL

► Transcriptional Repressor

► Transcriptional Termination

► Transcriptome

► Transcytosis

► Transdifferentiation

► Transdominant Genetic Effect

► Transduction

► Transepidermal Water Loss

► Trans-Epithelial Electrical Resistance

► Transesterification

► Transfected Cell Array

► Transfection

► Transfer RNA

► Transferrin

► Transferrin Receptor

► Transformation

► Transforming Growth Factor- β

► Transforming Protein

► Transgene

► Transgene Expression

► **Transgene Silencing**

ERIC MILOT, JAMES ELLIS

► Transgenesis

► Transgenic

► **Transgenic and Knock-Out Animals**

STEFAN BRITSCH

► Trans-Golgi Network

► Transition

► Transition State

► Translation

► Translation Elongation Factors

► Translation Initiation Factor

► Translation Start Codon

► Translational Bypassing

► **Translational Control in Eukaryotes**

MARTINA MUCKENTHALER,
THOMAS PREISS

► **Translational Frameshifting, Non-Standard
Reading of the Genetic Code**

PHILIP FARABAUGH

► Translational Inhibition

► Translational Pause

- ▶ Translationally Silent Mutation
- ▶ Translesion DNA Synthesis
- ▶ Translocation
- ▶ Transmembrane Proteins
- ▶ Transmissible Spongiform Encephalopathies
- ▶ Transmitters
- ▶ Transposable Element
- ▶ Transposition
- ▶ **Transposons**
ZOLTÁN IVICS, OLIVER WALISKO,
ZSUZSANNA IZSVÁK
- ▶ Trans-Splicing
- ▶ Transvection
- ▶ Transversion
- ▶ TRE
- ▶ Treadmilling
- ▶ Trichothiodystrophy
- ▶ Trinucleotide Repeat
- ▶ Trinucleotide Repeat Expansion Disease
- ▶ Triple Resonance NMR Experiment
- ▶ Trisomy
- ▶ **tRNA**
ESTHER LIZANO, MARIO MÖRL
- ▶ TRP Channel
- ▶ True Hermaphroditism
- ▶ Tryparedoxin
- ▶ TSA
- ▶ TSE
- ▶ TSG
- ▶ TSH
- ▶ t-SNARE
- ▶ TTD
- ▶ T-Tubulus
- ▶ **Tuberous Sclerosis**
DAVID J. KWIATKOWSKI
- ▶ Tubulin
- ▶ Tumor Angiogenesis
- ▶ Tumor Necrosis Factor
- ▶ Tumor Necrosis Factor Receptor-Associated Factor
- ▶ Tumor Necrosis Factor- α
- ▶ Tumor Promoter Genes
- ▶ Tumor Suppressor Gene
- ▶ **Tumor Suppressor Genes**
SUSAN WIJNHOFEN, ANNEMIEKE DE VRIES
- ▶ Tumor Viruses
- ▶ Tumorigenesis
- ▶ Tunicamycin
- ▶ Two Channel Microarrays
- ▶ Two-Dimensional Crystal
- ▶ **Two-Dimensional Crystallization of Membrane Proteins**
THOMAS BRAUN, THOMAS C. KAUFMANN,
HERVÉ-W. RÉMIGY, ANDREAS ENGEL
- ▶ **Two-Dimensional Gel Electrophoresis**
PETER R. JUNGBLUT
- ▶ Two-Dimensional Polyacrylamide Gel Electrophoresis
- ▶ Two-Hit Model
- ▶ **Two-Hybrid System**
CAROLINA S. VOLLERT, PETER UETZ
- ▶ Two-Photon Excitation
- ▶ Two-Site Immunoradiometric Assay
- ▶ Two-State Process
- ▶ Ty Retrotransposon
- ▶ Type B Reactions
- ▶ Type 1 Diabetes Mellitus
- ▶ Type 1 Pseudohypoaldosteronism
- ▶ Type 1A Topoisomerases
- ▶ Type 1B Topoisomerases
- ▶ Type 2 Diabetes Mellitus
- ▶ Type 2 Pseudohypoaldosteronism
- ▶ Type I Transmembrane Glycoprotein
- ▶ Type II Myosin
- ▶ Type IIA Topoisomerases
- ▶ Type III Hyperlipidemia
- ▶ Tyrosinase
- ▶ **Tyrosine Kinases**
UTE SCHAEFER, KATJA S. GROSSMANN
- ▶ **Tyrosine Sulfation of Proteins**
KEVIN L. MOORE
- ▶ Tyrosine-Kinase-Associated Receptors
- ▶ Tyrosylprotein Sulfotransferase
- ▶ UAP56
- ▶ UAS-Gal4 System
- ▶ UBF
- ▶ Ubiquitin
- ▶ Ubiquitin Chains
- ▶ Ubiquitin Proteasome Pathway
- ▶ **Ubiquitination**
THOMAS SOMMER, GUNNAR DITTMAR
- ▶ Unc

- ▶ Uncaging
- ▶ **Uncaging and Photoconversion/Activation**
VOLKER HAGEN
- ▶ Unconventional Myosins
- ▶ Unigene Library
- ▶ Unit Cell
- ▶ Uniparental Disomy
- ▶ Univalent
- ▶ Unsupervised Analysis
- ▶ Unsupervised Learning
- ▶ Untranslated
- ▶ 5' Untranslated Region
- ▶ UPD
- ▶ Upper Motor Neurons
- ▶ Upper to Lower Segment Ratio
- ▶ Upstream
- ▶ Uronic Acid
- ▶ UTR
- ▶ UV Signature Mutations
- ▶ V(D)J Recombination
- ▶ V2R
- ▶ Vaccine
- ▶ Validation of Therapeutic Targets
- ▶ Valve
- ▶ Van der Waals Interaction
- ▶ van't Hoff Equation
- ▶ Variegate Porphyria
- ▶ Vascular Morphogenesis
- ▶ Vasculitis
- ▶ Vasculogenesis
- ▶ Vasoactive Peptides
- ▶ Vasopressin
- ▶ vCJD
- ▶ Vector (Particle)
- ▶ VEGF
- ▶ Ventricular Zone
- ▶ Very Low Density Lipoprotein
- ▶ Very-Long-Chain Fatty Acids
- ▶ **Vesicular Traffic**
RAINER DUDEN
- ▶ VH
- ▶ VHH
- ▶ VHL-(von Hippel-Lindau) Disease
- ▶ VH-VL Pair
- ▶ Villin
- ▶ Vinblastine/Vincristine
- ▶ Vinca Alkaloids
- ▶ Vincristine
- ▶ Vinculin
- ▶ Viral Fitness
- ▶ **Viral Oncogenesis**
RALPH GRASSMANN,
BERNHARD FLECKENSTEIN
- ▶ Virion
- ▶ **Virtual Screening**
RONALD KÜHNE
- ▶ Virulent / Virulence
- ▶ Viscoelasticity
- ▶ Vitiligo
- ▶ Vitreous Ice
- ▶ VL
- ▶ VLCFA
- ▶ VLDL
- ▶ Voltage-Clamp Mode
- ▶ von Hippel-Lindau Disease
- ▶ von Willebrand Disease
- ▶ von Willebrand Factor
- ▶ v-SNARE
- ▶ vWF
- ▶ Walker A and B Motif
- ▶ WASp
- ▶ Watson-Crick DNA
- ▶ Wavelength of Light
- ▶ Wegener's Granulomatosis
- ▶ Weighted Back-Projection
- ▶ Werner's Syndrome
- ▶ Western Blotting
- ▶ White Sponge Naevus
- ▶ **Whole Mount ISH**
BERNHARD G. HERRMANN,
LORENZ NEIDHARDT
- ▶ Wide-Field Fluorescence Microscopy
- ▶ Wild Type
- ▶ William-Beuren Syndrome
- ▶ Winged Helix
- ▶ **Winged Helix Transcription Factors**
PATRICIA A. LABOSKY,
KLAUS H. KAESTNER
- ▶ WISH
- ▶ Wiskott Aldrich Syndrome Protein
- ▶ Withdrawal
- ▶ WNK

▶ **Wnt/ β -Catenin Signaling Pathway**

AJAMETE KAYKAS, RANDALL T. MOON

- ▶ Wnt / Wg
- ▶ Wolffian Duct
- ▶ Wolf-Hirschhorn Syndrome

▶ **Wound Healing**

SUSANNE BRAUN, SABINE WERNER

- ▶ WSC Domain
- ▶ WT
- ▶ X Chromosome
- ▶ XALD

▶ Xanthelasmata

▶ Xanthoma

▶ Xanthophore

▶ **X-Chromosome Inactivation**

BRYAN K. SUN, JEANNIE T. LEE

- ▶ X-Chromosome Silencing
- ▶ Xenobiotic Substances
- ▶ XenoMouse
- ▶ *Xenopus* as a Model Organism for Functional Genomics: Rich History, Promising Future

MATTHEW W. GROW,

MICHAEL W. KLYMKOWSKY

- ▶ *Xenopus Laevis*
- ▶ Xeroderma Pigmentosum
- ▶ X-Linked Adrenoleukodystrophy
- ▶ X-Linked Inheritance
- ▶ XML
- ▶ XP
- ▶ X-Ray Radiation
- ▶ X-Ray Crystallography
- ▶ **X-Ray Crystallography, Basic Principles**

RICHARD A. ENGH

▶ **X-Ray Data Collection from Macromolecular Crystals**

ANDREA SCHMIDT, VICTOR S. LAMZIN

- ▶ XRN1 and XRN2
- ▶ XX Male Syndrome
- ▶ XY Gonadal Dysgenesis
- ▶ **YAC and PAC Maps**
- ▶ LEONARD C SCHALKWYK
- ▶ YAC Transgene
- ▶ YAC/PAC/BAC Transgene
- ▶ YACs
- ▶ YAG
- ▶ Y-Box
- ▶ Yeast Artificial Chromosomes
- ▶ Yeast Two-Hybrid Analyses
- ▶ Z-Disc
- ▶ Zebrafish
- ▶ Zellweger Syndrome
- ▶ Zero Frame
- ▶ Zero-Length Cross-Linkers
- ▶ Zinc Finger Domain
- ▶ ZIP
- ▶ ZNF202
- ▶ ZO
- ▶ Zona Occludens
- ▶ Zona Pellucida
- ▶ Zona Reaction
- ▶ Zone of Polarizing Activity
- ▶ Zonula Adherens
- ▶ ZPA
- ▶ Zygote
- ▶ Zygotene
- ▶ Zygotic Transcript