

Springer Series
in Biophysics 10

José Luis R. Arrondo
Alicia Alonso (Eds.)

Advanced Techniques in Biophysics

 Springer

Springer Series in Biophysics 10

José Luis R. Arrondo Alicia Alonso (Eds.)

Advanced Techniques in Biophysics

With 93 Figures, 17 in Color and 2 Tables



Springer

Professor José Luis R. Arrondo

Professor Alicia Alonso

Unidad de Biofísica (Centro Mixto CSIC-UPV)
Departamento de Bioquímica y Biología Molecular
P.O. Box 644
48080 Bilbao
Spain

ISSN 0932-2353

ISBN 10 3-540-30700-1 Springer Berlin Heidelberg NewYork

ISBN 13 978-3-540-30700-6 Springer Berlin Heidelberg NewYork

Library of Congress Control Number: 2005938927

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

Springer is a part of Springer Science+Business Media

springeronline.com

© Springer-Verlag Berlin Heidelberg 2006

Printed in Germany

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

Editor: Dr. Sabine Schreck, Heidelberg

Desk Editor: Anette Lindqvist, Heidelberg

Production: LE-TeX Jelonek, Schmidt & Vöckler GbR, Leipzig

Typesetting: Satz-Druck-Service, Leimen

Cover Design: Design & Production, Heidelberg

Printed on acid-free paper

39/3100/YL

5 4 3 2 1 0

Preface

The present book of the Springer Series in Biophysics deals with some techniques that are being implemented nowadays. One of the motors that have driven the biosciences, like daily life, has been the technological boost produced by the advancement of microprocessor technology. A whole array of possibilities have been opened to develop the classical techniques that were used some years ago.

Abrahams and coworkers contribute with a chapter on protein nanocrystallography which deals with obtaining protein crystals in small, confined volumes, trying to overcome one of the setbacks in crystallography, the amount of material needed to obtain good samples for diffraction. This chapter is followed by one by Ibarra-Molero and Sanchez-Ruiz reviewing the recent advances of differential scanning calorimetry in the field of protein energetics and also in the energetic analysis of other biological systems. The following two chapters look at recent advances of IR spectroscopy. IR reflection-absorption spectroscopy (IRRAS) looks at the air-water interface of membranes and in the chapter by Mendelsohn and coworkers the general basis as well as the application to lipids and peptides or proteins are reviewed. Arrondo and coworkers address the analysis of IR spectra by a new approach called two-dimensional generalized spectroscopy, where information on protein changes after a perturbation is analysed by synchronous or asynchronous maps. This approach, essentially different from that of 2D-NMR spectroscopy, uses correlation analysis of the dynamic fluctuations caused by an external perturbation to enhance spectral resolution.

Three chapters are devoted to different technical developments of NMR. Szypersky deals with the principles of ultrafast NMR spectroscopy through the use of G-matrix Fourier transform (GFT) NMR as a technique for rapid sampling of multidimensional NMR data. Freeman and Kup e approach the problem of fast multidimensional NMR by outlining two radical new approaches, one using spatially encoded single-scan multidimensional NMR and the other using projection-reconstruction of multidimensional spectra. Size is one of the problems that NMR has to face in the study of proteins, Fernández and Wider analyse the use of transverse relaxation-optimized spectroscopy (TROSY) in combination with isotope-labelling techniques to extend applications of NMR spectroscopy in solution to much larger molecules, such as integral membrane proteins in detergent micelles, large proteins in monomeric form and in macromolecular complexes, and intermolecular interactions in large complexes.

Carrión-Vázquez and coworkers have addressed protein nanomechanics, a new multidisciplinary area of research to directly measure mechanical forces in single molecules, by applying atomic force microscopy (AFM). Large unilamellar

vesicles are the subject of the chapter by Bagatolli, who reviews the use of two-photon fluorescence microscopy in studying the lateral structure of compositionally simple vesicles and more complicated membranes. San Martín and Valle look at the three-dimensional organization and structural features of macromolecular assemblies, knowledge of which is indispensable for understanding their functions, by using three-dimensional electron microscopy.

This book constitutes a privileged observatory for reviewing novel applications of biophysical techniques that can help the reader utilize the efforts of the scientists contributing to the volume to enter an area where the technology is progressing quickly and where a comprehensive explanation is not always to be found.

Bilbao, June 2006

José Luis R. Arrondo
Alicia Alonso

Contents

CHAPTER 1

Protein Nanocrystallization

DILYANA GEORGIEVA, JAN PIETER ABRAHAMS, MAXIM E. KUIL

1.1	Introduction	1
1.2	Nucleation and Crystallization in Nanovolumes	2
1.3	Creating and Dispensing Small Liquid Volumes	5
1.3.1	Inkjet Technology	5
1.3.2	Acoustic Dispensing	7
1.3.3	Fast Solenoid Valve Technology	8
1.3.4	Pin-Transfer Technology	9
1.3.5	Comparison of Liquid Dispensing Methods	9
1.4	Droplet Evaporation	11
1.4.1	Evaporation of a Binary Mixture of Pure Liquids	12
1.4.2	Evaporation of a Solvent with a Solute	12
1.4.3	Evaporation from Microcompartments	12
1.4.4	Practical Approaches to Reduce Evaporation	12
1.5	Liquid Mixing in Small Volumes and Microfluidics	13
1.6	Design and Making of Nanostructures	14
1.6.1	Nanoarrays	14
1.6.2	Microfluidic Systems	15
1.7	Robotics	16
1.8	Crystal Recognition	18
1.9	Outlook	21
	References	22

CHAPTER 2

Differential Scanning Calorimetry of Proteins: an Overview and Some Recent Developments

BEATRIZ IBARRA-MOLERO, JOSE M. SANCHEZ-RUIZ

2.1	Introduction	27
2.2	What Is DSC?	28
2.3	What Is Heat Capacity?	29
2.4	Equilibrium Thermodynamics Analysis Versus Kinetic Analysis	30
2.5	The Calorimetric Criterion for Two-State Behaviour	34
2.6	Two-State Versus Continuous (Barrierless, Downhill) Transitions	36
2.7	Characterization of Ligand-Binding Effects	39

2.8	Absolute Heat Capacities and the Residual Structure in the Unfolded State	42
2.9	Calorimetrically Determined Denaturant m Values	44
	References	45

CHAPTER 3

**IR Reflectance–Absorbance Studies of Peptide Structure, Orientation,
and Conformational Flexibility in Langmuir Films:
Relevance for Models of Pulmonary Surfactant Action**

CAROL R. FLACH, PENG CAI, RICHARD MENDELSON

3.1	Introduction	49
3.2	Instrumentation	50
3.3	Information from IRRAS Measurements	50
3.4	Pulmonary Surfactant	57
3.4.1	Biochemistry and Models for Function	57
3.4.2	Application of IRRAS to the Pulmonary Surfactant	60
3.4.2.1	Squeeze-Out	60
3.4.2.2	Structure, Orientation, and Possible Mechanism of Action of Therapeutic Agents for RDS	61
3.5	Future Possibilities for IRRAS.	67
	References	69

CHAPTER 4

Two-Dimensional Infrared Correlation Spectroscopy

JOSÉ LUIS R. ARRONDO, IBON ILORO, MARCOS GARCIA-PACIOS,
FÉLIX M. GOÑI

4.1	Introduction	73
4.2	Description of the Technique	74
4.3	Spectral Simulations	76
4.3.1	Intensity Changes	76
4.3.2	Band Shifting	77
4.3.3	Bandwidth	79
4.4	D-IR Studies of Proteins	80
4.4.1	2D Maps of Homopolypeptides	80
4.4.2	Protein Denaturation	81
4.4.3	Protein Unfolding and Aggregation	84
4.5	Summary.	86
	References	87

CHAPTER 5

NMR Spectroscopy of Large Biological Macromolecules in Solution

CÉSAR FERNÁNDEZ, GERHARD WIDER

5.1	Introduction	89
5.2	Technical Background	90

5.2.1	The NMR Signal	90
5.2.2	NMR and Molecular Size	92
5.2.3	Isotope Labeling	93
5.2.4	Transverse Relaxation-Optimized Spectroscopy (TROSY)	93
5.2.4.1	The Foundations of TROSY	93
5.2.4.2	Field-Strength Dependence of TROSY for ^{15}N - ^1H Groups	95
5.2.4.3	Implementation of TROSY: 2D [^{15}N , ^1H]-TROSY	95
5.2.4.4	[^{13}C , ^1H]-TROSY	97
5.2.5	Cross-Correlated Relaxation-Induced Polarization Transfer for Studies of Very Large Structures	98
5.2.6	The Water Resonance	98
5.3	Isotope-Labeling Techniques	99
5.3.1	Uniform Isotope Labeling	99
5.3.2	Selective Isotope-Labeling Strategies	100
5.4	TROSY for NMR Studies of Large Biological Macromolecules	102
5.4.1	2D [^{15}N , ^1H]-TROSY	102
5.4.2	[^{13}C , ^1H] Correlation Experiments	104
5.4.3	TROSY for Resonance Assignments in Large Molecules	105
5.4.3.1	[^{15}N , ^1H]-TROSY for Sequential Assignment of Protein Backbone Resonances	105
5.4.3.2	[^{15}N , ^1H]-TROSY for Assignment of Protein Side-Chain Resonances	106
5.4.4	TROSY for Studies of Intermolecular Interactions and Drug Design	108
5.4.5	TROSY for Observation of Scalar Couplings Across Hydrogen Bonds	109
5.4.6	TROSY for Measurements of RDCs	109
5.4.7	TROSY for Studies of Dynamic Processes	111
5.4.8	TROSY in NOESY	111
5.4.9	Applications to Nucleic Acids	112
5.4.10	TROSY, CRIPT, and CRINEPT for Studies of Very Large Structures	112
5.5	Solution NMR Studies of Membrane Proteins	114
5.5.1	Resonance Assignments and Collection of Structural Constraints for Membrane Proteins	114
5.5.2	D Structure Determination	115
5.5.3	Studies of Intermolecular Interactions Between a Membrane Protein and Detergent Molecules in Micelles	116
5.5.4	Dynamic Properties of Membrane Proteins	118
5.6	Conclusion and Outlook	119
	References	120

CHAPTER 6

Emerging Techniques in Fast Multidimensional NMR

RAY FREEMAN, ERIKS KUPČE

6.1	Introduction	129
6.2	Spatially Encoded Single-Scan Multidimensional NMR	130
6.2.1	Evolution	130
6.2.2	Detection	132
6.2.3	Resolution	133
6.2.4	The F_2 Dimension	134
6.2.5	Signal-to-Noise Ratio	134
6.2.6	Application of Single-Scan Two-Dimensional Spectroscopy	136
6.3	Projection–Reconstruction of Multidimensional Spectra	136
6.3.1	Projection of NMR Spectra	138
6.3.2	Reconstruction from Projections	140
6.3.3	Application of Projection–Reconstruction NMR	142
6.3.4	Related Methods	142
6.4	Conclusions	144
	References	145

CHAPTER 7

Principles and Application of Projected Multidimensional NMR Spectroscopy – G-matrix Fourier Transform NMR

THOMAS SZYPERSKI

7.1	Introduction	147
7.2	Background	147
7.2.1	RD and Accordion NMR Spectroscopy	148
7.2.2	RD NMR Spectroscopy	148
7.2.2.1	Joint Sampling of Two Shift Evolution Periods	148
7.2.2.2	Time-Proportional Phase Incrementation in RD NMR	150
7.2.2.3	Editing Peak Pairs in RD NMR	151
7.2.2.4	Peak Pattern in RD NMR Spectra	151
7.2.2.5	Double RD NMR	151
7.2.2.6	Use of Heteronuclear Magnetization	151
7.2.2.7	Application of RD NMR in Structural Genomics	152
7.3	GFT NMR	152
7.3.1	Theory of GFT NMR	152
7.3.2	Sensitivity of GFT NMR	156
7.3.3	Precision of Measurements in GFT NMR	156
7.3.4	Survey: GFT NMR and Comparison with FT NMR	158
7.3.5	Application of GFT NMR	159
7.3.5.1	Protein Resonance Assignment	159
7.3.5.2	G-matrix Transformation for Measurement of Residual Dipolar Couplings	159
7.3.6	GFT projection-reconstruction	160
	References	160

CHAPTER 8

**Protein Nanomechanics – as Studied
by AFM Single-Molecule Force Spectroscopy**

MARIANO CARRIÓN-VÁZQUEZ, ANDRÉS F. OBERHAUSER, HÉCTOR DÍEZ,
RUBÉN HERVÁS, JAVIER OROZ, JESÚS FERNÁNDEZ, DAVID MARTÍNEZ-MARTÍN

8.1	Introduction	163
8.2	Biological Machines, Single-Molecule Manipulation and Protein Mechanics	164
8.2.1	Protein Machines	164
8.2.2	Range of Relevant Forces in Biology	166
8.2.3	Mechanical Proteins and Mechanical Nanomachines	167
8.2.4	Single-Molecule Techniques	170
8.2.5	Single-Molecule Manipulation Techniques	171
8.3	SMFS of Proteins: Physical Principles and Methodology	172
8.3.1	Mechanical Force Transduction by AFM	174
8.3.1.1	Cantilever and Tip Choices	174
8.3.1.2	Calibration of the AFM	175
8.3.1.3	Obtaining the Final Force–Distance Curve	177
8.3.2	Establishing a “Mechanical Circuit”: Protein Attachment and Functionalization	178
8.3.3	Mechanical Unfolding of Proteins	178
8.3.3.1	Length-Clamp (Standard) SMFS	179
8.3.3.2	Force-Clamp SMFS	206
8.3.3.3	Computer Simulations of Mechanical Unfolding	206
8.3.4	Mechanical Refolding of Proteins.	207
8.4	Intramolecular Interactions	208
8.4.1	Model System: I27 Module, an Immunoglobulin Domain from Titin	211
8.4.2	Other Protein Folds: from “Mechanical” and “Nonmechanical” Proteins	214
8.4.3	Supramolecular Mechanical Properties of Protein Complexes	216
8.4.4	Achievements	217
8.4.4.1	Single-Molecule Mechanics vs. Bulk Biochemical Assays	217
8.4.4.2	Detection of Rare Misfolding Events	218
8.4.4.3	Directional Aspects of the Force: Shear vs. Zipper Configurations and “Achilles Heels”	219
8.4.4.4	Molecular Determinants of the Mechanical Stability of Proteins	220
8.4.4.5	Muscle Elasticity, a Macroscopic Biological Property, Reduced to the Single-Molecule Level	222
8.5	Intermolecular Interactions: Findings and Limitations	223
8.5.1	Model System: Biotin–(Strept)avidin.	224
8.5.2	Other Protein–Biomolecule Pairs.	225
8.6	Limitations and Perspectives	226
8.6.1	Technical Limitations	226
8.6.2	Methodological Limitations	231

8.6.3	SMFS and Protein Mechanics in Vivo	232
8.7	Conclusions	233
	References	235

CHAPTER 9

Multiphoton-Excitation Fluorescence Microscopy and Membranes

LUIS A. BAGATOLLI

9.1	Introduction	247
9.2	Model Systems.	248
9.3	Fluorescence Microscopy and Membrane Domains in GUVs.	249
9.4	Fluorescent Probes	250
9.5	Two-Photon-Excitation Microscopy	251
9.6	LAURDAN Probe: the Tips	252
9.7	Membrane Lateral Structure in Artificial Lipid Mixtures and Natural Lipid Extracts as Seen by LAURDAN	255
9.7.1	The Importance of Visual Information to Ascertain Lateral Structure in Compositionally Complex Mixtures	257
9.7.2	LAURDAN in Cell Membranes and Tissues.	258
9.8	Concluding Remarks	260
9.9	Summary.	261
	References	262

CHAPTER 10

Three-Dimensional Electron Microscopy. The Coming of Age of a Versatile Structural Biology Technique

CARMEN SAN MARTÍN, MIKEL K. VALLE

10.1	Introduction	267
10.2	Basis of 3D-EM	268
10.3	Recent Evolution of the 3D-EM Field.	270
10.4	The Cutting Edge	270
10.5	Dynamic Structures and the Fourth Dimension in Cryo-EM.	272
10.6	Cryo-electron Tomography	273
10.7	Concluding Remarks	276
	References	276

Subject Index	279
--------------------------------	-----

List of Contributors

Abrahams, Jan Pieter

Section Biophysical Structure
Chemistry, Leiden Institute of
Chemistry, Leiden University, PO
Box 9502, 2300 RA Leiden, The
Netherlands

Arrondo, José Luis R.

Unidad de Biofísica (Centro Mixto
CSIC-UPV/EHU) and Departamento
de Bioquímica, Universidad del País
Vasco, Apdo. 644, 48080 Bilbao, Spain

Bagatolli, Luis A.

MEMPHYS – Center for
Biomembrane Physics, Department of
Biochemistry and Molecular Biology,
University of Southern Denmark.
Campusvej 55, 5230 Odense M,
Denmark

Cai, Peng

Department of Chemistry, Newark
College of Arts and Science, Rutgers
University, Newark, NJ 07102, USA

Carrión-Vázquez, Mariano

Instituto Cajal/CSIC, Avda. Doctor
Arce 37, 28002 Madrid, Spain

Díez, Héctor

Instituto Cajal/CSIC, Avda. Doctor
Arce 37, 28002 Madrid, Spain

Fernández, César

Novartis Institutes for Biomedical
Research, 4002 Basel, Switzerland

Fernández, Jesús

Instituto Cajal/CSIC, Avda. Doctor
Arce 37, 28002 Madrid, Spain

Flach, Carol R.

Department of Chemistry, Newark
College of Arts and Science, Rutgers
University, Newark, NJ 07102, USA

Freeman, Ray

Jesus College, University of
Cambridge, Cambridge CB5 8BL, UK

García-Pacios Marcos

Unidad de Biofísica (Centro Mixto
CSIC-UPV/EHU) and Departamento
de Bioquímica, Universidad del País
Vasco, Apdo. 644, 48080 Bilbao, Spain

Georgieva, Dilyana

Section Biophysical Structure
Chemistry, Leiden Institute of
Chemistry, Leiden University, PO
Box 9502, 2300 RA Leiden, The
Netherlands

Goñi, Félix M.

Unidad de Biofísica (Centro Mixto
CSIC-UPV/EHU) and Departamento
de Bioquímica, Universidad del País
Vasco, Apdo. 644, 48080 Bilbao, Spain

Hervás, Rubén

Instituto Cajal/CSIC, Avda. Doctor
Arce 37, 28002 Madrid, Spain

Ibarra-Molero, Beatriz

Facultad de Ciencias, Departamento de Química Física, Universidad de Granada, Fuentenueva s/n, 18071 Granada, Spain

Iloro, Ibon

Unidad de Biofísica (Centro Mixto CSIC-UPV/EHU) and Departamento de Bioquímica, Universidad del País Vasco, Apdo. 644, 48080 Bilbao, Spain

Kuil, Maxim E.

Section Biophysical Structure Chemistry, Leiden Institute of Chemistry, Leiden University, PO Box 9502, 2300 RA Leiden, The Netherlands

Kupče, Eriks

Varian Ltd., 6 Mead Road,, Oxford Industrial Park, Yarnton, Oxford OX5 1QU, UK

Martínez-Martín, David

Instituto Cajal/CSIC, Avda. Doctor Arce 37, 28002 Madrid, Spain

Mendelsohn, Richard

Department of Chemistry, Newark College of Arts and Science, Rutgers University, Newark, NJ 07102, USA

Oberhauser, Andrés F.

Department of Neuroscience and Cell Biology, University of Texas Medical Branch, Galveston, TX 77555, USA

Oroz, Javier

Instituto Cajal/CSIC, Avda. Doctor Arce 37, 28002 Madrid, Spain

Sanchez-Ruiz, Jose M.

Facultad de Ciencias, Departamento de Química Física, Universidad de Granada, Fuentenueva s/n, 18071 Granada, Spain

San Martín, Carmen

Biocomputing Unit, Centro Nacional de Biotecnología (CNB-CSIC), Darwin 3, 28049 Madrid, Spain

Szyperski, Thomas

Departments of Chemistry and Structural Biology, State University of New York at Buffalo, Buffalo, NY 14260, USA

Valle, Mikel K.

Biocomputing Unit, Centro Nacional de Biotecnología (CNB-CSIC), Darwin 3, 28049 Madrid, Spain

Wider, Gerhard

Institut für Molekularbiologie und Biophysik, ETH Zürich, 8093 Zurich, Switzerland

Protein Nanocrystallization

DILYANA GEORGIEVA, JAN PIETER ABRAHAMS, MAXIM E. KUIL

1.1 Introduction

There is no theory that allows us to predict when or where proteins will crystallize. However, for several reasons the problem is a very pertinent one, especially when we consider crystallization of proteins that are physically confined within a very small volume.

There is also a practical reason for studying protein crystallization in small, confined volumes: crystals are required for determining three-dimensional protein structures by X-ray crystallography. As crystallization conditions can only be found through trial and error, current practice requires simultaneous testing of many different conditions. The obvious idea that minimizing the volume of single tests maximizes the number of different conditions that can be screened with a given quantity of protein prompted the development of high-throughput nanocrystallization systems (Stevens 2000; Rupp 2003a, b; Bard et al. 2004).

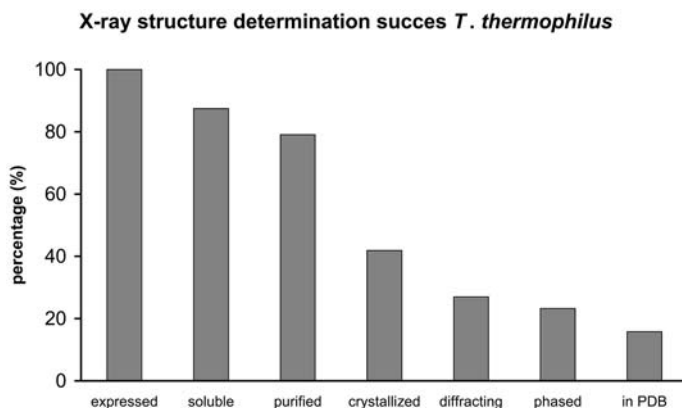


Fig. 1.1. The success rate of high-throughput crystallization. The overall success of the different stages in the high-throughput approach used by the RIKEN consortium is shown. The numerical data were presented at the ICCBM10 conference in Beijing by S. Yokoyama and represent the throughput obtained using expression in *Thermus thermophilus*. The high overall success rate in this example is not typical and expression in higher organisms shows a lower success rate

Although nanocrystallization is quickly becoming a mainstream method, the crystallization step remains the major bottleneck in the structure production process (Blundell and Patel 2004). This is illustrated by recent data from a large structural genomics initiative, indicating that the least successful step in going from sequence to structure is the one from purified protein to crystal. Note that the overall trend illustrated in Fig. 1.1 is not very different from a report predating the widespread use of nanocrystallization (Chayen and Saridakis 2002; Chayen 2004). Probably micro-heterogeneity of the proteins is the prime cause of this bottleneck.

Constructing genetic variants and developing more advanced means of protein production and purification might increase the success rate. Nevertheless, advances in nanocrystallization should also accompany this, as nanocrystallization favors throughput whilst substantially reducing demands on large-scale production and purification platforms.

Here we focus on miniaturization aimed at increasing the probability of finding crystallization conditions when the amount of protein available is limited. First we will review current understanding of nucleation and crystallization of proteins, and focus mainly on those aspects affected by the volume of the mother liquor. Subsequently we will review in detail the major practical obstacles typical of protein nanocrystallization. Problems typically associated with nanovolumes (500 nL or less) concern their dispensing, evaporation and mixing¹. We also discuss the limits imposed by the design of substrates suitable for storing liquid arrays, the robotic accuracy of dispensing strategies, and strategies for scoring nanocrystallization trials.

1.2 Nucleation and Crystallization in Nanovolumes

Naively, one might think that the protein concentration determines the level of supersaturation regardless of the volume. However, this may not be the case, considering that in tiny droplets the surface tension forces become relevant and below a certain volume even predominant. Inside a small nanodroplet the pressure can be substantially higher than the ambient pressure and can be calculated using the Young–Laplace equation (for a review see de Gennes 1985; Blokhuis 2004). However, these effects are less likely to influence protein crystallization in the microliter range. The pressure difference between the inside of a water droplet of 100- μm radius and the gas phase for a surface tension of 72 mN/m is only equal to 1.44 kPa (kN/m^2). Giegé and coworkers studied the influence of external hydrostatic pressure on the nucleation and growth of lysozyme crystals and reported that increasing the pressure from 0.1 MPa (atmospheric pressure) to 250 MPa leads to reduction of the size and number of lysozyme crystals. Moreover a transition to urchinlike particles made of crystalline needles progressively occurs (Lorber et al. 1996; Kadri et al. 2003).

These considerations are obviously irrelevant when the protein is confined within a lipid membrane and thus do not apply for proteins dissolved in the cytoplasm

¹ Classical numerical rounding separates the nanoliter from the microliter range: less than 0.5 is rounded to zero, if one wants to define the nanoliter regime its upper boundary is 500 nL.

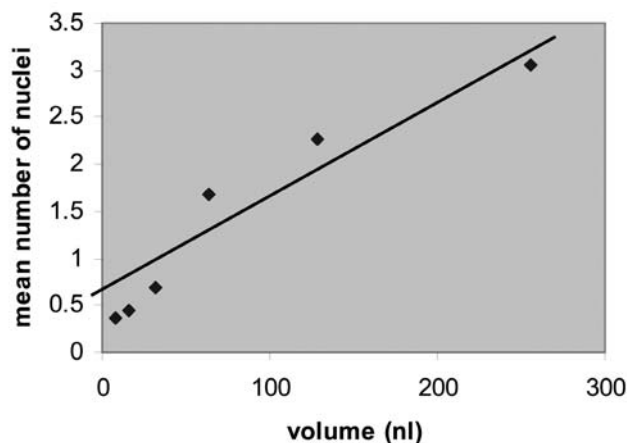


Fig. 1.2. Heterogeneous nucleation in submicroliter volumes. The average number of tetragonal crystals per droplet detected 24 h after mixing as a function of the volume of the droplet. Each data point is the count obtained from 16 droplets. In the smaller droplets needlelike crystals showed a higher relative abundance. (From Bodenstaff et al. 2002)

of living cells. The pressure inside a living cell is well regulated and partially determined by the presence of surrounding tissue. In plant cells the turgor or intracellular pressure can reach several atmospheres at most (Tomos and Leigh 1999).

For practical purposes it is more important that the homologous nucleation rate in protein crystallization is theoretically determined by the level of supersaturation, and it is independent of the volume of the mother liquor. If at a certain level of supersaturation it takes on average a full day to form a stable nucleus that grows into a macroscopic protein crystal in say 1 μL , then it would take 50 days on average for a similar event to occur in a volume of 20 nL. If the nucleation rate per unit volume is constant, reduction of the crystallization volume therefore results in a reduced chance of finding crystals. In other words, one has to increase the level of supersaturation in nanoliter crystallization trials in order to observe rare nucleation events. The relation between the crystallization volume in submicroliter volumes and the observed number of crystals is shown in Fig. 1.2 and indicates that there is a dependence on the droplet volume (Bodenstaff et al. 2002). The relation appears to be linear, but does not go through the origin, indicating that a basic assumption of the homogeneous nucleation theory is not satisfied. This suggests that heterogeneous nucleation plays an important role in low volumes. Vekilov et al. report that despite precautions, heterogeneous nucleation is always observed in their experiments and led to a nonzero intercept of the linear dependence of N (mean number of observed crystals) as a function of the induction time, Δt , in a volume of 700 nL (Galkin and Vekilov 1999; Chernov 2003; Vekilov and Galkin 2003).

Note that although the probability of finding a crystal is very low, a nucleus can always be formed owing to a spontaneous (homogeneous) nucleation event because of density fluctuations (ten Wolde and Frenkel 1997) At this point two types of heterogeneous nucleation should be distinguished: heterogeneous nucleation that de-

depends on nuclei that float in the bulk volume and heterogeneous nucleation that is somehow related to the surface of the mother liquor. In the first case, homogeneous and heterogeneous nucleation cannot be distinguished by changing the crystallization volume. In the latter case reduction of the crystallization volume would increase the relative contribution of heterogeneous nucleation. On the basis of the experimental results it can be argued that there may exist a certain (very low) volume below which heterogeneous nucleation will be the dominant nucleation mechanism (Galkin and Vekilov 1999; Bodenstaff et al. 2002). The early stages of crystallization have been probed using fluorescence energy transfer (Pusey and Nadarajah 2002), but the mechanism of nucleation (homogeneous or heterogeneous) remains poorly understood. Most of the atomic force microscopy work has focused on crystal growth (McPherson et al. 2001, 2003) with the notable exception of work from the Vekilov group (Yau and Vekilov 2001). To induce nucleation or to reduce the induction time of crystallization, different engineered and natural seeding materials have been tested, but they turned out to be successful only for certain proteins. This indicates that, probably there is no “universal nucleating surface,” so finding a suitable substrate is another process of trial and error in the quest for crystals (Chayen et al. 2001; Pechkova and Nicolini 2001; Sanjoh et al. 2001; Bergfors 2003; d’Arcy et al. 2003).

The critical radius, r^* , of the crystal nucleus is the same for homogeneous and heterogeneous nucleation. Following the notation and arguments given by Veessler and Boistelle, we can express the critical radius as (Veessler and Boistelle 1999)

$$r^* = \frac{2\gamma_1 V_m}{kT \ln \beta} . \quad (1.1)$$

The supersaturation is given by the ratio of the actual concentration, C , and the equilibrium saturation concentration, C_s , k is Boltzmann’s constant, T the absolute temperature and γ_1 is the interfacial free energy of the nucleus with respect to the solution. The volume of one molecule in the nucleus is V_m . At the critical size r^* the nucleus is in a very labile equilibrium. If it gains one molecule such that $r > r^*$, it will continue to grow. But if it loses one molecule such that $r < r^*$, it will spontaneously dissolve. If a cap-shaped nucleus with radius r is formed on a surface it contains fewer molecules than a sphere with the same radius in bulk solution. In heterogeneous nucleation three surface free energies play a role: γ_1 between the nucleus and the solution, γ_a between the nucleus and the substrate and γ_o between the substrate and the solution. Depending on the values of these energies the probability of a nucleation event may increase. The substrate can induce nucleation at even lower supersaturation as less energy is required to form the nucleus on the surface (Veessler and Boistelle 1999). As below a certain volume homogeneous nucleation becomes highly improbable, introducing heterogeneous nucleation sites could be an attractive approach to induce crystallization in a controlled manner in very small volumes. Although the chance of finding crystals decreases with decreasing volume, protein nanocrystallization has been shown to be a viable approach. A relatively small increase in supersaturation can easily compensate for the decreasing chance of finding crystals in the screening phase. The important optimization of the crystal growth phase can only be started after the identification of suitable nucleation conditions. In this respect the use of heterogeneous surfaces may help us to develop even small

assays to find these nucleation conditions. Carefully designed growth strategies are subsequently needed to provide us with X-ray diffraction quality crystals needed for successful structural biology.

1.3 Creating and Dispensing Small Liquid Volumes

The controlled dispensing of very small liquid volumes was first demonstrated by Elmqvist (in the context of printing) in the Siemens–Elema Minograf recording mechanism (US patent 2,566,443, issued September 1951). Important factors in the dispensing of small liquids volumes are:

- Dynamic range of the dispensed volume
- Dispensing frequency (determines throughput)
- Precision and accuracy
- Linearity
- Reliability
- Ease of operation and maintenance
- General compatibility of surfaces and liquids (compatible with labile compounds)

The preferred size range for (protein) droplets is between 20 pL and 20 nL, as the total trial volume should be low to realize significant advantage over classical methods. The manual, classical, dispensing of small volumes is by pipetting, but below a volume of roughly 200 nL pipetting becomes notably inaccurate and unreliable. Although manual dispensing can be used for small volumes, convenience and accuracy rules out their use in high-throughput experimentation. Low-volume manual dispensing in protein crystallization was reported by Yeh for drops above 100 nL using a handheld nanoject pipettor with an error of the order 5–9%. For drops smaller than 100-nL volume the error rises rapidly (Yeh 2003). For most applications a standard error of 5% is considered the upper limit (Rose 1999). As manual dispensing is neither accurate nor convenient at volumes below 100 nL, especially when variation in droplet composition is essential for the experiment, different methods are clearly needed. Three established methods used in the field that can dispense in the nanoliter and picoliter ranges are the inkjet, electrospray and pin-transfer methods.

1.3.1 Inkjet Technology

Several dispensing systems in protein nanocrystallization have been described in the literature (Stevens 2000; Bodenstaff et al. 2002; Howard and Cachau 2002; Krupka et al. 2002; Kuil et al. 2002; Santesson et al. 2003; Blundell and Patel 2004). Inkjet nanodispensing involves application of a force – electrical, thermal or acoustic – that generates a pressure wave through the fluid. The liquid stream created is allowed to escape through a small orifice. When the liquid passes through the ori-

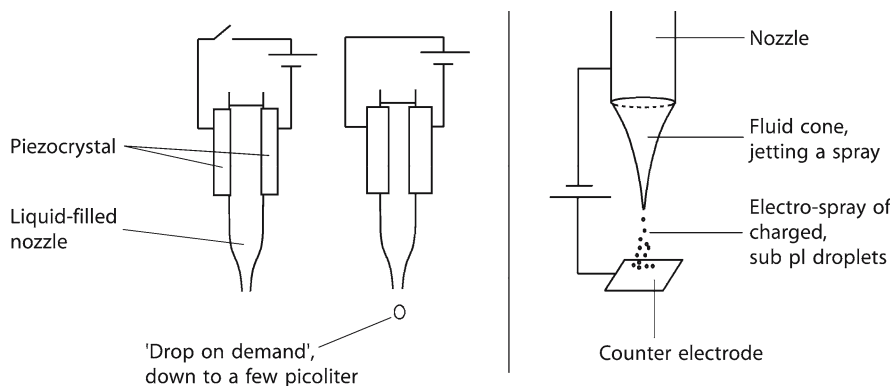


Fig. 1.3. Piezoelectric dispensing versus electro-spray dispensing. In the piezoelectric dispensing method (*left*) a pressure wave is generated that allows detachment and propelling of the droplet from the fluid body. Electro-spray dispensing uses a different principle: an electric potential difference pulls droplets out of a fluid body. (From Bodenstaff et al. 2002)

Since the pressure difference allows the stream to overcome the surface tension forces and to be ejected as a drop. In a continuous inkjet the liquid supply is pressurized sufficiently to create a jet. The breakup of the jet can be synchronized by applying a periodic modulation of the velocity of the fluid exiting the nozzle. One of the most successful applications of the continuous inkjet technology is the Hertz continuous mist inkjet. In this method charged drops with volume of about 3 pL are produced at very high drop repetition frequency. Although this method is very fast and allows for high throughput, it has not been applied for protein crystallization. In drop-on-demand inkjet methods the forces to overcome surface tension forces and emit a drop or a cluster of drops are generated in response to a signal. The liquid supply is not sufficiently pressurized to form a (continuous) jet. The liquid is held in a nozzle, forming a meniscus, and remains in place until some force overcomes the inherent surface tension that keeps the liquid together. The commonest approach is to suddenly raise the pressure on the liquid, propelling it from the nozzle to the surface. It is also possible to pull the liquid out of a nozzle by an attractive force overcoming the surface tension (Pond 2000). For protein nanocrystallization the precise positioning and timing of the droplet deposition is of great importance; therefore, the drop-on-demand method of dispensing appears more suitable than the Hertz technologies. In view of the required drop volume, the precise positioning and the nature of the dispensed liquid, the relevant technologies are piezoelectric, electrostatic and acoustic drop-on-demand dispensing. To dispense solutions containing proteins that are possibly heat-sensitive, the bubble jet technology is considered less suitable. In piezoelectric dispensing a piezocrystal changes its shape in response to an electrical pulse. This results in “squeezing” a glass capillary and thus creates a pressure difference, either by opening a valve leading out of the pressurized container or by pushing against the fluid. As a result a drop is created “on demand.” This type of piezoelectric dispensing is known as “squeeze mode.” Various other methods such as the “bend”, “shear” and “push” modes have been developed and differ in aspects

less relevant for our purpose (Pond 2000). On-demand piezoelectric dispensers can typically create single drops in the picoliter size range and have been successfully used to dispense liquids and solutions with various properties in volumes as small as 0.3 pL (Howard and Cachau 2002). With present-day technology it is possible to very reliably produce droplets with a volume of 25 pL or more. These can be produced at a rate of at least 1,000 droplets per second. The range of viscosities that can be dispensed with piezoelectric inkjet technology ranges from 0.4 to 100 mPa s, e.g., 100 times the viscosity of water at 20°C. This somewhat limited viscosity range is extended by the use of electrospray dispensing, where an electric potential difference is used to pull droplets out of a fluid body (for a review see Rohner et al. 2004). The viscosity of the fluids and the presence of detergents are less relevant for electrospray dispensing compared with piezoelectric dispensing. However, there are currently no commercial dispensing stations that make use of drop-on-demand electrospray methods although electrospray dispensing has significant advantages for the dispensing of very viscous liquids. The use of pulsed electrospray is relatively new (Wei et al. 2002), and recently we constructed several prototypes that can successfully dispense viscous liquids on flat substrates. A schematic comparison between the setup for piezoelectric dispensing and electrospray dispensing is given in Fig. 1.3.

A major problem in the reliable operation of any dispensing system is the problem of clogging, e.g., accumulation of solid material, given the small size of the orifice and the nature of the components to be dispensed. Clogging of the orifice starts and often results in drop misdirection with a disastrous effect on the overall reliability. Careful design of the chamber and orifice, including the proper choice of materials (and their wetting properties), results in reliable operation. For prolonged reliable operation good cleaning procedures and the purity and stability of the liquids to be dispensed are crucial.

1.3.2 Acoustic Dispensing

A novel technology that avoids the clogging problem is acoustic drop-on-demand dispensing where the free surface of a liquid is disrupted by a strong acoustic field. If the acoustic energy is well focused into a small volume of liquid near the surface, a drop of variable size can be ejected. The droplet size depends on the acoustic energy field, the acoustic frequency in combination with the liquid properties. In acoustic droplet ejection (ADE), nanoliter or picoliter droplets are ejected from a conventional microplate by means of the acoustic energy generated by a piezoelectric transducer. The energy is focused via acoustic lenses on the surface of the liquid, causing a droplet of precise volume to be ejected without any physical contact between the acoustic device and the liquid being dispensed. Drops are collected on another surface (e.g., a microplate) positioned in the path of the droplet. The “nozzleless” ADE method avoids the reliability issues associated with the small orifice such as clogging followed by drop misdirection. As acoustic ejection is a noncontact method, cross-contamination between samples caused by the transfer device – a common problem in liquid handling – is largely avoided. A smallest drop volume of 2 pL is possible and the volume can continuously be adjusted up to 40 pL. The acoustic drop-on-de-

mand dispensing technology can generate variable drop volumes at megahertz rates. ADE can dispense a wide variety of solvents, DNA, proteins and even live mammalian cells without detectable loss of activity or viability (Ellson 2003). Although acoustic dispensing has not been applied so far in protein nanocrystallization, the high speed and contactless mode of operation are substantial advantages. The major advantage of ADE is the fast and powerful transfer of liquids. It is less clear how suitable the method is when the total amount of material is very limited and has to be dispensed in nanoliter volumes as in protein crystallography.

1.3.3 Fast Solenoid Valve Technology

The fast solenoid valve microdispensing method couples the accuracy of a stepper-motor-driven syringe pump with the high-speed actuation of a microsolenoid valve. The syringe creates a steady hydraulic pressure within the system fluid, essential to obtain consistent and accurate droplet sizes. For a given pressure the desired

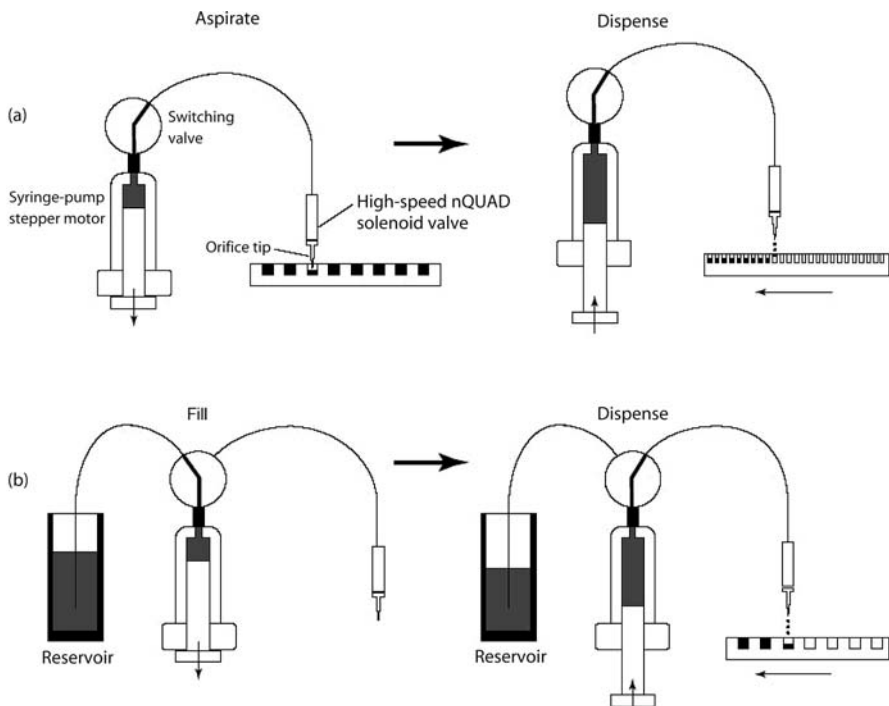


Fig. 1.4. Fast solenoid-syringe method. **a** In the aspirate dispensing mode liquid is collected from microvials and redistributed into nanovials. This mode is used for small volume rearrangements. **b** In the bulk dispensing mode a larger volume is aspirated and distributed into the microvials. This step is used to rearrange moderate volumes. As shown in Fig. 1.7 the Cartesian instrument shows good linearity in both dispensing modes. (From Rose 1999)

drop volume is obtained by choosing an appropriate valve opening time. The technology can be used in two different ways, aspirate dispensing and bulk dispensing. Using the syringe pump, either the sample can be aspirated from a source and then deposited into a destination, or the sample can be introduced through the entire fluidic path and dispensed in a continuous fashion. A schematic view of the bulk and aspirate fast solenoid dispensing mechanisms is given in Fig. 1.4 (Rose 1999). The system is robust and adequate for setting up crystallization in the nanoliter range. In this method the potential of cross-contamination of the dispensed liquids makes a washing step between dispensing different liquids essential. The use of parallel dispensing nozzles each dedicated to a single fluid reduces such demands. As in most dispensing technologies, calibration of the correct parameters in the dispensing process is needed for each solution.

1.3.4 Pin-Transfer Technology

Pin-transfer technology transfers liquid from a source to a location using a solid pin. The process requires dipping a pin into a sample and taking it out. A small volume of liquid remains on the tip of the pin and by placing the pin on a solid surface the liquid is dispensed. The major advantage of this technology is that the pins are simple and relatively inexpensive, while volumes in the low-nanoliter range can be transferred. The pin method has limited flexibility as the pin dispenses a fixed small volume that depends on the tip properties and geometry in combination with the liquid properties. This limits the use in applications requiring variable volumes. Another disadvantage of this technology is that clogging can occur, especially when suspended particles are relatively large with respect to the tip or the gap in the pin. If throughput and cost are more important parameters than high precision and accuracy, pin-transfer technology is a possible alternative to noncontact dispensing. When a screen using fixed volumes of liquids optimized for protein crystallization has been identified, a very fast system with disposable pins could be mass-produced.

1.3.5 Comparison of Liquid Dispensing Methods

Important aspects of the dispensing techniques related to protein crystallization are summarized in Table 1.1. Some highly relevant factors for protein crystallization cannot be compared easily for all techniques and all strategies. As an example, the dead volume of a dispensing system is not a problem for most liquids but for the component of most interest, the protein to be studied, it is highly relevant. No matter how accurate and reliable a dispensing system is, if it is necessary to inject some hundred microliters of precious protein the system it will be of little to no value. Evaluation and reduction of the dead volume is an issue that is not widely addressed in the field of nanodispensing, which mostly focuses on throughput and accuracy. If one would like to compare the amount of protein needed per trial between methods, the dead volume should be included. Equally relevant for all dispensing techniques

Table 1.1. Comparison of nanodispensing techniques. Some important aspects of different dispensing techniques related to protein crystallization are compared. Electro spray has a great potential in protein crystallization, but it is not yet commercially available

Property	Pin method	Piezoelectric inkjet	Thermal inkjet	Acoustic inkjet	Electrospray	Fast solenoid valve column
Driving force	Mechanical	Pressure	Heat	Sound wave pressure	Electric	Pressure (release)
Biocompatible	Yes	Yes	No	Yes	Yes	Yes
Viscous solutions	No	No	Yes	No	Yes	No
Accurate, reproducible	No	Yes	Yes	Yes	Yes	Yes
Small volumes (nanoliter and subnanoliter)	Yes	Yes	Yes	Yes	Yes	No
Low energy transfer to sample	Yes	Yes	No	Yes	Yes	Yes

is the (cross-) contamination of the dispensing system that ultimately results in unreliable operation and failure. The problem of tip contamination has been addressed in the Mosquito dispensing system where an automated system simply discards the tips after use for a single compound. An elegant and rapid way of cleaning and drying for nondisposable tips and equipment is still to be found. In this respect the ADE method seems a viable method as it eliminates the need for disposing and/or washing steps altogether.

1.4 Droplet Evaporation

Langmuir (1918) reached the amazing conclusion that the evaporation rate of a droplet is proportional to the droplet radius. As the droplet volume is proportional to the radius cubed the study of small droplets will eventually be hampered by evaporation. Although droplet evaporation has been studied extensively it is not yet completely understood. Three different situations, most relevant for protein crystallization, can be ranked in complexity:

1. Free droplet in a supporting gas (liquid–gas interface)
2. Sessile or hanging droplets on a supporting solid (liquid–gas–solid interface)
3. Droplet inside a small container (liquid–gas–solid and geometry of container)

The general evaporation problem is complex and many aspects are of importance; especially the evaporation of sessile droplets with dissolved particles differs fundamentally from pure solvent evaporation. The factors that influence the evaporation of a fluid are:

- The droplet volume and shape
- The temperature of the fluid and surrounding vapor
- The composition of the fluid
- The external pressure and the partial vapor pressure

In addition, for a droplet in contact with a solid surface, the surface roughness, composition, temperature and wetting properties as well as the geometry are of importance.

As examples, ingenious experiments have been performed on the evaporation of flying droplets of pure solvents trapped in electric or acoustic fields. In these experiments the presence and influence of a support is eliminated. Beauchamp and coworkers studied the evaporation of charged droplets in the so-called ping-pong configuration (Grimm and Beauchamp 2002; Smith et al. 2002). They used droplets generated by electrospray dispensing and controlled the motion of the charged droplets by reversing the electric field direction when the droplets moved outside the observation volume. In the thesis of Eberhardt (1999), the evaporation of levitated droplets trapped in an acoustic field was studied. Both studies confirmed Langmuir's findings and show that the droplet diameter d_p as a function of time is given by $d_p^2(t) = d_p^2(0) + Ct$, where $d_p(t)$ represents the radius at time t and $d_p(0)$ the initial radius of a pure liquid droplet, with C (a negative constant) representing the evaporation.

1.4.1

Evaporation of a Binary Mixture of Pure Liquids

Eberhardt (1999) and Sefiane et al. (2003) have studied the evaporation of binary mixtures from free and sessile droplets respectively. They both observed the evaporation of the most volatile component first, as expected, but in the presence of a supporting surface there appears an intermediate phase related to the wetting properties of the surface. The addition of a less volatile component to the droplets will slow down the overall evaporation. This observation may allow the reduction of the evaporation of very small droplets of mother liquor too.

1.4.2

Evaporation of a Solvent with a Solute

The evaporation of a solvent with a (nonevaporating) solute proceeds significantly differently from the case discussed before. The most general case is the evaporation of a solvent with a solute without the presence of a substrate as used in the crystallization in acoustically and/or electrostatically levitated droplets (Chung and Trinh 1998; Santesson et al. 2003; Knezic et al. 2004). In the presence of a surface due to the pinning of the contact line, e.g., the rim of the droplet on the surface, there is a significant flow inside the droplet to supply solvent to the surface of the droplet where the evaporation takes place (as in levitated droplets). The flow of solvent causes the relative accumulation of solute close to the contact line. The effect is illustrated by looking at the ring left by a spilled coffee droplet (Deegan et al. 1997). The solute flow in levitated droplets has been visualized but is difficult to control to improve the crystallization (Chung and Trinh 1998).

1.4.3 Evaporation from Microcompartments

In a series of publications the liquid evaporation from silicon microwells showed solute accumulation as a result of contact pinning in very small wells (Hjelt et al. 2000; Young et al. 2003; Rieger et al. 2003). The accumulation of solute particles and their dynamics has been studied by fluorescence spectroscopy; some characteristic phases in the evaporation process are shown in Fig. 1.5. The white lines represent the trajectory of the solute during the evaporation. These results are relevant for protein crystallogenesis, as they suggest that a concentration gradient builds up during the initial stages of the crystallization experiment.

1.4.4

Practical Approaches to Reduce Evaporation

The aforementioned considerations suggest the following precautions to reduce evaporation. The most important factor (if not interfering with the crystallization process) is the reduction of the temperature of the droplet (and container). Covering

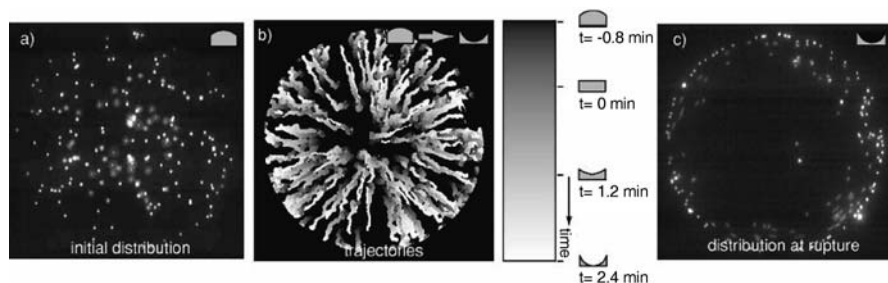


Fig. 1.5. Contact line pinning in nanovials. In this experiment the trajectories of small fluorescent spheres dissolved in a water–glycerol mixture contained in a nanovial are recorded. The *inserts* in grey reflect the shape of the meniscus of the evaporating liquid. It is clear that nonevaporation solutes such as the fluorescent beads travel to the liquid boundary and accumulate owing to contact line pinning in nanovials with a radius of 100 μm . (From Rieger et al. 2003)

the droplet with an immiscible liquid with a high boiling point can further reduce the evaporation. This approach has been successfully applied in batch crystallization where the droplets are covered with oil. By choosing the appropriate oil–solvent combination, the evaporation can be controlled to a certain extent. Evaporation can be prevented also by controlling the surrounding vapor pressure. During the filling of microarrays the evaporation can hardly be avoided. In our own experiments we use a contact-cooled microarray to prevent evaporation. After dispensing, the array can be covered with oil and sealed with adhesive transparent tape. Cooling the microarray has the additional advantage of reversing the convective flow and thereby reducing the solvent transport from the array (Bodenstaff et al. 2002). For most applications in microbatch crystallization it is also possible to fill the (nano) wells first with oil that has lower density than the volatile component(s) (Chayen et al. 1992; Kuil et al. 2002). Subsequently added components sink to the bottom of the well and are covered with a layer of oil. However, sinking droplets do not always merge and mix if this approach is used. Mayer and Köhler (1997) studied the effect of droplet evaporation in microarrays and reported a linear dependence of the evaporation rate on the opening surface of the container. However, more recently a linear dependence of the evaporation rate on the radius of cylindrical microcompartments was reported and it was concluded that the evaporation is diffusion-limited (Rieger et al. 2003).

1.5

Liquid Mixing in Small Volumes and Microfluidics

Two consecutive processes dominate the mixing of liquids: fusion of the liquid boundaries and diffusion of the components. Although both processes are similar in bulk fluid and microfluidics, their outcome is significantly different; unexpected mixing results in microfluidics were reported. Two liquid streams can flow alongside

in a tube a few micrometers wide over a period of time without mixing, almost as if they were separated by glass (Knight 2002). The fusion step is not only the first step, but is also the rate-limiting step in the mixing process. Macroscopically stirring can speed up fusion, as turbulence increases the interfacial area between the liquids, but in small channels it is almost impossible to produce such a turbulent flow. Mechanical forces as in shaking or thermal forces inducing convective flows are less effective and more difficult to apply in very small volumes. A recent approach to accelerate mixing of small volumes is electrosmosis, where components are displaced by electric fields. For the aim of protein nanocrystallization mixing of protein droplets on a small scale can best be achieved during the dispensing phase. For example, in the Microdrop robot system droplets are shot from the nanodispenser with a linear velocity of 3–5 m/s and they dive in the bulk solution without splashing as there is not enough energy for splashes to be formed – droplets with a diameter less than 100 μm have more surface energy than kinetic energy at the speeds generated (see <http://www.microdrop.de>).

When two droplets meet on a solid surface they usually fuse if the liquids involved are miscible. Little is known about what happens next: the subsequent diffusion of dissolved components (e.g., proteins) in high concentrations within the fused nanovolumes. Techniques like dynamic light scattering and fluorescence correlation spectroscopy probe diffusion in (very) small volumes but are always conducted in a total volume of a few microliters, far away from interfaces (Nijman et al. 2001; Schmauder et al. 2002). Although usually homogeneous mixing is aimed for, the lack of homogeneous mixing can sometimes also be an advantage as demonstrated in free-interface diffusion methods used in protein crystallization without evaporation (Hansen et al. 2002; Zheng et al. 2004).

1.6 Design and Making of Nanostructures

1.6.1 Nanoarrays

Making small droplets of mother liquor is not enough: they should also be stored for prolonged periods without evaporation. For this purpose, nanoarrays are used. In robotic dispensing systems the positional accuracy is mostly better than 5 μm and repositioning within 10–20 μm is possible. The latter value limits the minimal size of the wells at present to about 40 μm (and subsequently the volumes to roughly 40 $\mu\text{m} \times 40 \mu\text{m} \times 40 \mu\text{m}$). Apart from size considerations a practical geometry for the containers, which minimizes the evaporation, should be chosen. Irrespective of the method of dispensing, the evaporation is reduced when the opening surface is minimal, implying the height should be as large as possible. This suggests that a tall, straight, cylindrical structure is optimal. However, because the wells also have to be monitored for crystal growth, in practice other geometries are chosen. For example, we designed our wells as truncated inverted pyramids with a bottom and top width of 300 and 800 μm , respectively. For manufacturing nanoarrays, the choice of the material is important and compatibility with proteins is vital in this application; sur-

faces that denature proteins should be avoided. Other essential issues are transparency in the visible part of the spectrum, noncrystalline structure and high uniformity at the length scales studied. If birefringence is used as a detection or recognition technique, the material should in addition also be nonbirefringent. In an early approach we used the elastomer poly(dimethylsiloxane), PDMS, to produce the microwells (Sia and Whitesides 2003). It meets all the requirements mentioned earlier; moreover, the time to produce a prototype is short and the material is not prohibitively expensive. A brass master was micro-machined using a CNC cutting machine to produce deep wells of the desired size. Alternatively, multilayered photolithography can be used to manufacture a master with higher accuracy. Currently, we use polyurethane to make a rigid and more-solvent-resistant array, as shown in Fig. 1.6. The outside dimensions are 29 mm \times 78 mm and the array contains 720 wells. There are numerous high-density microplates on the market that can be used for protein nanocrystallization. Hanging and sitting drop vapor diffusion methods are possible with some plates. Other plates can be exposed to X-rays in order to check the quality of the crystals in the droplet directly. The plates having a low birefringence are suitable if polarization techniques are used for crystal recognition. As the current trend in the high-throughput-screening market is to reduce costs, this reduction is accomplished not only with low-volume but also with high-density plates such as the 6,144-well format made in PDMS used by King at Harvard Medical School and even higher densities such as the 24,576 array available from BioTrove.

1.6.2 Microfluidic Systems

Hansen and Quake introduced a completely different approach to protein nanocrystallization. They designed nanofluidic networks with incorporated valves that could be used to bring protein and precipitant solution in contact in nanoliter volumes. The solutions are introduced using excess pressure forcing the liquids into the fluidic network. The elastomer used is permeable to gas and the channels can be completely filled with liquid. The permeability is also a disadvantage, as crystalliza-

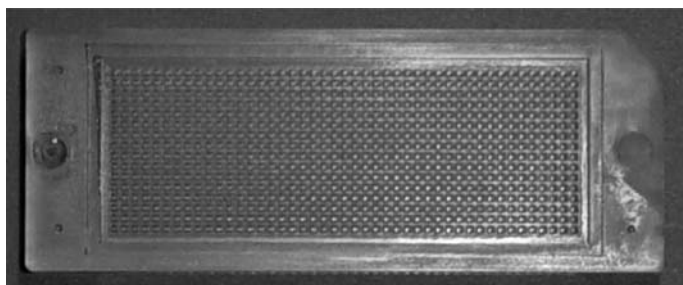


Fig. 1.6. Polurethane array. The array consists of 720 wells (45 \times 16) each with a square opening of 800 μ m and a square bottom 300- μ m wide. The well depth is 2 mm. The array is made by polymerizing polyurethane in a micromachined metal mould

tion trials eventually can dry out or equilibrate with neighboring trials. By applying a pressure to a second network that controls the valves, the protein and precipitant solutions can be pumped around, mixing and/or sealing them (Hansen et al. 2002). The network and channels are made of a very transparent elastomer, allowing a good visualization of the crystallization process. A particular advantage is the use of rectangular chambers for crystallization, allowing for an even illumination of the entire volume when inspecting crystallization trials. The total use of protein is 150 μL for 6,048 crystallization conditions; the time needed to grow crystals was reported to be shorter than by vapor diffusion. Moreover, diffraction-quality crystals can be recovered from the system, indicating that the method is not only suited for screening. The equipment for fluid handling and the nanofluidic chips are commercially available as the Topaz Crystallization System from the Fluidigm Company. The LabChip system from Caliper Life Sciences is another microfluidic system that is used for protein-crystallization experiments.

1.7 Robotics

There are four important aspects in the robotics of protein nanocrystallization:

1. Motion during the dispensing
2. Motion involved in filling/refilling and cleaning
3. Dispensing control
4. Motion involved in inspection of plates

The last aspect, inspection of plates, is not discussed at length as most of the critical steps are found in the first three aspects. Several commercial systems storing, retrieving and moving the plates with reduced vibration, accurate alignment and good temperature control are available (e.g., the Bruker Nonius Crystal Farm).

Robotics required for dispensing should be fast in order to reduce evaporation and increase throughput. Massive parallel-dispensing systems, such as 1,536-nozzle inkjet systems, have almost eliminated the speed problem as all compounds can be dispensed in one step (Gast and Fiehn 2003). Cocktails can be freshly prepared in microliter to milliliter quantities using existing technology and can be “uploaded” to a “multinozzle” pipette and multiple 1,536 arrays can be processed with a very high throughput. The overall speed is determined by the dispensing speed of the protein – the most precious component. As the amount of protein is often limited, the amount of protein needed for dispensing should be minimal. To reduce dead-volume losses, the optimal dispensing strategy is to dispense the protein sequentially from a highly reliable single nozzle system, which will then take most of the processing time. This time can only be reduced by accurate in-flight dispensing. For example, to optimize the protein stability, a chilled protein holder has been developed. One microliter can be dispensed to all wells of a 96-well plate in approximately 15 s (Cartesian synQUAD technology). Most pipetting stations use standard size microtitre plates and the spacing and sizes of future-generation plates can easily be extrapolated. The plate height has not been considered too much in the early design considerations. When the standard footprint size is kept constant we will likely be

faced with very flat plates that will easily bend, and as the assay volumes continue to decrease the plates will inevitably become lower. The manipulation of plates is at present mainly mechanical, but advances in magnetic materials could lead to plates that are picked up by electromagnets covered by a small elastic buffer to dampen the vibrations upon contact. Another approach could be the use of vacuum tweezers if the arrays become really small.

As screening operations tend to use lower volumes and thus higher-density plates, not only reproducible reagent dispensing but also rapid and highly accurate positioning are required. The accuracy needed scales with the radius of the droplets e.g., with $V^{1/3}$ and there are many robotic systems that achieve the desired resolution, sometimes at amazing speeds (and accelerations). In comparing different high-throughput concepts, measuring the average timing of individual dispensing steps is insufficient. For example, although most inkjet-type dispensing is very fast, this speed is only realized for multiple dispensing of a small number of components (four to eight colors). If washing is required between each transfer, the throughput is substantially reduced. Often, intermediate washing steps or the change of tips is rate-limiting for the overall speed of the dispensing process. For example in the Cartesian dispensing system the dispensing head is composed of either 4, 8, or 16 independent channels that each can be used to dispense a different reagent with a different volume and therefore there is no need to wash the channels during the dispensing process if the number of components is not too high.

The precise positioning of the droplets in the nanowells as well as the reproducibility of the system can be checked using fluorescence. If the liquid to be dispensed is labeled with fluorophores, the trajectory, final location and the dispensed volume can be verified. Using the intensity of the fluorescence signal, we can calibrate the amount of liquid that has been deposited in each well.

Moving the dispensing head over the well, stopping, dispensing the drop into the well and then moving to the next well in high-density plates is the usual practice in syringe–solenoid technology. In-flight dispensing, which increases throughput, as no stopping and starting of the dispensing head relative to the plate is required, is more problematic. Here, the actuation frequency and the table speed must be adjusted to deliver the correct number of drops to each well. Nevertheless, dispensing, a high-density plate using multiple nozzles can be completed in tens of seconds. This is more than sufficient to keep up with the protein purification steps. Data showing linearity, accuracy and dynamic range for syringe–solenoid dispensing technology are given in Fig. 1.7; the insert shows the linearity for low volumes. Note that the linearity is good both in the submicroliter and in the microliter range. Accuracy tests for nanodispensing have been described by Rose (1999) and Walter et al. (2003). With increasing viscosity of the liquids, the reproducibility of dispensing decreases. In the Microdrop system this has been remedied by equipping the dispensing head with a nozzle heater controlled by a built-in temperature sensor that limits temperature variations to less than 1°C. The nozzle heater has a dual function: in the range 20–100 mPa s it keeps the temperature constant to avoid viscosity changes by variations of ambient temperature; above 100 mPa s it is necessary to reduce the viscosity below 100 mPa s by increasing the liquid temperature (see <http://www.microdrop.de>). Protein solutions cannot be heated, but since these rarely have a viscosity in this high range, this is not a limitation.

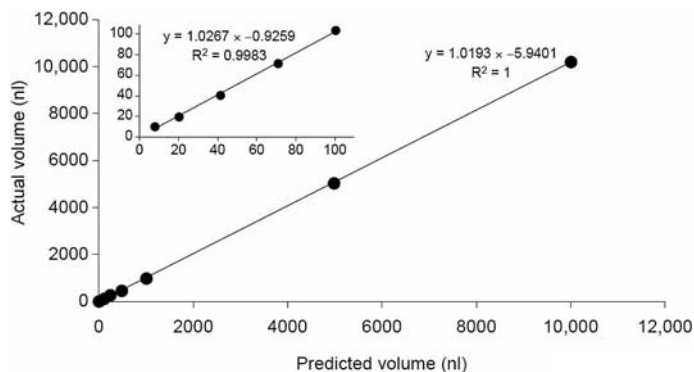


Fig. 1.7. Dispensing linearity. The linearity, accuracy and dynamic range of a syringe–valve system (PixSys 3200, Cartesian Technologies). Note the two different ranges that can be used in this system which allows the dispensing of both the protein component and the reservoir solution for vapor-diffusion crystallization experiments. (From Rose 1999)

1.8 Crystal Recognition

When several thousands of crystallization conditions need to be screened daily, robotic inspection systems that can reliably identify clear or empty droplets (most cases) and easily recognize crystals become essential. Human experts can then handle a reduced number of more difficult and possibly interesting cases that are selected automatically from an image database. Several groups involved in (ultra-) high-throughput experimentation recognized the data congestion problem very early (Jurisica et al. 2001; Cumbaa et al. 2003; Luft et al. 2003).

To optimize crystal recognition, the properties of (protein) crystals have to be compared with those of all other (often highly unexpected) structures found in crystallization trials. Currently the classical method of detecting crystals using a microscope is still most frequently used. As the crystal size for a diffraction experiment is in the 10–100- μm range, the optical microscope has ample spatial resolution to detect crystals of a suitable size. Protein crystals often have poor contrast because they contain 30–80% solvent. The study of such highly transparent samples is best done with advanced microscopic techniques to enhance the weak contrast. To avoid missing crystals growing at different heights in the droplet in most cases the focal plane needs to be varied. The most popular microscopy methods can be classified into three types: bright-field, polarized-light and fluorescence methods. Protein crystals come in a great variety of forms: microcrystals, microneedles, needlelike crystals and larger faceted crystals. As many crystals show straight edges, crystal detection often involves recognition of such edges, although in the case of microcrystals the image magnification may not be sufficient to detect them. Edge detection is widely used in image processing and exploits the correlation between neighboring pixels in an image. When a pixel's grey-level value is similar to the values in pixels around it, there is probably not an edge at that point. If, on the other hand, a pixel

has neighbors with widely varying grey levels, it may represent an edge point. In other words, an edge is identified by a discontinuity in grey-level (intensity) values. Ideally, an edge separates two distinct objects, but in practice apparent edges are caused by changes in color, texture, lighting conditions or detection efficiency during the image acquisition process. After all edge points have been marked, they are merged to form lines and object outlines. As microcrystals are important indicators of successful nucleation conditions special methods including correlation filter techniques have been devised (Cumbaa et al. 2003). Because a crystallization drop is not optically uniform, as surface tension forces it into a lenslike structure, the outside edges tend to be dark and distorted, and the detection contrast is not uniform in droplets. A further complication comes from foreign bodies in the droplet that often tend to be flagged as crystals. Also inhomogeneous protein precipitates which form immediately upon mixing lead to sharp regular boundaries that are often misinterpreted as crystal edges. Scratches introduced on the plastic surface where the drop sits, either during manufacture or handling, are also likely to be flagged as possible crystals. Hence, the major problem remains the elimination of false-positive images, although often the degree of false-negatives (missed crystals) is often considered to be even more important.

Crystals can be classified optically into three categories: cubic, uniaxial and biaxial (Born and Wolf 2002). In the last two cases crystals can also be detected using their birefringence in polarized light. Throughout the history of crystallography, polarization microscopy has been a valuable method for crystallographers. The polarization microscope is still one of the most used tools to detect and characterize crystals prior to X-ray analysis. Polarizing microscopy may be used to detect structures that are birefringent. Note that birefringent material does not need to be crystalline in three dimensions. Classically, birefringence is measured when the sample is placed between two crossed polarizers. Light will only be transmitted when the sample influences the polarization. We constructed an automatic setup for massive parallel crystal detection by birefringence in nanoarrays, which is schematically depicted in Fig. 1.8. It uses two linear polarizers that are crossed and that rotate in tandem, giving a dark background if no depolarization occurs. Wherever crystalline material is present the polarized light is depolarized so the analyzing polarization filter no longer extinguishes it. During a full rotation of the polarizer pair, two maxima and minima separated by a 90°-phase difference are observed. The amount of transmitted light is integrated using a CCD camera. Crystalline material shows up as light patches in an otherwise dark background. In our setup the total evaluation time is less than 1 min for 360 samples and is able to detect crystals smaller than 10 μm (Bodenstaff et al. 2002). Using lysozyme and glucose isomerase, Echaliier et al. (2004) show that rotating-polarization microscopy is a technique capable of detecting protein crystallites in what appears to be an amorphous precipitate.

An obvious requirement for setups identifying crystals on the basis of their ability to polarize light is that the well plates containing the crystallizing solution are optically homogeneous and nonbirefringent. The advantage of the method is that it measures the transmittance of polarized light, so there is no need for extensive image processing of high-resolution images of individual crystals.

The major disadvantage is that methods based on polarization will certainly miss crystals, i.e., crystals with cubic symmetry are not birefringent and will not

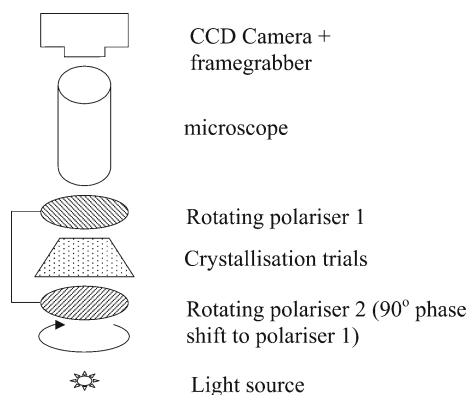


Fig. 1.8. Parallel optical scanner for microarray crystallization. The array is inserted between the two polarizers that are in the crossed position. No light is transmitted unless a depolarizing component, such as a crystal, is present in the array. The system can detect (non-cubic) crystals that are about 10 μm or larger. (From Bodenstaff et al. 2002)

be detected. The problem of missing cubic crystals might be resolved by using edge birefringence, which is observed near boundaries of transparent isotropic and nonisotropic objects (Oldenbourg 1991). Edge birefringence appears at all specimen boundaries when there is a sharp gradient or discontinuity in the refractive index, independent of the nature of the phases at the boundary. Solid–liquid, solid–gas, liquid–liquid and liquid–gas boundaries all give rise to edge birefringence. Although edge birefringence can detect very small cubic crystals it has, as far as we are aware, not yet been used for crystal recognition in protein crystallization.

Fluorescence microscopy is not routinely used as a crystal recognition technique but by exploiting the affinity of fluorescent labels for specific amino acids it may become an important option. By using coupling reactions specific for amino acids, we can visualize only proteins and when the labeling is integrated in the purification protocol unspecific binding of fluorophores can be reduced (Sumida et al. 2001). Recent results indicate that many different labels can be bound to proteins and they incorporate easily into crystals (results of Pusey et al. presented at the ICCBM10, Beijing, 2004). Accumulation of (labeled) proteins in crystals can be determined by quantitative fluorescence detection if labeled proteins are not preferentially located in the bulk solution. Fluorescence intensity measurements can be used to detect protein crystallites in an amorphous precipitate and to distinguish between protein and salt crystals, often a problem in crystallization.

Apart from the detection of crystals automatic classification has to be considered in processing high-throughput protein-crystallization trials. High-throughput robotic protein-crystallization systems are now capable of testing thousands of protein-cocktail combinations per day and are flooding databases with images. This problem is also relevant even for lower-throughput systems, but in the case of high throughput it is a rate-limiting step in the search for favorable protein-crystallization conditions, if human experts have to visually assess the outcome of each trial. Several recent publications discuss automatic classification of images (Spraggon et al. 2002; Wilson 2002; Cumbaa et al. 2003). Some groups classify the state of the entire droplet only, while others characterize all individual objects inside each droplet (Wilson 2002). Some of the approaches obtain a binary set of outcomes (crystals or no crystals) (Cumbaa et al. 2003). For the automatic evaluation they use a “loopy

probabilistic graphical model with a two-layered grid topology using edge detection for crystal recognition” and achieve 85% accuracy in classifying images, assuming human-classified images as ground truth (Cumbaa et al. 2003). To exclude the false evaluation of the well boundaries and to detect the edges of the drop they developed a strategy with several stages: well registration, droplet segmentation and feature extraction to extract descriptive features used during image classification. In total they reduce each image to a vector of 23 features: 20 measuring microcrystal features, two measuring the presence of straight edges detected within a droplet and one measuring the smoothness of the contents of the droplet. In the automatic classification of protein-crystallization images acquired from a high-throughput vapor-diffusion system of Bern et al., edge detection is used and the images are classified in five categories – “empty,” “clear,” “precipitate,” “microcrystal hit” and “crystal.” On test data, this classifier gives about 12% false-negatives (true crystals called “empty,” “clear” or “precipitate”) and about 14% false-positives (true clears or precipitates called “crystal” or “microcrystal hit”) (Bern et al 2004). Although there are still significant problems in the field of automatic crystal recognition in the nanocrystallization environment, there is significant progress in recognition and classification.

1.9 Outlook

Nanocrystallization is becoming a robust means of identifying crystallization conditions and producing seedlings for macroscopic crystals required for structure determination by X-ray crystallography. As the mechanism of nucleation remains enigmatic and finding the proper conditions is still a process of trial and error, testing as many as conditions as possible is at present the only one way out of increasing the chance of getting crystals. The additional costs involved in screening for a single crystal outweigh the costs and efforts involved in protein production and purification. A strict and formal assessment of the success rate of the different screens could largely improve the success of the crystallization trials both in high and low throughput.

Automation and miniaturization also allows us to investigate the statistical aspects of the crystallization process reliably. The use of standardized illumination and detection software allows for a direct comparison between the vast range of protein-crystallization methods and leads to a better understanding of the rationale behind each approach. Nanocrystallization is not just a simple miniaturization of bulk crystallization. Going from the macro to the nano world, the methods described here can only be significantly improved if our understanding of the nucleation mechanism is increased.

Ultimately, this practical consideration will also guide our understanding of an important aspect of living matter. All natural proteins originate from the cytoplasm, which is so concentrated that most of its protein would quickly precipitate if it were a bulk solution. However, the cytoplasm is not a bulk solution: it is confined within cells, it is further compartmentalized within organelles and has a substructure. The confinement of the cytoplasm is a major contributor to the stability of its solvated state. Functional crystal-like structures – the cytoskeleton, chromatin – do exist in-

side cells, but their dynamics of nucleation, growth and dissolution are tightly controlled. When such control fails, nonfunctional crystal-like aggregates disrupt cellular processes and can even result in prion-type diseases. Full understanding of these fundamental processes of life is incomplete without insight into the physical chemistry of nucleation and growth of such (semi) crystalline structures within confined environments like those encountered inside cells (Giegé et al. 1995).

Acknowledgements. The authors acknowledge support from STW-CW (NWO), IOP (Senter-Novem) and the Softlink (FOM) program to conduct this research.

References

- Bard J, Ercolani K, Svensson A, Olland A, Somers W (2004) Automated systems for protein crystallization. *Methods* 34:329–347
- Bergfors T (2003) Seeds to crystals. *J Struct Biol* 142:66–76
- Bern M, Goldberg D, Stevens RC, Kuhn P (2004) Automatic classification of protein crystallization images using a curve-tracking algorithm. *J Appl Crystallogr* 37:279–287
- Blokhuis EM (2004) Liquid drops at surfaces. In: Hartland S (ed) *Surface and interfacial tension measurement, theory, and applications*. Dekker, New York, pp 149–193
- Blundell TL, Patel S (2004) High-throughput X-ray crystallography for drug discovery. *Curr Opin Pharmacol* 4:490–496
- Bodenstaff ER, Hoedemaeker FJ, Kuil ME, de Vrind HPM, Abrahams JP (2002) The prospects of protein nanocrystallography. *Acta Crystallogr D* 58:1901–1906
- Born M, Wolf E (2003) *Principle of optics*. Cambridge University Press, Cambridge
- Chayen NE (2004) Turning protein crystallisation from an art into a science. *Curr Opin Struct Biol* 14:577–583
- Chayen NE, Saridakis E (2002) Protein crystallization for genomics: towards high-throughput optimization techniques. *Acta Crystallogr D* 58:921–927
- Chayen NE, Stewart PDS, Blow DM (1992) Microbatch crystallization under oil – a new technique allowing many small-volume crystallization trials. *J Cryst Growth* 122:176–180
- Chayen NE, Saridakis E, El Bahar R, Nemirovsky Y (2001) Porous silicon: an effective nucleation-inducing material for protein crystallization. *J Mol Biol* 312:591–595
- Chernov AA (2003) Protein crystals and their growth. *J Struct Biol* 142:3–21
- Chung SK, Trinh EH (1998) Containerless protein crystal growth in rotating levitated drops. *J Cryst Growth* 194:384–397
- Cumbaa CA, Lauricella A, Fehrman N, Veatch C, Collins R, Luft J, DeTitta G, Jurisica I (2003) Automatic classification of sub-microlitre protein-crystallization trials in 1536-well plates. *Acta Crystallogr D* 59:1619–1627
- D’Arcy A, Mac Sweeney A, Haber A (2003) Using natural seeding material to generate nucleation in protein crystallization experiments. *Acta Crystallogr D* 59:1343–1346
- De Gennes PG (1985) Wetting: statics and dynamics. *Rev Mod Phys* 57:827–863
- Deegan RD, Bakajin O, Dupont TF, Huber G, Nagel SR, Witten TA (1997) Capillary flow as the cause of ring stains from dried liquid drops. *Nature* 389:827–829
- Eberhardt, R (1999) *Entwicklung einer Ultraschall-Levitations-Apparatur für die Probenvorbereitung in der Mikro- und Spurenanalyse*. PhD thesis, Philipps-Universität Marburg

- Echalier A, Glazer RL, Fulop V, Geday MA (2004) Assessing crystallization droplets using birefringence. *Acta Crystallogr D* 60:696–702
- Ellson R (2003) Nanoliter and picoliter liquid handling for life science applications. In: *Proceedings of the nanotech and biotech convergence conference 2003*
- Galkin O, Vekilov PG (1999) Direct determination of the nucleation rates of protein crystals. *J Phys Chem B* 103:10965–10971
- Gast FU, Fiehn H (2003) The development of integrated microfluidic systems at GeSiM. *Lab Chip* 3:6N–10N
- Giegé R, Drenth J, Ducruix A, McPherson A, Saenger W (1995) Crystallogeneses of biological macromolecules. Biological, microgravity and other physicochemical aspects. *Prog Cryst Growth Ch* 30:237–281
- Grimm RL, Beauchamp JL (2002) Evaporation and discharge dynamics of highly charged droplets of heptane, octane, and p-xylene generated by electrospray ionization. *Anal Chem* 74:6291–6297
- Hansen CL, Skordalakes E, Berger JM, Quake SR (2002) A robust and scalable microfluidic metering method that allows protein crystal growth by free interface diffusion. *Proc Natl Acad Sci USA* 99:16531–16536
- Hjelt KT, van den Doel R, Lubking W, Vellekoop MJ (2000) Measuring liquid evaporation from micromachined wells. *Sens Actuators A* 85:384–389
- Howard EI, Cachau RE (2002) Ink-jet printer heads for ultra-small-drop protein crystallography. *Biotechniques* 33:1302–1306
- Jurisica I, Rogers P, Glasgow JI, Collins RJ, Wolfley JR, Luft JR, DeTitta GT (2001) Improving objectivity and scalability in protein crystallization. *IEEE Intell Syst* 16:26–34
- Kadri A, Damak M, Jenner G, Lorber B, Giegé R (2003) Investigating the nucleation of protein crystals with hydrostatic pressure. *J Phys Condens Matter* 15:8253–8262
- Knezic D, Zaccaro J, Myerson AS (2004) Thermodynamic properties of supersaturated protein solutions. *Cryst Growth Des* 4:199–208
- Knight J (2002) Microfluidics: honey, I shrunk the lab. *Nature* 418:474–475
- Krupka HI, Rupp B, Segelke BW, Lakin TP, Wright D, Wu HC, Todd P, Azarani A (2002) The high-speed Hydra-Plus-One system for automated high-throughput protein crystallography. *Acta Crystallogr D* 58:1523–1526
- Kuil ME, Bodenstaff ER, Hoedemaeker FJ, Abrahams JP (2002) Protein nano-crystallogeneses. *Enzyme Microb Technol* 30:262–265
- Langmuir I (1918) The evaporation of small spheres. *Phys Rev* 12:368–370
- Lorber B, Jenner G, Giegé R (1996) Effect of high hydrostatic pressure on nucleation and growth of protein crystals. *J Cryst Growth* 158:103–117
- Luft JR, Collins RJ, Fehrman NA, Lauricella AM, Veatch CK, DeTitta GT (2003) A deliberate approach to screening for initial crystallization conditions of biological macromolecules. *J Struct Biol* 142:170–179
- Mayer G, Köhler JM (1997) Micromechanical compartments for biotechnological applications: fabrication and investigation of liquid evaporation. *Sens Actuators A* 60:202–207
- McPherson A, Malkin AJ, Kuznetsov YG, Plomp M (2001) Atomic force microscopy applications in macromolecular crystallography. *Acta Crystallogr D* 57:1053–1060
- McPherson A, Kuznetsov YG, Malkin A, Plomp M (2003) Macromolecular crystal growth as revealed by atomic force microscopy. *J Struct Biol* 142:32–46

- Nijman EJ, Merkus HG, Marijnissen JCM, Scarlett B (2001) Simulations and experiments on number fluctuations in photon-correlation spectroscopy at low particle concentrations. *Appl Opt* 40:4058–4063
- Oldenbourg R (1991) Analysis of edge birefringence. *Biophys J* 60:629–641
- Pechkova E, Nicolini C (2001) Accelerated protein crystal growth by protein thin film template. *J Cryst Growth* 231:599–602
- Pond SF (2000) Inkjet technology and product development strategies. Torrey Pines Research, Calsbad, CA
- Pusey ML, Nadarajah A (2002) A model for tetragonal lysozyme crystal nucleation and growth. *Cryst Growth Des* 2:475–483
- Rieger B, van den Doel LR, van Vliet LJ (2003) Ring formation in nanoliter cups: quantitative measurements of flow in micromachined wells. *Phys Rev* 68(3):036312
- Rohner TC, Lion N, Girault HH (2004) Electrochemical and theoretical aspects of electrospray ionisation. *Phys Chem Chem Phys* 6:3056–3068
- Rose D (1999) Microdispensing technologies in drug discovery. *Drug Discov Today* 4:411–419
- Rupp B (2003a) Maximum-likelihood crystallization. *J Struct Biol* 142:162–169
- Rupp B (2003b) High-throughput crystallography at an affordable cost: the TB Structural Genomics Consortium Crystallization Facility. *Acc Chem Res* 36:173–181
- Sanjoh A, Tsukihara T, Gorti S (2001) Surface-potential controlled Si-microarray devices for heterogeneous protein crystallization screening. *J Cryst Growth* 232:618–628
- Santesson S, Cedergren-Zeppezauer ES, Johansson T, Laurell T, Nilsson J, Nilsson S (2003) Screening of nucleation conditions using levitated drops for protein crystallization. *Anal Chem* 75:1733–1740
- Schmauder R, Schmidt T, Abrahams JP, Kuil ME (2002) Screening crystallisation conditions using fluorescence correlation spectroscopy. *Acta Crystallogr D* 58:1536–1541
- Sefiane K, Tadrist L, Douglas M (2003) Experimental study of evaporating water-ethanol mixture sessile drop: influence of concentration. *Int J Heat Mass Transfer* 46:4527–4534
- Sia SK, Whitesides GM (2003) Microfluidic devices fabricated in poly(dimethylsiloxane) for biological studies. *Electrophoresis* 24:3563–3576
- Smith JN, Flagan RC, Beauchamp JL (2002) Droplet evaporation and discharge dynamics in electrospray ionization. *J Phys Chem A* 106:9957–9967
- Spraggon G, Lesley SA, Kreis A, Priestle JP (2002) Computational analysis of crystallization trials. *Acta Crystallogr D* 58:1915–1923
- Stevens RC (2000) High-throughput protein crystallization. *Curr Opin Struct Biol* 10:558–563
- Sumida JP, Forsythe EL, Pusey ML (2001) Preparation and preliminary characterization of crystallizing fluorescent derivatives of chicken egg white lysozyme. *J Cryst Growth* 232:308–316
- ten Wolde PR, Frenkel D (1997) Enhancement of protein crystal nucleation by critical density fluctuations. *Science* 277:1975–1978
- Tomos AD, Leigh RA (1999) The pressure probe: A versatile tool in plant cell physiology. *Annu Rev Plant Physiol Plant Mol Biol* 50:447–472
- Veesler S, Boistelle R (1999) Diagnostic of pre-nucleation and nucleation by spectroscopic methods and background on the physics of crystal growth. In: Ducruix A, Giegé R (eds) *Crystallization of nucleic acids and proteins – a practical approach*. IRL, Oxford
- Vekilov PG, Galkin O (2003) On the methods of determination of homogeneous nucleation rates of protein crystals. *Colloids Surf A* 215:125–130

- Walter TS, Diprose J, Brown J, Pickford M, Owens RJ, Stuart DI, Harlos K (2003) A procedure for setting up high-throughput nanolitre crystallization experiments. I. Protocol design and validation. *J Appl Crystallogr* 36:308–314
- Wei JF, Shui WQ, Zhou F, Lu Y, Chen KK, Xu GB, Yang PY (2002) Naturally and externally pulsed electrospray. *Mass Spectrom Rev* 21:148–162
- Wilson J (2002) Towards the automated evaluation of crystallization trials. *Acta Crystallogr D* 58:1907–1914
- Yau ST, Vekilov PG (2001) Direct observation of nucleus structure and nucleation pathways in apoferritin crystallization. *J Am Chem Soc* 123:1080–1089
- Yeh JI (2003) A manual nanoscale method for protein crystallization. *Acta Crystallogr D* 59:1408–1413
- Young IT, Moerman R, Van den Doel LR, Iordanov V, Kroon A, Dietrich HRC, Van Dedem GWK, Bossche A, Gray BL, Sarro L, Verbeek PW, Van Vliet LJ (2003) Monitoring enzymatic reactions in nanolitre wells. *J Microsc* 212:254–263
- Zheng B, Tice JD, Roach LS, Ismagilov RF (2004) A droplet-based, composite PDMS/glass capillary microfluidic system for evaluating protein crystallization conditions by microbatch and vapor-diffusion methods with on-chip X-ray diffraction. *Angew Chem* 43:2508–2511

Differential Scanning Calorimetry of Proteins: an Overview and Some Recent Developments

BEATRIZ IBARRA-MOLERO, JOSE M. SANCHEZ-RUIZ

2.1 Introduction

This volume is devoted to emerging techniques in biophysics. Differential scanning calorimetry (DSC) has a long history of interesting and useful applications in the field of protein energetics (and also in the energetic analysis of other biological systems). In fact, DSC studies on protein thermal denaturation have played a central role in the development of current views about the factors that determine protein stability. Reviews on the various aspects of this technique are available in the literature (Privalov 1979, 1982, 1989; Sanchez-Ruiz and Mateo 1987; Freire et al. 1990; Freire 1994, 1995; Makhatadze and Privalov 1995; Sanchez-Ruiz 1995; Makhatadze 1998). It may appear, therefore, that DSC cannot be considered as an “emerging technique” in any sense of the word. However, there have been several recent developments in DSC of proteins that merit special consideration. First and foremost, recent studies (in which DSC has been instrumental) have provided evidence for the downhill (barrierless) character of the folding process for some small proteins. Needless to say, the existence of downhill folding (and the possibility that it may be detected by DSC) may have a huge influence in future studies on protein folding and stability. In addition, studies published in the last few years have made apparent that DSC can be used to probe complex processes involving, for instance, interactions of proteins with other molecules or cooperative changes in protein denatured states. In this chapter we focus on these and other “emerging applications” of DSC (although we also provide a short, general introduction to the technique).

Sooner or later, the scientist interested in using DSC will need to confront some classical physical chemistry (including thermodynamics, statistical thermodynamics and even some chemical kinetics) together with the unavoidable mathematics. We believe, however, that many basic concepts in DSC can be explained at a qualitative level without using mathematics and that many readers will find this approach useful, at least on first exposure to the technique. Therefore, no equation is to be found in this chapter. We, of course, make throughout appropriate references to more technical work in the field, where the interested reader may find the physico-chemical and mathematical details.

2.2 What Is DSC?

DSC is a powerful technique to characterize temperature-induced conformational changes in proteins and other biological macromolecules. A differential scanning calorimeter basically consists of two cells: the reference cell and the sample cell (for a detailed technical description of this type of calorimeter, see Privalov 1980; Plotnikov et al. 1997). This section is concerned with the case in which the sample is a diluted protein solution and the reference is the pure solvent (buffer). Both cells are simultaneously heated at a constant heating (i.e., scan) rate and the instrument output is (proportional to) the difference in the heat capacity between the two cells. A typical DSC profile (instrument output versus temperature) for a protein solution is shown in Fig. 2.1. Usually, the profile shows one peak (transition), which is attributed to the heat absorption associated with the denaturation of the protein (for “complex” proteins, several more or less overlapping transitions are often observed), while the pretransition and posttransition values of the output signal reflect the partial heat capacities of the native and unfolded states of the protein, respectively. Also shown in Fig. 2.1 is the instrumental baseline obtained from a DSC scan with the two cells filled with buffer. This “buffer–buffer” baseline is commonly subtracted from the protein–buffer profile, in order to correct for instrumental distortions. Note that, within most of the recorded temperature range, this subtraction would yield negative values. This indicates that the heat capacity of the reference is higher than that of the sample, which is the expected consequence of the lower water content in the latter (the protein “displaces” some water and the heat capacity of liquid water is higher than that of the protein). Therefore, as the heat capacity of the protein

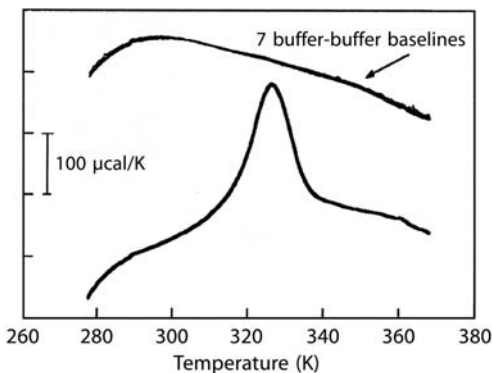


Fig. 2.1. Original differential scanning calorimetry (DSC) thermogram for a solution of apo-flavodoxin (concentration 1.82 mg/mL; 50 mM 3-(*N*-morpholino)propanesulfonic acid buffer pH 7; scan rate 0.5 K/min) and the corresponding buffer–buffer baselines. Six buffer–buffer baselines were obtained before the protein run, and one additional buffer–buffer baseline was recorded afterwards. Recording from the authors’ laboratory using a VP-DSC calorimeter from MicroCal (Northampton, MA, USA). (Reprinted from *J Mol Biol.* Vol 306, Irun et al., Native hydrogen bonds in a molten globule: the apo-flavodoxin thermal intermediate, 877–88, Copyright (2001), with permission from Elsevier)

is measured on the background of high heat capacity of water, it is essential that the thermal history (signal stability) of the instrument be carefully established before any reliable data are collected; for practical issues on DSC experiments, see Makhatadze (1998).

In order to check the reversibility of the processes, DSC experiments on protein denaturation usually include a “reheating run”; that is, a second scan carried out after cooling the protein solution (inside the calorimetric cell) to room temperature. Typically, a process is considered calorimetrically reversible if 85–90% of the endotherm is recovered in the reheating run. The relation between the degree of reversibility and the DSC data analysis is discussed in some detail in Sect. 2.4.

2.3 What Is Heat Capacity?

A DSC calorimeter provides a profile of heat capacity versus temperature for a protein solution. In any textbook of thermodynamics or physical chemistry the reader will find a definition of heat capacity as the temperature-derivative of the enthalpy (or energy) function. This definition admits a simple and intuitive interpretation: heat capacity is a measure of the capacity of the system to store energy. That is, whenever there is in the system an efficient energy storage mechanism, the energy of the system will increase rapidly with temperature, thus giving rise to a high value for the system heat capacity. It is well known, for instance, that the heat capacity of liquid water, per unit of mass, is higher than that for most liquids (in fact, the high heat capacity of liquid water has a considerable influence on the climate of the planet). It is clear that a very efficient energy storage mechanism must be operational in liquid water. If we assume a “mixture model” for liquid water (a “mixture” of formed and broken hydrogen bonds), the mechanism could be described in simplistic terms as “hydrogen-bond breaking”. That is, as temperature is increased, the system (liquid water) stores energy in “broken hydrogen bonds”.

This example also helps us to understand why the differences in heat capacity between protein states (native and unfolded states, for instance) are usually interpreted in terms of hydration contributions from the residues (often classified as “polar” or “apolar”) which are exposed to the solvent in the given states (Gomez et al. 1995; Sanchez-Ruiz 1995; Robertson and Murphy 1997; Luque and Freire 1998; Cooper 2005). This simply recognizes that one of the main sources of heat capacity for proteins in aqueous solution must be the effect of protein molecules on the properties of the surrounding water.

Besides the native and unfolded heat capacity levels (the pretransition and post-transition baselines), a DSC thermogram shows at least one “peak” (transition) that signals the denaturation process (Figs. 2.1, 2.2). The existence of a transition can now be easily understood at a qualitative level. In the temperature range in which unfolding occurs, the system (protein solution) has available an additional energy storage mechanism: since unfolding by increasing temperature requires energy, the system stores energy in unfolded protein molecules. Once essentially all protein molecules are unfolded, that mechanism is no longer operative and the heat capac-

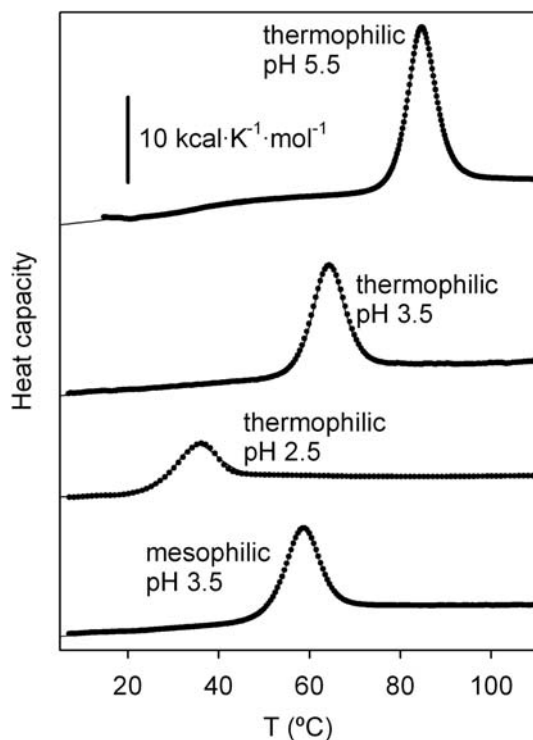


Fig. 2.2. Representative examples of DSC thermograms (plots of heat capacity versus temperature). The proteins used are a thermophilic (*Thermus thermophilus*) and a mesophilic (*Escherichia coli*) ribonuclease H at the indicated pH values. Circles are the experimental heat capacity data obtained after correcting for the instrumental baseline and normalizing to a mole of protein. The thin lines are the best fits of the two-state model to the data. The profiles shown have been shifted in the y-axis for display purposes. (Reprinted from *J Mol Biol*, Vol 329, Guzman-Casado et al., Energetic evidence for formation of a pH-dependent hydrophobic cluster in the denatured state of *Thermus thermophilus* ribonuclease H, 731–743, Copyright (2003), with permission from Elsevier)

ity returns to a comparatively low value (that corresponding to the unfolded protein: the posttransition baseline).

It must be noted, finally, that statistical thermodynamics provides an additional interpretation of heat capacity as a measure of the size of the energy fluctuations in the system (Cooper 1976). Recently (Muñoz and Sanchez-Ruiz 2004; Naganathan et al. 2005), such an interpretation has been found useful in the interpretation of the heat capacity of small proteins in relation to the possibility of downhill folding (Sect. 2.6).

2.4 Equilibrium Thermodynamics Analysis Versus Kinetic Analysis

Possibly the most basic issue regarding DSC data analysis is whether the experimental heat capacity values reflect an equilibrium denaturation process or not. By “equilibrium process” we mean that thermodynamic equilibrium between the significantly populated protein states (native, unfolded, partly unfolded, etc.) is established throughout the temperature scan. In other words, the concentrations of the populated states at any given temperature are determined by values of the relevant equilibrium constants at that temperature. Of course, equilibrium constants

do change with temperature according to simple thermodynamic relationships (i.e., according to the van't Hoff equation and the process enthalpy change), thus leading to the shift in populations of states with temperature that results in transitions ("peaks") in the DSC thermogram. For instance, for an equilibrium two-state process with a sufficiently high denaturation enthalpy value, the denaturation equilibrium constant changes from a low value that favours the native state to a high value that favours the denatured state in a comparatively narrow range. It must be noted that processes that occur through a series of equilibrium states are usually called "reversible" (or quasistatic) in thermodynamics.

The analysis of equilibrium DSC thermograms is well established and has been described in detail in several reviews (Freire et al. 1990; Freire 1994, 1995; Sanchez-Ruiz 1995). In principle, it is possible to use textbook thermodynamics to obtain the number and the thermodynamic parameters (values of free energy, enthalpy, entropy and heat capacity) for all the significantly populated states. Furthermore, this detailed energetic description of the denaturation process can be interpreted in structural terms using structure–energetics correlations that have been developed and refined over the years (Robertson and Murphy 1997; Luque and Freire 1998). More recent advances in the equilibrium analysis of DSC thermograms (such as those related to the characterization of ligand effects and barrierless folding) will be described in subsequent sections of this review.

Powerful though it is, the equilibrium thermodynamics analysis of DSC thermograms cannot be applied to DSC thermograms that are distorted by the occurrence of processes of a kinetic character (Sanchez-Ruiz 1992, 1995; Plaza del Pino et al. 2000). The reason is that the concentrations of states involved in kinetic processes are determined by rate equations (not by a temperature-dependent equilibrium constant). The two common sources of kinetic distortions in DSC of proteins are "slow equilibrium" and the occurrence of irreversible protein alterations (aggregation, autolysis in the case of proteolytic enzymes, etc.)

Irreversible alterations take the protein to a "final" state which is unable to fold back to the native state (under the conditions used in the DSC experiment). Irreversible alterations are, therefore, essentially irreversible processes and must be described in terms of rate equations. In experimental DSC, the occurrence of irreversible alterations is easily detected by the absence of a calorimetric transition in a reheating run (i.e., a second DSC scan performed after cooling from the first one, without removing the protein sample from the calorimetric cell; Fig. 2.3). It is, of course, important to stop the first scan immediately after the transition has been completed to avoid detecting the effect of irreversible alterations that may occur at higher temperatures.

A DSC transition that is not reproduced in the reheating run is said to be calorimetrically irreversible (or just "irreversible"). Calorimetric irreversibility suggests that the DSC transition may be kinetically distorted to a significant extent. A more detailed assessment may be derived from DSC experiments carried out at different scan rates. Since kinetic processes are by definition time-dependent processes, the state of a kinetically determined system at a given temperature will depend on the time required to reach that temperature. It follows that kinetically distorted DSC transitions must show a significant scan-rate dependence (Sanchez-Ruiz 1992, 1995; Plaza del Pino et al. 2000). In fact, most calorimetrically irreversible DSC transitions

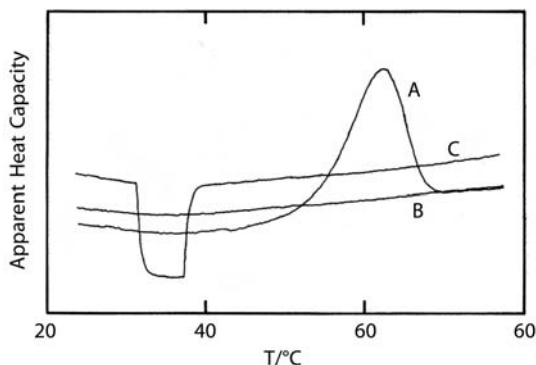


Fig. 2.3. Experimental example of a calorimetrically irreversible DSC transition. *A* Original DSC recording for the denaturation of porcine pancreas procarboxipeptidase B at 2 K/min and pH 9.0; protein concentration 1.13 mg/mL; Zn^{2+} concentration 0.37 mM. *B* Reheating run showing no sign of a transition. *C* Buffer–buffer baseline (showing a 25- μ W calibration mark). (Reprinted with permission from Conejero-Lara et al., 1991 *Biochemistry* 30: 2067–2072, Copyright (1991), American Chemical Society)

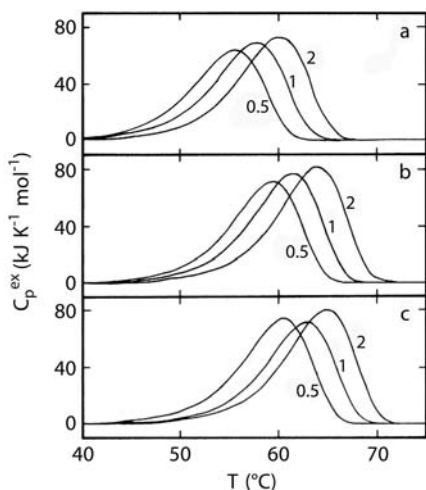


Fig. 2.4. Scan-rate effect on the DSC transitions for the thermal denaturation of porcine pancreas procarboxipeptidase B at pH 9.0 and in the presence of several Zn^{2+} concentrations: *A* 0.26 mM; *B* 1.48 mM; *C* 2.96 mM. Note that the transitions shown are “excess heat capacity” profiles that have been corrected for the instrumental baseline and the so-called chemical baseline (a smooth sigmoidal-like baseline that connects the pretransition and posttransition heat capacity levels). The *numbers* alongside the transitions stand for the scan rate in kelvins per minute. (Reprinted with permission from Conejero-Lara et al., 1991 *Biochemistry* 30: 2067–2072, Copyright (1991), American Chemical Society)

studied so far have been found to be strongly scan-rate dependent (Fig. 2.4). Furthermore, in many cases, these irreversible DSC transitions obey closely a two-state irreversible model, which assumes that only the native and the final irreversible-denatured states are significantly populated during the DSC scan and that the conversion from the native to the final state is described by a first-order rate equation (Sanchez-Ruiz 1992). This simple model is actually a limiting case of more complex situations and it is reached when the irreversible alteration is very fast and the population of states other than native and final becomes negligible. If a DSC transition follows the two-state irreversible model, equilibrium thermodynamics analysis is not possible and information about the equilibrium denaturation mechanism is not available from the experimental DSC profile.

In a few cases, irreversible DSC transitions have been found to show a negligible scan-rate effect (Vogl et al. 1997; Thórólfsson et al. 2002). The likely interpretation is that the irreversible alteration occurs with little thermal effect at temperatures somewhat above those of the calorimetric transition or, perhaps, overlapping slightly with the high-temperature side of the transition. In these cases, equilibrium thermodynamics analysis is acceptable with some obvious precautions. For instance, a reliable value for the denaturational change in heat capacity cannot be possibly obtained, since the denatured, high-temperature baseline is almost certainly distorted by the irreversible alteration, even if the transition is not.

Overall, the DSC studies on irreversible protein denaturation have contributed to the realization that protein stability has both thermodynamic and kinetic aspects. In particular, it has made clear that thermodynamic stability (a positive value for the denaturation free energy at physiological temperature) does not guarantee that the protein will remain in the native state during a given time scale, since irreversible alterations (even if they occur through lowly populated unfolded or partially unfolded states) may deplete the native state in a time-dependent manner (Plaza del Pino et al. 2000). It is clear, therefore, that many proteins (in particular, complex protein systems) must be “designed by evolution” (i.e., naturally selected) to have significant kinetic stability when confronted with the destabilizing effect of the numerous irreversible processes that may occur *in vivo*. Provided that irreversible processes occur mainly from non-native states (unfolded, partially-unfolded, etc.), the required kinetic stability may be achieved through the design of a sufficiently high free-energy barrier for unfolding (the barrier that “separates” the native state from the non-native states). It must be noted that, in recent years, the stability of many complex protein systems of interest has been shown to be of kinetic origin (Schnyrov et al. 1997; Cunningham et al. 1999; Persikov and Brodsky 2000; Jaswal et al. 2002; Mehta et al. 2003; Forrer et al. 2004; Lynch et al. 2004; Manning and Colon 2004; Jaswal et al. 2005; Jaramayan et al. 2005a, b) and that some emerging molecular approaches to the inhibition of amyloidogenesis focus on the increase of the kinetic stability of the protein native state (Foss et al. 2005; Petrassi et al. 2005; Wiseman et al. 2005), as we suggested a few years ago (Plaza del Pino et al. 2000).

When the DSC transition is reproduced with significant amplitude in the reheat-run we say that the denaturation process is calorimetrically reversible. This situation is usually found with “small” proteins (i.e., the kind of “model” proteins often used in fundamental studies on folding and stability). Strictly speaking, calorimetric reversibility does not necessarily imply thermodynamic reversibility (i.e., that the

process occurs through equilibrium states, so that equilibrium is always established during the DSC scan). The reason is that there is the possibility that the rates of the folding and unfolding processes could be comparatively slow in such a way that the scanning (at the usual scan rates) is too fast for equilibrium to be established. This situation is known as “slow equilibrium” (Freire et al. 1990). Slow equilibrium does not appear to be a serious problem for most comparatively small proteins, which are expected to fold and unfold fast at the transition temperature [note that the folding of small proteins is usually fast compared with the time-scale of the DSC experiment (Jackson 1998) and that, at the transition temperature and assuming a two-state process, the rates of folding and unfolding must be equal, so that both processes will be fast]. However, DSC studies carried out in the presence of concentrations of denaturants (urea, guanidine) close to those of the “bottom” of the kinetic chevron plot may show significant slow-equilibrium distortion, since the rates of folding–unfolding may be slow in those conditions (Plaza del Pino et al. 1992). In the absence of denaturants, slow-equilibrium distortions are expected to be small (for a possible exception, see Kaushik et al. 2002) and calorimetric reversibility constitutes good evidence for thermodynamic reversibility. Of course, if in a given particular case it is suspected that the rates of folding–unfolding may be unusually slow, it is advisable to carry out DSC experiments at different scan rates (even if the denaturation process is calorimetrically reversible). Obviously, the slow-equilibrium, kinetic distortion will become less important at the lowest scan rates.

The analyses described in the following sections assume equilibrium thermodynamics is applicable.

2.5

The Calorimetric Criterion for Two-State Behaviour

The two-state equilibrium model assumes that only two states of the protein are significantly populated at equilibrium: the native and the unfolded (reversibly denatured) states. The unfolded state (and, to some extent, the native state as well) is to be considered as an ensemble of conformations in equilibrium. For this reason, these states are often referred to as “macrostates”.

The two-state equilibrium model has been extensively used in studies on protein folding and stability. In fact, it has become so commonplace that it is often taken for granted and two-state behaviour is sometimes assumed with little supporting evidence. Actually, finding “hard” evidence for two-state behaviour is not as straightforward as it might seem at first. For a two-state denaturation, agreement between the experimental transition profiles obtained using different physical probes (fluorescence, far-UV circular dichroism, near-UV circular dichroism, etc.) is expected. However, while a clear disagreement between such profiles effectively rules out two-state behaviour, the agreement does not constitute definitive evidence in support, since there is always the possibility that an additional transition profile (based on a different physical probe) could significantly differ.

Scanning calorimetry, on the other hand, provides a more reliable experimental two-state criterion (Jackson and Brandts 1970). The reason is that the denaturation enthalpy (essentially, the difference in energy between the native and the un-

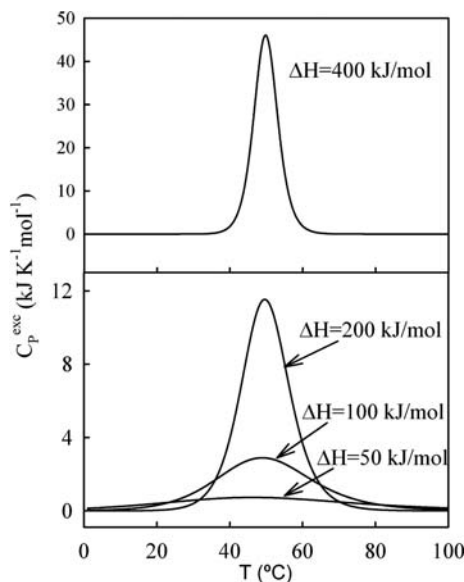


Fig. 2.5. For an equilibrium two-state transition, the unfolding enthalpy determines both the area under the DSC transition and its width. This is illustrated by the excess heat capacity (see legend to Fig. 2.4) versus temperature profiles shown, which were calculated for a transition temperature of 50°C and the values for the unfolding enthalpy shown alongside the transitions

folded macrostates) is determined in two different ways from the DSC experiment (Fig. 2.5). Thus, it can be calculated from the area under the heat capacity peak and, in this case, we calculate the so-called calorimetric enthalpy. In addition, it can be obtained from the shape (width) of the transition, since the unfolding enthalpy determines the temperature dependence of the unfolding equilibrium constant. A high value for the unfolding enthalpy implies that the unfolding equilibrium constant shows a high-temperature-dependence so that it changes from favouring the native state (approximately 100% native) to favouring the unfolded state (approximately 100% unfolded) in a narrow temperature range: the unfolding DSC transition is narrow. Conversely, when the unfolding enthalpy value is low, the unfolding equilibrium constant changes more slowly with temperature and the DSC transition is wide. The unfolding enthalpy value calculated from the shape of the DSC transition is known as the van't Hoff enthalpy (after the van't Hoff equation that gives the temperature dependence of any equilibrium constant). The calculation of the calorimetric enthalpy is essentially model-independent, while the calculation of the van't Hoff enthalpy relies on the two-state model. Agreement between the two values constitutes evidence for two-state behaviour.

The calorimetric criterion based in the comparison between the calorimetric and van't Hoff enthalpies has been successfully used in many studies to ascertain the validity of the two-state model. However, an important point of its application is often overlooked. Small proteins have necessarily low values of the denaturation enthalpy and, therefore, give rise to very broad DSC transitions in which the baselines of the native and unfolded states (the approximately linear pretransition and posttransition heat capacity levels) are not clearly apparent. In these cases, good fits to the experimental data with calorimetric enthalpy equal to the van't Hoff enthalpy can almost certainly be obtained, provided that the baselines of the native and unfolded

states are adequately “chosen” by the person who analyses the data or by the fitting program (Zhou et al. 1999; Kaya and Chan 2000). Obviously, these fits are not proof of two-state behaviour without evaluation of the baselines involved in the fitting. In particular, crossing of the baselines of the native and unfolded states at a temperature in the middle of the transition is clearly unphysical (it would imply that the unfolding heat capacity change becomes negative at that temperature). Recently, such crossings have been detected in several experimental cases and interpreted indeed as evidence of deviation from two-state behaviour (Irun et al. 2001; Dragan and Privalov 2002; Garcia-Mira et al. 2002; Naganathan et al. 2005).

2.6

Two-State Versus Continuous (Barrierless, Downhill) Transitions

Deviations from two-state behaviour have been routinely interpreted in terms of the presence of significantly populated intermediate states. This interpretation is certainly plausible for complex protein systems. For instance, if a protein has two clearly defined structural domains, partially folded intermediates (one domain folded, one domain unfolded) are expected to be populated. Actually, in these cases DSC provides a very convenient way to characterize the energetics of domain–domain interactions (Brandts et al. 1989; Ramsay and Freire 1989; Freire et al. 1992). However, recent studies on small, fast-folding proteins support a different source of deviation from two-state behaviour, associated with the potentially low (or nonexistent in thermodynamic terms) folding–unfolding barrier in some of these proteins (Garcia-Mira et al. 2002; Oliva and Muñoz 2004; Muñoz and Sanchez-Ruiz 2004; Naganathan et al. 2005).

It must be noted that the two-state model (as well as its obvious generalization: a series of equilibria linking the native, intermediate and unfolded states) is actually rooted in the tradition of describing chemical processes in terms of reaction schemes. However, protein folding is not a chemical reaction. While in an organic chemistry reaction a few strong covalent bonds are broken and formed, protein folding involves a myriad of weak, noncovalent interactions (hydrogen bonding, van der Waals interactions, hydrophobic effect, etc.). We may expect an organic chemistry reaction to occur through very well defined intermediate states separated by high energy barriers (since breaking a covalent chemical bond may require energies on the order of hundreds of kilojoules per mole). For protein folding, on the other hand, two-state behaviour only means that microstates of intermediate degree of unfolding have higher free energies, in such a way that they are not significantly populated. A higher free energy for states of intermediate degree of unfolding is equivalent to a thermodynamic free-energy barrier for the folding–unfolding process (consequently, the free energy versus the degree of unfolding profile would show two minima corresponding to the native and unfolded macrostates). There is no physical reason, however, for this barrier to be high (since covalent bonds are not formed/broken). In fact, the barrier could be rather small in fast-folding proteins (otherwise, it is not clear how they could manage to fold so fast). Actually, if the thermodynamic barrier is on the order of the thermal energy (a few kilojoules per mole), microstates of all degrees of unfolding may become populated during denaturation at equilibrium.

That is, the free energy versus degree of unfolding profile would essentially show a single minimum that would shift upon changing temperature and the folding–unfolding process would be continuous and involve a single macrostate. Note also that in a free-energy profile with only one minimum the folding–unfolding relaxation is always downhill (since no barrier needs to be overcome to reach the minimum of the profile).

Recent work has shown that the size of the thermodynamic folding–unfolding barrier for small proteins (and, hence, the two-state or continuous nature of the process) can be ascertained from the experimental DSC thermograms using a suitable data analysis approach (Muñoz and Sanchez-Ruiz 2004). An enthalpy scale (taking the native state as a reference) is used as a measure of the degree of unfolding (i.e., it is used as an order parameter for the unfolding process). Following Landau theory of critical transitions, a free-energy functional is defined as an expansion in even powers of the enthalpy and the expansion is truncated at the fourth power. Such an expansion produces a barrierless free-energy profile with a single minimum (continuous transition) or a profile with a barrier and two minima (two-state transition) depending on the value of a single parameter that describes the size of the barrier (Fig. 2.6). The functional can then be implemented in a nonlinear, least-squares fitting procedure and the size of the barrier can be determined from the fitting to the DSC profile (Fig. 2.7). Indeed, such analysis yielded a barrierless profile for the protein BBL (Fig. 2.8), the first downhill folder identified (Garcia-Mira et al. 2002)

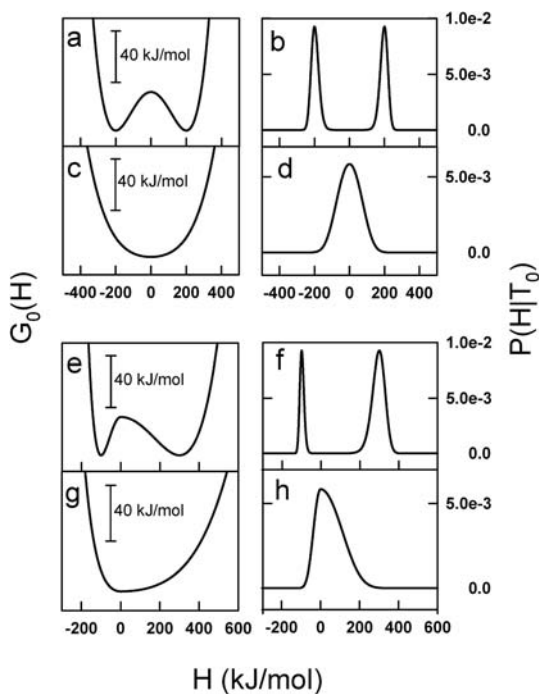


Fig. 2.6. Illustrative free energy versus enthalpy profiles (**a, c, e, g**) and the corresponding probability distributions (**b, d, f, h**). The free-energy profiles in **a** and **e** exemplify two-state behaviour (i.e., a large barrier separates the low-enthalpy and high-enthalpy macrostates). The profiles in **c** and **g** correspond to the barrierless, single-macrostate case. (From Muñoz V and Sanchez-Ruiz JM (2004) Exploring protein-folding ensembles: a variable-barrier model for the analysis of equilibrium unfolding experiments. *Proc Natl Acad Sci USA* 101: 17646–17651, Copyright (2004), National Academy of Sciences, U.S.A.)

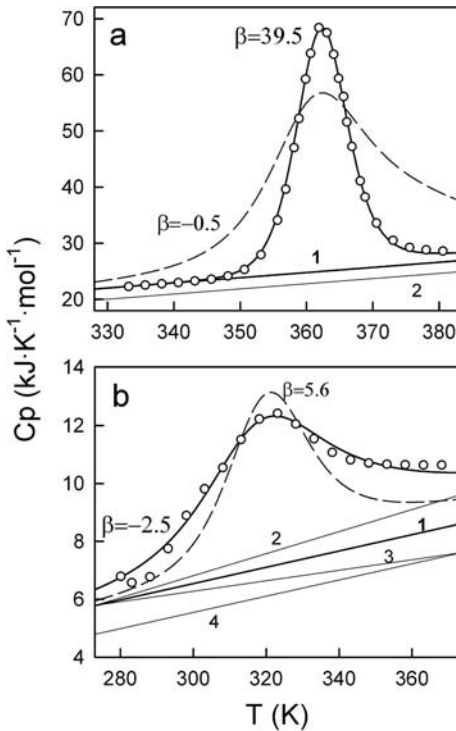


Fig. 2.7. Fitting of a “variable-barrier” model to the absolute heat capacity profiles for the thermal denaturation of *E. coli* thioredoxin (a) and BBL (b). In each case, the best fit is shown with a *continuous line* and the corresponding barrier size (β in kilojoules per mole) is given *alongside the line*. For illustration, we also show (*broken lines*) the fits obtained with values of β of different sign than those for the optimum fit. Note that the optimum fit for BBL (b) is obtained with negative value of β , indicating a barrierless unfolding process. (From Muñoz V and Sanchez-Ruiz IM (2004) Exploring protein-folding ensembles: a variable-barrier model for the analysis of equilibrium unfolding experiments. Proc Natl Acad Sci USA 101: 17646-17651, Copyright (2004), National Academy of Sciences, U.S.A.)

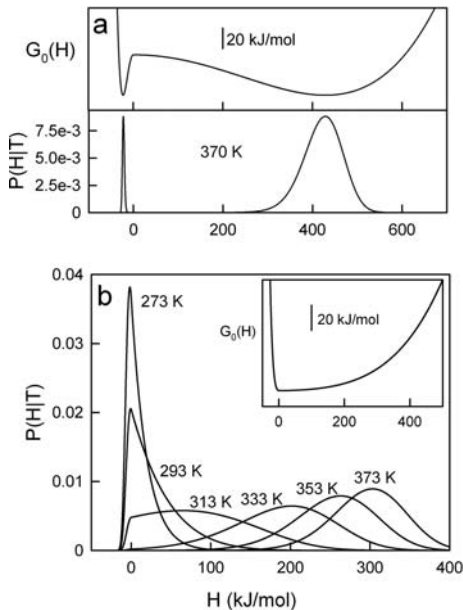


Fig. 2.8. Free energy versus enthalpy and probability distributions for *E. coli* thioredoxin (a) and BBL (b) derived from the fittings shown in Fig. 2.7. Note that for BBL the probability distributions are unimodal, indicating a single macrostate. (From Muñoz V and Sanchez-Ruiz IM (2004) Exploring protein-folding ensembles: a variable-barrier model for the analysis of equilibrium unfolding experiments. Proc Natl Acad Sci USA 101: 17646-17651, Copyright (2004), National Academy of Sciences, U.S.A.)

and promises to be an efficient tool in the identification of other potential instances of barrierless, downhill folding.

2.7 Characterization of Ligand-Binding Effects

Equilibrium thermodynamics analysis of DSC thermograms for protein solutions can, in principle, provide a complete energetic description of the protein denaturation process. Furthermore, information regarding ligand-binding energetics can also be derived from the analysis of ligand effects on DSC profiles. Significant advances in this regard have been reported recently (Rosgen and Hinz 2001; Rosengarth et al. 2001; Luque et al. 2002; Thóroífsson et al. 2002). In particular, a new theoretical approach, based on the binding partition function (binding polynomial), has been proposed to describe the ligand effects on DSC transition temperatures. This is actually a powerful tool as it provides a general, model-independent procedure to analyse data for complex systems (multidomain proteins with multiple and potentially interacting binding sites, for instance). In fact, this approach may allow information regarding localization of binding sites to be derived as well as different binding models to be tested. The potential of this analysis is clearly shown in the case of human phenylalanine hydroxylase (hPAH) (Thóroífsson et al. 2002).

hPAH is a tetrameric enzyme that catalyses the hydroxylation of L-phenylalanine (L-Phe) to L-tyrosine; a dysfunction of this enzyme causes phenylketonuria. Each subunit in hPAH contains an N-terminal regulatory domain, a catalytic domain and an oligomerization domain. Figure 2.9 shows an illustrative example of DSC thermograms of hPAH in the absence and in the presence of L-Phe. Two partially overlapping transitions affected by addition of ligand are apparent.

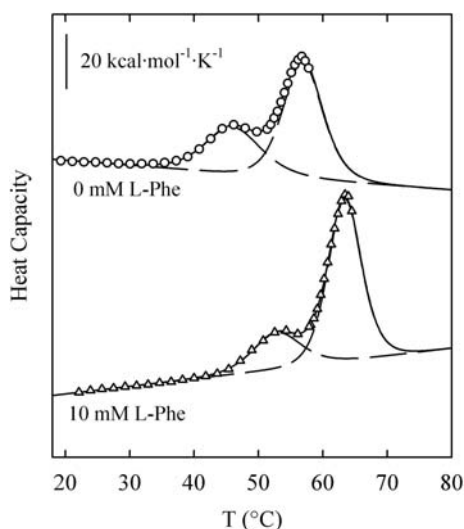


Fig. 2.9. Heat capacity versus temperature profiles for the thermal denaturation of wild-type phenylalanine hydroxylase in the absence and presence of 10 mM L-phenylalanine (L-Phe). Open circles and triangles, experimental heat capacity data. The continuous lines represent the best fits to the sum of two non-two-state transitions; the individual profiles have been displaced in the y-axis for the sake of clarity. (Reprinted with permission from Thóroífsson et al., 2002, *Biochemistry* 41: 7573–7585, Copyright (2002), American Chemical Society)

Additional experiments on truncated hPAH forms have revealed that the low-temperature transition corresponds to the denaturation of the regulatory domains, while the high-temperature transition is due to the denaturation of the catalytic domains.

The thermal denaturation of hPAH was always irreversible (in the absence and in the presence of ligand), as no transition could be detected in reheating runs (not even when the first run had been stopped immediately after the end of the second transition). Furthermore, aggregation was evident in the sample at the end of the experiment. As seen in Sect. 2.4, the analysis of DSC thermograms according to the equilibrium thermodynamics is seriously hampered in such cases of irreversible denaturation. Nevertheless, in a few cases, the analysis of the scan-rate effect on the DSC transitions has supported the equilibrium thermodynamics analysis of irreversible DSC thermograms (Vogl et al. 1997; Thórólfsson et al. 2002). In fact, Fig. 2.10 suggests that at the highest scan-rate kinetic distortions become negligible and the equilibrium thermodynamic analysis is feasible in the case of hPAH. In view of this, the ligand effect on the transition temperatures was studied using the aforementioned binding polynomial formalism. Before describing the results obtained, however, we must note an important result demonstrated, for equilibrium transitions, by Sturtevant and coworkers many years ago (Fukada et al. 1983; Manly et al. 1985): for denaturation processes that involve ligand dissociation, the plot of transition temperature versus ligand concentration does not show a plateau, but the

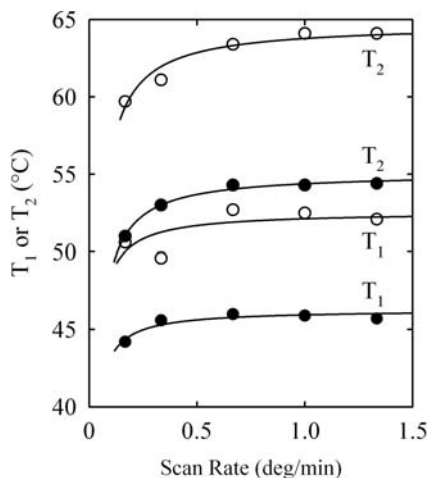


Fig. 2.10. Scan-rate effect on the transition temperatures for the two transitions observed in the DSC thermograms for the thermal denaturation of wild-type phenylalanine hydroxylase (see Fig. 2.9). *Closed symbols*, transition temperatures in the absence of L-Phe. *Open symbols*, transition temperatures in the presence of 22 mM L-Phe. Note that there appears to be little or no scan-rate effect on the transition temperatures within the range 0.6–1.3 K/min, which suggests that, for those scan rates, kinetic distortions become negligible and the equilibrium thermodynamic analysis is feasible. (Reprinted with permission from Thórólfsson et al., 2002, *Biochemistry* 41: 7573–7585, Copyright (2002), American Chemical Society)

transition temperature keeps increasing with ligand concentration even at concentrations for which binding sites are essentially fully occupied. The reason behind this has to do with the contribution to the relevant Gibbs energy change that arises from the increase in ligand translational entropy that takes place upon ligand dissociation, a contribution that is proportional to the logarithm of free-ligand concentration. With this in mind, the analysis of ligand effects on DSC transitions (Fig. 2.11) based upon the general binding polynomial formalism yielded the following relevant results: (1) L-Phe dissociation takes place upon catalytic domain denaturation, consistent with the presence of binding sites in the catalytic domains; (2) there are no binding sites in the regulatory domains and, therefore, the observed effect on the transition temperature is due to the interaction of the regulatory domain with the catalytic domain.

Beyond the application to this particular system, the theoretical approach based on the binding polynomial formalism described by Thórólfsson et al. (2002) provides a general procedure to obtain relevant information regarding ligand-binding processes in complex protein systems.

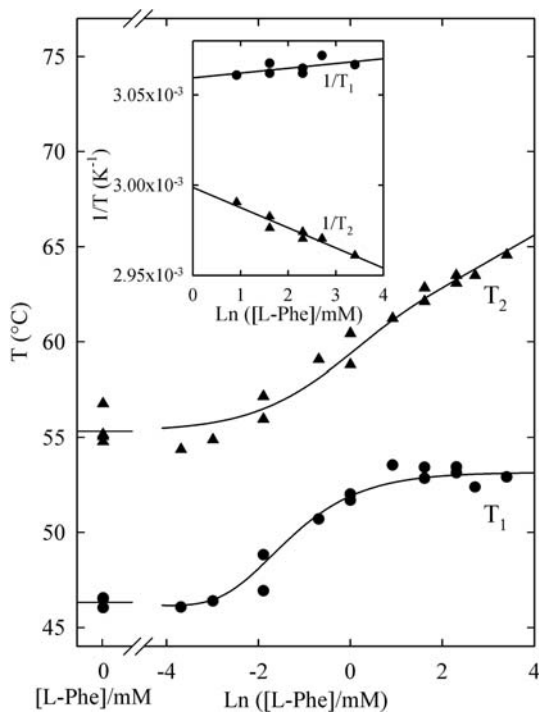


Fig. 2.11. Effect of L-Phe concentration on the transition temperatures for the two transitions observed in the DSC thermograms for the thermal denaturation of wild-type phenylalanine hydroxylase (see Fig. 2.9). Note that while the T_1 value shows saturation (plateau) at high ligand concentrations, the T_2 value does not. (Reprinted with permission from Thórólfsson et al., 2002, *Biochemistry* 41: 7573–7585, Copyright (2002), American Chemical Society)

2.8 Absolute Heat Capacities and the Residual Structure in the Unfolded State

A DSC calorimeter is a differential instrument and, as such, the measured signal is actually related to the difference in heat capacity between the sample cell (protein solution) and the reference cell (buffer). Such “apparent” heat capacity includes a contribution from water displacement by the protein in the sample cell and, in some cases, may even have a negative value (owing to the presence of protein, there is less water in the sample cell than in the reference cell and the heat capacity of water is larger than that of the protein). Correction for this water-displacement effect (and normalization to a mole of protein) yields the so-called absolute heat capacities, which are bound to be positive numbers (a negative value for an absolute heat capacity would be in conflict with the second law of thermodynamics). Possibly, the most accurate way to do such a calculation is to follow a suggestion by Kholodenko and Freire (1999) and perform a series of DSC experiments at different protein concentrations. Absolute heat capacities are then obtained from the concentration dependence of the apparent heat capacity values (Fig. 2.12). This procedure has the additional advantage of allowing estimates of the associated errors to be obtained.

The existence of residual structure in protein unfolded states has been recognized for many years. Furthermore, recent studies have emphasized that long-range interactions and cooperative phenomena can occur in denatured proteins (Shortle and Ackerman 2001; Baldwin 2002; Klein-Seetharaman et al. 2002; Guzman-Casa-

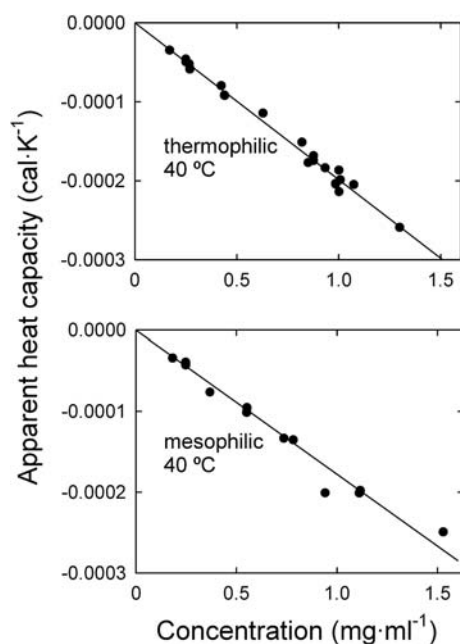


Fig. 2.12. Apparent heat capacity versus protein concentration for a thermophilic (*T. thermophilus*) and mesophilic (*E. coli*) ribonucleases H. Each data point comes from an independent DSC experiment. The continuous lines are the best fits of zero-intercept straight lines. Absolute heat capacity values and their associated errors can be calculated from the slopes of those lines. (Reprinted from *J Mol Biol*, Vol 329, Guzman-Casado et al., Energetic evidence for formation of a pH-dependent hydrophobic cluster in the denatured state of *Thermus thermophilus* ribonuclease H, 731–743, Copyright (2003), with permission from Elsevier)

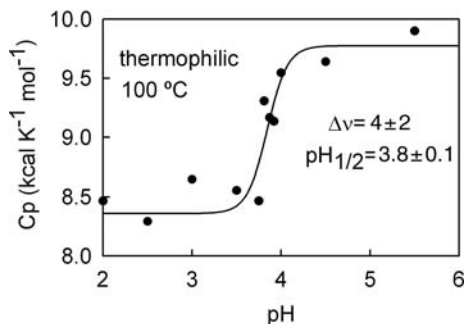


Fig. 2.13. The pH-dependence of the heat capacity of thermally unfolded ribonuclease H from *T. thermophilus*. Analysis of this dependence indicates a cooperative uptake of about four protons in the unfolded state. (Reprinted from *J Mol Biol*, Vol 329, Guzman-Casado et al., Energetic evidence for formation of a pH-dependent hydrophobic cluster in the denatured state of *Thermus thermophilus* ribonuclease H, 731–743, Copyright (2003), with permission from Elsevier)

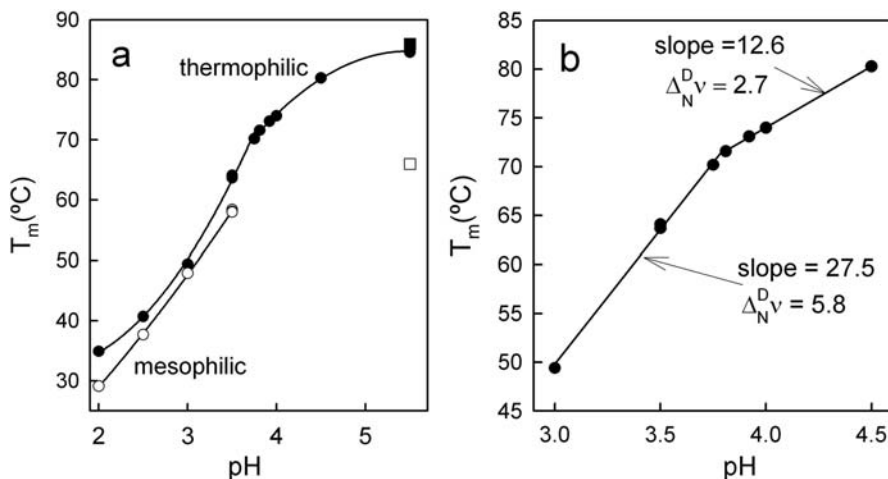


Fig. 2.14. a Effect of pH on the denaturation temperature for (*closed symbols*) thermophilic (*T. thermophilus*) and (*open symbols*) mesophilic (*E. coli*) ribonucleases H. **b** is a blowup of the data for the thermophilic protein. The break at pH 3.8 reflects the heat-capacity-detected change in the unfolded state of the thermophilic protein (see Fig. 2.13). (Reprinted from *J Mol Biol*, Vol 329, Guzman-Casado et al., Energetic evidence for formation of a pH-dependent hydrophobic cluster in the denatured state of *Thermus thermophilus* ribonuclease H, 731–743, Copyright (2003), with permission from Elsevier)

do et al. 2003). Absolute heat capacities determined by DSC provide an excellent approach to characterize such complex processes in unfolded states. In a recent study on a thermophilic ribonuclease H (Guzman-Casado et al. 2003), a sharp decrease in the denatured state heat capacity upon lowering of the pH was detected (Fig. 2.13).

This decrease was found to take place in a very narrow pH range, indicating a cooperative change in residual structure triggered by the protonation of several carboxylic acid residues. The change possibly involves formation of hydrophobic clusters, which may be favoured by the less hydrophilic character of protonated carboxylic acids. Significantly, this cooperative change in residual structure has direct consequences for protein stability, as shown by the break (sudden change in slope) that is seen in a plot of denaturation temperature versus pH (Fig. 2.14) at the same pH value of the heat-capacity-detected transition in the unfolded protein. In connection with these results, it is of interest that the value of the denaturation heat capacity change for the thermophilic ribonuclease H has been found to be smaller than that for the corresponding mesophilic protein. A lower denaturation heat capacity implies a flatter stability curve (a flatter unfolding free energy versus temperature profile) and allows for a high denaturation temperature without an excessively high unfolding free energy at physiological temperature (which might perhaps compromise the degree of flexibility required for function). Indeed, a role for unfolded-state residual structure in fine-tuning the stability of thermophilic proteins has been suggested (Robic et al. 2003).

2.9

Calorimetrically Determined Denaturant m Values

Thermodynamic stability of proteins is often characterized on the basis of chemical denaturation studies, in which unfolding is induced by urea or guanidine. In contrast with the DSC approach, the analysis of chemical denaturation is to a large extent empirical and based upon extrathermodynamic assumptions (for instance, assuming that the denaturant-concentration dependence of the unfolding free energy is linear outside the narrow transition region). It must be recognized, however, that one of the parameters derived from chemical denaturation profiles has found widespread application in protein folding studies: the slope of the plot of folding free energy versus denaturant concentration, referred to as the m value (Greene and Pace 1974), which has been found empirically to be related to the amount of protein surface exposed to the solvent upon denaturation (Myers et al. 1995).

Recent work (Ibarra-Molero et al. 2004; Perez-Jimenez et al. 2004) shows, nevertheless, that denaturant m values can also be calculated from DSC experiments in a simple and, to a large extent, model-independent manner. The calculation requires knowledge (derived from DSC experiments) of the effect of denaturant concentration on denaturation temperatures and enthalpies. It does not require, however, values of the denaturation heat capacity change.

One of the main advantages of the calorimetrically determined m values is that they are obtained as a function of denaturant concentration, thus providing detailed information about the denaturant-concentration dependence of the unfolding free energy. In particular, calorimetrically determined m values have been found to change abruptly at low denaturant concentrations when the denaturant is a salt (guanidinium chloride, guanidinium thiocyanate), reflecting very likely the salt-induced screening of interactions between the protein charged groups (Ibarra-Molero

et al. 2004; Perez-Jimenez et al. 2004) and providing, therefore, a convenient approach to probe those interactions.

Acknowledgements. Research in the authors' laboratory is supported by Spanish Ministry of Education and Science Grant BIO2003-02229 and Feder Funds.

References

- Baldwin RL (2002) Making a network of hydrophobic clusters. *Science* 295:487–489
- Brandts JF, Hu CQ, Lin LN (1989) A simple model for proteins with interacting domains. Applications to scanning calorimetry data. *Biochemistry* 28:8588–8596
- Conejero-Lara F, Mateo PL, Aviles FX, Sanchez-Ruiz JM (1991) Effect of Zn^{2+} on the thermal denaturation of carboxypeptidase B. *Biochemistry* 30:2067–2072
- Cooper A (1976) Thermodynamic fluctuations in protein molecules. *Proc Natl Acad Sci USA* 73:2740–2741
- Cooper A (2005) Heat capacity effects in protein folding and ligand binding: a re-evaluation of the role of water in biomolecular thermodynamics. *Biophys Chem* 115:89–97
- Cunningham EL, Jaswal SS, Sohk JL, Agard DA (1999) Kinetic stability as mechanism for protein longevity. *Proc Natl Acad Sci USA* 96:11008–11014
- Dragan AI, Privalov PL (2002) Unfolding of a leucine zipper is not a simple two-state transition. *J Mol Biol* 321:891–908
- Forrer P, Chang C, Ott D, Wlodawer A, Plückthun A (2004) Kinetic stability and crystal structure of the viral capsid protein SHP. *J Mol Biol* 344:179–193
- Foss TR, Kelker MS, Wiseman RL, Wilson IA, Kelly JW (2005) Kinetic stabilization of the native state by protein engineering: implications for inhibition of transthyretin amyloidogenesis. *J Mol Biol* 347:841–854
- Freire E (1994) Statistical thermodynamics analysis of differential scanning calorimetry data: structural deconvolution of heat capacity function of proteins. *Methods Enzymol* 240:502–530
- Freire E (1995) Differential scanning calorimetry. *Methods Mol Biol* 40:191–218
- Freire E, van Osdol WW, Mayorga OL, Sanchez-Ruiz JM (1990) Calorimetrically determined dynamics of complex unfolding transitions in proteins. *Annu Rev Biophys Biophys Chem* 19:159–188
- Freire E, Murphy KP, Sanchez-Ruiz JM, Galisteo ML, Privalov PL (1992) The molecular basis of cooperativity in protein folding. Thermodynamic dissection of interdomain interactions in phosphoglycerate kinase. *Biochemistry* 31:250–256
- Fukada H, Sturtevant JM, Quioco F (1983) Thermodynamics of the binding of L-arabinose and of D-galactose to the L-arabinose-binding protein of *Escherichia coli*. *J Biol Chem* 258:13193–13198
- Garcia-Mira MM, Sadqui M, Fischer N, Sanchez-Ruiz JM, Muñoz V (2002) Experimental identification of downhill folding. *Science* 298:2191–2195
- Gomez J, Hilser VJ, Xie D, Freire E (1995) The heat capacity of proteins. *Proteins* 22:404–412
- Greene RF, Pace CN (1974) Urea and guanidinium chloride denaturation of ribonuclease, α -chymotrypsin, and β -lactoglobulin. *J Biol Chem* 249:5388–5393

- Guzman-Casado M, Parody-Morreale A, Robic S, Marqusee S, Sanchez-Ruiz JM (2003) Energetic evidence for formation of a pH-dependent hydrophobic cluster in the denatured state of *Thermus thermophilus* ribonuclease H. *J Mol Biol* 329:731–743
- Ibarra-Molero B, Perez-Jimenez R, Godoy-Ruiz R, Sanchez-Ruiz JM (2004) Linkage between temperature and chemical denaturant effects on protein stability: the interpretation of calorimetrically determined m values. In: Doyle M, Ladbury J (eds) *Biocalorimetry II: applications of calorimetry in the biological sciences*. Wiley, New York, pp 203–214
- Irun MP, Garcia-Mira MM, Sanchez-Ruiz JM, Sancho J (2001) Native hydrogen bonds in a molten globule: the apoflavodoxin thermal intermediate. *J Mol Biol* 306:877–888
- Jackson SE (1998) How do small single-domain proteins fold? *Folding Des* 3:R81–R91
- Jackson WM, Brandts JF (1970) Thermodynamics of protein denaturation. A calorimetric study of reversible denaturation of chymotrypsinogen and conclusions regarding accuracy of two-state approximation. *Biochemistry* 9:2294–2301
- Jaswal SS, Sohl JL, Dans JH, Agard DA (2002) Energetic landscape of alpha-lytic protease optimizes longevity through kinetic stability. *Nature* 415:343–346
- Jaswal SS, Truhlar SM, Dill KA, Agard DA (2005) Comprehensive analysis of protein folding activation thermodynamics reveals a universal behaviour violated by kinetically stable proteins. *J Mol Biol* 347:355–366
- Jayaraman S, Gantz D, Gursky O (2005a) Structural basis for thermal stability of human low-density lipoprotein. *Biochemistry* 44:3965–3971
- Jayaraman S, Gantz D, Gursky O (2005b) Kinetic stabilization and function of apolipoprotein A-2-DMPC disks: comparison with apo A-1 and apo C-1. *Biophys J* 88:2907–2918
- Kaushik JK, Ogasahara K, Yutani K (2002) The unusually slow relaxation kinetics of the folding-unfolding of pyrrolydione carboxyl peptidase from a hyperthermophile, *pyrococcus furiosus*. *J Mol Biol* 316:991–1003
- Kaya H, Chan HS (2000) Polymer principles of protein calorimetric two-state cooperativity. *Proteins* 40:637–661
- Kholodenko V, Freire E (1999) A simple method to measure the absolute heat capacity of proteins. *Anal Biochem* 270:336–338
- Klein-Seetharaman J, Oikawa M, Grimshaw SB, Wimer J, Duchart E, Ueada T et al (2002) Long-range interactions in a non-native protein. *Science* 295:1719–1722
- Luque I, Freire E (1998) Structure-based prediction of binding affinities and molecular design of peptide ligands. *Methods Enzymol* 295:100–127
- Luque I, Leavitt SA, Freire E (2002) The linkage between protein folding and functional cooperativity: two sides of the same coin? *Annu Rev Biophys Biomol Struct* 31:235–256
- Lynch SM, Boswell SA, Colon W (2004) Kinetic stability of Cu/Zn superoxide dismutase is dependent on its metal ligands: implications for ALS. *Biochemistry* 43:16525–16531
- Makhatadze GI (1998) *Measuring protein thermostability by differential scanning calorimetry. Current protocols in protein science* 7.9.1–7.9.14. Wiley, New York
- Makhatadze GI, Privalov PL (1995) Energetics of protein structure. *Adv Protein Chem* 47:307–425
- Manly SP, Mathews KS, Sturtevant JM (1985) Thermal denaturation of the core protein of lac repressor. *Biochemistry* 24:3842–3846
- Manning M, Colon W (2004) Structural basis of protein kinetic stability: resistance to sodium dodecyl sulphate suggests a central role for rigidity and a bias toward beta-sheet structure. *Biochemistry* 43:11248–11254

- Mehta R, Gantz DL, Guskys O (2003) Human plasma high-density lipoproteins are stabilized by kinetic factors. *J Mol Biol* 328:183–192
- Muñoz V, Sanchez-Ruiz JM (2004) Exploring protein-folding ensembles: a variable-barrier model for the analysis of equilibrium unfolding experiments. *Proc Natl Acad Sci USA* 101:17646–17651
- Myers JK, Pace CN, Scholtz JM (1995) Denaturant *m* values and heat capacity changes: relation to changes in accessible surface areas of protein unfolding. *Protein Sci* 4:2138–2148
- Naganathan AN, Perez-Jimenez R, Sanchez-Ruiz JM, Muñoz V (2005) Robustness of downhill folding: guidelines for the analysis of equilibrium folding experiments on small proteins. *Biochemistry* 44:7435–7449
- Oliva FY, Muñoz V (2004) A simple thermodynamic test to discriminate between two-state and downhill folding. *J Am Chem Soc* 126:8596–8597
- Perez-Jimenez R, Godoy-Ruiz R, Ibarra-Molero B, Sanchez-Ruiz JM (2004) The efficiency of different salts to screen charge interactions in proteins: a Hofmeister effect? *Biophys J* 86:2414–2429
- Persikov AV, Brodsky B (2000) Unstable molecules from stable tissues. *Proc Natl Acad Sci USA* 99:1101–1103
- Petrassi MM, Johnson SM, Purkey HE, Chiang KP, Walkup T, Jinag X, Powers ET, Kelly JW (2005) Potent and selective structure-based dibenzofuran inhibitors of transthyretin amyloidogenesis: kinetic stabilization of the native state. *J Am Chem Soc* 127:6662–6671
- Plaza del Pino IM, Pace CN, Freire E (1992) Temperature and guanidinium chloride dependence of the structural stability of ribonuclease T1. *Biochemistry* 31:11196–11202
- Plaza del Pino IM, Ibarra-Molero B, Sanchez-Ruiz JM (2000) Lower kinetic limit to protein thermal stability: a proposal regarding protein stability in vivo and its relation with misfolding diseases. *Proteins* 40:58–70
- Plotnikov VV, Brandts JM, Lin LN, Brandts JF (1997) A new ultrasensitive scanning calorimeter. *Anal Biochem* 250:237–244
- Privalov PL (1979) Stability of proteins: Small globular proteins. *Adv Protein Chem* 33:167–241
- Privalov PL (1980) Scanning microcalorimeters for studying macromolecules. *Pure Appl Chem* 52:479–497
- Privalov PL (1982) Stability of proteins. Proteins which do not present a single cooperative system. *Adv Protein Chem* 35:1–104
- Privalov PL (1989) Thermodynamic problems of protein structure. *Annu Rev Biophys Biophys Chem* 18:47–69
- Ramsay G, Freire E (1989) Linked thermal and solute perturbation analysis of cooperative domain interactions in proteins. Structural stability of diphtheria toxin. *Biochemistry* 29:8677–8683
- Robertson A, Murphy KP (1997) Protein structure and the energetics of protein stability. *Chem Rev* 97:1251–1268
- Robic S, Guzman-Casado M, Sanchez-Ruiz JM, Marqusee S (2003) Role of residual structure in the unfolded state of a thermophilic protein. *Proc Natl Acad Sci USA* 100:11345–11349
- Rosgen J, Hinz HJ (2001) Folding energetics of ligand binding proteins. I. Theoretical model. *J Mol Biol* 306:825–835
- Rossgenarth A, Rosgen J, Hinz HJ, Gerke V (2001) Folding energetics of ligand binding proteins. II. Cooperative binding of Ca²⁺ to annexin. *J Mol Biol* 306:825–835
- Sanchez-Ruiz JM (1992) A theoretical analysis of Lumry–Eyring models in differential scanning calorimetry. *Biophys J* 61:921–935

- Sanchez-Ruiz JM (1995) Differential scanning calorimetry of proteins. *Subcell Biochem* 24:133–176
- Sanchez-Ruiz JM, Mateo PL (1987) Differential scanning calorimetry of membrane proteins. *Revis Biol Celular* 11:15–45
- Schnyrov VL, Zhadan GG, Cobaleda C, Muñoz-Barroso I, Villar E (1997) A differential scanning calorimetric study of Newcastle disease virus: identification of proteins involved in thermal transitions. *Arch Biochem Biophys* 341:89–97
- Shortle D, Ackerman MS (2001) Persistence of native-like topology in a denatured protein in 8 M urea. *Science* 293:487–489
- Thórólfsson M, Ibarra-Molero B, Fojan P, Petersen SB, Sanchez-Ruiz JM, Martínez A (2002) L-Phenylalanine binding and domain organization in human phenylalanine hydroxylase: a differential scanning calorimetry study. *Biochemistry* 41:7573–7585
- Vogl T, Jattzke C, Hinz HJ, Benz J, Huber R (1997) Thermodynamic stability of annexin VE17G: equilibrium parameters from an irreversible unfolding transition. *Biochemistry* 36:1657–1668
- Wiseman RL, Johnson SM, Kelker MS, Foss T, Wilson IA, Kelly JW (2005) Kinetic stabilization of an oligomeric protein by a single ligand event. *J Am Chem Soc* 127:5540–5541
- Zhou YQ, Hall CK, Karplus M (1999) The calorimetric criterion for a two-state process revisited. *Protein Sci* 8:1064–1074

IR Reflectance – Absorbance Studies of Peptide Structure, Orientation, and Conformational Flexibility in Langmuir Films: Relevance for Models of Pulmonary Surfactant Action

CAROL R. FLACH, PENG CAI, RICHARD MENDELSON

3.1 Introduction

Lipid/protein monolayers at the air/water interface (Langmuir films) are widely used to mimic biological interfaces such as half a lipid bilayer and the air/alveolar lining of the mammalian lung. The advantages of monolayers as models for these interfaces arise primarily from the ease with which experimental variables such as lateral pressure, surface area, and domain structure may be controlled and compared with those of bulk phases or with films prepared on solid substrates.

Until some 20 years ago, although a variety of physical techniques had been used to study monolayer thermodynamics and rheology, only limited information about the molecular structure and interactions of monolayer constituents could be acquired. At that time, Dluhy et al. (Dluhy 1986; Dluhy and Cornell 1985; Dluhy and Mendelsohn 1988; Mitchell and Dluhy 1988) demonstrated the feasibility of acquiring IR spectra from Langmuir films of fatty acids and phospholipids in situ at the air/water interface. The technique used was external IR reflection–absorption spectroscopy (IRRAS), in which an IR beam impinges on the aqueous surface and is partially reflected to a detector. The reflected radiation contains the spectrum of the monolayer constituents. A variety of studies of lipid chain conformation and phase behavior ensued. The second generation of IRRAS applications involved determination of peptide or protein secondary structure in Langmuir films (Cornut et al. 1996; Flach et al. 1994; Pastrana-Rios et al. 1994). Technically, these experiments required the development of methods for eliminating interference from water vapor absorption (Buffeteau et al. 2000; Flach et al. 1994).

The basic idea underlying IRRAS is straightforward. When IR radiation illuminates an aqueous film, a small fraction of the incident light is reflected. The reflected intensity depends on the experimental geometry (angle of incidence and state of polarization of the radiation), the optical constants (real and imaginary parts of the refractive indices) of the film and subphase, and the orientation of the transition moments. The sample vibrational modes (i.e., the IR spectrum of the film constituents) appear in the spectrum of reflected light. IRRAS spectra are presented as plots of reflectance–absorbance (RA) versus wavenumber (per centimeter). RA is defined as $-\log_{10}(R/R_0)$, where R is the reflectivity of the film covered surface and R_0 is the reflectivity of the subphase. The subphase is generally D_2O for proteins, since the bending mode of H_2O near $1,640\text{ cm}^{-1}$ overlaps the conformation-sensitive amide I mode of proteins ($1,610\text{--}1,690\text{ cm}^{-1}$). In normal transmission IR spectroscopy, absorbance values are positive numbers. In contrast, RA's may take on positive or

negative values depending on the experimental parameters. Previous publications provide more details regarding early advances in IRRAS (Mendelsohn et al. 1995; Mendelsohn and Flach 2002a, b).

3.2 Instrumentation

A schematic of the IRRAS instrument currently in use at Rutgers University is shown in Fig. 3.1. While details of design vary among the dozen or so groups around the world utilizing this technique, a minimal set of component requirements is represented in the figure. Thus, a Langmuir trough for containing surface films, a Wilhelmy plate or equivalent device for surface-pressure measurements, and an IR polarizer are all required in addition to a high-end IR instrument. Other useful ancillary equipment includes devices for control of the relative humidity of the environment and the temperature of the subphase. The Langmuir trough must be sufficiently large so as to provide a sampling area lacking a meniscus. For Teflon troughs, this linear dimension is at least 6 cm.

3.3 Information from IRRAS Measurements

IRRAS measurements provide the two types of information traditionally acquired from molecular IR absorption spectroscopy, namely, the frequencies of molecular vibrations and band intensities. There are several levels of sophistication for interpretation of vibrational frequencies. For small molecules, IR spectral analysis is based on an assumed molecular geometry. Spectra are interpreted as completely as possible using symmetry considerations, spectroscopic selection rules, and additional theoretical arguments. Each observed molecular vibration is a normal mode, i.e., a linear combination of the internal motions of small groups of atoms (internal coordinates). Since many thousands of small molecules have been analyzed using the aforementioned methods, a large database is available to provide a basis for understanding spectra of large molecules. For peptides, proteins, and lipid assemblies, the “normal coordinate problem” is substantially underdetermined, even with extensive isotopic labeling. That is, there are many fewer observed vibrational frequencies than there are force constants to describe them. Less rigorous, more empirical levels of analysis are therefore required, based on the assumption that spectra–structure correlations established for small molecules may be directly transferred to lipid assemblies or proteins. The added notion that must be considered is the occurrence of vibrational coupling between (nearly) identical chemical units in regular structures such as polymers or proteins, as discussed by Zbinden (1964), Krimm and Abe (1972), Krimm and Bandekar (1986), and Brauner et al. (2005). For example, the peptide bond amide I (mostly C=O stretch) vibrations of an antiparallel β -sheet structure couple due to through-space and through-bond interactions (Miyazawa and Blout 1961; Brauner et al. 2000, 2005) so that an initially isolated single amide

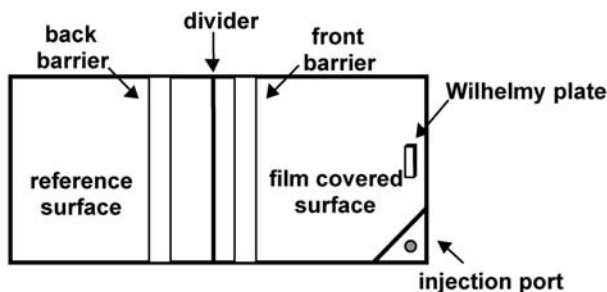
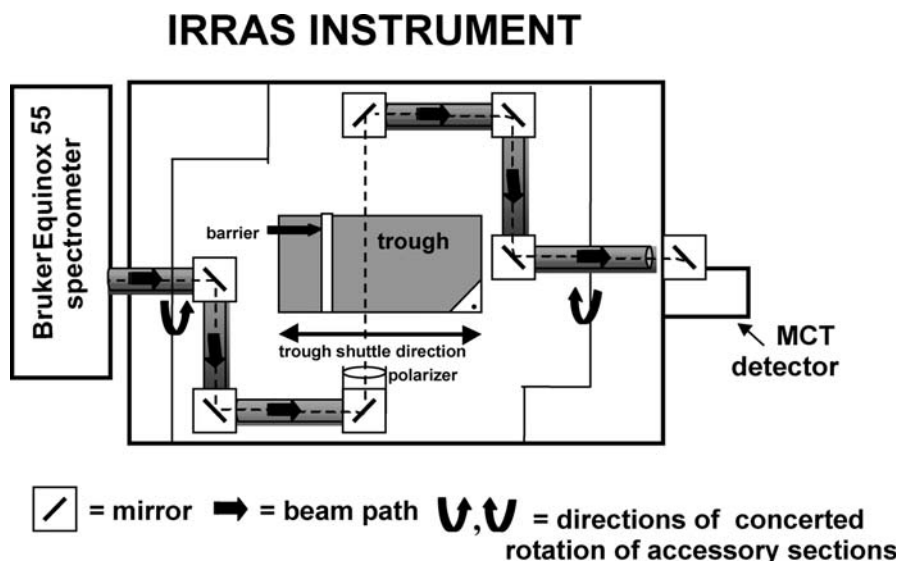


Fig. 3.1. Block diagram of the IR reflection-absorption spectroscopy (IRRAS) accessory supplied by Bruker Instruments and modified for current use at Rutgers University. The IR beam exits the spectrometer, reflects off three mirrors, passes through a polarizer, and impinges on the aqueous surface. The reflected radiation follows an equivalent path to a final mirror which directs the radiation to the mercury cadmium telluride (MCT) detector. The angle of incidence is controlled by concerted rotation of the incident and collection mirror sections. Details of the Langmuir trough are given at the *bottom*. The front barrier divides the trough into a film-covered surface and a reference (aqueous) surface, while the back barrier sits on the divider in experiments where proteins or ions are injected into the subphase. The back barrier/divider combination limits the diffusion of proteins from the subphase under the film to the reference subphase. Interferograms are collected from the reference surface and the entire trough is shuttled so that the film-covered surface can then be sampled by the IR radiation

I mode (near $1,650\text{ cm}^{-1}$) splits and results in an IR spectrum with a strong peak at about $1,625\text{--}1,635\text{ cm}^{-1}$ and a weaker peak at about $1,690\text{ cm}^{-1}$.

Due primarily to anomalous dispersion effects in the real part of the refractive index, the absorption frequencies measured from IRRAS data collected at angles of incidence close to the Brewster angle may differ slightly from those observed in transmission studies of the same material. These effects are usually small (less than $1\text{--}2\text{ cm}^{-1}$). In addition, for angles of incidence within about 5° of the Brewster angle (about 52° for water) the band shapes for spectra acquired with p-polarized radiation may be dramatically distorted by the appearance of derivative-like features.

For both lipids and proteins, some vibrational modes are quite sensitive to molecular conformation. This sensitivity provides unique information about monolayer structure and is the primary *raison d'être* for the IRRAS experiment. As an example, the symmetric and asymmetric methylene stretching vibrations from phospholipid or fatty acid acyl chains may be confidently assumed to occur at about $2,850$ and $2,920\text{ cm}^{-1}$, respectively. Typical IRRAS data for the methylene stretching region of 1,2-dipalmitoylphosphatidylcholine (DPPC) as a function of surface pressure are shown in Fig. 3.2a. The observed signal-to-noise ratios of 300:1 are easily achievable in a couple of minutes of scanning. The band intensities increase with increasing surface pressure (Π), mostly due to greater numbers of molecules in the IR beam. Generally, the surface pressure decreases slightly (less than $1\text{--}4\text{ mN m}^{-1}$) during the course of a measurement. The range of Π for each of the spectra shown in Fig. 3.2a is indicated. When lipid acyl chains undergo conformational disordering, as induced, for example, by increasing temperatures or decreasing surface pressures, these frequencies increase by $2\text{--}6\text{ cm}^{-1}$. Such a variation is evident in the spectra shown in Fig. 3.2a. Although the variation is small, the precision of the frequency measurement (0.1 cm^{-1}) is such that the change in frequency is a useful measure of lipid chain conformational order, and provides structural information not easily measured otherwise in monolayers.

The details of the Π -induced phase transition are shown in Fig. 3.2b, c. The Π -A isotherm in Fig. 3.2b reveals the liquid-expanded to liquid-condensed (LE \rightarrow LC) transition as a near-horizontal plateau at around 8 mN m^{-1} . This isotherm was generated during the IRRAS experiment reported in Fig. 3.2a, and the slight drops in Π when the barrier is stopped during IRRAS data acquisition are evident in the

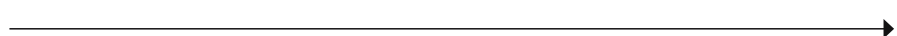
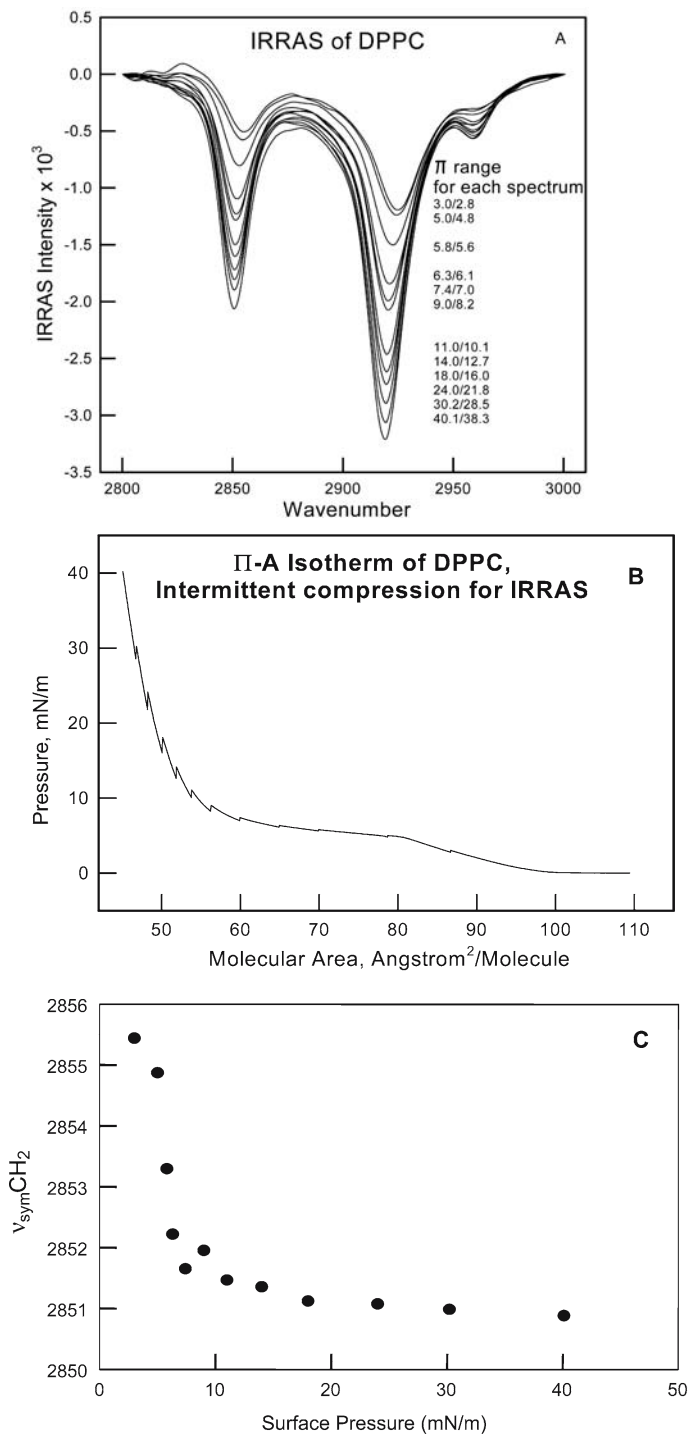


Fig. 3.2. A IRRAS spectra of the C–H stretching region of a 1,2-dipalmitoylphosphatidylcholine (DPPC) monolayer collected at various surface pressures. The pressure range (initial and final pressures) during spectral acquisition is marked. Slight frequency decreases in both the symmetric (around $2,850\text{ cm}^{-1}$) and asymmetric (around $2,920\text{ cm}^{-1}$) methylene stretching modes are evident as the surface pressure is increased, indicative of chain conformational ordering. **B** The surface pressure vs. area isotherm acquired during the experiment whose results are presented in a. The discontinuities in the curve reflect the pressure drops observed while the barrier motion was arrested for IRRAS data collection. The liquid-expanded to liquid-condensed transition is clearly observable as a near plateau in the isotherm at a surface pressure of around 8 mN m^{-1} . **C** The frequency of the CH_2 symmetric stretching mode of DPPC as a function of surface pressure. The liquid-expanded to liquid-condensed transition is marked by a frequency decrease of around 3.5 cm^{-1}



isotherm. Changes in the frequency of the CH_2 symmetric stretching mode during a compression curve are shown in Fig. 3.2c. The mode exhibits a significant decrease from around 2,855 to around 2,851.5 cm^{-1} during the (LE \rightarrow LC) transition, revealing a conformational ordering of the DPPC acyl chains.

The secondary structure information available from the amide I vibration of proteins is revealed in Fig. 3.3. α -helical proteins absorb strongly at 1,650–1,655 cm^{-1}

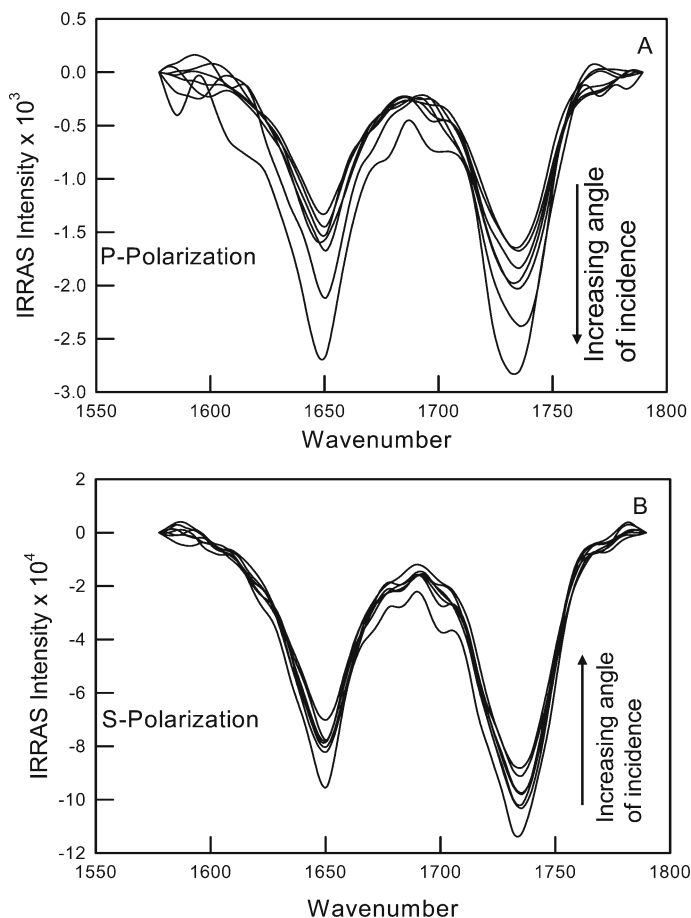


Fig. 3.3. IRRAS spectra of the pulmonary surfactant protein SP-C in a film with DPPC/1,2-dipalmitoylphosphatidylglycerol (DPPG)/cholesterol. The concentration of SP-C by weight with respect to total lipid is 4 mol%, while the molar ratio of DPPC to DPPG is 4/1 and the molar ratio of cholesterol with respect to the phospholipids is 7%. The angles of incidence range from 34 to 46° in increments of 2°. The peptide bond amide I vibration of this α -helical protein appears near 1,650 cm^{-1} , while the phospholipid C=O stretching modes appear near 1,730 cm^{-1} . The surface pressure was 25 mN m^{-1} . **A** Spectra acquired with p-polarized radiation, increasing incident angle from *top* to *bottom*. **B** Spectra acquired with s-polarized radiation, increasing incident angle from *bottom* to *top*

while antiparallel β -sheets exhibit both a strong band at 1,625–1,635 cm^{-1} and a weak feature near 1,690 cm^{-1} . A series of spectra of the hydrophobic lung surfactant protein SP-C in a mixed monolayer with DPPC/DPPG/cholesterol and acquired with s- and p-polarized radiation, are shown in Fig. 3.3a and b, respectively. The strong feature near 1,650 cm^{-1} reveals the prevalence of the α -helical conformation. The peak near 1,730 cm^{-1} arises from the ester carbonyls of DPPC.

The amide I region of the spectra of Val-Glu-Val-Orn-Val-Glu-Val-Orn-Val-Glu-Val-Orn-Val-OH (Orn stands for *N* δ -trifluoroacetyl-l-ornithine), a peptide designed as an amphipathic β -sheet, is shown in Fig. 3.4 for s- and p-polarization at angles of incidence both above and below the Brewster angle. The amide I mode appears split, with a strong band near 1,620 cm^{-1} , and a weaker feature (possibly a doublet) near 1,680–1,690 cm^{-1} . The sensitivity to peptide secondary structure of the mode is thus revealed in Figs. 3.3 and 3.4.

The second type of information provided by IRRAS measurements is derived from band intensities. By consideration of the reflection properties of light from a three-layer (air/film/water) interface, it is feasible to determine the angle of orientation of transition moments for particular vibrational modes with respect to the surface normal. The various theoretical approaches have been reviewed elsewhere (Mendelsohn et al. 1995). We utilize the equations of Kuzmin et al. (Kuzmin and Michailov 1981; Kuzmin et al. 1992). For calculations of lipid acyl chain orientation in lipid or lipid/protein films, the average tilt angles obtained from IRRAS agree closely with those deduced from X-ray reflection.

Models based on uniaxial orientation distributions, which provide for free rotation of the molecule about the molecular axis and the normal to the surface (*Z*-axis), also successfully predicted the tilt of α -helical proteins. A typical example is shown in Fig. 3.5, based on the IRRAS data for SP-C plotted in Fig. 3.3a and b. The method of analysis currently used plots the ratio of the intensity of the amide I band acquired with p-polarized radiation to that acquired with s-polarized radiation as a function of the angle of incidence. The experimental data (plotted as circles) are compared with calculated tilt angles of the helix axis with respect to the surface normal. The overall sensitivity of the measured intensity ratio with respect to tilt is shown in the inset. The experimental precision currently available suggests that the tilt angle of SP-C is around 75° with an uncertainty of $\pm 5^\circ$. The peptide is thus determined to be lying nearly parallel to the aqueous surface. Simple geometric considerations show that the hydrophobic thickness of the tilted helix perpendicular to the surface matches that of the acyl chain thickness in the DPPC monolayer into which the protein has been inserted.

In the simulations just described uniaxial symmetry about the *Z*-axis (perpendicular to the subphase surface) and the molecular axis is assumed. A recent study from this laboratory (Brauner et al. 2003) of behenic acid methyl ester, which took advantage of the high quality of spectra currently available, shows that this assumption may not be exactly correct for all monolayers. The analysis indicated that the plane of the zigzag carbon backbone has a slight preferential alignment perpendicular to the direction of compression.

The extension of theoretical models and simulations to determine β -sheet orientation in peptides required additional geometric variables, φ and χ , which measure the degree of strand rotation about the *Z*-axis and the molecular axis, respectively

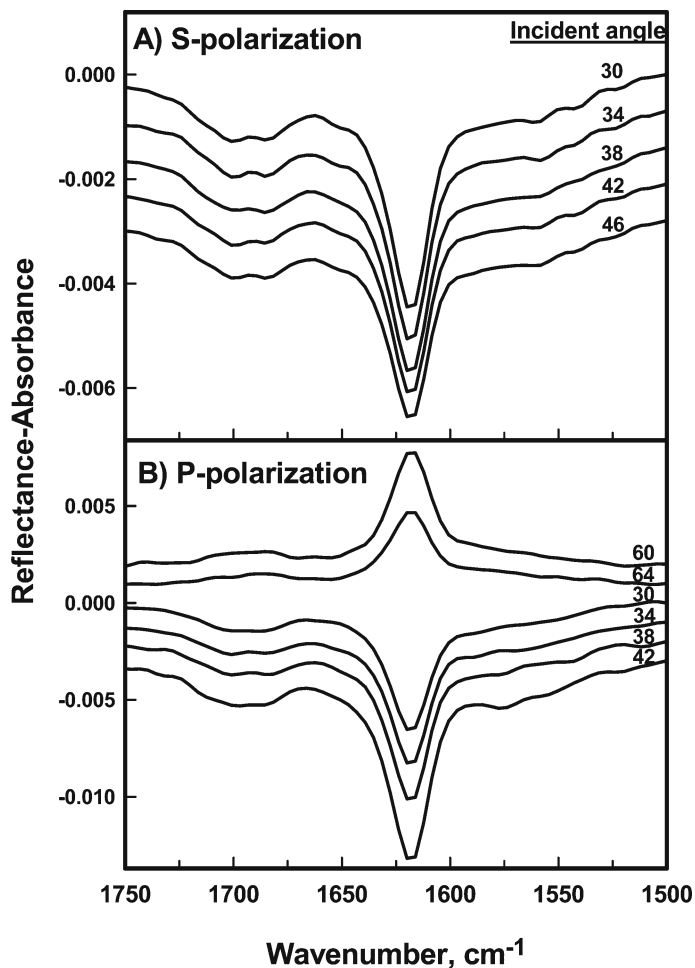


Fig. 3.4. The amide I region of the spectra of Val-Glu-Val-Orn-Val-Glu-Val-Orn-Val-Glu-Val-Orn-Val-OH (*Orn* stands for N^{δ} -trifluoroacetyl-L-ornithine), a peptide designed as an amphiphilic β -sheet, acquired at $II=3 \text{ mN m}^{-1}$. *A* Spectra acquired under s-polarization at the indicated angles of incidence. *B* Spectra acquired under p-polarization at the indicated angles of incidence. Note the reversal in the sign of the IRRAS signal, which occurs for this polarization at angles of incidence greater than the Brewster angle. The splitting of the amide I mode is discussed in the text.

(Xu et al. 2004). Details of the coordinate system used in the current studies are shown in Fig. 6a. IRRAS spectra for the peptide selected to test this model, Val-Glu-Val-Orn-Val-Glu-Val-Orn-Val-Glu-Val-Orn-Val-OH, are shown in Fig. 3.4. A priori, the most likely orientation of the peptide is a model in which the sheet lies flat on the aqueous surface, i.e., both $\theta=90^\circ$ and $\chi=90^\circ$. It was found that the $1,620 \text{ cm}^{-1}$ component of the amide I contour oscillates completely in the X-Y plane ($\theta=90^\circ$

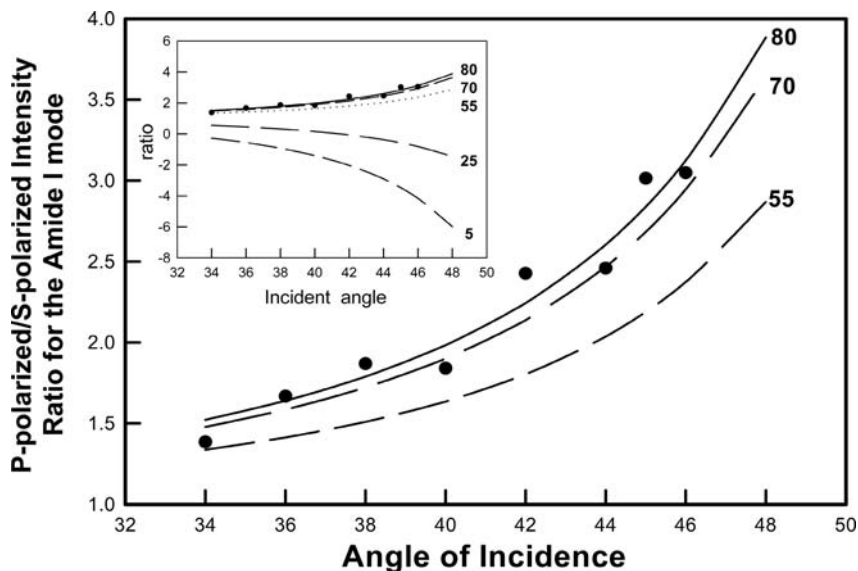


Fig. 3.5. Determination of the orientation of the helical pulmonary surfactant protein SP-C from IRRAS intensities. The observed experimental data points of the ratio of intensity of the amide I band acquired with p-polarized radiation to that acquired with s-polarized radiation (p/s intensity ratio) for the amide I mode derived from the IRRAS spectra in Fig. 3.3a and b, are plotted as circles. Theoretical curves corresponding to various assumed tilt angles are plotted as a function of the tilt angle as shown. The *inset* shows the same data but for a wider range of assumed tilt angles, and reveals that the sensitivity of the measured ratio to the tilt angle is a nonlinear function of the tilt angle. The greatest sensitivity occurs at low tilt angles, i.e., with the helix nearly parallel to the surface normal. See text for details.

and $\chi=90^\circ$), consistent with this model. Comparison of experimental data with simulations (Fig. 3.6b) at two surface pressure values also reveals a slight deviation from uniaxial symmetry. The best fit for the experimental data is found for $\varphi=43.5^\circ$ at $\Pi=3 \text{ mN m}^{-1}$ and $\varphi=42^\circ$ at $\Pi=18 \text{ mN m}^{-1}$, not $\varphi=45^\circ$, which is the equivalent of uniaxial symmetry.

3.4 Pulmonary Surfactant

3.4.1 Biochemistry and Models for Function

The biophysical utility of IRRAS measurements will be demonstrated in the current chapter through studies of the structure, orientation, and conformational flexibility of peptides important for understanding the pulmonary surfactant system. Some background is provided in the following for reference.

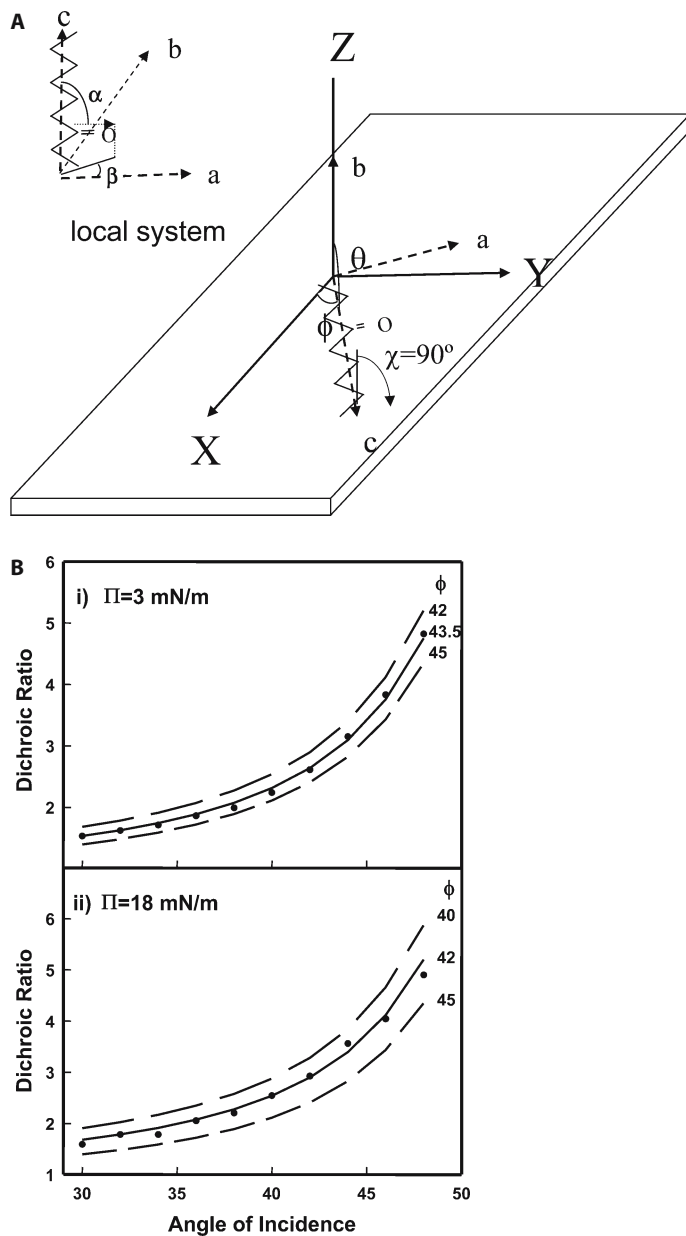


Fig. 3.6. A Coordinate systems for the IRRAS simulations. *Top left:* A local right-handed Cartesian coordinate system abc is attached to the β -strand such that the c -axis is in the direction of the strand, the a -axis is in the plane of the sheet perpendicular to the c -axis, and the b -axis is perpendicular to both the a -axis and the c -axis. The orientation of the transition moment vector, associated with the amide I band in the local system is described by a polar angle α and

Pulmonary surfactant is a mixture of lipids and proteins that forms a monolayer film at the air/alveolar interface in the terminal airway of the mammalian lung. The major function of surfactant is to lower surface tension, thereby reducing the work required to increase lung volume and preventing alveolar collapse. The pathological consequences of a deficiency in surfactant are severe. Respiratory distress syndrome (RDS) in premature infants and adult RDS (ARDS) are two common conditions.

The biochemical composition of pulmonary surfactant consists mainly of lipids (90% by weight) and proteins (10% by weight). The lipid fraction is predominantly DPPC and phosphatidylglycerols (PGs) along with a significant proportion of unsaturated phosphatidylcholines (PCs) and PGs, other phospholipid classes, and cholesterol. There are four surfactant-associated proteins, SP-A, SP-B, SP-C, and SP-D, named in order of their discovery, of which two, SP-B and SP-C, are very hydrophobic and are implicated in controlling the surface properties of surfactant. In particular these proteins have been shown to facilitate in vitro adsorption and spreading of lipids across air/water interfaces (Johansson and Curstedt 1997). In vivo, the physiological importance of SP-B is demonstrated by the observation that neonatal SP-B-deficient mice die at birth from respiratory failure (Lin et al. 1999).

Biophysical studies have tended to focus on two issues, namely (1) the molecular mechanism by which surfactant lipids and proteins interact to facilitate spreading across the interface and (2) the rational design of therapeutic agents. The former is a nontrivial matter as surfactant in vivo must possess two apparently contradictory attributes. It must be able to form stable films at the high surface pressures (around 70 mN m^{-1}) that form upon exhalation; and it must also be able to spread sufficiently rapidly across the air/alveolar interface to keep up with breathing rates. DPPC is the only major surfactant constituent that can form stable monolayers at 70 mN m^{-1} under compressive forces; however, it spreads too slowly during surface area expansion to be effective in vivo. To address this issue, Goerke and Clements (1986) proposed that for films to be stable, a process must occur during which the surface concentration of DPPC was enhanced at high pressures. The process, not defined in molecular structure terms, was labeled “squeeze-out.”

←

an azimuthal angle β . α is the angle between the transition moment and the c -axis known to be 90° and β is the angle of the projection of the transition moment in the ab plane from the a -axis set to 0° . *Center:* A laboratory Cartesian coordinate system is selected such that the positive Z -axis is along the surface normal, the X -axis is in the direction of the barrier movement in the Langmuir trough, and the YZ plane is the plane of incidence. The orientation of the local system in the laboratory frame is described by the Euler angles, θ , φ , and χ . θ and φ are the polar coordinates of the c -axis in the XYZ system. θ is the angle from OZ to Oc , where O is the origin, φ is the angle in the XY plane from OX to the projection of Oc on the XY plane, and χ is an angle in the ab plane measuring the rotation clockwise about the c -axis (from the Z -axis). **B** Comparison of the experimental (*circles*) and theoretical values of the IRRAS p/s intensity ratio for various assumed values of φ , at a monolayer surface pressure of 3 mN m^{-1} and a monolayer surface pressure of 18 mN m^{-1} . In this instance the best values of φ are around 43.5° and 42° , for surface pressures of 3 and 18 mN m^{-1} , respectively. Values less than 45° indicate that the strand has a slight preferential alignment parallel to the compression direction

3.4.2 Application of IRRAS to the Pulmonary Surfactant

3.4.2.1 Squeeze-Out

An early triumph of IRRAS occurred about a decade ago, at which time the phenomenon of enrichment of mixed monolayers in a particular component at high pressures was observed. Binary lipid mixtures such as DPPC/1-palmitoyl-2-oleoylphosphatidylglycerol (POPG) or DPPC/1,2-dioleoylphosphatidylglycerol (DOPG) (Pastrana-Rios et al. 1994) in which one component (the PG) was more structurally disordered and thus more likely to be unstable in monolayers at high surface pressures were studied. Typical data are shown in Fig. 3.7 for DPPC- d_{62} /DOPG. Acyl chain perdeutereration of the DPPC enabled the separate detection of the meth-

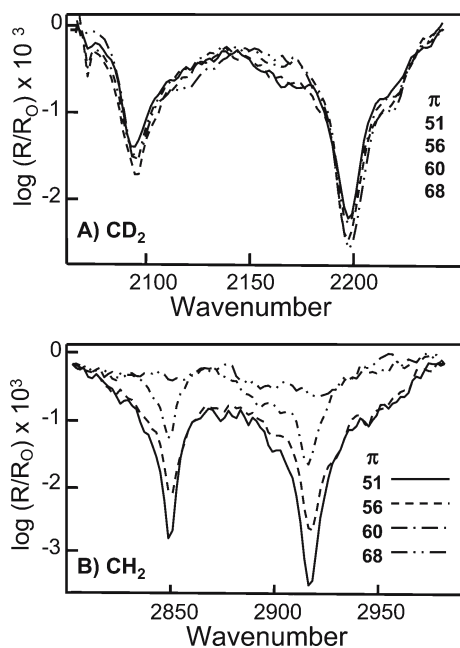


Fig. 3.7. IRRAS spectra for binary lipid mixtures of DPPC- d_{62} with 1,2-dioleoylphosphatidylglycerol. Acyl chain perdeutereration of the DPPC enables the separate detection of the methylene vibrations of each lipid species. Spectra are shown as a function of surface pressure from 51 to 68 mN m^{-1} . **A** The symmetric and asymmetric CD₂ stretching modes from the phosphatidylcholine component near 2,090 and 2,190 cm^{-1} remain relatively constant as surface pressure is increased. **B** The symmetric and asymmetric CH₂ stretching modes from the phosphatidylglycerol component near 2,850 and 2,920 cm^{-1} show progressively diminishing intensity with increased surface pressure, to the point where signals are barely detectable at 68 mN m^{-1} .

ylene vibrations of each lipid species. As the surface pressure was increased from 51 to 68 mN m^{-1} , the intensities of the symmetric and asymmetric CD_2 stretching modes from the PC component near 2,090 and 2,190 cm^{-1} remained more or less constant (Fig. 3.7a). In contrast, the intensities of the symmetric and asymmetric CH_2 stretching modes from the PG component near 2,850 and 2,920 cm^{-1} progressively diminished with increased surface pressure, to the point where signals were barely detectable at 68 mN m^{-1} (Fig. 3.7b). These data provided strong evidence for surface-pressure-induced enrichment of the monolayer in the PC component. Subsequent experiments (Pastrana-Rios et al. 1995) evaluated the extent of squeeze-out of SP-B and SP-C from the surface in lipid/protein films.

A deficiency in the squeeze-out model of surfactant action was that the mechanism of surfactant respreading was not addressed. Recent data, both from intact lung and from transferred films, have suggested that surface-associated multilayers form reversibly at high surface pressures during contraction–expansion cycles (von Nahmen et al. 1997; Schürch et al. 1998; Takamoto et al. 2001). These experiments suggest that the excluded material remains near the interface at high pressures, leading to the interesting proposal that reversible, protein-facilitated, 2D-to-3D transformations of the surface-active material take place during compression–expansion cycles. The 3D structures form reservoirs from which respreading occurs during successive cycles. In a very recent study from this laboratory (Wang et al. 2005), we evaluated the seminal role of the pulmonary surfactant protein SP-C in the establishment of multilayers at high surface pressures. A combination of atomic force microscopy and IRRAS provided evidence that SP-C orientation is changed during multilayer formation.

3.4.2.2

Structure, Orientation, and Possible Mechanism of Action of Therapeutic Agents for RDS

Other applications of IRRAS follow from the expectation that understanding the molecular mechanism of surfactant function will improve rational design of therapeutic agents for respiratory distress syndrome. Johansson et al. (2001) recently reviewed the use of artificial surfactants based on analogues of SP-B and SP-C. For example, peptides (with at least 17 amino acids) corresponding to the C-terminal part of SP-B accelerate surfactant spreading and improve lung compliance in fetal rabbits. Also the N-terminal peptide 1-25 tends to mimic most of the *in vitro* properties of native SP-B. The most studied protein analogue is termed KL_4 (sequence KLLLLKLLLLKLLLLKLLLLK), whose sequence is based on the hydrophobic/hydrophilic domains of SP-B (Cochrane and Revak 1991). The therapeutic form, termed KL_4 -surfactant, is a mixture of this peptide with DPPC, POPG, and phosphatidic acid. This surfactant replacement preparation has performed well in clinical trials although the mechanism of its action and its biophysical relationship to SP-B and SP-C are unclear (Cochrane et al. 1996, 1998).

The utility of IRRAS for assessing the validity of models for the mechanism of action of putative therapeutic agents will be demonstrated through studies of the structure, orientation, and conformational flexibility of the peptide KL_4 in monolayer-

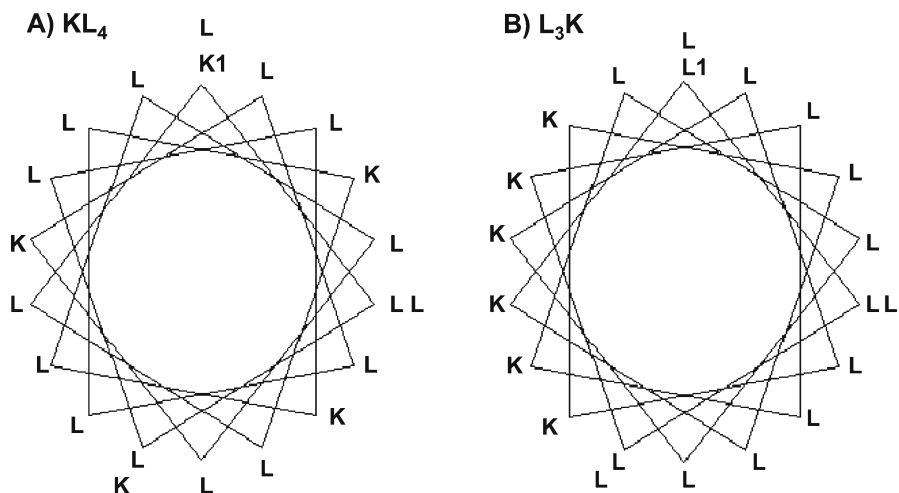


Fig. 3.8. Helical wheel plots of **A** KL_4 (KLLLLKLLLLKLLLLKLLLL) and **B** L_3K (LLLKLLKLLLLKLLKLLKLLL)

ers with both zwitterionic and anionic phospholipids. The role of the spatial distribution of peptide charges is addressed through corresponding IRRAS studies of the KL_4 peptide and a related peptide with the sequence LLLKLLKLLLLKLLKLLKLLL, termed L_3K . Helical wheel plots of both peptides (Fig. 3.8a, b, respectively) show very different charge distributions when the peptides adopt α -helical structures. Early reports investigating the mechanism of action of KL_4 in lipid monolayers proposed that the charged lysine residues interact with the polar headgroups of lipids and the leucine side chains penetrate into the hydrophobic acyl chain region (Cochrane and Revak 1991). Subsequent investigations using attenuated total reflectance (ATR)-IR revealed that KL_4 forms an α -helix oriented parallel to lipid acyl chains in two different zwitterionic/anionic lipid mixtures (Gustafsson et al. 1996; Nilsson et al. 1998). Similar ATR-IR measurements were conducted for an amphipathic α -helical peptide ($KL_{2,3}$) quite similar in sequence to L_3K , and the helix axis was found to be perpendicular to the lipid acyl chains (Nilsson et al. 1998). It was suggested that the action of KL_4 was more closely related to that of native SP-C, known to exist as a transmembrane α -helix in lipid bilayers. It is not easy to reconcile the transmembrane helical orientation reported for KL_4 with the charge distribution shown in Fig. 3.8a. Additionally, the experiments from which the structure and orientation of both peptides were determined were performed in a lipid bilayer environment as opposed to a lipid monolayer where the putative mechanisms of action for the native proteins and their analogues are thought to be critical to lung function.

Figure 3.9, panels A and B show the dependence of KL_4 secondary structure on lipid headgroup charge and surface pressure. IRRAS spectra of the lipid carbonyl (around $1,735\text{ cm}^{-1}$) and the protein amide I region for a DPPC/1,2-dipalmitoylphosphatidylglycerol (DPPG) (7/3 molar ratio) monolayer containing 2.5 mol% KL_4 are shown in Fig. 3.9, panel A as the monolayer is compressed and subsequently

expanded. The frequency of the major amide I band component at $1,620\text{ cm}^{-1}$ with a very weak component at around $1,685\text{ cm}^{-1}$ are indicative of antiparallel β -sheet structure. There is a second very weak amide I component at around $1,650\text{ cm}^{-1}$ due to helical secondary structure, which significantly decreases in relative intensity at $\Pi > 5\text{ mN m}^{-1}$. The pressure-driven conversion of the peptide structure from mixed helix/sheet to essentially all β -sheet is irreversible. Equivalent behavior was observed for KL_4 in a pure DPPC monolayer (for more detail about the KL_4 IRRAS experiments, see Cai et al. 2003). When the anionic lipid composition of the monolayer exceeds a certain value, the balance of secondary structure elements changes. Figure 3.9, panel B displays IRRAS spectra of a DPPG monolayer with 5 mol% KL_4 . The relative intensity of the α -helical amide I component ($1,652\text{ cm}^{-1}$) remains greater than that of the β -sheet component band at $1,620\text{ cm}^{-1}$ at all pressure values sampled during compression and subsequent expansion. The observation of various amounts of β -sheet structure was not anticipated on the basis of the aforementioned reports. Different behavior is observed for the L_3K peptide where an α -helical structure dominates in both pure DPPG and DPPC monolayers at all surface pressures studied (Fig. 3.9, panels C, D, respectively).

Additional experiments were conducted to investigate the lipid composition/charge dependence of KL_4 secondary structure. To evaluate the role of electrostatics in the interaction, Ca^{2+} ions were introduced into the subphase and IRRAS spectra of a DPPG/ KL_4 monolayer, at $\Pi = 13\text{ mN m}^{-1}$, were acquired over a 2-h period (Fig. 3.10a). A transition from predominantly helical structure (around $1,650\text{ cm}^{-1}$) to β -sheet (around $1,620\text{ cm}^{-1}$) is observed. Results from curve-fitting the experimental amide I contour with three bands ($1,650$, $1,635$, and $1,620\text{ cm}^{-1}$) as a function of time are shown in Fig. 3.10b. Indeed, the fractional area of the helical component is observed to decrease, as the β -sheet component increases and a component at $1,630\text{ cm}^{-1}$ remains relatively the same for the 2-h time period. A control experiment was conducted where the same solution without calcium was injected into the subphase and IRRAS spectra were acquired over a 2-h time period. As is evident in Fig. 3.10c, the amide I contour does not change over this time period. Thus, Ca^{2+} ions disrupt the electrostatic interaction between the negative charge of the DPPG headgroup and the KL_4 peptide, which appears to be critical for stabilization of helical secondary structure.

To further explore the peptides' interaction with lipid at the air/water interface, the orientation of the helix axis and the β -strand axis were determined in appropriate lipid monolayers for the two secondary structures. For this quantitative evaluation, as discussed before, IRRAS spectra are acquired using s- and p-polarized radiation at a variety of incident angles and the intensity of the appropriate component within the amide I band contour is measured after curve-fitting. The experimental dichroic ratios are plotted as a function of incident angle in Fig. 3.11, panel A for a DPPG/ KL_4 monolayer where the spectra were acquired at the physiologically relevant surface pressure of 48 mN m^{-1} . In the figure, the experimental data are compared with calculated dichroic ratios obtained at several helix tilt angles. The experimental dichroic ratios for the DPPG/ KL_4 monolayer (Fig. 3.11, panel A) are best fit by simulations conducted using helix tilt angles of 65 – 75° . The orientation of the helix with respect to the surface normal is much different in monolayers than that reported for KL_4 in lipid bilayers (transmembrane orientation). This discrepancy may at first

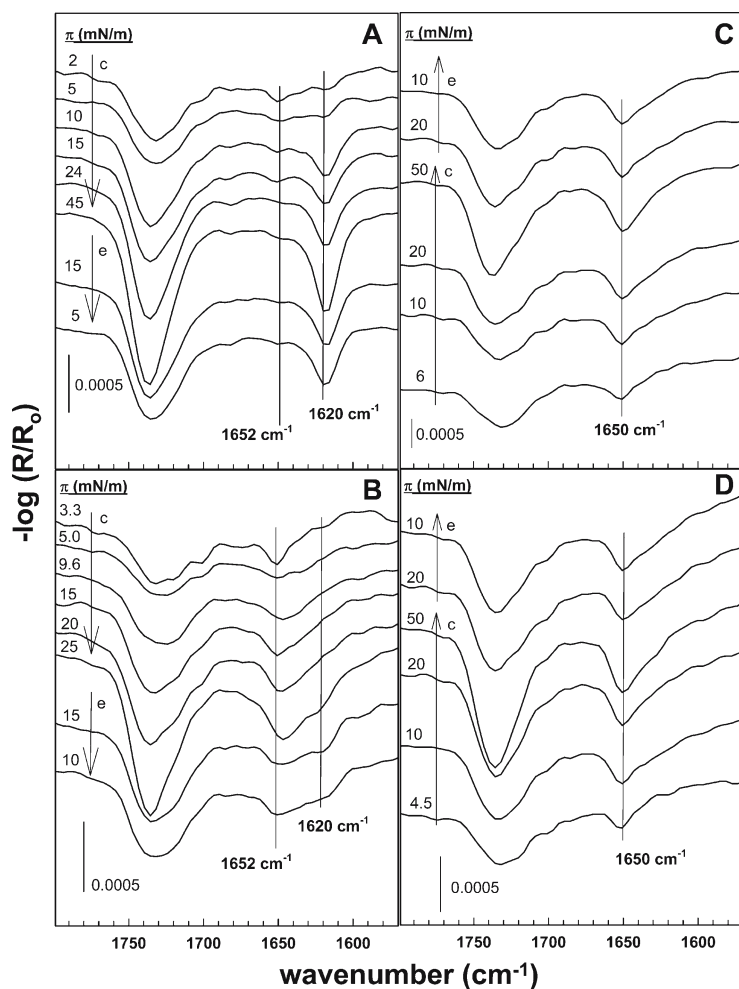


Fig. 3.9. IRRAS spectra for the lipid carbonyl (around $1,735\text{ cm}^{-1}$) and protein amide I region of various lipid/peptide mixtures acquired using *s*-polarized radiation at a 50° angle of incidence. **A** IRRAS spectra for a DPPC/DPPG (7/3 molar ratio) monolayer containing 2.5 mol% KL_4 . The monolayer was first compressed through a sequence of pressures 2, 5, 10, 15, 24, and 45 mN m^{-1} then expanded to 15 and 5 mN m^{-1} . **B** IRRAS spectra for a DPPG monolayer with 5 mol% KL_4 . The monolayer was first compressed through a sequence of pressures 3.3, 5, 9.6, 15, 20, and 25 mN m^{-1} then expanded to 15 and 10 mN m^{-1} . **C** IRRAS spectra for a DPPG monolayer containing 5 mol% L_3K . The monolayer was first compressed through a sequence of pressures 6, 10, 20, and 50 mN m^{-1} then expanded to 20 and 10 mN m^{-1} . **D** IRRAS spectra for a DPPC monolayer containing 5 mol% L_3K . The monolayer was first compressed through a sequence of pressures 4.5, 10, 20, and 50 mN m^{-1} , then expanded to 20 and 10 mN m^{-1} .

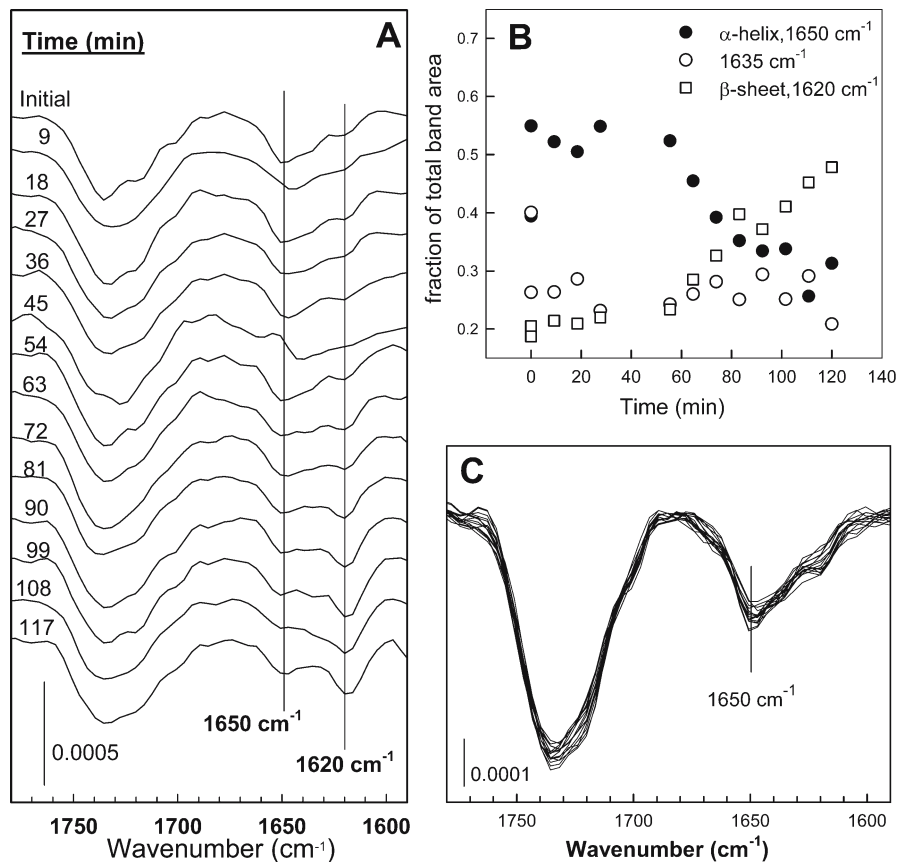


Fig. 3.10. **A** IRRAS spectra of a DPPG monolayer containing 5 mol% KL₄, at $\Pi=13$ mN/ m⁻¹, acquired over a 2-h period after Ca²⁺ ions (approximately 100 μ M in the subphase) had been introduced into the subphase. **B** Fractional areas from curve-fitting the experimental amide I contours shown in (A) with three bands (1,650, 1,635, and 1,620 cm^{-1}). **C** IRRAS spectra of a DPPG/KL₄ monolayer, at $\Pi=13$ mN/ m⁻¹, acquired over a 2-h period, after an identical solution (as in (B)) without calcium had been introduced into the subphase. The amide I contour is virtually unchanged over this time period.

seem to be due to the obvious difference in the lipid acyl chain hydrophobic thickness between the two systems; however, the secondary structure of the peptide in monolayers is governed by the electrostatic interaction between the peptide and the anionic headgroup of DPPG. It seems reasonable to suggest that in monolayers the helical peptide resides in the interfacial region interacting with both the PG headgroup and the subphase since the charged lysine residues are distributed about the circumference of the helix.

The same approach was used to evaluate the helix tilt angle for L₃K in both DPPG and DPPC monolayers at $\Pi=55$ mN m⁻¹. The results for the DPPG/L₃K monolayer

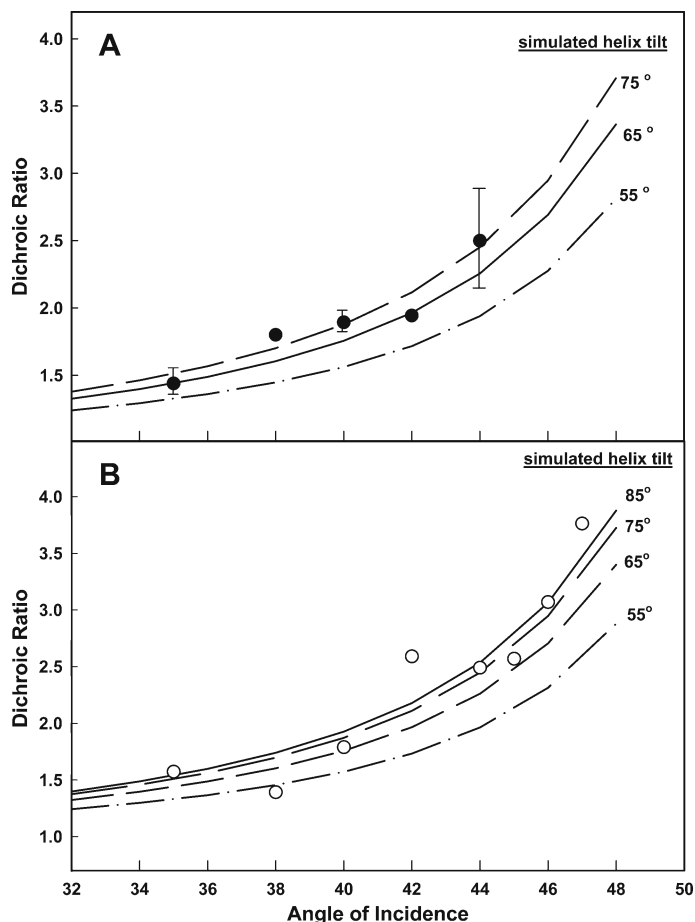


Fig. 3.11. **A** Experimental dichroic ratios (*circles*) as a function of incident angle for a DPPG/KL₄ monolayer. The surface pressure was 48 mN m⁻¹. The experimental data are compared with calculated dichroic ratios obtained at several helix tilt angles. **B** Experimental dichroic ratios at a surface pressure of 55 mN m⁻¹ (*circles*) plotted as a function of incident angle for a DPPG/L₃K monolayer. The experimental data are compared with calculated dichroic ratios obtained at several helix tilt angles.

along with simulations at various helix tilt angles are shown in Fig. 3.11, panel B. The experimental data are best fit with a calculated helix tilt of around 85°, i.e., the helix axis is essentially parallel to the water surface. A slightly smaller tilt angle (65–75°) was found to best fit the experimental data for the helical L₃K peptide in DPPC monolayers (not shown).

A recent extension of the IRRAS simulation allows the determination of β -sheet orientation (as discussed in Sect.3.3). The same procedure was applied to determine the orientation of the KL₄ β -sheet in DPPC monolayers. Experimental dichroic ra-

tios from IRRAS spectra are plotted with calculated values for a few different χ angles in Fig. 3.12. The angle χ is a measure of the rotation of the sheet out of the plane of the surface, with 90° being equivalent to the sheet lying flat on the surface (Fig. 3.6a). The experimental data were best fit with $\chi=90^\circ$ and $\varphi=35^\circ$, in which the strand axis shows a preferential alignment parallel to the barrier compression direction in the Langmuir trough. These two parameters, χ and φ , are needed in the simulations of β -sheet orientation given that the system does not have uniaxial symmetry. When KL₄ forms a β -sheet, the lysine and leucine R groups are distributed above and below the plane of the sheet. Similar to KL₄ in a helical structure, the sheet most likely resides in the interfacial region, i.e., interacting with the polar headgroup of DPPC and the water subphase.

3.5 Future Possibilities for IRRAS

Currently, IRRAS spectra may be acquired with good signal-to-noise ratios (up to about 1,000:1) from stable Langmuir films in a reasonable data collection time (arbitrarily chosen to be less than 10 min). Bands with RAs of 0.0001 may be readily detected and frequency shifts of 0.2 cm^{-1} may be monitored. This sensitivity may be utilized in several ways. One application at the limit of current feasibility is isotope labeling of peptides. Substitution of ^{13}C in the peptide bond C=O group is a useful technique to locate particular secondary structure elements. The approach is illustrated in Fig. 3.13. The peptide (similar to residues 9–36 in SP-B) WLARALIK RIQAMIPKGA*LA*VA*VA*QVCR, was synthesized in two variants, one with normal $^{12}\text{C}=\text{O}$ bonds, the second with $^{13}\text{C}=\text{O}$ bonds at the indicated alanine residues (Flach et al. 2003). The IRRAS spectra of both peptides are shown in the figure. The spectrum of the unlabeled peptide shows two major amide I components at around $1,626$ and $1,645\text{ cm}^{-1}$, due to antiparallel β -sheet structure and a mixture of α -helical and random coil, respectively. Isotope labeling enables us to locate the β -structure within the sequence. Thus, in the spectrum of the labeled peptide, the presence of an amide I component at $1,597\text{ cm}^{-1}$, coupled with the disappearance of the $1,626\text{ cm}^{-1}$ band, directly identifies the location of the β -sheet (i.e., at the indicated sites of the

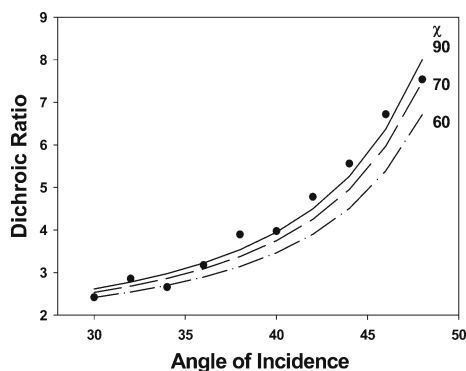


Fig. 3.12. Experimental dichroic ratios from IRRAS spectra of KL₄ (β -sheet) in DPPC monolayers (circles) are compared with calculated values for $\chi = 90$, 70 , and 60° . The experimental data were best fit with $\chi = 90^\circ$ and $\varphi = 35^\circ$

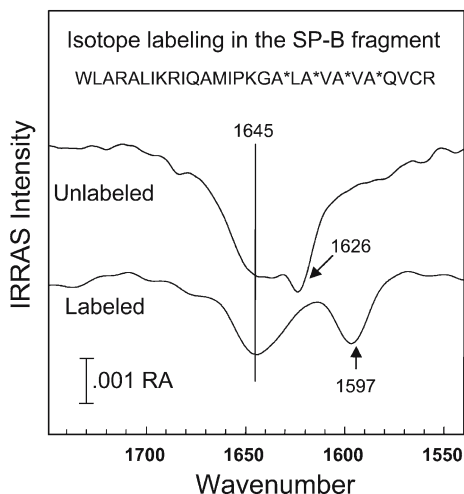


Fig. 3.13. IRRAS spectra of two forms of the peptide, WLARALIKRIQAMIPKGA*LA*VA*VA*QVCR, unlabeled with normal $^{12}\text{C}=\text{O}$ bonds, and labeled with $^{13}\text{C}=\text{O}$ bonds at the indicated alanine residues

labeling). To extend this approach, increases in sensitivity will enable the detection of spectral changes in larger proteins labeled at a single site.

The prospects for enhanced sensitivity center around improved hardware. IR sources are currently standard broadband IR sources, e.g., globars of sintered silicon carbide. On the horizon, the routine availability of tunable IR diode lasers is expected in the next few years. These will provide a source with several orders of magnitude more photons at each frequency. There are several experiments which will benefit from the anticipated increased spectral quality from the improved sources. The availability of spectra at better signal-to-noise ratios will permit detection of smaller spectral changes than is currently feasible. However, this in itself is a limited gain, since the interpretation of such spectral changes is sparse. Improved theoretical understanding of spectra–structure correlations will also be required.

A second appealing possibility is the incorporation of array detectors into the IRRAS instrumentation to more rapidly acquire IRRAS spectra of high quality. Rabolt, Chase, and their coworkers (Pellerin et al. 2004; D.B. Chase, personal communication) have demonstrated the feasibility of this powerful technique by using a conventional source and spreading a grating-dispersed IRRAS spectrum onto an IR array detector. Usable IRRAS data have been acquired in as little as 8.7 ms. The prospect of combining the ability to rapidly acquire IRRAS spectra with the spatial resolution that planar array detectors are known for will provide opportunities for new investigations.

Acknowledgment. The work from Rutgers University described in this chapter was carried out with the generous support of a grant to R.M. from the National Institutes of Health (GM-29864).

References

- Brauner JW, Dugan C, Mendelsohn R (2000) ^{13}C isotope labeling of hydrophobic peptides. Origin of the anomalous intensity distribution in the infrared amide I spectral region of β -sheet structures. *J Am Chem Soc* 122:677–683
- Brauner JW, Flach CR, Xu Z, Bi X, Lewis RNAH, McElhane RN, Gericke A, Mendelsohn R (2003) Quantitative functional group orientation in Langmuir films by infrared reflection-absorption spectroscopy: C=O groups on behenic acid methyl ester and sn2- ^{13}C -DSPC. *J Phys Chem B* 107:7202–7211
- Brauner JW, Flach CR, Mendelsohn R (2005) A quantitative reconstruction of the amide I contour in the IR spectra of globular proteins: from structure to spectrum. *J Am Chem Soc* 127:100–109
- Buffeteau T, Le Calvez E, Castano S, Desbat B, Blaudez D, Dufourcq EJ (2000) Anisotropic optical constants of α -helix and β -sheet secondary structures in the infrared. *J Phys Chem B* 104:4537–4544
- Cai P, Flach CR, Mendelsohn R (2003) An infrared reflection-absorption spectroscopy study of the secondary structure in $(\text{KL}_4)_4\text{K}$, a therapeutic agent for respiratory distress syndrome, in aqueous monolayers with phospholipids. *Biochemistry* 42:9446–9452
- Cochrane CG, Revak SD (1991) Pulmonary surfactant protein B (SP-B): structure-function relationships. *Science* 254:566–568
- Cochrane CG, Revak SD, Merritt TA, Heldt GP, Hallman M, Cunningham MD, Easa D, Pramanik A, Edwards DK, Alberts MS (1996) The efficacy and safety of KL_4 -Surfactant in preterm infants with respiratory distress syndrome. *Am J Respir Crit Care Med* 153:404–410
- Cochrane CG, Revak SD, Merritt TA, Schraufstatter IU, Hoch RC, Henderson C, Andersson S, Takamori H, Oades ZG (1998) Bronchoalveolar lavage with KL_4 -surfactant in models of meconium aspiration syndrome. *Pediatr Res* 44:705–715
- Cornut I, Desbat B, Turlet JM, Dufourcq J (1996) In situ study by polarization modulated Fourier transform infrared spectroscopy of the structure and orientation of lipids and amphiphathic peptides at the air-water interface. *Biophys J* 70:305–312
- Dluhy RA (1986) Quantitative external reflection infrared spectroscopic analysis of insoluble monolayers spread at the air-water interface. *J Phys Chem* 90:1373–1379
- Dluhy RA, Cornell DG (1985) In situ measurement of the infrared spectra of insoluble monolayers at the air-water interface. *J Phys Chem* 89:3195–3197
- Dluhy RA, Mendelsohn R (1988) Emerging techniques in biophysical FT-IR. *Anal Chem* 60:269A–278A
- Flach CR, Brauner JW, Taylor JW, Baldwin RC, Mendelsohn R (1994) External reflection FTIR of peptide monolayer films in situ at the air/water interface – experimental design, spectra-structure correlations, and effects of hydrogen-deuterium exchange. *Biophys J* 67:402–410
- Flach CR, Cai P, Dieudonné D, Brauner JW, Keough KMW, Stewart J, Mendelsohn R (2003) Location of structural transitions in an isotopically labeled lung surfactant SP-B peptide by IRRAS. *Biophys J* 85:340–349
- Goerke J, Clements JA (1986) Alveolar surface tension and lung surfactant. In: Mackelme PT, Mead J (eds) *Handbook of physiology: the respiratory system III*. American Physiology Society, Washington, DC, pp 247–261

- Gustafsson M, Vandenbussche G, Curstedt T, Ruyschaert JM, Johansson J (1996) The 21-residue surfactant peptide (LysLeu₄)₄Lys(KL₄) is a transmembrane α -helix with a mixed non-polar/polar surface. *FEBS Lett* 384:185–188
- Johansson J, Curstedt T (1997) Molecular structures and interactions of pulmonary surfactant components. *Eur J Biochem* 244:675–693
- Johansson J, Curstedt T, Robertson B (2001) Artificial surfactants based on analogues of SP-B and SP-C. *Pediatr Pathol Mol Med* 20:501–518
- Krimm S, Abe Y (1972) Intermolecular interaction effects in the amide I vibrations of β polypeptides. *Proc Natl Acad Sci USA* 69:2788–2792
- Krimm S, Bandekar J (1986) Vibrational spectroscopy and conformation of peptides, polypeptides, and proteins. *Adv Protein Chem* 38:181–364
- Kuzmin VL, Michailov AV (1981) Molecular theory of light reflection and applicability limits of the macroscopic approach. *Opt Spectrosc* 51:383–385
- Kuzmin VL, Romanov VP, Michailov AV (1992) Reflection of light at the boundary of liquid systems and structure of the surface layer: a review. *Opt Spectrosc* 73:1–26 (translated from *Opt Spektrosk* 73:3–47, 1992)
- Lin S, Na CL, Akinbi HT, Apsley KS, Whitsett JA, Weaver TE (1999) Surfactant protein B (SP-B) $-/-$ mice are rescued by restoration of SP-B expression in alveolar type II cells but not Clara cells. *J Biol Chem* 274:19168–19174
- Mendelsohn R, Flach CR (2002a) Infrared reflection-absorption spectrometry of monolayer films at the air-water interface. In: Chalmers JM, Griffiths PR (eds) *Handbook of vibrational spectroscopy*, vol 2. Wiley, Chichester, pp 1028–1041
- Mendelsohn R, Flach CR (2002b) Infrared reflection-absorption spectroscopy of lipid, peptides, and proteins in aqueous monolayers. In: Simon SA, McIntosh TJ (eds) *Current topics in membranes: peptide-lipid interactions*, vol 52. Academic, San Diego, pp 57–88
- Mendelsohn R, Brauner JW, Gericke A (1995) External infrared reflection absorption spectrometry of monolayer films at the air-water interface. *Annu Rev Phys Chem* 46:305–334
- Mitchell ML, Dluhy RA (1988) In situ FT-IR investigation of phospholipid monolayer phase transitions at the air-water interface. *J Am Chem Soc* 110:712–718
- Miyazawa T, Blout ER (1961) The infrared spectra of polypeptides in various conformations: amide I and II bands. *J Am Chem Soc* 83:712–719
- Nilsson G, Gustafsson M, Vandenbussche G, Veldhuizen E, Griffiths WJ, Sjövall J, Haagsman HP, Ruyschaert J-M, Robertson B, Curstedt T, Johansson J (1998) Synthetic peptide-containing surfactants: evaluation of transmembrane versus amphipathic helices and surfactant protein C poly-valyl to poly-leucyl substitution. *Eur J Biochem* 255:116–124
- Pastrana-Rios B, Flach CR, Brauner JW, Mautone AJ, Mendelsohn R (1994) A direct test of the “squeeze-out” hypothesis of lung surfactant function – external reflection FT-IR at the air/water interface. *Biochemistry* 33:5121–5127
- Pastrana-Rios B, Taneva S, Keough KMW, Mautone AJ, Mendelsohn R (1995) External reflection absorption infrared spectroscopy study of lung surfactant proteins SP-B and SP-C in phospholipid monolayers at the air/water interface. *Biophys J* 69:2531–2540
- Pellerin C, Snively CM, Chase DB, Rabolt JF (2004) Performance and application of a new planar array infrared spectrograph operating in the mid-infrared (2000–975 cm^{-1}) fingerprint region. *Appl Spectrosc* 58:639–646
- Schürch S, Green FHY, Bachofen H (1998) Formation and structure of surface films: captive bubble surfactometry. *Biochim Biophys Acta* 1408:180–202

- Takamoto DY, Lipp MM, van Nahmen A, Lee KYC, Waring AJ, Zasadzinski JA (2001) Interaction of lung surfactant proteins with anionic phospholipids. *Biophys J* 81:153–169
- Von Nahmen A, Schenk M, Sieber M, Amrein M (1997) The structure of a model pulmonary surfactant as revealed by scanning force microscopy. *Biophys J* 72:463–469
- Wang L, Cai P, Galla H-J, He H, Flach CR, Mendelsohn R (2005) Monolayer–multilayer transitions in a lung surfactant model: IR reflection–absorption spectroscopy and atomic force microscopy. *Eur Biophys J* 54:243–254
- Xu Z, Brauner JW, Flach CR, Mendelsohn R (2004) Orientation of peptides in aqueous monolayer films. Infrared reflection-absorption spectroscopy studies of a synthetic amphipathic β -sheet. *Langmuir* 20:3730–3733
- Zbinden R (1964) *Infrared spectroscopy of high polymers*. Academic Press, New York

Two-Dimensional Infrared Correlation Spectroscopy

JOSÉ LUIS R. ARRONDO, IBON ILOORO, MARCOS GARCIA-PACIOS,
FÉLIX M. GOÑI

4.1 Introduction

Proteins exert their action as 3D entities. A unique folding pattern characterizes the action of a single protein giving rise to one of the more important biochemical paradigms: the structure–function relationship. The techniques that provide detailed information on protein 3D structures are X-ray crystallography and NMR. They have limitations either in the size of the protein or in the requirement of protein crystals. It should also be noted that resolution of a protein structure does not immediately lead to understanding its molecular mechanism. Therefore, there is ample field in many aspects of protein structure and function to be covered by complementary, lower-resolution techniques that can provide useful information, even in proteins with known 3D structures.

IR spectroscopy is acknowledged as a choice method, either as an alternative option or as a complement to the high-resolution techniques in protein studies, because of the wealth of information provided and also because of the sensitivity of the technique to changes in the environment of the protein. IR spectroscopy was first applied to proteins as early as in 1952 (Sutherland 1952) before any detailed X-ray results were available. The principles of applying the study of the amide I band in H₂O and D₂O media to proteins were established in 1967 (Susi et al. 1967). However, this work was limited to a qualitative approach to the secondary structure. The complexity of the amide I band and the presence of water (a strong IR absorber) as a physiological solvent made it difficult to extract the information from the IR spectrum. Its use in proteins in physiological environments in general has only been possible since the revolution in instrumentation produced by the development of microcomputers that allowed the design of instruments based on the Michelson interferometer and the fast Fourier transform (Griffiths and de Haseth 1986). The appearance in the market of instruments based on interferometric devices with a high signal-to-noise ratio (known as Fourier transform IR spectrometers) changed the scene, allowing the measurement of spectra in water and the application of data-treatment techniques that helped to discern the information contained in the spectrum. The increase in the signal-to-noise ratio obtained with these instruments allows the subtraction of the aqueous buffer, thus giving spectra whose bands contain information on the protein in its native environment and that are free from spectral interference by the solvent. The structural analysis usually implies a mathematical approach in order to extract the information contained in the conformationally sensitive composite bands, designated in IR spectroscopy as “amide bands”, obtained

from vibrations of the peptide bond. Commonly used methods of analysis imply narrowing the intrinsic bandwidths to visualize the overlapping band components and then decomposing the original band contour into these components by means of an iterative process. The various components are finally assigned to protein or subunit structural features (Arrondo and Goñi 1999).

The idea of 2D spectroscopy was put forward in NMR spectroscopy, where by spreading spectral peaks over a second dimension, the visualization of complex spectra consisting of many overlapped peaks could be simplified. Lately, different forms of 2D spectroscopy applicable to many other types of spectroscopic techniques have been developed. The concept of perturbation-based 2D spectroscopy was introduced by Noda in 1986 (for a review see Noda et al. 2000). This mathematical tool emphasizes in-phase and out-of-phase correlations between spectral intensity variations obtained from the spectra before and after an external perturbation is applied to the system. This approach, essentially different from 2D-NMR spectroscopy, uses correlation analysis of the dynamic fluctuations caused by an external perturbation to enhance spectral resolution without assuming any lineshape model for the bands. The technique was intended for the study of polymers and liquid crystals, but it has recently been applied to proteins. Proteins are a good target for this method, since changes induced by temperature (Fabian et al. 1999) or by the presence of lipids (Torrecillas et al. 2003; Shanmukh and Dluhy 2004) or other external ligands (Pastrana-Rios et al. 2002) can be studied in more detail than with the conventional IR approach. The power of the 2D correlation approach results primarily in an increase of the spectral resolution by dispersal of the peaks along a second dimension that also reveals the order of the events induced by the perturbation (Sasic et al. 2001). Correlation between bands is found through the so-called synchronous and asynchronous spectra that correspond to the real and imaginary parts of the cross-correlation of spectral intensity at two wavenumbers. In a synchronous 2D map, the peaks located in the diagonal (autopeaks) correspond to changes in the intensity and are always positive. The cross-relation peaks indicate a relationship between the two bands involved. Asynchronous maps show not-in-phase cross-relation between the bands and give an idea of the order of the events produced by the perturbation.

4.2

Description of the Technique

The power of the 2D correlation approach results primarily in an increase of the spectral resolution by dispersal of the peaks along a second dimension that also reveals the time course of the events induced by the perturbation. Correlations between bands are found through the so-called synchronous (Φ) and asynchronous (Ψ) spectra that correspond to the real and imaginary parts of the cross-correlation of spectral intensity at two wavenumbers, i.e. that two vibrations of the protein characterized by two different wavenumbers (ν_1 and ν_2) are affected at the same time (synchronous) or that the vibrations of the functional groups corresponding to the different wavenumbers each change at a different time (asynchronous). In a synchronous 2D map, the peaks located in the diagonal (autopeaks) correspond to

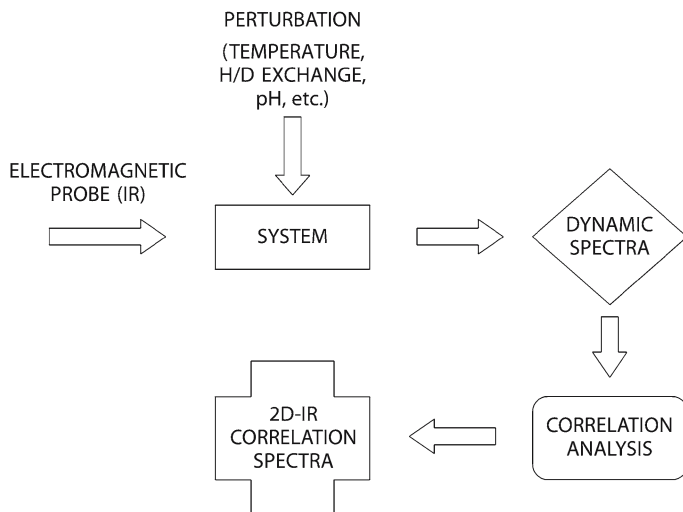


Fig. 4.1. General scheme for obtaining 2D-IR spectra. From the response of the system to perturbation, a dynamic spectrum is obtained. The correlational analysis gives two components, the real one corresponding to the synchronous spectrum and the imaginary one corresponding to the asynchronous one

changes in intensity induced by the perturbation, and are always positive. The cross-correlation peaks indicate an in-phase relationship between two bands involved. Asynchronous maps show not-in-phase cross-correlation between the bands and this gives an idea of the time course of the events produced by the perturbation.

To obtain 2D-IR correlation spectra the formalism proposed by Noda is usually followed. From the response of the system to the perturbation, a dynamic spectrum is obtained. The type of physical information contained in a dynamic spectrum is determined by the perturbation method. Once the dynamic spectrum has been obtained as a matrix formed by the spectra ordered according to the change produced by the perturbation, the Fourier transform gives two components, the real one corresponding to the synchronous spectrum and the imaginary one corresponding to the asynchronous one (Fig. 4.1). In terms of calculation, instead of using a Fourier transform that would require large computation times, an adequate numerical evaluation of 2D correlation intensity is used. Thus, the synchronous 2D correlation intensity expressed as

$$\Phi(v_1, v_2) = \frac{1}{m-1} \sum_{j=1}^m \bar{y}_j(v_1) \cdot \bar{y}_j(v_2). \quad (4.1)$$

$\bar{y}_j(v_j)$ is the dynamic spectra calculated from the spectral intensities as a deviation from a reference spectrum at a point of physical variable t_j .

The computation of asynchronous 2D correlation intensity is somewhat more complicated. Two approaches can be used, (1) using the Hilbert transform and (2)

a direct procedure, giving similar results (for a detailed discussion on the asynchronous calculation, see Noda and Ozaki 2004).

4.3 Spectral Simulations

Most studies using 2D-IR correlation spectroscopy have interpreted the synchronous and asynchronous maps in terms of changes in intensity. However, shifts in band position, which can be associated with protein environmental changes, or changes in bandwidth can affect the maps, mainly the asynchronous one. Also, it has been demonstrated that spectral effects aside from true asynchronism can generate peaks in the correlation maps (Lefevre et al. 2004). Therefore, spectral simulations are needed to understand the changes observed in complex processes such as those observed in proteins.

To carry out the simulations, Gaussian peaks are generated using the general Gaussian equation:

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} e^{-(x-\mu)^2/(2\sigma^2)}. \quad (4.2)$$

In all simulations each dynamic spectrum consists of 13 spectra with 130 points equivalent to an amide I band in the region 1,700–1,600 cm^{-1} with a nominal resolution of 2 cm^{-1} . The changes are assumed to be linear. The maps are obtained assuming that one or two peaks are changing.

4.3.1 Intensity Changes

Changes in intensity were observed in spectra composed of one or two peaks. The corresponding correlation maps are represented in Fig. 4.2. The bottom and top maps correspond respectively to spectra generated from changes in one or two peaks. Left-hand traces correspond to synchronous spectra, while right-hand traces are from asynchronous spectra. Negative peaks are represented as shaded areas. Asynchronous maps, if only band intensity is changed, are only noise. Similar results are obtained in simulations of 2D electronic correlation spectra (Kim and Jeon 2001) or using Lorentzian bandshapes (Czarnecki 2000). The lack of bands in the asynchronous spectra has been observed in protein thermal denaturation after aggregation has taken place (Paquet et al. 2001), and is not a noise artefact due to the decrease in spectral intensity after aggregation, because artificial spectra are noise-free, and yet the asynchronous map consists only of low-amplitude noise. Thus, the lack of bands is a characteristic of the denatured proteins once the aggregation has been completed. On the other hand, considering that synchronous spectra reflect in-phase changes, differences in peak intensity are expected to produce changes. It can be seen that the complexity of the synchronous spectra changes as a function of the number of peaks involved. If only one peak is changing, one autopeak is seen. If two or more peaks are involved, the respective cross-peaks show in which way

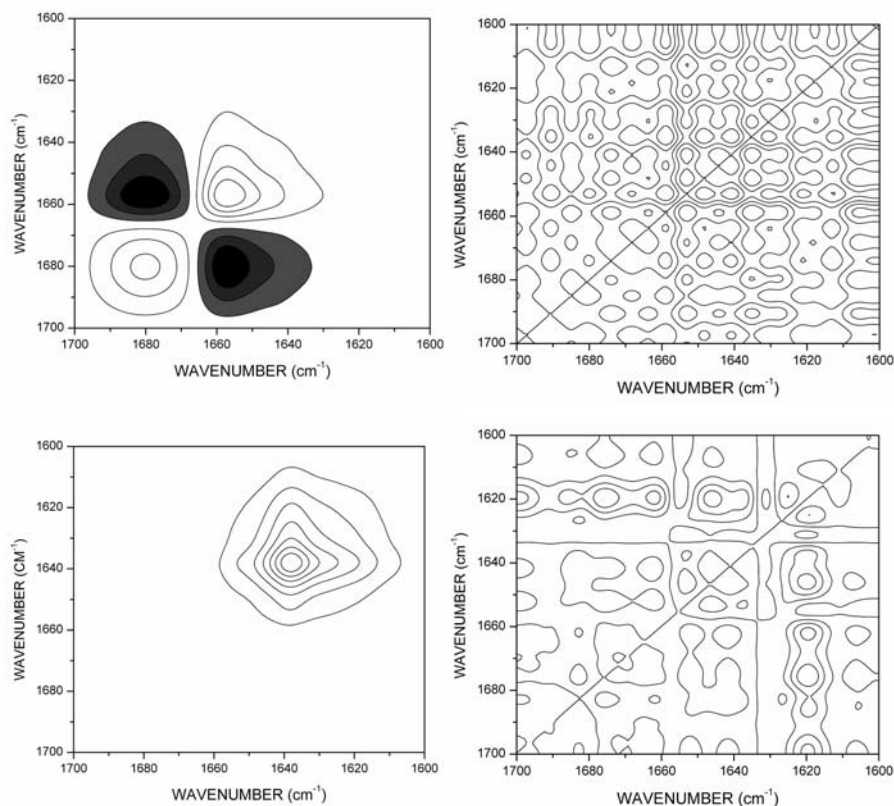


Fig. 4.2. 2D correlation maps corresponding to changes in the intensity of a band composed of one peak (*bottom*) or two peaks (*top*). Synchronous maps are located to the *left* and asynchronous ones to the *right*. The *X*-axis and the *Y*-axis correspond to the number of the point in the artificial curve and the *Z*-axis is the correlational intensity. Roughly they represent a protein amide I band in a D_2O buffer. Negative peaks are *shaded*

intensities are changing. If either increases or decreases simultaneously, the cross-peaks are positive, and if one peak is increasing and the other one is decreasing they are negative.

4.3.2 Band Shifting

Changes in band position were analysed in a similar way. Simulated bands were constructed with one or two components changing simultaneously, and the synchronous and asynchronous maps are represented in Fig. 4.3. In all the synchronous

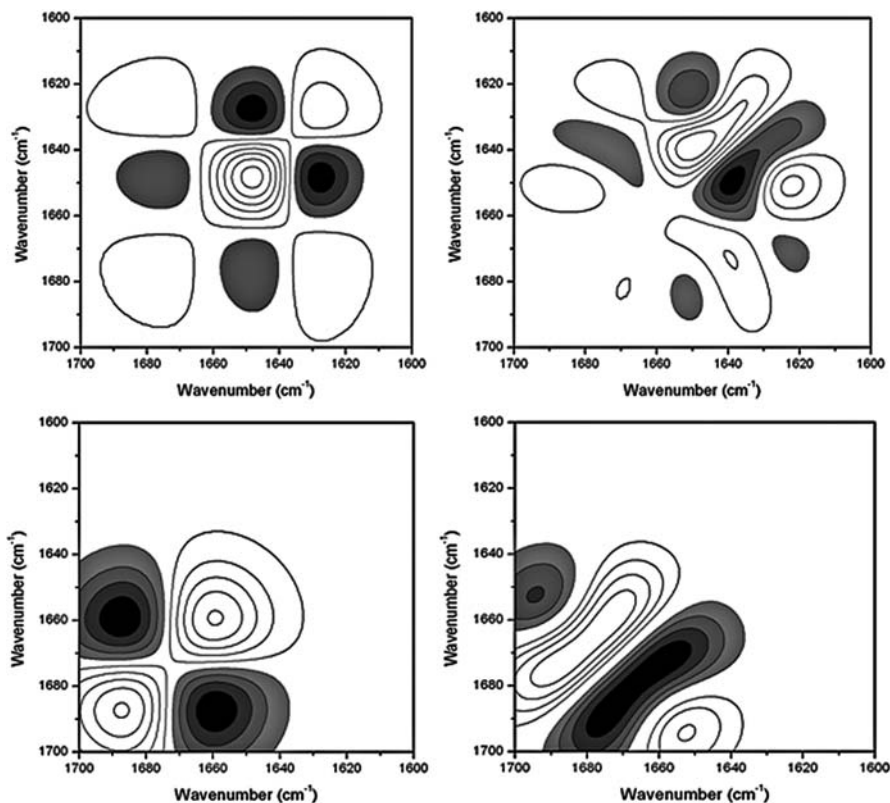


Fig. 4.3. 2D correlation maps corresponding to changes in position (band shifting) of a band composed of one peak (*bottom*) or two peaks (*top*). Synchronous maps are located to the *left* and asynchronous ones to the *right*. The axes are as in Fig. 4.2

maps, autopeaks are located at the initial and final positions of the maxima during band shifting. This is the reason why two peaks are observed in the synchronous map corresponding to a change of band position in one peak. Cross-peaks are positioned next to the equivalent autopeaks with a negative intensity since the intensity at a given frequency at the beginning of the shift decreases, whereas the intensity at a different frequency will increase after the shift. The characteristic shape in the asynchronous maps for a band shift is a butterfly (Gericke et al. 1996) or a banana, depending on whether we consider the full symmetric map or only one side of the diagonal. The peaks include the frequencies through which the maximum has moved throughout the band shifting. As in the synchronous spectra, there are also cross-peaks of the opposite sign between the beginning and the end of the “bananas”.

4.3.3 Bandwidth

The same approach as in the two previous models was used for band broadening. In this case the pattern obtained is more complex. It has to be taken into account that changes in bandwidth in proteins are not usually a result of a phase transition as in the case of lipids (Arrondo and Goñi 1997), but they are rather the consequence of changes in conformation and subsequent variation in the area of the band associated with that conformation. Hence, changes in bandwidth can be discerned because a more complicated correlation map is produced, but they are not easily assigned (see later). In the case of the models studied, the resulting synchronous maps (Fig. 4.4) show patterns with several autopeaks located at the peak inflection points, since these are the frequencies at which greater intensity variations occur throughout the band broadening. These autopeaks are associated with positive cross-peaks because

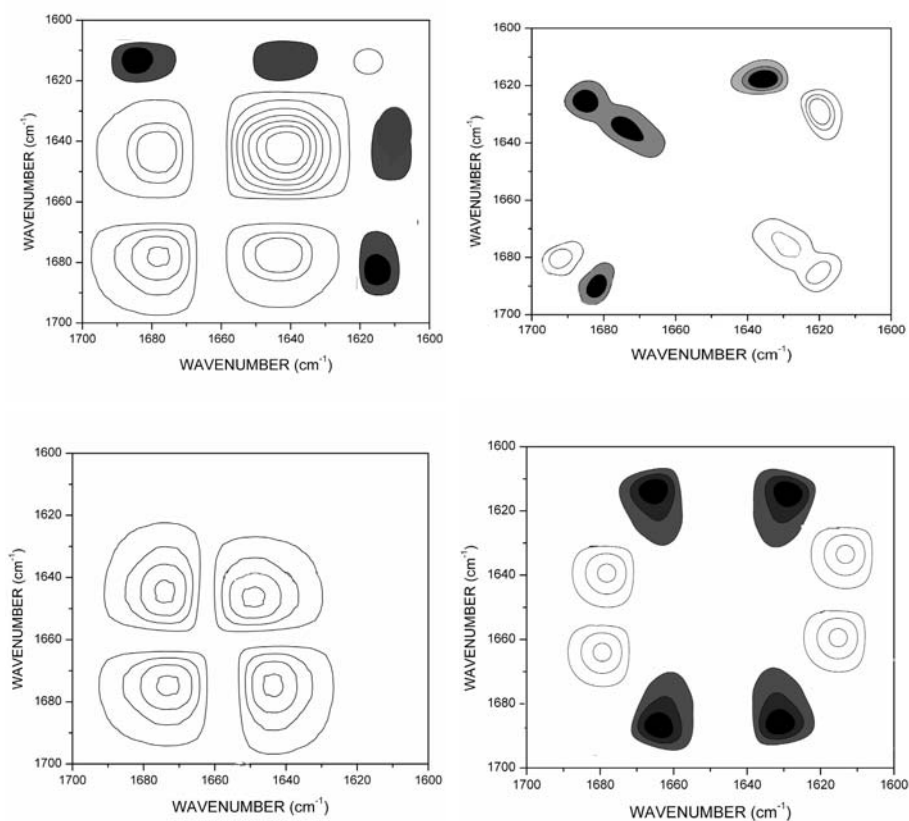


Fig. 4.4. 2D correlation maps corresponding to changes in the width (bandwidth) of a band composed of one peak (*bottom*) or two peaks (*top*). Synchronous maps are located to the *left* and asynchronous to the *right*. The axes are as in Fig. 4.2

both points are changing their intensity in the same direction. The asynchronous spectrum is much more complicated, showing an increased number of peaks with different sizes and intensities.

4.4 2D-IR Studies of Proteins

2D-IR correlation spectroscopy has become a popular tool in the analysis of protein spectra since they gives rise to broad bands and the application of an external perturbation produces complex phenomena. It has been used to study the correlation of the amide bands and the assignment of the different band components to secondary structure (Arrondo et al. 2004). It has also been applied to analyse the kinetics of hydrogen–deuterium exchange in order to study solvent accessibility and the flexibility of the protein (Nabet and Pezolet 1997; Meskers et al. 1999). Protein unfolding and aggregation are also targets for the use of 2D-IR correlation spectroscopy since besides the analysis provided by synchronous maps, the asynchronous maps give an idea of the hierarchical order of events.

4.4.1 2D Maps of Homopolypeptides

Proteins are very complex entities that can undergo complex changes after being exposed to an external perturbation. Homopolypeptides have been used extensively to assign IR bands to secondary structures (Susi 1969). They are used because of their ability to adopt regular-canonical conformations without containing tertiary structure. Polyproline is a particular example since its inability to form hydrogen bonds with other peptide bonds produces the so-called polyproline I and polyproline II (PPII) helices. Between them, the PPII helix, left-handed with dihedral angles closer to β -sheet than to α -helix is a special secondary structure. Temperature produces a single transition in this helix that can be studied by 2D-IR spectroscopy. Since no tertiary structure is involved, the transition in a D_2O buffer goes from the PPII helix to an unordered conformation without residual structure as can be seen in Fig. 4.5. The perturbation produces a change in intensity of the peak together with a shift in band position. In Fig. 4.6 the 2D-IR maps are shown. The synchronous spectrum represents two autopeaks centred around $1,624$ and $1,642\text{ cm}^{-1}$, corresponding to the PPII bands that vary in intensity. As expected from the simulation corresponding to changes in bandwidth, the asynchronous map gives not only noise. Accordingly, the temperature-induced unfolding of the PPII helix corresponds to a protein with a perturbation involving changes in the intensity of two components with variations in bandwidth.

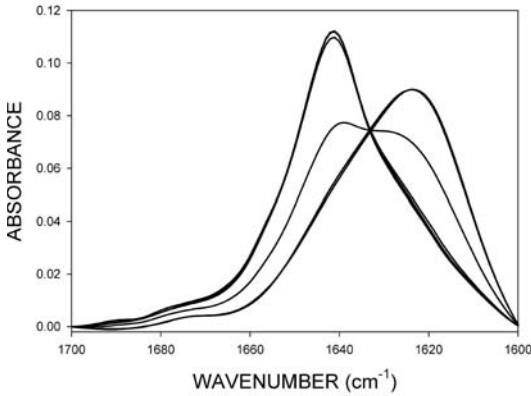


Fig. 4.5. Amide I band of polyproline in a D_2O solution. The temperature goes from 30 to 70°C. The polyproline II helix is located at 1620 cm^{-1} and the band corresponding to an unordered structure at 1642 cm^{-1}

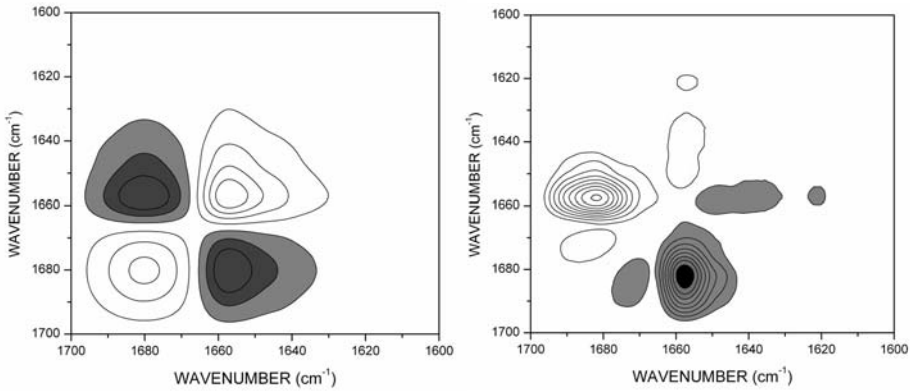


Fig. 4.6. Polyproline 2D-IR correlation maps corresponding to the spectra shown in Fig. 4.4. Synchronous (*left*) and asynchronous (*right*) maps are represented. Negative peaks are shaded

4.4.2 Protein Denaturation

In some simple proteins, denaturing agents, e.g. urea, guanidinium chloride and heat, induce a two-state transition, from native to random conformation (Barone et al. 1997; Mayne and Englander 2000). However, most proteins, and in particular integral membrane proteins, can be regarded as complex systems containing independent domains that do not necessarily form cooperative structures (Arrondo and Goñi 1999; Kumar and Yu 2004). Urea is a chaotropic agent that causes protein unfolding/denaturation by disrupting hydrogen bonds which hold the protein in its unique structure. Urea may also disrupt hydrophobic interactions by promoting the solubility of hydrophobic residues in aqueous solutions (Tanford 1968; Monera et al. 1994; Soloaga et al. 1998). In principle, urea cannot be used in IR studies, because

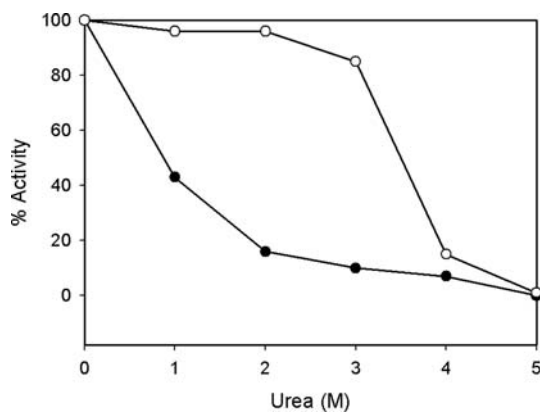


Fig. 4.7. Reversible and irreversible effects of urea on sarcoplasmic reticulum ATPase activity. Enzyme activity was assayed on sarcoplasmic reticulum vesicles treated with varying concentrations of urea, before (closed circles) and after (open circles) washing with urea-free buffer. Average values of two very similar experiments

of its strong absorbance in the spectral region corresponding to the protein amide I band. For this reason ^{13}C -urea, whose IR spectrum does not interfere so strongly with the protein amide I band (Reinstädler et al. 1996), has to be used. Sarcoplasmic reticulum (SR) ATPase activity was assessed in SR vesicles in the presence of urea. As seen in Fig. 4.7, 2 M urea causes a dramatic decrease in activity; higher concentrations produce further decreases. When urea is washed away after treatment, the activity is almost fully recovered for 2 and 3 M urea; however, 4 and 5 M urea appear to cause irreversible damage under our conditions. The conformational changes associated with the activity changes can be studied by 2D-IR correlation spectroscopy using temperature as a perturbation. The 2D-IR study whose results are summarized in Fig. 4.8 was designed primarily to understand the changes taking place during the “intermediate state” of thermal denaturation, whose temperature is 45°C (Echabe et al. 1998). For this reason the maps in Fig. 4.8 were obtained by correlating spectra between 20 and 45°C. The control sample spectral map in the absence of urea is shown in Fig. 4.8a. Several autopeaks are seen, at 1,620, 1,632, 1,640, 1,654, 1,664, 1,675 and 1,682 cm^{-1} , and also a number of cross-peaks, showing an important redistribution of the secondary structures. The fact that 4 M and higher concentrations of urea lead to irreversible structural and functional changes in SR ATPase indicates a profound modification of protein conformation, including likely the transmembrane domain. However, in the presence of urea no differences are seen in this temperature region, indicating that the intermediate state (characterized by band changes at around 1,640 cm^{-1}) is lost under these conditions. The recovery of the correlation maps in 3 M urea but not in 4 M urea points to an interaction with the intramembranal domain of the ATPase. Figure 4.8b corresponds to the sample in 3 M urea. There is only background noise, and the characteristic shift of the urea band “tail”. Thus, 2D-IR correlation spectroscopy confirms the lack of an intermediate transition during SR ATPase thermal denaturation in the presence of 3 M urea. The reversibility of the 3 M urea effect is apparent in Fig. 4.8c, corresponding to the SR vesicles treated with 3 M urea, and then washed. The original pattern (Fig. 4.8a) is almost quantitatively recovered. Also in agreement with our previous observations, treatment with 4 M urea prevents the oligomer–monomer transition

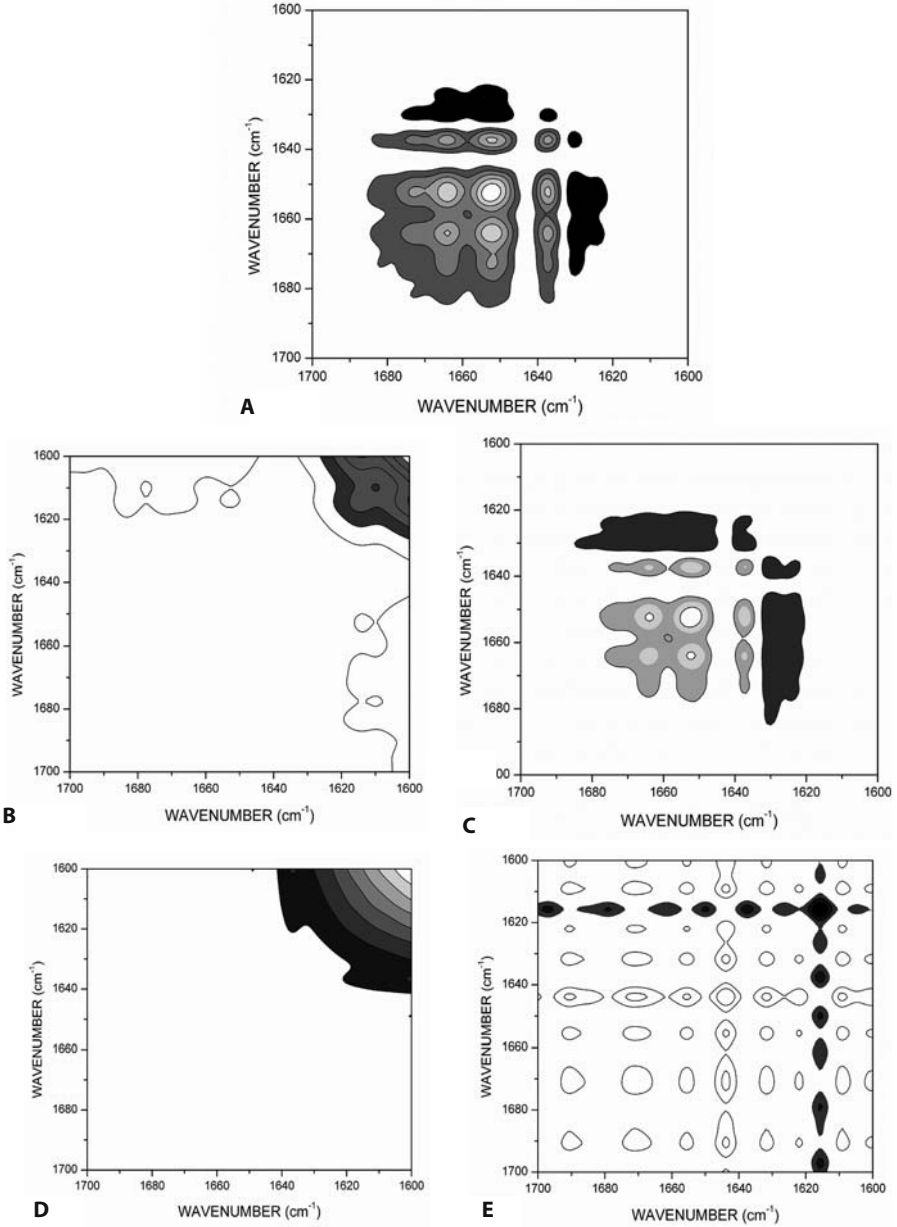


Fig. 4.8. 2D-IR synchronous correlation spectr. The spectra were obtained by correlating data in the interval 20–45°C (T_m of the intermediate state). **A** Control, non-urea-treated, **B** 3 M urea, **C** 3 M urea, and then washing twice in urea-free buffer, **D** 4 M urea, **E** 4 M urea, and then washing twice in urea-free buffer

of the enzyme, so that the corresponding spectral map is featureless (Fig. 4.8d), and this effect cannot be reverted by urea removal (Fig. 4.8e).

4.4.3 Protein Unfolding and Aggregation

IR spectroscopy has been extensively used to study protein unfolding produced using heat as an external perturbation. The reason is that usually this unfolding can lead to aggregation and IR is particularly suitable for measuring aggregation. Aggregation is not only related to temperature, but is also found in amyloidosis and body inclusion formation. 2D-IR correlation spectroscopy is very suitable for studying both phenomena (Arrondo and Goñi 1999; Fabian et al. 1999; Arrondo et al. 2004; Lefevre et al. 2004). Protein aggregation is characterized in IR spectroscopy by the appearance of bands around $1,620$ and $1,685\text{ cm}^{-1}$ pointing to aggregation being related to intermolecular contacts. The time sequence of unfolding and aggregation processes can be visualized by looking at spectral parameters or other intrinsic protein properties, such as enzyme activity. Figure 4.9 correlates enzyme activity and the changes induced by temperature in the amide I lower-frequency band attributed to aggregation in methionine adenosyltransferase (MAT). The parallelism between the loss in activity and the aggregation band should be noted. This time sequence can also be observed using other bands, e.g. the $1,652\text{-cm}^{-1}$ due to the α -helix or the tryptophan intrinsic fluorescence (Iloro et al. 2004).

The temperature profile usually defines three regions (Fabian et al. 1999; Paquet et al. 2001; Iloro et al. 2004): region I, of full enzyme activity ($20\text{--}37^\circ\text{C}$); region II, of activity loss with concomitant changes in the activity and IR spectra ($37\text{--}55^\circ\text{C}$); and region III (above 55°C), in which activity is zero and no further changes in intrinsic

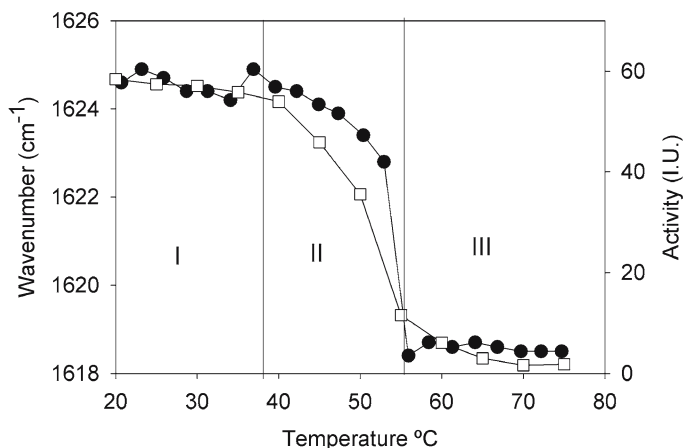


Fig. 4.9. Thermal profile of methionine adenosyltransferase (MAT). Traces corresponding to the IR $1,625\text{-cm}^{-1}$ band (circles) assigned to the oligomer interactions in MAT measured in D_2O buffer, and profiles corresponding to protein activity (squares)

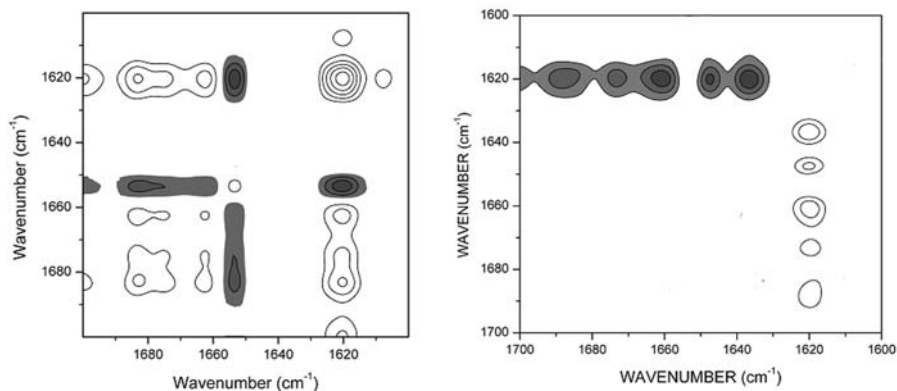


Fig. 4.10. Synchronous (*left*) and asynchronous (*right*) correlation map contour in the 1,700–1,600- cm^{-1} region of MAT IR spectra in the interval 25–37°C, in which full enzyme activity is maintained. *White peaks* correspond to positive correlations and *grey peaks* to negative ones

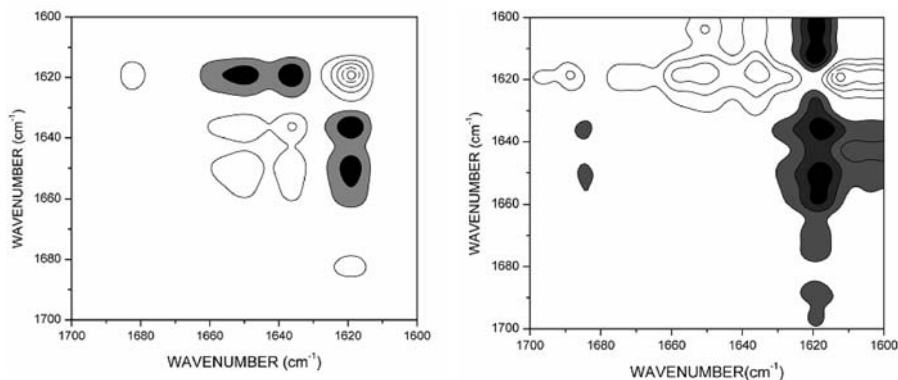


Fig. 4.11. Synchronous (*left*) and asynchronous (*right*) correlation map contour in the 1,700–1,600- cm^{-1} region of MAT in the interval 37–55°C, in which enzyme activity is being lost. *White peaks* correspond to positive correlations and *grey peaks* to negative ones

fluorescence or IR bands are detected. T_m for the loss of activity is 51°C, and the corresponding value for the decrease in intrinsic fluorescence is 50°C. The 2D synchronous (left-hand) and asynchronous (right-hand) maps corresponding to region I are depicted in Fig. 4.10. The synchronous plot shows autopeaks at 1,622, 1,652, 1,668 and 1,682 cm^{-1} indicating temperature-dependent intensity changes in these peaks. Note that no autopeak corresponding to the 1,635- cm^{-1} band (β -sheet) is present, whereas the α -helix band is clearly seen. Positive cross-correlation peaks in this re-

gion, indicating that both peaks either increase or decrease in intensity, are found for the band pairs $1,622\text{ cm}^{-1}/1,644\text{ cm}^{-1}$, $1,622\text{ cm}^{-1}/1,662\text{ cm}^{-1}$, $1,622\text{ cm}^{-1}/1,682\text{ cm}^{-1}$ and $1,662\text{ cm}^{-1}/1,682\text{ cm}^{-1}$. The negative band pairs are $1,622\text{ cm}^{-1}/1,652\text{ cm}^{-1}$, $1,652\text{ cm}^{-1}/1,662\text{ cm}^{-1}$ and $1,652\text{ cm}^{-1}/1,680\text{ cm}^{-1}$, indicating that these peak intensities follow the opposite behaviour, i.e. one increases while the other decreases. The asynchronous maps are symmetrical; the asynchronous plot corresponding to this region shows correlation peaks corresponding to the pairs $1,621\text{ cm}^{-1}/1,645\text{ cm}^{-1}$, $1,621\text{ cm}^{-1}/1,662\text{ cm}^{-1}$ and $1,621\text{ cm}^{-1}/1,675\text{ cm}^{-1}$. The three pairs are negative and their intensity is similar, indicating that they are the earliest events that take place during protein unfolding and that the change taking place at the second wavenumber of each pair occurs earlier than that occurring at the first wavenumber. Therefore, the change of the bands at $1,645$, $1,662$ and $1,675\text{ cm}^{-1}$ would be the first event in MAT thermal unfolding, indicating the loss of the oligomer interaction band at $1,624\text{ cm}^{-1}$ and the rise in the $1,617\text{-cm}^{-1}$ aggregation band. These results confirm the low-temperature start of the $1,624\text{-cm}^{-1}$ band change (Fig. 4.9). Region II comprises the major changes in activity and amide I bandshape. The correlation maps corresponding to region II are shown in Fig. 4.11. The synchronous plot shows autopeaks at $1,617$, $1,635$ and $1,652\text{ cm}^{-1}$. Positive cross-peaks are located at $1,617\text{ cm}^{-1}/1,682\text{ cm}^{-1}$ and $1,635\text{ cm}^{-1}/1,652\text{ cm}^{-1}$. Negative ones are at $1,617\text{ cm}^{-1}/1,635\text{ cm}^{-1}$ and $1,617\text{ cm}^{-1}/1,652\text{ cm}^{-1}$, indicating that now both structures, α -helix and β -sheet, are affected by aggregation/denaturation. The asynchronous map shows the main correlation peaks at $1,617\text{ cm}^{-1}/1,635\text{ cm}^{-1}$ and $1,617\text{ cm}^{-1}/1,652\text{ cm}^{-1}$, indicating again that the major event taking place in region II is the rise of the $1,617\text{-cm}^{-1}$ aggregation band, simultaneous with changes in the α and β components. Region III comprises the highest temperature range, once denaturation has taken place. The correlation maps are less complex since no further major changes in protein conformation are taking place. The synchronous map shows an autopeak at $1,617\text{ cm}^{-1}$, negative cross-correlation peaks at $1,617\text{ cm}^{-1}/1,635\text{ cm}^{-1}$ and $1,617\text{ cm}^{-1}/1,652\text{ cm}^{-1}$ and a faint positive cross-peak at $1,617\text{ cm}^{-1}/1,682\text{ cm}^{-1}$. The asynchronous spectrum does not show peaks above the noise level, a typical behaviour when only changes in intensity without bandshifts or variations in bandwidth are produced. 2D-IR correlation spectroscopy show that unfolding starts with α -helical segments and turns, located in the outer part of the protein, while an extended structure (including a β -sheet), associated with subunit contacts, unfolds at higher temperatures.

2D-IR spectroscopy has also been used for the study of the amyloidogenesis process, where the appearance of a band at a specific wavenumber associated with β -sheet structure can be correlated with the structures present in the native protein (Fabian et al. 1993, 1999; Paquet et al. 2001; Arrondo et al. 2004; Iloro et al. 2004).

4.5 Summary

2D-IR spectroscopy can be of great help in extracting more information from the spectra. 2D-IR correlation maps obtained from artificial bands with known changes in band parameters give rise to characteristic patterns. Biological systems of varied complexity will give rise to 2D-IR correlation maps that reflect the changes

induced by the perturbation and can be interpreted as a composite of the band simulation maps. The maps can be studied along the full perturbation process correlating changes in the overall pattern with changes in the properties of the protein, e.g. enzyme activity or intrinsic fluorescence, or the perturbation process can be dissected in several steps to detect early changes in the perturbation process.

Acknowledgements. This work was supported in part by grant no. BMC2002-01438 from the Ministerio de Educación y Ciencia, and grant no. 00042.310-13552/2001 from the University of the Basque Country, Spain. I.I. was a predoctoral student supported by the Basque Government.

References

- Arrondo JLR, Goñi FM (1997) Infrared spectroscopic studies of membrane lipids. In: Vergoten G, Theophanides T (eds) *Biomolecular structure and dynamics*, vol 342. Kluwer, Dordrecht, pp 229–242
- Arrondo JLR, Goñi FM (1999) Structure and dynamics of membrane proteins as studied by infrared spectroscopy. *Prog Biophys Mol Biol* 72:367–405
- Arrondo JLR, Iloro I, Aguirre J, Goñi FM (2004) A two-dimensional IR spectroscopic (2D-IR) simulation of protein conformational changes. *Spectroscopy* 18:49–58
- Barone G, Catanzano F, DelVecchio P, Giancola C, Graziano G (1997) Thermodynamics of protein stability: a family of ribonucleases. *Pure Appl Chem* 69:2307–2313
- Czarnecki MA (2000) Two-dimensional correlation spectroscopy: effect of band position, width, and intensity changes on correlation intensities. *Appl Spectrosc* 54:986–993
- Echabe I, Dornberger U, Prado A, Goñi FM, Arrondo JLR (1998) Topology of sarcoplasmic reticulum Ca²⁺-ATPase: an infrared study of thermal denaturation and limited proteolysis. *Protein Sci* 7:1172–1179
- Fabian H, Szendrei GI, Mantsch HH, Otvos L Jr (1993) Comparative analysis of human and Dutch-type Alzheimer β -amyloid peptides by infrared spectroscopy and circular dichroism. *Biochem Biophys Res Commun* 191:232–239
- Fabian H, Mantsch HH, Schultz CP (1999) Two-dimensional IR correlation spectroscopy: sequential events in the unfolding process of the lambda Cro-V55C repressor protein. *Proc Natl Acad Sci USA* 96:13153–13158
- Gericke A, Gadaleta SJ, Brauner JW, Mendelsohn R. (1996) Characterization of biological samples by two-dimensional infrared spectroscopy: simulation of frequency, bandwidth, and intensity changes. *Biospectroscopy* 2:341–351
- Griffiths PR, de Haseth JA (1986) *Fourier transform infrared spectrometry*. Wiley, Toronto
- Iloro I, Chehin R, Goñi FM, Pajares MA, Arrondo JLR (2004) Methionine adenosyltransferase alpha-helix structure unfolds at lower temperatures than beta-sheet: a 2D-IR study. *Biophys J* 86:3951–3958
- Kim H, Jeon SJ (2001) Simulations of two-dimensional electronic correlation spectra. *Bull Korean Chem Soc* 22:807–815
- Kumar TK, Yu C (2004) Monitoring protein folding at atomic resolution. *Acc Chem Res* 37:929–936
- Lefevre T, Arseneault K, Pezolet M (2004) Study of protein aggregation using two-dimensional correlation infrared spectroscopy and spectral simulations. *Biopolymers* 73:705–715

- Mayne L, Englander SW (2000) Two-state vs. multistate protein unfolding studied by optical melting and hydrogen exchange. *Protein Sci* 9:1873–1877
- Meskers S, Ruysschaert JM, Goormaghtigh E (1999) Hydrogen-deuterium exchange of streptavidin and its complex with biotin studied by 2D-attenuated total reflection Fourier transform infrared spectroscopy. *J Am Chem Soc* 121:5115–5122
- Monera OD, Kay CM, Hodges RS (1994). Protein denaturation with guanidine-hydrochloride or urea provides a different estimate of stability depending on the contributions of electrostatic interactions. *Protein Sci* 3:1984–1991
- Nabet A, Pezolet M (1997) Two-dimensional FT-IR spectroscopy: a powerful method to study the secondary structure of proteins using H-D exchange. *Appl Spectrosc* 51:466–469
- Noda I, Ozaki Y (2004) Two-dimensional correlation spectroscopy. Wiley, Chichester
- Noda I, Dowrey AE, Marcott C, Story GM, Ozaki Y (2000) Generalized two-dimensional correlation spectroscopy. *Appl Spectrosc* 54:236a–248a
- Paquet MJ, Laviolette M, Pezolet M, Auger M (2001) Two-dimensional infrared correlation spectroscopy study of the aggregation of cytochrome c in the presence of dimyristoyl-phosphatidylglycerol. *Biophys J* 81:305–312
- Pastrana-Rios B, Ocana W, Rios M, Vargas GL, Ysa G, Poynter G, Tapia J, Salisbury JL (2002) Centrin: its secondary structure in the presence and absence of cations. *Biochemistry* 41:6911–6919
- Reinstädler D, Fabian H, Backmann J, Naumann D (1996) Refolding of thermally and urea-denatured ribonuclease A monitored by time-resolved FTIR spectroscopy. *Biochemistry* 35:15822–15830
- Sasic S, Muszynski A, Ozaki Y (2001) New insight into the mathematical background of generalized two-dimensional correlation spectroscopy and the influence of mean normalization pretreatment on two-dimensional correlation spectra. *Appl Spectrosc* 55:343–349
- Shanmukh S, Dluhy RA (2004) kv correlation analysis. A quantitative two-dimensional IR correlation method for analysis of rate processes with exponential functions. *J Phys Chem A* 108:5625–5634
- Soloaga A, Ramirez JM, Goñi FM (1998) Reversible denaturation, self-aggregation, and membrane activity of *Escherichia coli* alpha-hemolysin, a protein stable in 6 M urea. *Biochemistry* 37:6387–6393
- Susi H (1969) Infrared spectra of biological macromolecules and related systems. In: Timasheff SN, Stevens L (eds) *Structure and stability of biological macromolecules*. Dekker, New York, pp 575–663
- Susi H, Timasheff SN, Stevens L (1967) Infrared spectra and protein conformations in aqueous solutions. I. The amide I band in H₂O and D₂O solutions. *J Biol Chem* 242:5460–5466
- Sutherland GBBM (1952) Infrared analysis of the structure of amino acids, polypeptides and proteins. *Adv Protein Chem* 7:291–318
- Tanford C (1968) Protein denaturation. *Adv Protein Chem* 23:121–282
- Torrecillas A, Corbalan-Garcia S, Gomez-Fernandez JC (2003) Structural study of the C2 domains of the classical PKC isoenzymes using infrared spectroscopy and two-dimensional infrared correlation spectroscopy. *Biochemistry* 42:11669–11681

NMR Spectroscopy of Large Biological Macromolecules in Solution

CÉSAR FERNÁNDEZ, GERHARD WIDER

5.1 Introduction

NMR spectroscopy is a well-established tool in structural biology. NMR constitutes an alternative technique to X-ray crystallography to obtain structural and dynamic information at atomic resolution and to study intermolecular interactions of biological macromolecules at near-physiological solution conditions (Wüthrich 1986; Ferentz and Wagner 2000; Wider 2000). Considerable efforts are being devoted to extend applications of solution NMR to larger molecular systems, because a large number of biologically important macromolecules and macromolecular complexes have molecular masses beyond the practical range amenable to conventional NMR spectroscopy in solution. Increasing this size limit allows, for instance, structure determinations of proteins that cannot be crystallized, including integral membrane proteins, investigations of intermolecular interactions involving large molecules and macromolecular assemblies, and the structure determination of larger oligonucleotides and their complexes with proteins.

During the past 2 decades, solution NMR studies of biological macromolecules have been limited to relatively small structures. This is directly reflected by the size distribution of the NMR structure entries in the Protein Data Bank, where most of the NMR structures are in the range 2–25 kDa, with an average around 10 kDa (Güntert 1998). In the 1980s, when the first 3D protein structures were elucidated by NMR (Braun et al. 1983; Williamson et al. 1985; Wüthrich 1986; Wagner et al. 1987), one of the main bottlenecks when investigating “larger” proteins was the daunting analysis of their crowded homonuclear proton–proton 2D NMR spectra. In the early 1990s, biochemical labeling of proteins with stable ^{15}N and ^{13}C isotopes (LeMaster 1994; Kainosho 1997; Gardner and Kay 1998) became a widespread solution for this problem, particularly for the structure determination of biological macromolecules with molecular masses above approximately 10 kDa. Isotope labeling in combination with heteronuclear multidimensional NMR spectroscopy, particularly the development of triple-resonance and $^{13}\text{C}/^{15}\text{N}$ -resolved 3D and 4D NMR experiments (Ikura et al. 1990; Bax and Grzesiek 1993; Sattler et al. 1999), was a significant step ahead enabling an increase of the molecular sizes amenable to NMR studies to about 20–25 kDa.

For proteins or protein complexes with molecular masses above 25–30 kDa the spectral quality rapidly deteriorates. A major limitation when working with these larger macromolecules arises from the high relaxation rates of the NMR signal, causing severe line broadening, which translates into poor spectral resolution and low

signal-to-noise ratios. Substantial improvements of the quality of NMR spectra of biological macromolecules with molecular masses above approximately 25 kDa can be obtained with deuteration, a technique that has been applied in biological NMR for more than 30 years. Combined with ^{15}N - and ^{13}C -labeling, ^2H -labeling experienced an impressive revival about 10 years ago and has become an essential tool for determining large structures in solution (for reviews, see LeMaster 1990, 1994; Gardner and Kay 1998; Goto and Kay 2000).

Deuteration extended the applicability of conventional NMR spectroscopy in solution to molecular sizes of up to 50 kDa. For even larger molecular structures, the spectral quality decreases again rapidly owing to relaxation, despite optimized partial deuteration. Full deuteration of biological macromolecules is not practical because the most valuable information comes from the hydrogen atoms. With the introduction of transverse relaxation-optimized spectroscopy (TROSY) (Pervushin et al. 1997; Wider and Wüthrich 1999; Pervushin 2000, 2003; Riek et al. 2000; Riek 2003; Wüthrich and Wider 2003), relaxation could be reduced to such an extent that satisfactory linewidths and sensitivity can be achieved in NMR experiments with very large molecules. TROSY works best with deuterated proteins and is especially suited for application to protonated amide groups. In concert with improved instrumentation, TROSY has greatly extended the size limit for macromolecules that can be studied by solution NMR. This technical progress has opened a wide range of new applications for solution NMR (Wüthrich 1998; Venters et al. 2002; Fernández and Wider 2003; Tugarinov et al. 2004) and, for instance, has made possible studies of molecular systems with molecular masses up to 1×10^6 Da (Fiaux et al. 2002; Riek et al. 2002),

In the following, the theoretical background of TROSY, important practical considerations regarding the implementation of TROSY, including isotope labeling of macromolecules, and numerous recent applications of TROSY for structural and functional studies of large biological macromolecules will be discussed.

5.2 Technical Background

5.2.1 The NMR Signal

NMR measures the response of nuclear spins in a large, homogenous magnetic field to perturbations caused by the irradiation of electromagnetic fields in the radio-frequency range (Wider 1998). In practice, a sequence of radiofrequency pulses is applied, with the pulses being separated by interpulse time periods. The response to such an NMR pulse sequence (Fig. 5.1) is the sum of the radiofrequencies that have been emitted by the nuclei.

The NMR signal decays exponentially with a characteristic time constant, the transverse relaxation time T_2 (Fig. 5.1, case A). For the analysis, the signal is Fourier transformed into a spectrum containing resonance lines that represent the various emitted radiofrequencies. The width of the resonance lines in the spectrum is inversely proportional to T_2 . The transverse relaxation time T_2 depends on the size of

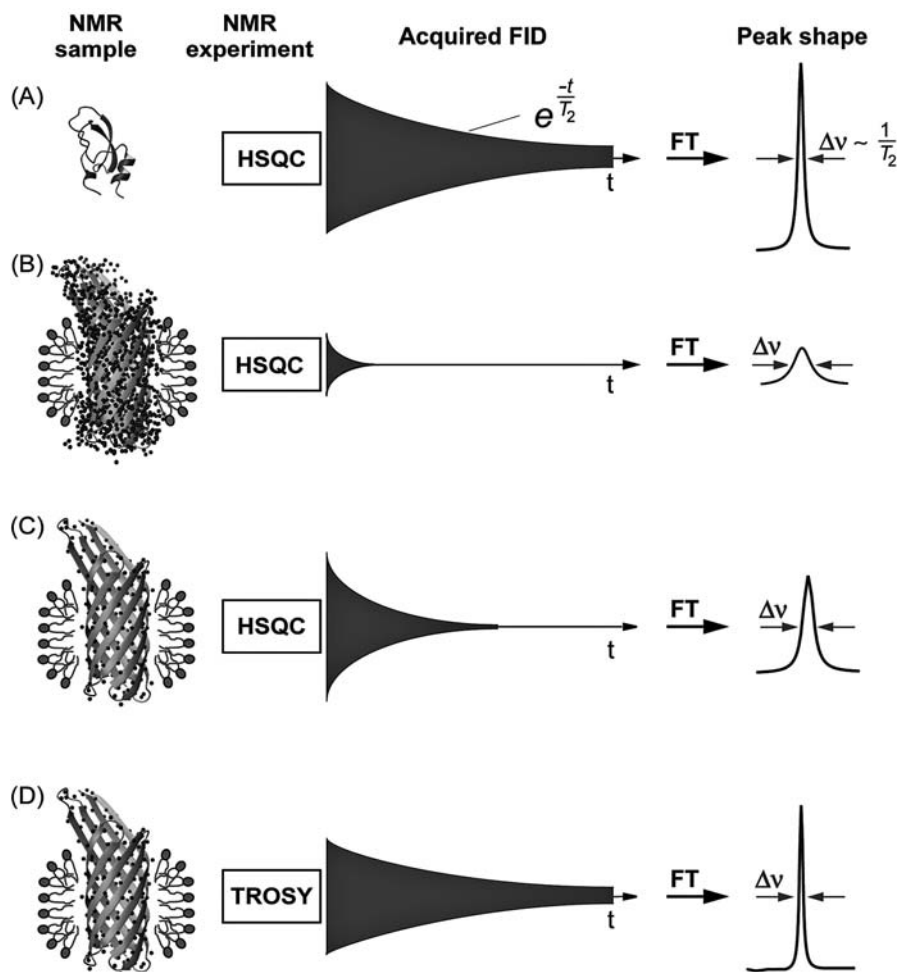


Fig. 5.1. Solution NMR spectroscopy with small and large molecules. **A** The NMR signal obtained from small molecules in solution relaxes slowly; it has a long transverse relaxation time T_2 . A large T_2 value translates into narrow linewidths ($\Delta\nu$) in the NMR spectrum after Fourier transformation of the NMR signal. *HSQC* stands for the pulse sequence used (see text). **B** For larger molecules the decay of the NMR signal is faster (T_2 is smaller). This results in a weaker signal measured after the NMR pulse sequence and in broad lines in the spectra. In the schematic structure on the *left-hand side*, the high density of protons is indicated by *black dots*. **C** By deuteration of the macromolecule (reduced number of protons indicated by a smaller number of *black dots* in the schematic structure), the transverse relaxation can be substantially reduced, which results in improved spectral resolution and improved sensitivity for large molecules. **D** With transverse relaxation-optimized spectroscopy (*TROSY*) pulse sequences on deuterated macromolecular samples the transverse relaxation can be further reduced, thereby increasing considerably the molecular mass amenable to NMR

the molecule: for increasing molecular masses, T_2 becomes shorter (fast transverse relaxation) and consequently the lines in the spectrum broaden (Fig. 5.1, case B). Transverse relaxation is the major reason for the limitation to relatively small molecular masses in standard solution NMR.

5.2.2 NMR and Molecular Size

When studying large molecules by solution NMR methods, usually three major difficulties arise (1) signal overlap, (2) fast transverse relaxation, and (3) limited solubility:

1. The large number of resonances causes signal overlap, which can make analysis of the spectra a very difficult task. The overlap of signals in the NMR spectra can be alleviated by a proper choice of isotope-labeling schemes (Kainosho 1997; Gardner and Kay 1998; Yamazaki et al. 1998; Cowburn and Muir 2001). Uniform ^{15}N - and ^{13}C -labeling enables the application of heteronuclear multidimensional NMR experiments, which reduce signal overlap by spreading out the proton resonances in additional dimensions. Moreover, selective isotope labeling techniques offer the possibility to selectively observe only particular groups of signals, which can simplify the NMR spectra considerably.
2. The transverse magnetization in larger molecules relaxes faster, which leads to line broadening and poor sensitivity in the spectra, and eventually no NMR signals can be detected at all (Fig. 5.1, case B). This effect becomes especially bothersome when performing long and relatively complex heteronuclear multidimensional NMR experiments. In such experiments, multiple nuclei are correlated by their mutual scalar couplings over several chemical bonds. For larger systems, fast transverse relaxation (equivalent to short transverse relaxation times) reduces the signal intensities beyond the detection limit before the desired signal can be measured.
3. With increasing molecular mass of the macromolecules, their solubility decreases, reducing the signal per unit of volume. New developments in NMR techniques and instrumentation that increase sensitivity, e.g., the development of cryogenic probes and spectrometers with higher magnetic fields, are contributing to alleviate this issue.

This short discussion indicates that the limitations caused by nuclear transverse relaxation poses the severest technical challenge for studies of larger biological macromolecules in solution. The combination of deuteration techniques and TROSY can alleviate the deleterious effects of transverse relaxation in such systems and can overcome, at least partially, the molecular size limitations (Fig. 5.1, cases C and D). These approaches will be discussed in the following sections, and will be the main topics presented in this review.

5.2.3 Isotope Labeling

A major source for transverse relaxation are the protons in the omnipresent hydrogen atoms. Via dipole–dipole (DD) interactions they efficiently relax NMR-active nuclei, e.g., ^1H , ^{13}C , or ^{15}N , in macromolecules. The strength of the DD interaction depends on the type of nucleus and can be greatly reduced by replacing protons by deuterons (deuteration), because of the significantly lower gyromagnetic ratio γ of ^2H compared with that of ^1H ($\gamma[^2\text{H}]/\gamma[^1\text{H}]=0.15$) (Gardner and Kay 1998). Particularly, for ^{13}C – ^1H groups deuteration causes substantial reduction of transverse ^{13}C relaxation rates, which results in increased resolution and significant sensitivity gains in experiments based on transfer of magnetization through ^{13}C nuclei (Grzesiek et al. 1993). Deuteration has clear advantages with regard to relaxation (LeMaster 1990, 1994; Kushlan and LeMaster 1993; Gardner and Kay 1998); however, there are drawbacks since protons contribute a major part of the structural information and produce the most sensitive NMR signal. Thus, a compromise has to be found and measuring completely deuterated proteins is not an option, for example, C–H moieties in macromolecules are often deuterated only to a certain extent, e.g., to 70%, or protons are selectively reintroduced into otherwise highly deuterated molecules (practical details are given in the next section).

Reintroducing protons into amide groups of a highly deuterated protein is especially simple and generally requires only dissolving the protein in H_2O . The exchangeable ^{15}N – ^1H groups will become protonated and the strategically important amide groups in the polypeptide backbone of proteins will be accessible to NMR experiments. Although deuteration of the C–H groups in biological macromolecules reduces significantly DD interactions between protons in ^{15}N – ^1H moieties and remote protons, i.e., the protons outside the ^{15}N – ^1H group (Fig. 5.1, case C), the ^{15}N – ^1H DD interactions are omnipresent in any NMR experiment with ^{15}N -labeled samples dissolved in H_2O . It has been shown that these ^{15}N – ^1H DD interactions can be reduced by spectroscopic means using TROSY (Pervushin et al. 1997) (Fig. 5.1, case D).

5.2.4 Transverse Relaxation-Optimized Spectroscopy (TROSY)

5.2.4.1 The Foundations of TROSY

The TROSY technique (Pervushin et al. 1997) is based on the interference of at least two different relaxation pathways that contribute to the relaxation of a particular nucleus. The interference can be additive or subtractive, resulting in increased or reduced relaxation, respectively. In addition to the omnipresent relaxation due to DD coupling, chemical shift anisotropy (CSA) of ^1H , ^{15}N , and ^{13}C can be a significant source of transverse relaxation at the high magnetic fields typically used for studies of biological macromolecules. This effect can be nicely illustrated in a correlation spectrum of ^{15}N and ^1H nuclei of amide groups in a polypeptide backbone (Fig. 5.2). Each ^1H nucleus couples to its directly attached ^{15}N nucleus by scalar

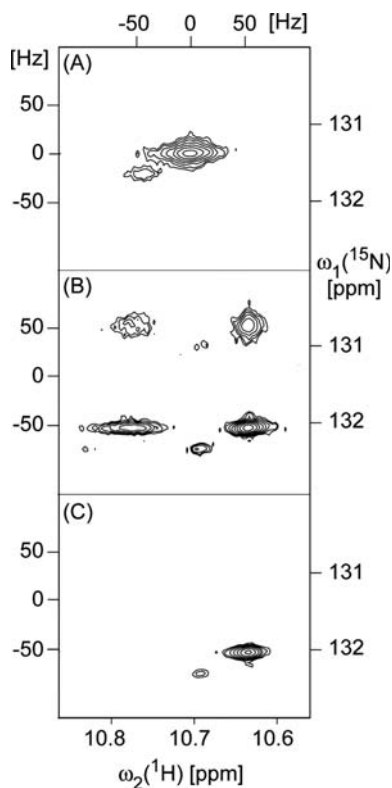


Fig. 5.2. Contour plots of a backbone amide ^{15}N - ^1H correlation peak extracted from three different types of 2D [^{15}N , ^1H] correlation experiments. **A** Conventional broadband decoupled [^{15}N , ^1H]-HSQC. **B** Same as **A** without any decoupling during the experiment. **C** [^{15}N , ^1H]-TROSY spectrum recorded with the scheme of Fig. 5.3a. (Adapted from Pervushin et al. 1997 with permission)

coupling. The ^1H NMR spectrum of such an amide moiety thus consists of two lines representing the protons attached to ^{15}N nuclei with spin up and those protons attached to ^{15}N nuclei with spin down relative to the external magnetic field. The same effect is observed for the ^{15}N NMR spectrum of an amide group: it consists of two lines representing the ^{15}N spins attached to protons with spin up or with spin down. Therefore, in a 2D correlation experiment without decoupling a four-line fine structure is observed (Fig. 5.2, panel B). In the spectrum of a large protein, the four lines have different linewidths, which is a consequence of the different relaxation rates of the four individual multiplet components owing to relaxation interference between DD interaction and CSA.

In conventional NMR experiments the four multiplet lines are collapsed by a technique called “decoupling,” however, at the cost of averaging the relaxation rates (Fig. 5.2, panel A). For smaller molecules, all four multiplet components have almost identical linewidths and decoupling results in a simplified spectrum with improved sensitivity. However, for large molecules the multiplet components have largely different linewidths (Fig. 5.2, panel B) and the decoupled signal (Fig. 5.2, panel A) may be very much attenuated owing to the contribution of the fast relaxing resonance lines.

The TROSY technique selects exclusively the slowest-relaxing component of the four-line pattern (narrowest signal in Fig. 5.2, panel B; Fig. 5.2, panel C) eliminating the faster-relaxing multiplet components. Thus, TROSY neglects part of the potential signal, which is, however, more than compensated in large molecules by the much slower relaxation during the pulse sequence and the acquisition (Fig. 5.1, case D). Generally, superior resolution and sensitivity are readily achieved with TROSY when working with molecular sizes larger than 15–20 kDa at magnetic field strengths corresponding to a proton resonance frequency of at least 700 MHz.

5.2.4.2

Field-Strength Dependence of TROSY for ^{15}N - ^1H Groups

For the amide proton of a ^{15}N - ^1H moiety, the two interfering relaxation mechanisms used in TROSY are DD relaxation between the proton and nitrogen spins and CSA relaxation of the proton. In general, DD and CSA relaxation are not equal in magnitude and cannot compensate each other. However, the DD interaction does not depend on the strength of the static magnetic field, whereas the CSA relaxation increases with larger magnetic fields. The optimal TROSY effect for one doublet component can thus be obtained by choosing the appropriate field strength, where its relaxation rate will be near zero. For amide protons in polypeptides, this “magic field” is about 23.5 T, corresponding to a proton resonance frequency of approximately 1,000 MHz. The ^{15}N nucleus in an amide moiety shows a similar interference between ^{15}N - ^1H DD interaction and its CSA, with a minimal transverse relaxation rate at magnetic field strengths corresponding to a proton resonance frequency of approximately 900 MHz, surprisingly very close to the value for the amide protons.

In practice, some deviations are expected from the “magic field” calculated for an isolated two-spin system, since the CSA varies slightly for different amide moieties. Further, residual DD coupling of amide group spins (especially of amide protons) with remote protons gives rise to relaxation that cannot be compensated by the TROSY effect. The residual relaxation can be minimized in uniformly deuterated proteins, for which remote DD couplings are limited to nearby amide protons. In general, one approaches the optimal TROSY effect for peptide ^{15}N - ^1H groups, manifested in optimal resolution and sensitivity of NMR spectra, at the highest presently available ^1H frequencies of 900 MHz (Pervushin et al. 1997) using perdeuterated samples in aqueous solutions.

5.2.4.3

Implementation of TROSY: 2D [^{15}N , ^1H]-TROSY

For the implementation of TROSY into experimental schemes it is essential that the scalar couplings between participating nuclei are not decoupled, and that only the most slowly relaxing multiplet component is selected for detection, while all other components are suppressed (Fig. 5.2). Here, we discuss the simplest scheme for the application of the TROSY principle: the 2D heteronuclear single-quantum correlation (HSQC) experiment [^{15}N , ^1H]-TROSY (Fig. 5.3a) which is based on the

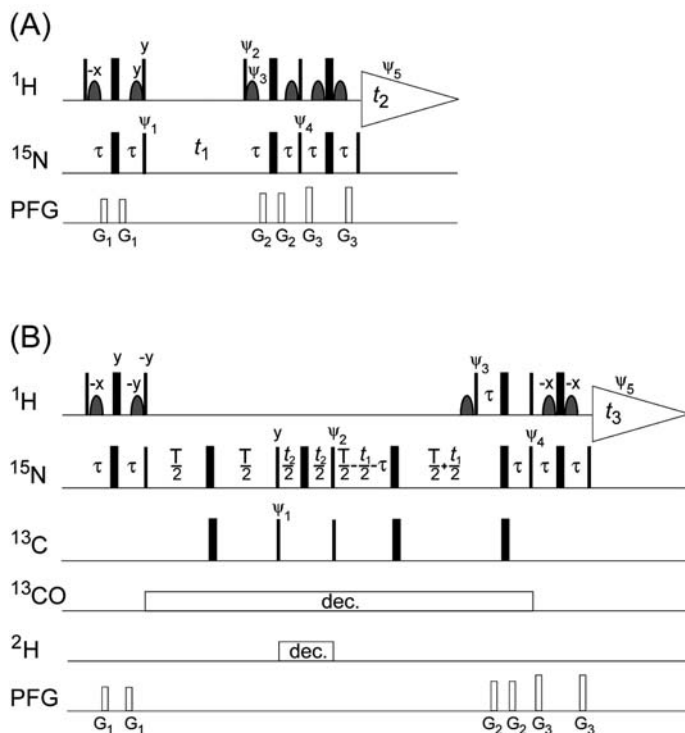


Fig. 5.3. Experimental scheme for two TROSY-type NMR experiments: **A** 2D $[^{15}\text{N}, ^1\text{H}]$ -TROSY correlation experiment; **B** 3D TROSY-HNCA triple-resonance experiment. On the lines marked ^1H , ^{13}C , and ^{15}N , narrow bars and wide bars stand for nonselective 90° and 180° radiofrequency pulses, respectively. Water saturation is minimized by keeping the water magnetization along the positive z -axis during the entire experiment, using water-selective 90° pulses (indicated by curved shapes on the line ^1H). The phases of the pulses are set to x unless otherwise indicated above the pulse. The line marked PFG indicates the pulsed magnetic field gradients applied along the z -axis. In both experiments, the ^{15}N chemical shift evolution period is denoted t_1 and the ^1H detection period is represented by a triangle. The time periods τ are set to 2.7 ms. **A** The phases for the pulses are $\psi_1=\{y, -x\}$, $\psi_2=\{-y\}$, $\psi_3=\{y\}$, $\psi_4=\{-y\}$, $\psi_5(\text{receiver})=\{y, -x\}$, and a complex interferogram is obtained by recording a second free induction decay for each increment of the t_1 time period with $\psi=\{y, -x\}$, $\psi_2=\{y\}$, $\psi_3=\{-y\}$, $\psi_4=\{y\}$. **B** The ^{13}C evolution period is t_2 , periods of broadband decoupling are represented by rectangles marked *dec.*, $T=22$ ms; the phases for the pulses are $\psi_1=\{x, x, -x, -x\}$, $\psi_2=\{y, -y, -x, x\}$, $\psi_3=\{y\}$, $\psi_4=\{y\}$, $\psi_5(\text{receiver})=\{y, -y, -x, x, -y, y, x, -x\}$. A phase-sensitive spectrum is obtained as described in Salzmann et al. (1999c)

conventional 2D $[^{15}\text{N}, ^1\text{H}]$ -HSQC (Bodenhausen and Ruben 1980). $[^{15}\text{N}, ^1\text{H}]$ -TROSY does not contain any heteronuclear decoupling, neither during the ^{15}N chemical shift evolution period (t_1), nor during the signal acquisition period (t_2). Without decoupling, the four transitions that correspond to the multiplet components with dif-

ferent relaxation rates (Fig. 5.2) are not mixed. The selection of the slowest-relaxing component and the concomitant elimination of the remaining three components is achieved by phase cycling of the radiofrequency pulses; pulsed field gradients are used to improve the spectral quality (Fig 5.3).

The ostensible signal loss by the selection of one out of four multiplet components in [^{15}N , ^1H]-TROSY can be reduced by the use of a novel polarization transfer element, ST2-PT (Pervushin et al. 1998c; Salzmann et al. 1999c). This element is applied before data acquisition (Fig. 5.3) and allows 50% of the original proton polarization to be retained rather than only 25%. However, most importantly, a higher signal intensity is detected when working with large biological macromolecules owing to the slower relaxation of the detected component. Further, without decoupling the steady-state heteronuclear magnetization can be used to enhance the sensitivity of the experiments (Riek 2001). Overall, when working with molecular sizes above 20 kDa, a higher signal-to-noise ratio is readily obtained with TROSY when compared with the corresponding conventional experiments.

5.2.4.4 [^{13}C , ^1H]-TROSY

The application of the TROSY principle is not limited to ^{15}N - ^1H groups in biological macromolecules. For example, [^{13}C , ^1H]-TROSY can be implemented in experiments with aromatic rings and methyl groups, where cross-correlated relaxation effects are also observed. In aromatic spin systems, the relaxation mechanisms of interest for optimizing ^1H and ^{13}C transverse relaxation are ^{13}C - ^1H -DD coupling and ^{13}C CSA (Pervushin et al. 1998b; Pervushin 2000). In TROSY experiments with aromatic rings, intrinsically large CSA values for ^{13}C provide efficient compensation of ^{13}C transverse relaxation by dipolar coupling to the attached proton. In contrast, the small CSA of aromatic protons is not suitable for the application of TROSY. Therefore, TROSY is used only with ^{13}C and the protons are decoupled from ^{13}C during acquisition. For aromatic ^{13}C - ^1H groups the optimal TROSY effect is observed at a ^1H frequency of 600 MHz. For ^{13}C -labeled biological macromolecules of any molecular mass, significant sensitivity enhancement in ^{13}C - ^1H correlation spectra of aromatic spin systems can readily be obtained with the use of TROSY, when compared with the widely used conventional HSQC-based experimental schemes.

Another [^{13}C , ^1H]-TROSY effect can be exploited in methyl groups, where the TROSY effect is based on different dipolar interactions that compensate each other in large molecules. The dominant relaxation of a methyl carbon originates in the dipolar coupling with the three methyl protons. These three protons have eight spin states with different combinations of spin up and spin down. Six of the eight states contain a spin-up/spin-down combination, the elements of which compensate each other with regard to dipolar interactions with the carbon (Ollerenshaw et al. 2003; Tugarinov et al. 2003); however, the relaxation due to the remaining proton is still active. For a full compensation, the remaining proton has to be brought into a double quantum state with the carbon, which eliminates their mutual dipolar interaction; for half of the carbon nuclei the other two attached protons have opposite spin states and their dipolar fields at the carbon nuclei cancel each other in

large molecules. Thus, the standard [$^{13}\text{C}, ^1\text{H}$] correlation experiment, the [$^{13}\text{C}, ^1\text{H}$] heteronuclear multiple-quantum correlation (HMQC) pulse sequence, is an optimized TROSY experiment for methyl carbons, which maintains 50% of the carbon magnetization in slowly relaxing states. For the methyl protons, the TROSY effect is much less efficient and carbon decoupling is applied during acquisition. For large molecules the 2D [$^{13}\text{C}, ^1\text{H}$]-HMQC experiment yields superior resolution and sensitivity compared with the [$^{13}\text{C}, ^1\text{H}$]-HSQC experiment, since [$^{13}\text{C}, ^1\text{H}$]-HSQC mixes fast and slowly relaxing transitions, deteriorating significantly the spectral quality. It is important to note that the dipolar interaction is field-independent and, therefore, the methyl TROSY can be applied at all field strengths for large molecules (Tugarinov and Kay 2004a).

5.2.5

Cross-Correlated Relaxation-Induced Polarization Transfer for Studies of Very Large Structures

In most of the heteronuclear NMR experiments, magnetization between the different nuclei is transferred on the basis of their scalar spin-spin couplings using insensitive nuclei enhanced by polarization transfer (INEPT) elements (Morris and Freeman 1979; Burum and Ernst 1980). During INEPT transfers, TROSY is not active since the slowly and fast relaxing transitions are mixed. For very large structures with molecular masses above approximately 150 kDa, rapid transverse relaxation during the INEPT periods leads to a virtually complete loss of intensity for most signals. This limitation can be alleviated by cross-correlated relaxation-enhanced polarization transfer (CRINEPT) (Riek et al. 1999, 2002). In this technique, INEPT and cross-correlated relaxation-induced polarization transfer (CRIPT) (Dalvit 1992) are combined. In contrast to INEPT, the transfer efficiency of CRIPT increases proportional to the size of the molecule, so that it becomes an efficient magnetization transfer mechanism for molecules with sizes above 200 kDa (Riek et al. 1999, 2002). Therefore, when studying very large structures significant gains in sensitivity can be achieved by substituting INEPT by CRINEPT or CRIPT. In addition to the improved transfer efficiency, the combination of TROSY and CRINEPT results in an experiment with transverse relaxation optimization throughout the entire pulse sequence. With this experimental setup ^1H - ^{15}N correlation spectra could be obtained for structures with molecular masses up to 900 kDa (Fiaux et al. 2002). In practice, CRINEPT and CRIPT have so far only been applied to obtain fingerprints of very large molecules. Methods that allow resonance assignments of macromolecules with molecular masses above approximately 150 kDa are currently not available.

5.2.6

The Water Resonance

NMR experiments with biological macromolecules are generally performed in aqueous solutions. A variety of techniques have been developed to prevent the extremely intense water resonance from contributing to the measured signal (Wider

1998). In NMR experiments with large macromolecules additional attention has to be given to the water suppression scheme. For example, an NMR experiment recorded with water presaturation can have up to 10 times less sensitivity than an experiment that conserves the water polarization by using a selective water flip-back pulse. Strong interactions between the water magnetization and the protein magnetization that affect the effective longitudinal proton relaxation rates are responsible for the dramatic influence of the water handling on the intensities of the protein signals (Riek et al. 2002). Therefore, it is strongly recommended to use water flip-back techniques (Grzesiek and Bax 1993) for water suppression in all experiments with large biological macromolecules, to keep the water magnetization along the $+z$ -axis during the entire duration of the measurement.

5.3 Isotope-Labeling Techniques

In NMR studies of biological macromolecules, uniform and selective isotopic labeling are applied. An overview of common isotope-labeling strategies for NMR structural and functional investigations with large proteins is given in the following sections.

5.3.1 Uniform Isotope Labeling

Uniform ^{13}C and ^{15}N isotope labeling is suitable for determining the overall structure of biological macromolecules larger than 10 kDa, whereas studies of molecules with molecular masses above 20–25 kDa require, in addition, high levels of deuteration. Isotope-labeled proteins are obtained from high-yield expression systems. Transformed *Escherichia coli* cells, yeast cells, or insect cells are used for the incorporation of stable isotopes into proteins. These cells are grown in media containing the desired isotopes in amino acid precursors and nutrients. Typically, minimal medium (Sambrook et al. 1989) containing ^{15}N ammonium salts as a source for ^{15}N , and ^{13}C -labeled glucose or ^{13}C -acetate as a ^{13}C source are used to express uniformly ^{13}C , ^{15}N -labeled proteins in *E. coli*. Alternatively, rich media containing labeled amino acids and nutrients can be purchased for high-yield protein expression in *E. coli* (e.g., Celtone, BioExpress, and Silantes-OD cell growth media). For uniform ^2H -labeling the growth medium is prepared in deuterated water, $^2\text{H}_2\text{O}$, instead of $^1\text{H}_2\text{O}$. The degree of protein deuteration can be adjusted, ranging from perdeuteration, where all carbon-bound protons are replaced by deuterons, to partial deuteration, typically in the range 50–80% deuteration. For perdeuteration, in addition to the use of $^2\text{H}_2\text{O}$ as a solvent, perdeuterated substances are used as carbon sources in the medium.

For most practical applications, ^{13}C and ^{15}N isotope incorporation rates of nearly 100% are desired in order to obtain maximal sensitivity and minimal influence of isotope effects in the NMR spectra. An exception is the application of partial ^{13}C -labeling for stereospecific assignments of methyl pairs in Val and Leu isopropyl

groups, where typically random 10% ^{13}C -labeling is used (Neri et al. 1989; Hilty et al. 2003). In contrast to this situation, deuteration levels should be tailored to the desired applications, mainly depending on the size of the protein studied and the types of experiment performed. There has to be a compromise between the decreased ^1H , ^{13}C , and ^{15}N transverse relaxation times and the decreased effective proton concentration in the samples. In perdeuterated samples only the exchangeable protons, such as the amide and hydroxyl protons, are potentially visible in the spectra. Although these samples are useful for resonance assignments in the polypeptide backbone, crucial structural information from the side chains of the proteins is not accessible. Random fractional deuteration has thus been used to obtain both backbone and side-chain information. Deuteration levels of about 50% have been found to provide a good compromise (Nietlispach et al. 1996), and seem to work well for both triple resonance experiments used for resonance assignments and nuclear Overhauser enhancement spectroscopy (NOESY) experiments used for collecting experimental distance constraints for structure determinations. However, the use is limited to proteins with molecular masses below 35–40 kDa. For larger molecules, the relatively high density of residual protons will cause efficient relaxation and eventually completely destroy the NMR signal before detection. One alternative to random deuteration is the use of methods that allow selective protonation of methyl groups, a technique discussed in the next section.

5.3.2 Selective Isotope-Labeling Strategies

Selective or specific labeling techniques are outstandingly useful tools for NMR studies of large proteins (Gardner and Kay 1998; Goto and Kay 2000). One important application is the reduction of the number of resonances in overcrowded NMR spectra, where only signals from an individual amino acid or from a group of amino acids in the sequence can then be observed. Several protocols have been developed for residue-specific labeling of proteins in a protonated background based on *E. coli* expression systems (Muchmore et al. 1989; McIntosh and Dahlquist 1990). Recently, an efficient protocol for residue-specific ^{15}N - and ^{13}C -labeling on a highly deuterated background has been proposed (Fiaux et al. 2004). This method meets the requirements of high-level deuteration, minimal cross-labeling, and high protein yield, which are crucial for NMR studies of structures with sizes above 150 kDa.

In cases where very high deuteration levels are required, selective labeling provides a means for obtaining structural information on side chains by introducing protons at strategically important positions in the structure. For example, selectively protonated methyl groups on a highly deuterated background have proven to be very useful. Several strategies for selective methyl group protonation have been developed, based on the metabolic pathways involved in synthesis of methyl-group-containing amino acids in *E. coli* (Rosen et al. 1996; Gardner and Kay 1997; Goto et al. 1999). A widely used protocol to selectively protonate methyl groups in Val, Leu, and Ile(δ 1) residues uses the isotope-labeled amino acid precursors α -ketoisovalerate and α -ketobutyrate, which are added to the growth medium before induction of protein overexpression (Goto et al. 1999). Another strategy allows protonation

of only one of the two methyl groups in Val and Leu residues (Tugarinov and Kay 2004a), and proved to be useful when working with very large molecules. Further applications of selective labeling for studies of larger biomolecules include the use of ^1H , ^{14}N -labeling of Phe, Tyr, Thr, Ile, and Val amides in a fully deuterated background (Kelly et al. 1999).

In the context of NMR screening in drug discovery, selective labeling strategies can be employed to extend the applicability to drug targets of higher molecular mass. Dual amino acid selective labeling and site directed non-native amino acid replacement (SNAAR) are labeling approaches that allow for the selective detection of NMR resonances of a specific amino acid residue (Weigelt et al. 2002a, b). In the case of dual amino acid selective labeling, amino acid X is labeled with ^{13}C and amino acid Y is labeled with ^{15}N . Provided that only one XY pair occurs in the amino acid sequence, only one signal in the HNCO spectrum (Sattler et al. 1999) will be observed. With this labeling strategy it is possible to screen selectively for binding to a selected epitope without the need of laborious sequence-specific resonance assignments.

Another approach to simplify crowded NMR spectra of large biological macromolecules is the application of segmental isotope-labeling techniques, which allow production of proteins labeled in one region or segment of their intact polypeptide sequence. NMR spectroscopy of the labeled segment of the protein, particularly TROSY- or CRINEPT-based NMR techniques, can then be performed without interference by resonances originating from the remaining parts of the polypeptide chain. By producing multiple protein samples with different labeled segments, structural studies can be performed on the intact protein in spite of its large size, which would otherwise produce very crowded NMR spectra. Segmental isotope labeling for protein NMR spectroscopy was first achieved using *trans*-splicing to combine protein domains (Yamazaki et al. 1998; Otomo et al. 1999). As an alternative, chemical ligation (Muir et al. 1998; Xu et al. 1999; Cowburn and Muir 2001; Camarero et al. 2002; Cowburn et al. 2004) has been used to assemble proteins from two or more independently generated fragments. The protein fragments are separately cloned and expressed in bacteria, and can thus be obtained independently with and without isotope labeling.

Recently, cell-free protein synthesis systems have been developed for large-scale production of protein samples for NMR. In vitro *E. coli* expression systems based on coupled transcription–translation (Yabuki et al. 1998; Kigawa et al. 1999; Kiga et al. 2002; Yokoyama 2003) have been successfully applied to produce milligram quantities of numerous biologically active proteins, including membrane proteins in detergent micelles (Klammt et al. 2004). More recently, a novel wheat germ cell-free protein synthesis system was developed that is especially suited for expression of eukaryotic proteins (Morita et al. 2003). Apart from being useful to produce uniformly ^2H , ^{13}C , ^{15}N -labeled proteins, one of the principal strengths of cell-free protein synthesis systems is the flexibility to apply any labeling pattern, including amino acid selective labeling, site-specific labeling, stereospecific deuteration of prochiral groups, and segmental isotope labeling. Furthermore, this technology may be suitable for automation and high-throughput optimization of expression conditions.

5.4

TROSY for NMR Studies of Large Biological Macromolecules

Although collection of high-quality data for structure determination of proteins with molecular masses up to approximately 100 kDa is now technically feasible, the complexity of the NMR spectra generally increases with the size of the molecule studied and the concomitant signal overlap may limit spectral analysis. Therefore, the preferred current use of relaxation-optimized NMR techniques is with large structures that yield relatively simple spectra compared with spectra of monomeric globular proteins of the same molecular size, such as homo-oligomeric proteins, individual small or medium-sized subunits in large molecular complexes, and membrane proteins in detergent micelles (Wüthrich 1998; Fernández and Wider 2003).

5.4.1

2D [^{15}N , ^1H]-TROSY

For large proteins, 2D [^{15}N , ^1H]-TROSY provides fingerprints with improved resolution and sensitivity compared with conventional experiments. This is illustrated in Fig. 5.4, panel A, which shows a [^{15}N , ^1H]-TROSY spectrum of the integral membrane protein OmpX in dihexanoylphosphatidylcholine (DHPC) micelles with an overall molecular mass of 60 kDa. 2D [^{15}N , ^1H] correlation spectra constitute a many-parameter NMR probe with high sensitivity to changes in the protein structure and environment (Hajduk et al. 1999; Pellecchia et al. 2002; Meyer and Peters 2003); therefore, they are widely used to study intermolecular interactions, either with low molecular mass ligands or with other biomolecules, such as peptides, proteins, and nucleic acids. For example, chemical shift mapping of the amide groups is used in NMR screening to detect binding of compounds that can be optimized to high-affinity ligands by a strategy named “SAR-by-NMR” (Shuker et al. 1996). 2D [^{15}N , ^1H] correlation spectra are also very useful in providing information about the quality of a protein preparation with regard to purity, folding state, stability, and aggregation properties (Woestenenk et al. 2003). In this way, NMR can be efficiently applied during the process of sample optimization at the onset of NMR studies and even X-ray crystallography studies, provided that ^{15}N -labeled and possibly deuterated protein can be produced. Moreover, 2D [^{15}N , ^1H] correlation spectra are the basis for studies of the dynamic properties of biomolecules based on spin relaxation experiments (Kay et al. 1989), and for the extraction of experimental data that yield structural information, e.g., based on amide proton exchange rates (Wüthrich 1986), coupling constants (Griesinger et al. 1999), or residual dipolar couplings (RDCs) (Prestegard et al. 2000). The introduction of the TROSY technique extends the applications based on 2D [^{15}N , ^1H] correlation experiments to much larger structures.

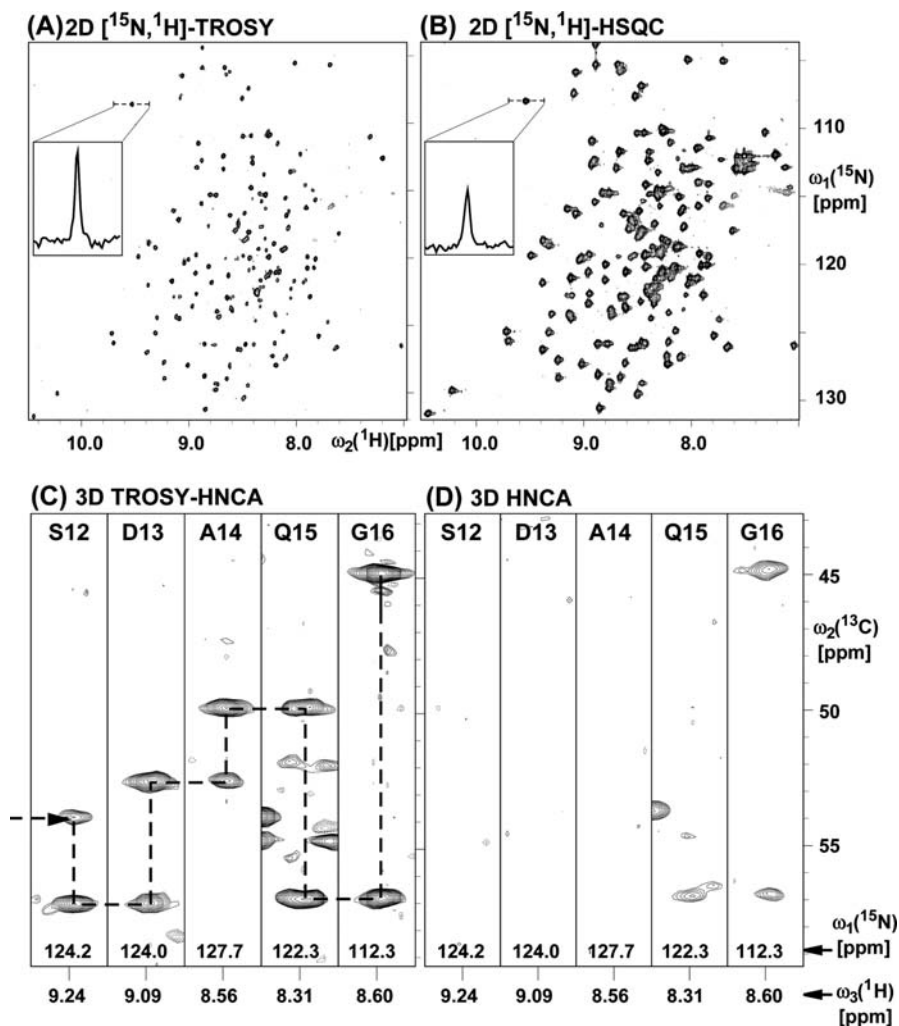


Fig. 5.4. Impact of TROSY on NMR spectra. The spectra were measured at a ^1H resonance frequency of 750 MHz with a sample of the uniformly $^2\text{H}, ^{13}\text{C}, ^{15}\text{N}$ -labeled integral membrane protein OmpX in dihexanoylphosphatidylcholine (DHPC) micelles, a 60-kDa complex. **A, B** ^{15}N - ^1H correlation spectra that were recorded and processed identically, except that TROSY was used in A only (experimental scheme in Fig. 5.3a). The *inserts* show cross sections that were taken parallel to the $\omega_2(^1\text{H})$ -axis at the position indicated by the *horizontal broken lines*. **C** Strips along the ^{13}C -dimension from a 3D $[^{15}\text{N}, ^1\text{H}]$ -TROSY-HNCA spectrum (experimental scheme in Fig. 5.3b). The strips were taken at the ^{15}N chemical shifts (indicated at the *bottom of the strips*) of the amino acid residues 12–16, and are centered on the corresponding amide proton chemical shifts, $\omega_3(^1\text{H}^N)$. *Horizontal and vertical broken lines* demonstrate the connectivities that can be obtained, leading to complete resonance assignments of backbone $^1\text{H}^N$, ^{15}N and $^{13}\text{C}^\alpha$ nuclei. **D** Corresponding strips to those in C extracted from a conventional 3D HNCA spectrum. (Reproduced from Fernández and Wider 2003 with permission)

5.4.2 [$^{13}\text{C},^1\text{H}$] Correlation Experiments

Aromatic protons provide important distance constraints for the structure determination of proteins and nucleic acids. The simplest experiment that exploits the ^{13}C - ^1H TROSY effect in aromatic systems is 2D constant-time [$^{13}\text{C},^1\text{H}$]-TROSY (Pervushin et al. 1998b), where a constant-time ^{13}C evolution period was implemented to increase resolution and to eliminate the effect of the $^1\text{J}(^{13}\text{C},^{13}\text{C})$ couplings. Figure 5.5 shows that a fourfold to tenfold signal enhancement has been achieved for resonances of the aromatic rings in an 18-kDa protein (Pervushin et al. 1998b). Similar signal enhancements have been obtained for RNA and DNA molecules (Brutscher et al. 1998; Brutscher and Simorre 2001). Based on the [$^{13}\text{C},^1\text{H}$]-TROSY building block, 3D pulse sequences can be developed that allow assignments of aromatic spin systems in proteins and nucleic acids, e.g., the constant-time TROSY-(H)CCH-correlation spectroscopy (COSY) experiment (Pervushin et al. 1998b).

TROSY can also be very profitably applied to methyl groups, which are very useful spectroscopic probes for protein structures and dynamic behavior. For example, in very large, perdeuterated biological macromolecules, NOEs with selectively protonated methyl groups constitute a valuable source of structural constraints for the calculation of the 3D structures (Gardner et al. 1997; Hilty et al. 2002; Fernández et al. 2004). Moreover, perturbations of methyl group chemical shifts can be used to qualitatively map ligand interactions and report on conformational changes upon ligand binding (Hajduk et al. 2000). For very large molecules the ^{13}C - ^1H TROSY effect in ^{13}C -labeled methyl groups can be exploited in 2D and 3D experiments for resonance assignments of methyl groups (see later) (Tugarinov and Kay 2003a), for

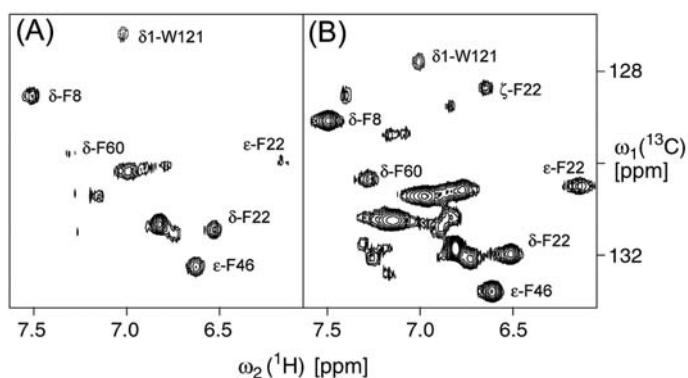


Fig. 5.5. NMR comparison of two different 2D [$^{13}\text{C},^1\text{H}$] correlation experiments performed with the uniformly ^{13}C -labeled 16-kDa protein cyclophilin A. The spectra were measured with a 1 mM protein sample, at 10°C and a ^1H -resonance frequency of 750 MHz. **A** conventional 2D [$^{13}\text{C},^1\text{H}$]-HSQC spectrum. **B** 2D constant-time [$^{13}\text{C},^1\text{H}$]-TROSY spectrum. The assignments of selected cross-peaks are indicated using the one-letter amino acid code and the sequence number. (Adapted from Pervushin et al. 1998b with permission)

stereospecific assignments of prochiral methyls and determination of the rotameric state of Val residues (Tugarinov and Kay 2004b), for studies of domain orientation based on RDCs and ligand binding (Tugarinov and Kay 2003b), and for studies of dynamic properties (Korzhnev et al. 2004). With a 2D [^{13}C , ^1H]-HMQC experiment with an inherent TROSY effect, sensitivity gains up to a factor of 3 compared with a 2D [^{13}C , ^1H]-HSQC experiment have been observed for a perdeuterated, selectively methyl- ^1H , ^{13}C -labeled protein with a rotational correlation time of 400–450 ns (equivalent to an approximately 800-kDa globular molecule at 37°C) (Tugarinov et al. 2003; Tugarinov and Kay 2004a), demonstrating the utility of methyl-TROSY for studies of very large proteins and supramolecular complexes (Kreishman-Deitrick et al. 2003).

5.4.3

TROSY for Resonance Assignments in Large Molecules

The assignment of resonances of a biological macromolecule, a procedure that attributes distinct resonance frequencies to individual nuclei, is the foundation for atomic resolution information from NMR data. Thus, after demonstrating that TROSY can provide high-quality spectra of large proteins, it was of keen interest to develop NMR experiments that can provide resonance assignments.

5.4.3.1

^{15}N , ^1H -TROSY for Sequential Assignment of Protein Backbone Resonances

Backbone resonance assignments of ^{13}C , ^{15}N -labeled proteins are usually based on triple-resonance experiments. Many of these experiments include the amide moiety and can easily be optimized for large molecules using TROSY. These experiments contain extended time periods with transverse amide proton or amide nitrogen magnetization (Fig. 5.3b) and yield a dramatic increase in signal intensities when using TROSY. The basis for most of the multidimensional TROSY experiments (e.g., Fig. 5.3b) were standard triple-resonance experiments (Ikura et al. 1990; Bax and Grzesiek 1993; Yamazaki et al. 1994; Sattler et al. 1999) that were used with ^2H , ^{13}C , ^{15}N -labeled proteins, typically up to molecular masses of about 30 kDa. For work with larger molecules, the 2D [^{15}N , ^1H]-TROSY described earlier (Fig. 5.3a) is introduced into these pulse sequences (Salzmann et al. 1998, 1999a, b; Konrat et al. 1999; Loria et al. 1999; Yang and Kay 1999b). Especially the ^{15}N constant-time evolution periods in the TROSY version of these experiments yield large sensitivity gains (Fig. 5.3b), which in turn can be used, for example, to improve spectral resolution in the ^{13}C dimension by using constant-time evolution times (Salzmann et al. 1999a). Further sensitivity improvements could be obtained by a variety of sensitivity enhancement schemes (Meissner and Sorensen 1999; Rance et al. 1999; Yang and Kay 1999a).

With TROSY techniques, resonance assignments have been obtained for much larger proteins than ever thought possible with conventional, non-TROSY NMR

techniques. This was first demonstrated with a homo-octameric 110-kDa protein (Salzmann et al. 2000), where 20–50 fold sensitivity gains were observed. For example, Fig. 5.4 shows the dramatic improvements in spectral quality that can be obtained with TROSY in a molecular complex of about 60 kDa. The HNCA triple-resonance spectrum shown in Fig. 5.4, panels C and D forms the basis for complete assignments of the backbone atoms in a polypeptide chain since it correlates the ^1H and ^{15}N nuclei of an amide moiety with the intraresidue and the preceding α -carbon nuclei. Whereas the TROSY version shows clear cross peaks, the conventional spectrum cannot at all yield the desired correlations. In general, the application of TROSY to triple-resonance experiments results in sensitivity gains of more than 1 order of magnitude with proteins in the molecular mass range from 25 to 150 kDa (Konrat et al. 1999; Salzmann et al. 1998, 1999b, 2000; Yang and Kay 1999b).

To date, sequence-specific backbone resonance assignments based on TROSY triple-resonance experiments have been described for numerous proteins. Important examples are the chemical shift assignments of the 723-residue monomeric protein malate synthase, which is the largest single polypeptide chain for which chemical shift assignments have been obtained (Tugarinov et al. 2002) (Fig. 5.6), and the NMR studies of the 91-kDa 11-meric TRAP protein (McElroy et al. 2002) and the 67-kDa dimeric form of the tumor suppression protein p53 (Mulder et al. 2000).

5.4.3.2

$[^{15}\text{N},^1\text{H}]$ -TROSY for Assignment of Protein Side-Chain Resonances

In addition to resonance assignments of nuclei in the polypeptide backbone, $[^{15}\text{N},^1\text{H}]$ -TROSY experiments have been developed for assignment of ^1H and ^{13}C chemical shifts of methyl groups in selectively methyl-protonated and otherwise deuterated large proteins. In one application, the membrane protein OmpX in 60-kDa DHPC micelles (Hilty et al. 2002) was selectively protonated at the Val, Leu, and Ile(δ 1) methyl groups (Goto et al. 1999). In a novel $[^{15}\text{N},^1\text{H}]$ TROSY-based $[^{13}\text{C},^{13}\text{C}]$ total correlation spectroscopy experiment the methyl groups were correlated with the backbone amide groups, yielding complete sequence-specific assignments of the protonated methyl groups (Hilty et al. 2002). These TROSY experiments resulted in sensitivity gains up to factor of 2.6 compared with the corresponding conventional schemes. Recently, nearly complete assignments of Val, Leu, and Ile(δ 1) methyl groups have been obtained for the 81-kDa protein malate synthase G using new labeling strategies in concert with new TROSY-type NMR experiments (Tugarinov and Kay 2003a, 2004a, b). The Val and Leu isopropyl groups were labeled with ^1H and ^{13}C in only one of the two methyl groups, and subsequently assigned using COSY-based TROSY experiments that correlate the methyl groups with the backbone amide groups. Despite the twofold dilution of methyl groups in the sample, the overall sensitivity gains were sizeable relative to those obtained with approaches in which both methyl groups are protonated.

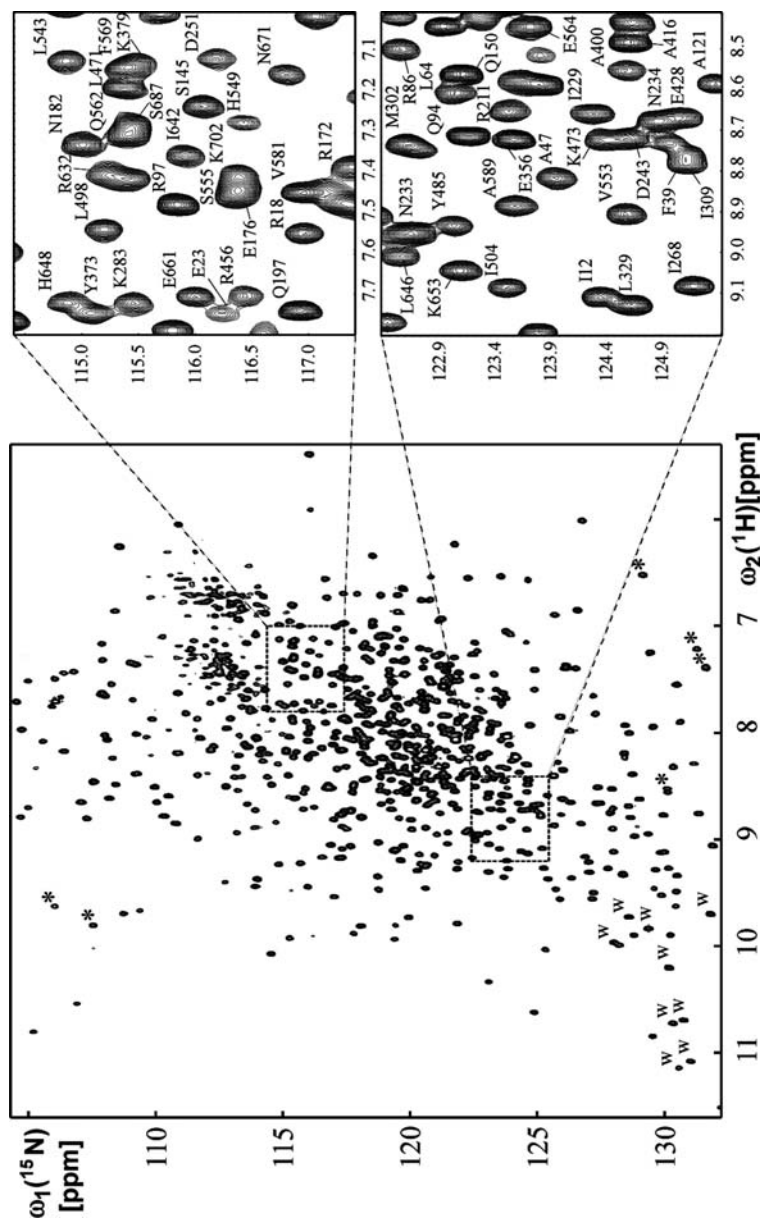


Fig. 5.6. 2D [^{15}N , ^1H]-TROSY spectrum of malate synthase G recorded at 37 °C at a ^1H resonance frequency of 800 MHz. Two regions of the spectrum are enlarged and labeled with residue numbers. Amide resonances from tryptophan side-chain resonances are labeled with *W*. Peaks aliased in the ^{15}N dimension are marked with *asterisks*. (Reproduced from Tugarinov et al. 2002 with permission)

5.4.4

TROSY for Studies of Intermolecular Interactions and Drug Design

Study of intermolecular interactions, including protein–protein, protein–nucleic acid, and protein–ligand interactions, can provide valuable information related to the physiological roles of a newly discovered protein. Moreover, these studies are of primary interest for drug discovery aiming at the design of high-affinity ligands for relevant biological molecules (Hajduk et al. 1999; Pellecchia et al. 2002; Meyer and Peters 2003). Several standard NMR experiments used to study intermolecular interaction can take advantage of TROSY, which substantially extends the molecular mass range of the targets that are amenable to NMR studies. The TROSY-type triple-resonance experiments open up the possibility to perform resonance assignments on larger protein targets, and 2D [^{15}N , ^1H]-TROSY experiments can then be used for screening. Further, TROSY can be introduced into pulse sequences used for intermolecular magnetization transfer experiments, spin-relaxation studies, or hydrogen–deuterium exchange measurements, which can be used for studies of intermolecular interactions (Pellecchia et al. 2002).

Intermolecular interactions can be detected by chemical shift mapping (Hajduk et al. 1999; Pellecchia et al. 2002; Meyer and Peters 2003) where chemical shift changes are measured upon binding of two molecules. If the assignments of the shifting resonances are known, putative contact regions in the complex can be identified and located on the 3D structure if it is available. The first application of TROSY in this context was the study of the protein–protein contacts in the 51-kDa complex formed between the type-1 pilus chaperone FimC and the pilus subunit FimH from *E. coli*, where the contact sites on FimC upon binding to FimH were identified (Pellecchia et al. 1999). A similar approach has been applied to a variety of macromolecular complexes. For example, identification of putative Ras interaction sites on the surface of the Ras binding domain in the protein kinase Byr2 in 35–40-kDa complexes (Gronwald et al. 2001), combined X-ray and TROSY-based NMR studies of the 38-kDa complexin–soluble *N*-ethylmaleimide-sensitive factor attachment protein receptor (SNARE) complex to elucidate complexin function (Chen et al. 2002), identification of the contact area of the P-domain of the lectin chaperone calreticulin and Erp57 in a 66.5-kDa complex (Frickel et al. 2002), studies of the p53 core domain bound in an approximately 200-kDa complex with Hsp90 (Rudiger et al. 2002), and TROSY-based NMR experiments to study the mechanism of ligand-mediated allosteric regulation in the 91-kDa 11-mer TRAP protein (McElroy et al. 2002). Recently, application of TROSY to malate synthase G from *E. coli*, an 81-kDa monomeric enzyme, yielded valuable quantitative information on ligand binding on the basis of chemical shift mapping, RDCs, amide proton exchange rates, and ^{15}N spin relaxation measurements (Tugarinov and Kay 2003b).

For the determination of the interfaces of large protein–protein complexes, it has been proposed to use a method based on concerted use of TROSY and isotope-labeling techniques to obtain sequential resonance assignments, and to use additional data from chemical shift mapping, from measurements of amide proton exchange rates, and from cross-saturation NMR (Takahashi et al. 2000). This approach was applied to a 64-kDa immunoglobulin complex with a domain of protein A to identify the binding sites of the complex. Another technique addresses the spectral over-

lap in larger molecular complexes, which becomes a limiting factor when studying intermolecular interactions. Since often intermolecular interactions mainly affect amide protons exposed to the solvent, a scheme selecting only these resonances has been proposed and was implemented in the experiment solvent-exposed amides with TROSY (SEA-TROSY) (Pellecchia et al. 2001). SEA-TROSY can, for example, be incorporated into standard triple-resonance experiments. The experiment has been applied to the 71-kDa, deuterium-labeled protein NADHP cytochrome P450(57–678).

5.4.5

TROSY for Observation of Scalar Couplings Across Hydrogen Bonds

Hydrogen bonds play a key role in the structure and function of biological macromolecules. Their presence is usually only indirectly inferred from experimental data, e.g., from NOEs and proton exchange rates, or from analysis of refined 3D structures based on geometrical considerations. Direct detection of hydrogen bonds in proteins and oligonucleotides in NMR spectra was enabled by the recent observation of scalar spin–spin couplings across hydrogen bonds (Dingley and Grzesiek 1998; Cordier and Grzesiek 1999; Pervushin et al. 1998a). These couplings can provide novel insights into the nature of hydrogen bonds and allow the determination of hydrogen-bond partners. The latter can be used to refine NMR structures, to study intermolecular interactions at atomic level, and to investigate biological mechanisms involving hydrogen-bonding interactions.

The sensitivity and spectral resolution in experiments used to detect hydrogen bonds can be substantially improved for large molecules when implementing TROSY in the experimental schemes. Applications of TROSY in this context include measurements of scalar couplings across hydrogen bonds in a ^{15}N , ^{13}C -labeled DNA duplex tetradecamer in a 17-kDa complex with the *Antennapedia* homeodomain (Fig. 5.7) (Pervushin et al. 1998a, 2000), in 25-kDa RNA oligonucleotides from a *Bacillus subtilis* tRNA^{Trp} A73 mutant (Yan et al. 2000), in the uniformly ^2H , ^{13}C , ^{15}N -labeled 30-kDa ribosome-inactivating protein MAP30 (Wang et al. 1999), in the 16-kDa monomeric protein superoxide dismutase (Banci et al. 2002), in the 147-residue flavoprotein riboflavin 5'-monophosphate (Löhr et al. 2000), where the scalar couplings were across NH...OP and OH...OP hydrogen bonds, and last but not least in the active site of the 44-kDa enzyme chorismate mutase (Eletsky et al. 2002). In this last example, which seems generally applicable for structural refinements, the measured scalar couplings provided unique information about the structure of the active site of the enzyme.

5.4.6

TROSY for Measurements of RDCs

In isotropic solutions, internuclear dipolar couplings average to zero owing to fast molecular tumbling. By choosing proper media that cause a very small alignment, dipolar couplings no longer average to zero and can readily be measured

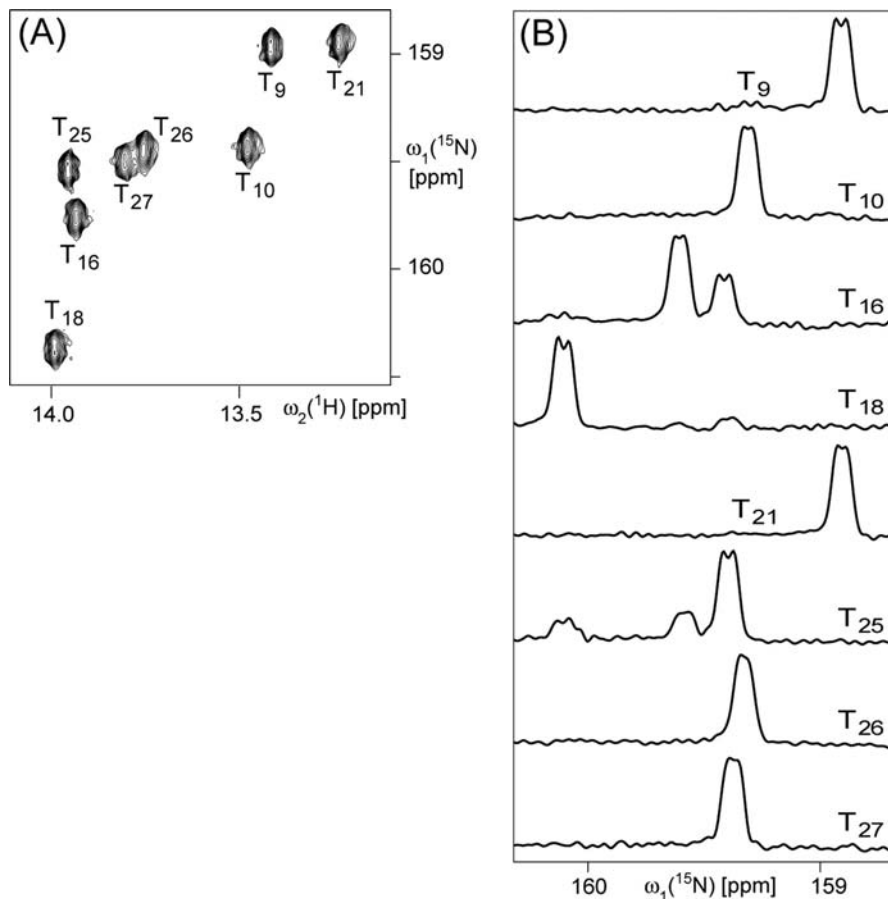


Fig. 5.7. NMR observation of ^{15}N – ^{15}N scalar couplings across hydrogen bonds in DNA, measured in an 18-kDa complex of the protein *Antennapedia* homeodomain with a uniformly ^{13}C , ^{15}N -labeled DNA duplex. **A** Contour plot from a 2D [^{15}N , ^1H]-TROSY spectrum showing the signals of the AT base pairs. **B** Cross sections along the dimension $\omega_1(^{15}\text{N})$ through the individual cross-peaks in the spectrum shown in A. The splitting of the resonance lines corresponds to scalar couplings across the Watson–Crick hydrogen bonds of about 6 Hz. (Reproduced from Pervushin et al. 2000 with permission)

(Prestegard et al. 2000, 2004; Lipsitz and Tjandra 2004). Frequently used aligning media in high-resolution NMR are, for example, phospholipid bicelles (Tjandra and Bax 1997; Tjandra et al. 1997), phage suspensions (Hansen et al. 1998, 2000), and polyacrylamide gels (Sass et al. 2000; Tycko et al. 2000). RDCs allow the determination of the relative orientation of interatomic vectors in a molecule and thus provide important structural restraints for obtaining global folds and for refining the 3D structure of proteins and oligonucleotides. This is especially important in large,

perdeuterated molecules, where only a very limited number of constraints can be obtained from NOEs (Mueller et al. 2000; Choy et al. 2001).

Various TROSY-based experimental schemes have been developed for measuring RDCs (for an overview, see Prestegard et al. 2004). Applications to the maltose-binding protein in a complex with β -cyclodextrin and to carbonic anhydrase II showed that precise RDCs can be obtained for proteins of 30–40-kDa molecular mass (Yang et al. 1999). Furthermore, RDCs in the amide groups of the protein chymotrypsin inhibitor 2 in lipid bicelles (Lerche et al. 1999) and in a ^2H , ^{15}N -labeled 53-kDa homomultimeric trimer from mannose-binding protein (Jain et al. 2003), as well as RDCs between α - and β -carbons in the 41-kDa maltose-binding protein (Evenäs et al. 2001) were measured using TROSY-based experiments. The latter application served to establish the relative orientation of the protein domains in solution. When resonance overlap in the 2D spectra becomes problematic, ^2H , ^{13}C , ^{15}N -labeled samples are required and a 3D TROSY-HNCO pulse sequence may become the method of choice, as in the case of a 723-residue enzyme, malate synthase G (Tugarinov and Kay 2003b).

5.4.7

TROSY for Studies of Dynamic Processes

In addition to structural data, NMR is able to provide information on dynamic processes at atomic resolution over a wide range of time scales (Kay et al. 1989; Palmer 2004), which potentially can help understanding structure and function relationships (Mulder et al. 2001; Eisenmesser et al. 2002). Key experiments for dynamic studies measure T_1 and T_2 relaxation times and heteronuclear $^{15}\text{N}\{^1\text{H}\}$ NOEs of the ^{15}N nuclei in amide groups. For these important experiments, pulse sequence using $[^{15}\text{N}, ^1\text{H}]$ -TROSY were developed (Zhu et al. 2000). Recently, these experiments were extended to three dimensions based on a 3D TROSY-HNCO sequence, which can be applied to resolve signal overlap in the 2D spectra (Xia et al. 2002). Please note that TROSY does not change the ^{15}N -spin relaxation parameters typically measured. A few examples of the application of $[^{15}\text{N}, ^1\text{H}]$ -TROSY experiments for the study of protein dynamics are given in Sect. 5.5.

Not only the $[^{15}\text{N}, ^1\text{H}]$ -TROSY effect can be used to obtain relaxation data, but also the methyl TROSY effect. Recently, a new experiment based on methyl-TROSY was proposed for studies of slow (millisecond) dynamic processes in high molecular mass proteins (Korzhnev et al. 2004) and was demonstrated with the 82 kDa enzyme malate synthase G, where the new approach resulted in a sensitivity gain of a factor of 5.

5.4.8

TROSY in NOESY

NOESY plays a crucial role in NMR studies of macromolecules. NOESY provides interproton distances, detects intermolecular interactions, and can support the sequential assignment procedure. Both, the ^{13}C - ^1H and the ^{15}N - ^1H TROSY effect

have been exploited in 3D ^{13}C -resolved and ^{15}N -resolved $[\text{}^1\text{H}, \text{}^1\text{H}]$ -NOESY experiments, respectively, to disperse ^1H - ^1H NOEs along the heteronuclear dimension (Brutscher et al. 1998; Zhu et al. 1999). In the pulse sequences of these experiments, TROSY-type chemical shift correlation schemes were used instead of the conventional HSQC or HMQC building blocks.

A novel TROSY effect on the ^{15}N - ^1H zero quantum state was used to develop fully relaxation optimized ^{15}N -resolved $[\text{}^1\text{H}, \text{}^1\text{H}]$ -NOESY, the 3D NOESY- $[\text{}^1\text{H}, \text{}^{15}\text{N}, \text{}^1\text{H}]$ -zero quantum-TROSY experiment (Pervushin et al. 1999). In addition to the TROSY compensation, the usually very intense diagonal peaks are almost completely suppressed. In such spectra, small resonances close to the diagonal become amenable for NMR analysis, alleviating a limitation of conventional NOESY. The utility of this approach was demonstrated for the 110-kDa protein aldolase (Pervushin et al. 1999).

5.4.9

Applications to Nucleic Acids

For NMR structural studies of nucleic acids, TROSY offers considerable advantages (for a review, see Mollova and Pardi 2000). Direct detection of hydrogen bonds and measurements of RDCs, topics that were discussed in the previous sections, are of considerable importance for the structure determination of nucleic acids, since in comparison with proteins, inherently fewer protons are available as sources for structural information. In addition, TROSY has been widely applied to increase the sensitivity and the resolution in correlation experiments and in triple-resonance experiments for ^{13}C , ^{15}N -labeled nucleic acids, increasing the range of their applicability to much larger oligonucleotides. Examples include the use of TROSY in correlation experiments (Brutscher et al. 1998) and in NOESY experiments (Brutscher et al. 2001) for the bases, in experiments that provide intrabase and sugar-to-base correlations (Fiala et al. 2000; Brutscher and Simorre 2001; Riek et al. 2001), and in an experiment that provides correlations between all carbon nuclei in the adenine base (Simon et al. 2001).

5.4.10

TROSY, CRIPT, and CRINEPT for Studies of Very Large Structures

With increasing molecular mass, most signal is lost during polarization transfer periods in heteronuclear NMR experiments where TROSY is not active. If the conventional polarization transfer element INEPT (Morris and Freeman 1979; Burum and Ernst 1980) is replaced by CRINEPT (Riek et al. 1999, 2002) and CRIPT (Dalvit 1992), the signal loss can be reduced. Using NMR experiments based on TROSY and CRINEPT or CRIPT, we can now investigate macromolecular systems with molecular masses up to 900 kDa by solution NMR (Fiaux et al. 2002).

A general strategy to assign signals in macromolecules with molecular masses above approximately 150 kDa is not (yet) available and therefore CRINEPT/CRIPT spectroscopy is usually applied to obtain fingerprints of very large structures. As

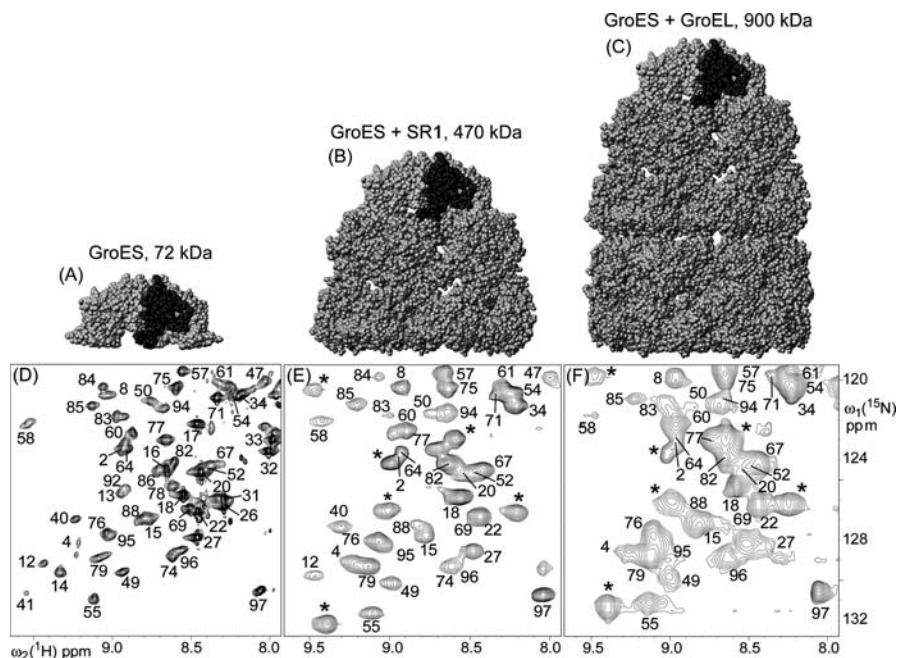


Fig. 5.8. NMR measurements of molecular complexes with very high molecular mass using TROSY. **A–C** All-atom representation of the structure of the molecules studied. **A** GroES, which is a 72-kDa homoheptameric protein (one subunit is shown in *dark grey*). **B** GroES in a complex with unlabeled SR1, with a molecular mass of 470 kDa. SR1 is a single-ring variant of GroEL. **C** GroES in a complex with unlabeled GroEL with a total molecular mass of almost 900 kDa. **D–F** 2D [^{15}N , ^1H] correlation spectra of uniformly ^2H , ^{15}N -labeled GroES in the macromolecular complexes shown in **A–C**. **D** 2D [^{15}N , ^1H]-TROSY spectrum of free GroES. **E** 2D [^{15}N , ^1H] cross-correlated relaxation-induced polarization transfer (CRIP/T) TROSY spectrum of GroES bound to SR1 in the presence of ADP. **F** 2D [^{15}N , ^1H]-CRIP/T-TROSY spectrum of GroES bound to GroEL. In **E** and **F**, the peaks that shifted significantly upon binding to SR1 or GroEL are marked with an *asterisk*. The *numbers* in **D–F** indicate the individual assignments of the resonances. (Adapted from Fiaux et al. 2002 with permission)

with TROSY-based experiments, the preferred current use of CRINEPT is to study relatively short polypeptide chains (up to approximately 100–200 amino acid residues) in large supramolecular assemblies, in complexes with large macromolecules, or in large detergent/lipid micelles. In these large structures, recording of ^{15}N - ^1H correlation spectra can be very useful for studies of intermolecular interactions and for NMR screening to search for low molecular mass ligands. For information at atomic resolution, sequence-specific resonance assignments are required. In favorable situations, assignment of the amide group chemical shifts may be obtained for smaller subunits in the absence of the whole complex, and these assignments can be transferred to the ^{15}N - ^1H correlation spectra of the whole complex (Fiaux et al.

2002). Without assignments, ^{15}N - ^1H correlation spectra still provide, for example, proof of binding events or information about dynamics of the polypeptide backbone. In both cases some of the information may be masked by spectral overlap and selective isotope-labeling techniques, e.g., segmental isotope labeling and selective amino acid labeling (see earlier), can help to simplify the spectra.

The potential of the CRINEPT and CRIPT experiments has recently been demonstrated for 900- and 500-kDa complexes formed by GroES with GroEL and SR1 (SR1 is a single-ring variant of GroEL), respectively (Fiaux et al. 2002; Riek et al. 2002) (Fig. 5.8). In this example, the sequence-specific assignments for free GroES (a heptamer of 72-kDa molecular mass) were obtained from TROSY triple-resonance experiments with a ^2H , ^{13}C , ^{15}N -labeled sample. Tentative assignments of GroES in the complexes with GroEL and SR1 were obtained by transferring assignments of free GroES to the 2D [^{15}N , ^1H]-CRIPT-TROSY spectra of GroES bound either to GroEL or to SR1. Comparison of the 2D [^{15}N , ^1H]-CRIPT-TROSY and 2D [^{15}N , ^1H]-TROSY spectra of GroES in the bound and in free form, respectively, provided structural and dynamical information on the 900-kDa GroEL-GroES complex.

5.5

Solution NMR Studies of Membrane Proteins

Membrane proteins take part in many important physiological functions, and constitute key drug targets. Structural studies of membrane proteins by X-ray crystallography or by NMR spectroscopy are much more difficult than for soluble proteins. Since real membrane systems are far too large for investigation by solution NMR experiments, membrane proteins are often reconstituted in detergent micelles. From these micellar systems, spectra can be obtained using TROSY (Arora and Tamm 2001; Fernández et al. 2001b; Fernández and Wüthrich 2003). Membrane proteins in detergent/lipid micelles yield fewer NMR resonances and thus less signal overlap than a globular protein of the corresponding molecular mass. Even though the detergent molecules may represent a large fraction of the large overall mass of the mixed micelles, proper isotope labeling, i.e., ^{13}C , ^{15}N -labeling of the protein and/or use of deuterated detergents, ensures that protein NMR signals can be detected with little or no interference from the signals of the detergent molecules.

5.5.1

Resonance Assignments and Collection of Structural Constraints for Membrane Proteins

All presently available NMR structures of larger integral membrane proteins have been determined using TROSY-based NMR techniques on uniformly ^2H , ^{13}C , ^{15}N -labeled proteins in detergent micelles (Arora and Tamm 2001; Arora et al. 2001; Fernández et al. 2001a, b, 2004; Hwang et al. 2002). For these large structures, the advantages of TROSY are particularly remarkable when performing triple-resonance experiments for backbone resonance assignment, where the application of TROSY results in sensitivity gains of more than 1 order of magnitude, as illustrated

in Fig. 5.4 for OmpX in 60-kDa DHPC micelles (OmpX/DHPC). Without the application of TROSY, only amide groups located in mobile loops in the OmpX structure were detected, e.g., Gly16 (Fig. 5.4, panel D), whereas no signals could be detected from structured regions of the protein.

On the one hand, uniform ^2H -labeling is required to make best use of the TROSY effect; on the other hand, extensive deuteration limits the obtainable structural information, e.g., NOEs can only be obtained from ^{15}N - ^1H groups. Thus, in the case of membrane proteins only low-precision structures can be obtained for β -barrel proteins (Fernández et al. 2001a), whereas for α -helical membrane proteins often only the secondary structure can be determined. With selective protonation of specific positions in the molecule, e.g., methyl groups, in a perdeuterated background, some of the NOEs can be recovered, resulting in greatly improved precision of the structure determination (Gardner et al. 1997; Gardner and Kay 1998). Sequence-specific assignment of the protonated groups can be obtained by through-bond correlation experiments to the assigned ^{15}N - ^1H resonances. This strategy was applied with deuterated [^{13}C , ^{15}N]-OmpX in DHPC micelles with protonation of the Val, Leu, and Ile($\delta 1$) methyl groups (Hilty et al. 2002). Subsequent analysis of the 3D ^{15}N -resolved [^1H , ^1H]-NOESY and ^{13}C -resolved [^1H , ^1H]-NOESY spectra yielded a fivefold increase of the number of NOE distance constraints and a concomitantly greatly improved precision of the NMR structure of OmpX/DHPC (Fernández et al. 2004).

5.5.2 3D Structure Determination

With availability of TROSY, the first NMR structures of larger integral membrane proteins were determined in the last few years (Arora and Tamm 2001; Arora et al. 2001; Fernández et al. 2001a, b, 2004; Hwang et al. 2002). The architecture of these three *E. coli* outer-membrane proteins consists of an eight-stranded antiparallel β -barrel, where sequentially successive β -strands are connected by loops on the extracellular and the periplasmic sides. The fold of the outer-membrane protein OmpX (148 residues) was obtained in DHPC micelles of about 60-kDa molecular mass (Fig. 5.9) (Fernández et al. 2001a, b, 2004; Hilty et al. 2002, 2003), the polypeptide backbone fold of the outer-membrane protein OmpA (177 residues) was determined in dodecylphosphocholine (DPC) micelles of 50-kDa molecular mass (Arora et al. 2001), and the backbone fold of the outer-membrane enzyme PagP (164 residues) was determined both in DPC and in *n*-octyl- β -D-glucoside micelles of size 50–60 kDa (Hwang et al. 2002).

The NMR spectral properties of α -helical membrane proteins are less favorable than those of the β -barrel proteins, and so far structure determination has been limited to smaller proteins (fewer than 100 amino acid residues) containing one transmembrane helix or short protein fragments. Furthermore, an appropriate folding protocol is often not available for α -helical membrane proteins. Although for the membrane-associated α -helical 29-residue polypeptide hormone glucagon in DPC micelles the secondary structure was determined by NMR in the early 1980s (Braun et al. 1983; Wider 2003), NMR structure determination of α -helical membrane

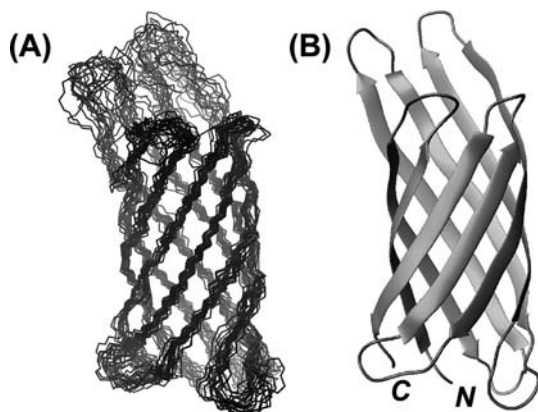


Fig. 5.9. Solution NMR structure of OmpX in DHPC micelles (Fernández et al. 2004). **A** Superposition of 20 energy-refined conformers representing the solution structure of OmpX/DHPC. **B** Ribbon representation of the solution structure of OmpX/DHPC. The figure was prepared with MOLMOL (Koradi et al. 1996)

proteins has so far yielded less complete results than for the β -barrel proteins discussed before. Examples of systems studied include the bacteriorhodopsin fragment comprising residues 1–71 (Pervushin et al. 1994), the 81-residue HIV membrane associated protein Vpu (Ma and Opella 2000), and the 80-residue membrane-transport protein from the bacterium MerF (Veglia and Opella 2000). With regard to future developments, recently reported results for the 39-kDa homotrimeric protein diacylglycerol kinase (DAGK) in micellar complexes with overall sizes larger than 100 kDa (Oxenoid et al. 2002, 2004; Sanders et al. 2002) suggest that NMR-based structure determination of membrane proteins as large and complex as some members of the G-protein-coupled receptor family may be feasible. Recently, partial backbone resonance assignments were reported for native bacteriorhodopsin in dodecylmaltoside micelles (Schubert et al. 2002).

5.5.3

Studies of Intermolecular Interactions Between a Membrane Protein and Detergent Molecules in Micelles

Studies of the solvation of the protein surface in OmpX/DHPC micelles were recently performed with NMR spectroscopy (Fernández et al. 2002; Hilty et al. 2004). In 3D ^{15}N - and ^{13}C -resolved $[^1\text{H}, ^1\text{H}]$ -NOESY spectra numerous intermolecular protein–detergent ^1H – ^1H NOEs could be identified (Fernández et al. 2002). Analysis of the intermolecular NOEs revealed that close contacts between the DHPC hydrophobic tails and the amide protons cover the surface of OmpX over a range of approximately 28 Å, centered about the middle of the β -barrel (Fig. 5.10, structure A), which coincides closely with the hydrophobic surface area of OmpX. These observations were confirmed by hydration studies using paramagnetic spin-labeled hydrophilic and lipophilic compounds (Hilty et al. 2004). In contrast, intermolecular NOEs of the polar moieties of DHPC with OmpX are limited to highly specific locations of the protein surface (Fig. 5.10, structure A). These polar interactions seem to be amino acid type specific, and such observations might eventually provide new

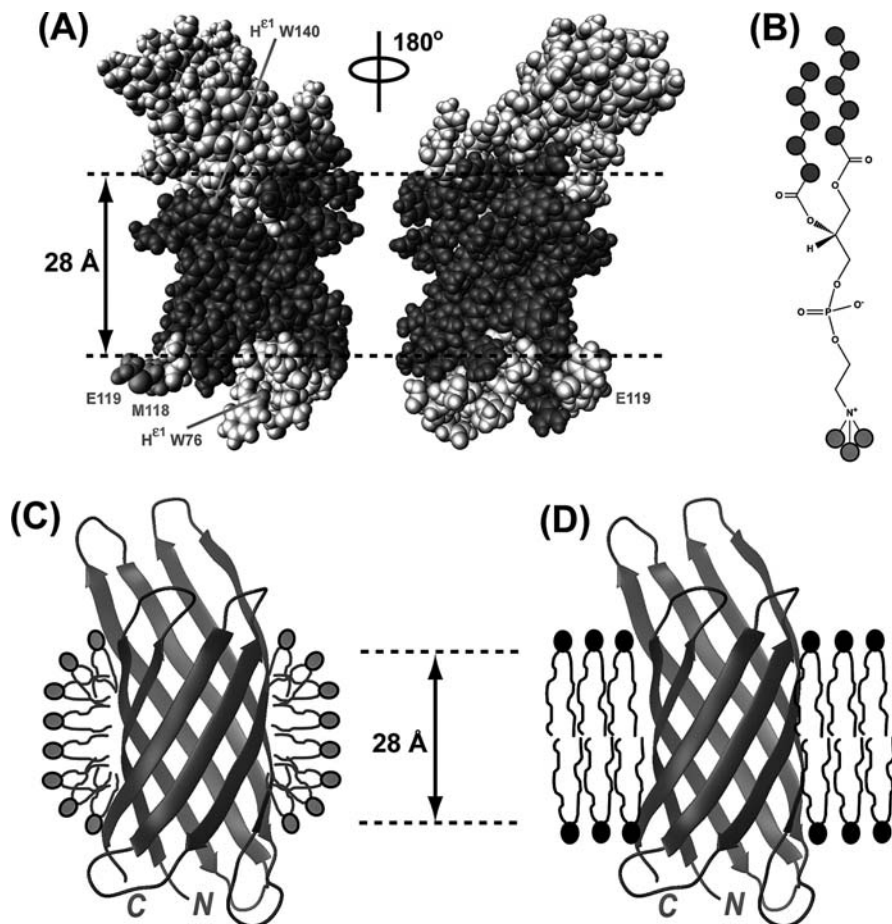


Fig. 5.10. NMR studies of lipid–protein interactions in DHPC micelles containing the integral membrane protein OmpX. **A** All-atom representation of the NMR structure of OmpX in DHPC micelles. The residues that showed nuclear Overhauser enhancements (NOEs) between protons of OmpX and the hydrophobic tails of DHPC are in dark grey. Residues with NOEs from the amide proton to the polar head methyl groups of DHPC are in light grey, and are identified with the one-letter amino acid symbol and the sequence position; for NOEs to side-chain amide protons, the atom position is also indicated. **B** Chemical structure of DHPC. Dark grey circles represent the CH_n groups of the hydrophobic tails, and light grey circles denote the polar head methyl groups. **C** Schematic drawing of OmpX/DHPC protein–lipid micelles, as inferred from the NMR data summarized in A. **D** Schematic drawing of the *Escherichia coli* outer membrane. The thickness of the hydrophobic phase in OmpX/DHPC micelles as inferred from A and the thickness of the hydrophobic phase of the *E. coli* outer membrane (Wimley 2002) are both about 28 Å

insights into the widely different behavior of membrane proteins in solutions with different detergents (Vinogradova et al. 1998; Sanders and Oxenoid 2000).

On the basis of the data for OmpX/DHPC micelles, the detergent molecules appear to form a cylindrical jacket of approximately 28 Å height around the hydrophobic surface (Fig. 5.10, structures A, C). These results are in line with previous model considerations on the embedding of membrane proteins in detergent micelles (Møller and le Maire 1993; Kleinschmidt et al. 1999). Moreover, the protein–detergent contacts provided information about the stoichiometry and size of the protein–detergent micelles. The observation that NOEs involving hydrophobic tails of DHPC form a continuous layer on the hydrophobic OmpX surface and the molecular mass estimation of the OmpX/DHPC micelles, which is in the range 50–70 kDa, confirm that the micelles contain only one OmpX molecule.

The comparison of the micellar model systems with the natural environment of the membrane proteins is very important, since membrane proteins will often need to be reconstituted in artificial milieus for 3D structure determinations. The NMR studies of OmpX/DHPC micelles show that, in spite of the different orientation of the lipid molecules on the OmpX hydrophobic surface area, a similar hydrophobic coating appears to be achieved in the two environments (Fig. 5.10, structures C, D). These observations indicate that the DHPC micellar system may be suitable also for functional studies of OmpX, since both ends of the barrel are freely accessible to the aqueous solvent. The experimental approach illustrated here with OmpX/DHPC should be applicable for structural and functional studies of other membrane proteins in micelles with different detergents or lipids.

5.5.4 Dynamic Properties of Membrane Proteins

Detailed information on dynamic properties of proteins forms an important link between static 3D structures and the biological function. For membrane proteins, TROSY-based spin relaxation experiments provided further insights into the importance of structure plasticity for membrane protein function. For example, as a peculiarity of OmpX, four of the eight β -strands extend beyond the barrel and form a protruding waving flag at the extracellular side of the β -barrel, which is believed to be involved in intermolecular interactions responsible for the biological functions of OmpX (Vogt and Schulz 1999). The NMR structure of OmpX (Fig. 5.9) shows that the protruding β -sheet is a structural feature of native OmpX in solution (Fernández et al. 2004). For a more detailed characterization of this structural feature, ^{15}N spin relaxation parameters and amide exchange rates were analyzed. These data supported the hypothesis that the plasticity and the dynamics of this β -sheet are important for optimal adaptation to the molecular targets, and may help in finding its interaction partners.

For the outer-membrane protein OmpA, TROSY-NMR relaxation data displayed a dynamic gradient extending from the well-structured central part of the barrel toward the highly mobile loops, which lead to the hypothesis that the conformational flexibility of the structure in solution may contribute to the membrane channel function of this protein (Arora and Tamm 2001; Arora et al. 2001; Tamm et

al. 2003). Finally, for the integral membrane enzyme PagP, results from NMR-based dynamic studies suggested that the protein can exist in two dynamically distinct states, and that the ability to switch between both states may be a key feature for the catalytic cycle of PagP (Hwang et al. 2004).

5.6 Conclusion and Outlook

Thanks to TROSY, CRINEPT, CRIPT, and isotope-labeling techniques, combined with recent advances in NMR instrumentation, solution NMR studies of biological macromolecules with molecular masses well beyond 100 kDa have become a reality. This has been demonstrated in numerous studies that tackle fundamental biological problems, extending from structural investigation of large proteins and the structure determination of the first larger integral membrane proteins in solution, to applications for intermolecular interactions and protein functional studies involving relatively large structures (Fiaux et al. 2002; Pellecchia et al. 2002; Fernández and Wider 2003; Fernández and Wüthrich 2003; Wüthrich and Wider 2003; Tugarinov et al. 2004). The ability to obtain resonance assignments of large biological macromolecules provides the basis for the determination of much larger 3D structures by NMR than thought possible just a few years ago. If crystal structures are available, the resonance assignment alone can be sufficient to carry out detailed studies of intermolecular interactions and investigations of dynamic processes, data that contribute important information to many interesting biological problems. However, for very large structures, where CRINEPT and CRIPT are the only means to obtain useful NMR spectra, general techniques for resonance assignments are still missing. Nevertheless, NMR fingerprints of proteins with molecular masses up to 900 kDa have been observed and interesting information could be derived based on the available crystal structure (Fiaux et al. 2002).

With the appropriate technical tools available, we expect that many new NMR structures of proteins and nucleic acids with molecular masses above 25 kDa will be elucidated. In this context, techniques that facilitate selective isotope labeling, such as cell-free protein expression (Yabuki et al. 1998; Kigawa et al. 1999; Kiga et al. 2002; Yokoyama 2003) and segmental isotope labeling (Muir et al. 1998; Yamazaki et al. 1998; Otomo et al. 1999; Xu et al. 1999; Cowburn and Muir 2001; Camarero et al. 2002; Cowburn et al. 2004), may become widely applied to solve larger structures. With the prospect of further advances in NMR techniques and instrumentation, and improved protein expression techniques, larger membrane proteins, such as G-protein coupled receptors, may become amenable to solution NMR studies. We are looking forward to many more applications of the techniques described in this review and to further progress, in the interest of answering challenging questions related to the structure and function of large biological molecules.

Acknowledgements. The Novartis Institutes for Biomedical Research (C.F.) and the Schweizerischer Nationalfonds (project 3100A0-100399) (G.W.) are gratefully acknowledged for continuous financial support. We kindly thank Vitali Tugarinov for contributing to Fig. 5.6.

References

- Arora A, Tamm LK (2001) Biophysical approaches to membrane protein structure determination. *Curr Opin Struct Biol* 11:540–547
- Arora A, Abildgaard F, Bushweller JH, Tamm LK (2001) Structure of outer membrane protein A transmembrane domain by NMR spectroscopy. *Nat Struct Biol* 8:334–338
- Banci L, Felli IC, Kummerle R (2002) Direct detection of hydrogen bonds in monomeric superoxide dismutase: biological implications. *Biochemistry* 41:2913–2920
- Bax A, Grzesiek S (1993) Methodological advances in protein NMR. *Acc Chem Res* 26:131–138
- Bodenhausen G, Ruben DJ (1980) Natural abundance ^{15}N -NMR by enhanced heteronuclear spectroscopy. *Chem Phys Lett* 69:185–189
- Braun W, Wider G, Lee KH, Wüthrich K (1983) Conformation of glucagon in a lipid-water interphase by ^1H nuclear magnetic resonance. *J Mol Biol* 169:921–948
- Brutscher B, Simorre JP (2001) Transverse relaxation optimized HCN experiment for nucleic acids: Combining the advantages of TROSY and MQ spin evolution. *J Biomol NMR* 21:367–372
- Brutscher B, Boisbouvier J, Pardi A, Marion D, Simorre JP (1998) Improved sensitivity and resolution in ^1H - ^{13}C NMR experiments of RNA. *J Am Chem Soc* 120:11845–11851
- Brutscher B, Boisbouvier J, Kupce E, Tisne C, Dardel F, Marion D, Simorre JP (2001) Base-type-selective high-resolution ^{13}C -edited NOESY for sequential assignment of large RNAs. *J Biomol NMR* 19:141–151
- Burum DP, Ernst RR (1980) Net polarization transfer via a J-ordered state for signal enhancement of low-sensitivity nuclei. *J Magn Reson* 39:163–168
- Camarero JA, Shekhtman A, Campbell EA, Chlenov M, Gruber TM, Bryant DA, Darst SA, Cowburn D, Muir TW (2002) Autoregulation of a bacterial sigma factor explored by using segmental isotopic labeling and NMR. *Proc Natl Acad Sci USA* 99:8536–8541
- Chen XC, Tomchick DR, Kovrigin E, Arac D, Machius M, Sudhof TC, Rizo J (2002) Three-dimensional structure of the complexin/SNARE complex. *Neuron* 33:397–409
- Choy WY, Tollinger M, Mueller GA, Kay LE (2001) Direct structure refinement of high molecular weight proteins against residual dipolar couplings and carbonyl chemical shift changes upon alignment: an application to maltose binding protein. *J Biomol NMR* 21:31–40
- Cordier F, Grzesiek S (1999) Direct observation of hydrogen bonds in proteins by interresidue $^3\text{h}J_{\text{NC}'}$ scalar couplings. *J Am Chem Soc* 121:1601–1602
- Cowburn D, Muir TW (2001) Segmental isotopic labeling using expressed protein ligation. *Methods Enzymol* 339:41–54
- Cowburn D, Shekhtman A, Xu R, Ottesen JJ, Muir TW (2004) Segmental isotopic labeling for structural biological applications of NMR. *Methods Mol Biol* 278:47–56
- Dalvit C (1992) Proton to ^{15}N polarization transfer via ^1H chemical-shift anisotropy- ^1H - ^{15}N dipole-dipole cross correlation. *J Magn Reson* 97:645–650
- Dingley AJ, Grzesiek S (1998) Direct observation of hydrogen bonds in nucleic acid base pairs by internucleotide $^2J_{\text{NN}}$ couplings. *J Am Chem Soc* 120:8293–8297
- Eisenmesser EZ, Bosco DA, Akke M, Kern D (2002) Enzyme dynamics during catalysis. *Science* 295:1520–1523
- Eletsky A, Heinz T, Moreira O, Kienhofer A, Hilvert D, Pervushin K (2002) Direct NMR observation and DFT calculations of a hydrogen bond at the active site of a 44 kDa enzyme. *J Biomol NMR* 24:31–39

- Evenäs J, Mittermaier A, Yang DW, Kay LE (2001) Measurement of $^{13}\text{C}^{\alpha}$ - $^{13}\text{C}^{\beta}$ dipolar couplings in ^{15}N , ^{13}C , ^2H -labeled proteins: application to domain orientation in maltose binding protein. *J Am Chem Soc* 123:2858–2864
- Ferentz AE, Wagner G (2000) NMR spectroscopy: a multifaceted approach to macromolecular structure. *Q Rev Biophys* 33:29–65
- Fernández C, Wider G (2003) TROSY in NMR studies of the structure and function of large biological macromolecules. *Curr Opin Struct Biol* 13:570–580
- Fernández C, Wüthrich K (2003) NMR solution structure determination of membrane proteins reconstituted in detergent micelles. *FEBS Lett* 555:144–150
- Fernández C, Adeishvili K, Wüthrich K (2001a) Transverse relaxation-optimized NMR spectroscopy with the outer membrane protein OmpX in dihexanoyl phosphatidylcholine micelles. *Proc Natl Acad Sci USA* 98:2358–2363
- Fernández C, Hilty C, Bonjour S, Adeishvili K, Pervushin K, Wüthrich K (2001b) Solution NMR studies of the integral membrane proteins OmpX and OmpA from *Escherichia coli*. *FEBS Lett* 504:173–178
- Fernández C, Hilty C, Wider G, Wüthrich K (2002) Lipid-protein interactions in DHPC micelles containing the integral membrane protein OmpX investigated by NMR. *Proc Natl Acad Sci USA* 99:13533–13537
- Fernández C, Hilty C, Wider G, Güntert P, Wüthrich K (2004) NMR structure of the integral membrane protein OmpX. *J Mol Biol* 336: 1211–1221
- Fiala R, Czernek J and Sklenar V (2000) Transverse relaxation optimized triple-resonance NMR experiments for nucleic acids. *J Biomol NMR* 16: 291–302
- Fiaux J, Bertelsen EB, Horwich AL, Wüthrich K (2002) NMR analysis of a 900K GroEL-GroES complex. *Nature* 418:207–211
- Fiaux J, Bertelsen EB, Horwich AL, Wüthrich K (2004) Uniform and residue-specific ^{15}N -labeling of proteins on a highly deuterated background. *J Biomol NMR* 29:289–297
- Frickel EM, Riek R, Jelesarov I, Helenius A, Wüthrich K, Ellgaard L (2002) TROSY-NMR reveals interaction between ERp57 and the tip of the calreticulin P-domain. *Proc Natl Acad Sci USA* 99:1954–1959
- Gardner KH, Kay LE (1997) Production and incorporation of ^{15}N , ^{13}C , ^2H ($^1\text{H}^{\delta 1}$ methyl) isoleucine into proteins for multidimensional NMR studies. *J Am Chem Soc* 119:7599–7600
- Gardner KH, Kay LE (1998) The use of ^2H , ^{13}C , ^{15}N multidimensional NMR to study the structure and dynamics of proteins. *Annu Rev Biophys Biomol Struct* 27:357–406
- Gardner KH, Rosen MK, Kay LE (1997) Global folds of highly deuterated, methyl-protonated proteins by multidimensional NMR. *Biochemistry* 36:1389–1401
- Goto NK, Kay LE (2000) New developments in isotope labeling strategies for protein solution NMR spectroscopy. *Curr Opin Struct Biol* 10:585–592
- Goto NK, Gardner KH, Mueller GA, Willis RC, Kay LE (1999) A robust and cost-effective method for the production of Val, Leu, Ile ($\delta 1$) methyl-protonated ^{15}N -, ^{13}C -, ^2H -labeled proteins. *J Biomol NMR* 13:369–374
- Griesinger C, Hennig M, Marino JP, Reif B, Richter C, Schwalbe H (1999) Methods for the determination of torsion angle restraints in biomacromolecules. In: Krishna RR, Berliner LJ (eds) *Biological magnetic resonance*. vol 16. Kluwer, New York, pp 259–367
- Gronwald W, Huber F, Grunewald P, Spörner M, Wohlgemuth S, Herrmann C, Kalbitzer HR (2001) Solution structure of the Ras binding domain of the protein kinase Byr2 from *Schizosaccharomyces pombe*. *Structure* 9:1029–1041

- Grzesiek S, Bax A (1993) The importance of not saturating water in protein NMR. Application to sensitivity enhancement and NOE measurements. *J Am Chem Soc* 115:12593–12594
- Grzesiek S, Anglister J, Ren H, Bax A (1993) ^{13}C line narrowing by deuterium decoupling in $^2\text{H}/^{13}\text{C}/^{15}\text{N}$ enriched proteins. Application to triple resonance 4D J connectivity of sequential amides. *J Am Chem Soc* 115:4369–4370
- Güntert P (1998) Structure calculation of biological macromolecules from NMR data. *Q Rev Biophys* 31:145–237
- Hajduk PJ, Meadows RP, Fesik SW (1999) NMR-based screening in drug discovery. *Q Rev Biophys* 32:211–240
- Hajduk PJ, Augeri DJ, Mack J, Mendoza R, Yang J, Betz SF, Fesik SW (2000) NMR-based screening of proteins containing ^{13}C -labeled methyl groups. *J Am Chem Soc* 122:7898–7904
- Hansen MR, Mueller L, Pardi A (1998) Tunable alignment of macromolecules by filamentous phage yields dipolar coupling interactions. *Nat Struct Biol* 5:1065–1074
- Hansen MR, Hanson P, Pardi A (2000) Filamentous bacteriophage for aligning RNA, DNA, and proteins for measurement of nuclear magnetic resonance dipolar coupling interactions. *Methods Enzymol* 317:220–240
- Hilty C, Fernández C, Wider G, Wüthrich K (2002) Side chain NMR assignments in the membrane protein OmpX reconstituted in DHPC micelles. *J Biomol NMR* 23:289–301
- Hilty C, Wider G, Fernández C, Wüthrich K (2003) Stereospecific assignments of the isopropyl methyl groups of the membrane protein OmpX in DHPC micelles. *J Biomol NMR* 27:377–382
- Hilty C, Wider G, Fernández C, Wüthrich K (2004) Membrane protein-lipid interactions in mixed micelles studied by NMR spectroscopy with the use of paramagnetic reagents. *Chembiochem* 5:467–473
- Hwang PM, Choy W, Lo EI, Chen L, Forman-Kay JD, Raetz CRH, Privé GG, Bishop RE, Kay LE (2002) Solution structure and dynamics of the outer membrane enzyme PagP by NMR. *Proc Natl Acad Sci USA* 99:13560–13565
- Hwang PM, Bishop RE, Kay LE (2004) The integral membrane enzyme PagP alternates between two dynamically distinct states. *Proc Natl Acad Sci USA* 101:9618–9623
- Ikura M, Kay LE, Bax A (1990) A novel approach for sequential assignment of ^1H , ^{13}C , and ^{15}N spectra of proteins: heteronuclear triple-resonance three-dimensional NMR spectroscopy. Application to calmodulin. *Biochemistry* 29:4659–4667
- Jain NU, Noble S and Prestegard JH (2003) Structural characterization of a mannose-binding protein-trimannoside complex using residual dipolar couplings. *J Mol Biol* 328:451–462
- Kainosho M (1997) Isotope labelling of macromolecules for structural determinations. *Nat Struct Biol* 4 (Suppl):858–861
- Kay LE, Torchia DA, Bax A (1989) Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease. *Biochemistry* 28:8972–8979
- Kelly MJ, Krieger C, Ball LJ, Yu Y, Richter G, Schmieder P, Bacher A, Oschkinat H (1999) Application of amino acid type-specific ^1H - and ^{14}N -labeling in a $^2\text{H},^{15}\text{N}$ -labeled background to a 47 kDa homodimer: potential for NMR structure determination of large proteins. *J Biomol NMR* 14:79–83
- Kiga D, Sakamoto K, Kodama K, Kigawa T, Matsuda T, Yabuki T, Shirouzu M, Harada Y, Nakayama H, Takio K, Hasegawa Y, Endo Y, Hirao I, Yokoyama S (2002) An engineered *Escherichia coli* tyrosyl-tRNA synthetase for site-specific incorporation of an unnatural amino acid into

- proteins in eukaryotic translation and its application in a wheat germ cell-free system. *Proc Natl Acad Sci USA* 99:9715–9720
- Kigawa T, Yabuki T, Yoshida Y, Tsutsui M, Ito Y, Shibata T, Yokoyama S (1999) Cell-free production and stable-isotope labeling of milligram quantities of proteins. *FEBS Lett* 442:15–19
- Klammt C, Lohr F, Schafer B, Haase W, Dotsch V, Ruterjans H, Glaubitz C, Bernhard F (2004) High level cell-free expression and specific labeling of integral membrane proteins. *Eur J Biochem* 271:568–580
- Kleinschmidt JH, Wiener MC, Tamm LK (1999) Outer membrane protein A of *E. coli* folds into detergent micelles, but not in the presence of monomeric detergent. *Protein Sci* 8:2065–2071
- Konrat R, Yang DW, Kay LE (1999) A 4D TROSY-based pulse scheme for correlating $^1\text{H}_i$, $^{15}\text{N}_i$, $^{13}\text{C}_i^\alpha$, $^{13}\text{C}_{i-1}$ chemical shifts in high molecular weight, ^{15}N , ^{13}C , ^2H labeled proteins. *J Biomol NMR* 15:309–313
- Koradi R, Billeter M, Wüthrich K (1996) MOLMOL: A program for display and analysis of macromolecular structures. *J Mol Graph* 14:51–55
- Korzhnev DM, Kloiber K, Kanelis V, Tugarinov V, Kay LE (2004) Probing slow dynamics in high molecular weight proteins by methyl-TROSY NMR spectroscopy: application to a 723-residue enzyme. *J Am Chem Soc* 126:3964–3973
- Kreishman-Deitrick M, Egile C, Hoyt DW, Ford JJ, Li R, Rosen MK (2003) NMR analysis of methyl groups at 100–500 kDa: model systems and Arp2/3 complex. *Biochemistry* 42:8579–8586
- Kushlan DM, LeMaster DM (1993) Resolution and sensitivity enhancement of heteronuclear correlation for methylene resonances via ^2H enrichment and decoupling. *J Biomol NMR* 3:701–708
- LeMaster DM (1990) Deuterium labelling in NMR structural analysis of larger proteins. *Q Rev Biophys* 23:133–174
- LeMaster DM (1994) Isotope labeling in solution protein assignment and structural analysis. *Prog NMR Spectrosc* 26:371–419
- Lerche MH, Meissner A, Poulsen FM, Sørensen OW (1999) Pulse sequences for measurement of one-bond ^{15}N - ^1H coupling constants in the protein backbone. *J Magn Reson* 140:259–263
- Lipsitz RS, Tjandra N (2004) Residual dipolar couplings in NMR structure analysis. *Annu Rev Biophys Biomol Struct* 33:387–413
- Löhr F, Mayhew SG, Ruterjans H (2000) Detection of scalar couplings across NH...OP and OH...OP hydrogen bonds in a flavoprotein. *J Am Chem Soc* 122:9289–9295
- Loria JP, Rance M, Palmer AG (1999) Transverse-relaxation-optimized (TROSY) gradient-enhanced triple-resonance NMR spectroscopy. *J Magn Reson* 141:180–184
- Ma C, Opella SJ (2000) Lanthanide ions bind specifically to an added „EF-hand“ and orient a membrane protein in micelles for solution NMR spectroscopy. *J Magn Reson* 146:381–384
- McElroy C, Manfreda A, Wendt A, Gollnick P, Foster M (2002) TROSY-NMR studies of the 91 kDa TRAP protein reveal allosteric control of a gene regulatory protein by ligand-altered flexibility. *J Mol Biol* 323:463–473
- McIntosh LP, Dahlquist FW (1990) Biosynthetic incorporation of ^{15}N and ^{13}C for assignment and interpretation of nuclear magnetic resonance spectra of proteins. *Q Rev Biophys* 23:1–38
- Meissner A, Sørensen OW (1999) The role of coherence transfer efficiency in design of TROSY-type multidimensional NMR experiments. *J Magn Reson* 139:439–442

- Meyer B, Peters T (2003) NMR spectroscopy techniques for screening and identifying ligand binding to protein receptors. *Angew Chem Int Ed Engl* 42:864–890
- Møller JV, le Maire M (1993) Detergent binding as a measure of hydrophobic surface-area of integral membrane proteins. *J Biol Chem* 268:18659–18672
- Mollova ET, Pardi A (2000) NMR solution structure determination of RNAs. *Curr Opin Struct Biol* 10:298–302
- Morita EH, Sawasaki T, Tanaka R, Endo Y, Kohno T (2003) A wheat germ cell-free system is a novel way to screen protein folding and function. *Protein Sci* 12:1216–1221
- Morris GA, Freeman R (1979) Enhancement of NMR signals by polarization transfer. *J Am Chem Soc* 101:760–762
- Muchmore DC, McIntosh LP, Russell CB, Anderson DE, Dahlquist FW (1989) Expression and ^{15}N labeling of proteins for proton and ^{15}N nuclear magnetic resonance. *Methods Enzymol* 177:44–73
- Mueller GA, Choy WY, Yang D, Forman-Kay JD, Venters RA, Kay LE (2000) Global folds of proteins with low densities of NOEs using residual dipolar couplings: application to the 370-residue maltodextrin-binding protein. *J Mol Biol* 300:197–212
- Muir TW, Sondhi D, Cole PA (1998) Expressed protein ligation: a general method for protein engineering. *Proc Natl Acad Sci USA* 95:6705–6710
- Mulder FA, Ayed A, Yang DW, Arrowsmith CH, Kay LE (2000) Assignment of $^1\text{H}^{\text{N}}$, ^{15}N , $^{13}\text{C}^{\alpha}$, ^{13}CO and $^{13}\text{C}^{\beta}$ resonances in a 67 kDa p53 dimer using 4D-TROSY NMR spectroscopy. *J Biomol NMR* 18:173–176
- Mulder FA, Mittermaier A, Hon B, Dahlquist FW, Kay LE (2001) Studying excited states of proteins by NMR spectroscopy. *Nat Struct Biol* 8:932–935
- Neri D, Szyperski T, Otting G, Senn H, Wüthrich K (1989) Stereospecific nuclear magnetic resonance assignments of the methyl groups of valine and leucine in the DNA-binding domain of the 434 repressor by biosynthetically directed fractional ^{13}C labeling. *Biochemistry* 28:7510–7516
- Nietispach D, Clowes RT, Broadhurst RW, Ito Y, Keeler J, Kelly M, Ashurst J, Oschkinat H, Dommelle PJ, Laue ED (1996) An approach to the structure determination of larger proteins using triple resonance NMR experiments in conjunction with random fractional deuteration. *J Am Chem Soc* 118:407–415
- Ollerenshaw JE, Tugarinov V, Kay LE (2003) Methyl TROSY: explanation and experimental verification. *Magn Reson Chem* 41:843–852
- Otomo T, Teruya K, Uegaki K, Yamazaki T, Kyogoku Y (1999) Improved segmental isotope labeling of proteins and application to a larger protein. *J Biomol NMR* 14:105–114
- Oxenoid K, Sönnichsen FD, Sanders CR (2002) Topology and secondary structure of the N-terminal domain of diacylglycerol kinase. *Biochemistry* 41:12876–12882
- Oxenoid K, Kim HJ, Jacob J, Sönnichsen FD, Sanders CR (2004) NMR assignments for a helical 40 kDa membrane protein. *J Am Chem Soc* 126:5048–5049
- Palmer AG III (2004) NMR characterization of the dynamics of biomacromolecules. *Chem Rev* 104:3623–3640
- Pellecchia M, Sebbel P, Hermanns U, Wüthrich K, Glockshuber R (1999) Pilus chaperone FimC-adhesin FimH interactions mapped by TROSY-NMR. *Nat Struct Biol* 6:336–339
- Pellecchia M, Meininger D, Shen AL, Jack R, Kasper CB, Sem DS (2001) SEA-TROSY (solvent exposed amides with TROSY): a method to resolve the problem of spectral overlap in very large proteins. *J Am Chem Soc* 123:4633–4634

- Pellecchia M, Sem DS, Wüthrich K (2002) NMR in drug discovery. *Nat Rev Drug Discov* 1:211–219
- Pervushin K (2000) Impact of transverse relaxation optimized spectroscopy (TROSY) on NMR as a technique in structural biology. *Q Rev Biophys* 33:161–197
- Pervushin K (2003) Transverse relaxation optimized spectroscopy. *Biol Magn Reson* 20:3–34
- Pervushin K, Orekhov VY, Popov AI, Musina LY, Arseniev AS (1994) Three-dimensional structure of (1-71)bacterioopsin solubilized in methanol/chloroform and SDS micelles determined by ^{15}N - ^1H Heteronuclear NMR-spectroscopy. *Eur J Biochem* 219:571–583
- Pervushin K, Riek R, Wider G, Wüthrich K (1997) Attenuated T_2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci USA* 94:12366–12371
- Pervushin K, Ono A, Fernández C, Szyperski T, Kainosho M, Wüthrich K (1998a) NMR scalar couplings across Watson–Crick base pair hydrogen bonds in DNA observed by transverse relaxation optimized spectroscopy. *Proc Natl Acad Sci USA* 95:14147–14151
- Pervushin K, Riek R, Wider G, Wüthrich K (1998b) Transverse relaxation-optimized spectroscopy (TROSY) for NMR studies of aromatic spin systems in ^{13}C -labeled proteins. *J Am Chem Soc* 120:6394–6400
- Pervushin K, Wider G, Wüthrich K (1998c) Single transition-to-single transition polarization transfer (ST2-PT) in $[\text{}^{15}\text{N}, \text{}^1\text{H}]$ -TROSY. *J Biomol NMR* 12:345–348
- Pervushin K, Wider G, Riek R, Wüthrich K (1999) The 3D NOESY- $[\text{}^1\text{H}, \text{}^{15}\text{N}, \text{}^1\text{H}]$ -ZQ-TROSY NMR experiment with diagonal peak suppression. *Proc Natl Acad Sci USA* 96:9607–9612
- Pervushin K, Fernández C, Riek R, Ono A, Kainosho M, Wüthrich K (2000) Determination of $^{\text{h}2}J_{\text{NN}}$ and $^{\text{h}1}J_{\text{HN}}$ coupling constants across Watson–Crick base pairs in the *Antennapedia* homeodomain-DNA complex using TROSY. *J Biomol NMR* 16:39–46
- Prestegard JH, Al-Hashimi HM, Tolman JR (2000) NMR structures of biomolecules using field oriented media and residual dipolar couplings. *Q Rev Biophys* 33: 371–424
- Prestegard JH, Bougault CM, Kishore AI (2004) Residual dipolar couplings in structure determination of biomolecules. *Chem Rev* 104: 3519–3540
- Rance M, Loria JP, Palmer AG (1999) Sensitivity improvement of transverse relaxation-optimized spectroscopy. *J Magn Reson* 136:92–101
- Riek R (2001) Enhancement of the steady-state magnetization in TROSY experiments. *J Biomol NMR* 21:99–105
- Riek R (2003) TROSY: transverse relaxation-optimized spectroscopy. *Meth Princ Med Chem* 16:227–241
- Riek R, Wider G, Pervushin K, Wüthrich K (1999) Polarization transfer by cross-correlated relaxation in solution NMR with very large molecules. *Proc Natl Acad Sci USA* 96:4918–4923
- Riek R, Pervushin K, Wüthrich K (2000) TROSY and CRINEPT: NMR with large molecular and supramolecular structures in solution. *Trends Biochem Sci* 25:462–468
- Riek R, Pervushin K, Fernández C, Kainosho M, Wüthrich K (2001) $[\text{}^{13}\text{C}, \text{}^{13}\text{C}]$ - and $[\text{}^{13}\text{C}, \text{}^1\text{H}]$ -TROSY in a triple resonance experiment for ribose-base and intrabase correlations in nucleic acids. *J Am Chem Soc* 123:658–664
- Riek R, Fiaux J, Bertelsen EB, Horwich AL, Wüthrich K (2002) Solution NMR techniques for large molecular and supramolecular structures. *J Am Chem Soc* 124:12144–12153
- Rosen MK, Gardner KH, Willis RC, Parris WE, Pawson T, Kay LE (1996) Selective methyl group protonation of perdeuterated proteins. *J Mol Biol* 263:627–636

- Rudiger S, Freund SMV, Veprincev DB, Fersht AR (2002) CRINEPT-TROSY NMR reveals p53 core domain bound in an unfolded form to the chaperone Hsp90. *Proc Natl Acad Sci USA* 99:11085–11090
- Salzmann M, Pervushin K, Wider G, Senn H, Wüthrich K (1998) TROSY in triple-resonance experiments: new perspectives for sequential NMR assignment of large proteins. *Proc Natl Acad Sci USA* 95:13585–13590
- Salzmann M, Pervushin K, Wider G, Senn H, Wüthrich K (1999a) [^{13}C]-constant-time [$^{15}\text{N},^1\text{H}$]-TROSY-HNCA for sequential assignments of large proteins. *J Biomol NMR* 14:85–88
- Salzmann M, Wider G, Pervushin K, Senn H, Wüthrich K (1999b) TROSY-type triple-resonance experiments for sequential NMR assignments of large proteins. *J Am Chem Soc* 121:844–848
- Salzmann M, Wider G, Pervushin K, Wüthrich K (1999c) Improved sensitivity and coherence selection for [$^{15}\text{N},^1\text{H}$]-TROSY elements in triple resonance experiments. *J Biomol NMR* 15:181–184
- Salzmann M, Pervushin K, Wider G, Senn H, Wüthrich K (2000) NMR assignment and secondary structure determination of an octameric 110 kDa protein using TROSY in triple resonance experiments. *J Am Chem Soc* 122:7543–7548
- Sambrook J, Fritsch EF, Maniatis T (1989) *Molecular cloning: a laboratory manual*. Cold Spring Harbor, New York
- Sanders CR, Oxenoid K (2000) Customizing model membranes and samples for NMR spectroscopic studies of complex membrane proteins. *Biochim Biophys Acta* 1508:129–145
- Sanders CR, Sonnichsen FD, Oxenoid K (2002) In: *Proceedings of the XXth international conference on magnetic resonance in biological systems*, Toronto, 25–30 August, p 65
- Sass HJ, Musco G, Stahl SJ, Wingfield PT, Grzesiek S (2000) Solution NMR of proteins within polyacrylamide gels: diffusional properties and residual alignment by mechanical stress or embedding of oriented purple membranes. *J Biomol NMR* 18:303–309
- Sattler M, Schleucher J, Griesinger C (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients. *Prog NMR Spectrosc* 34:93–158
- Schubert M, Kolbe M, Kessler B, Oesterhelt D, Schmiieder P (2002) Heteronuclear multidimensional NMR spectroscopy of solubilized membrane proteins: resonance assignment of native bacteriorhodopsin. *Chembiochem* 3:1019–1023
- Shuker SB, Hajduk PJ, Meadows RP, Fesik SW (1996) Discovering high-affinity ligands for proteins: SAR by NMR. *Science* 274:1531–1534
- Simon B, Zanier K, Sattler M (2001) A TROSY relayed HCCH-COSY experiment for correlating adenine H2/H8 resonances in uniformly ^{13}C -labeled RNA molecules. *J Biomol NMR* 20:173–176
- Takahashi H, Nakanishi T, Kami K, Arata Y, Shimada I (2000) A novel NMR method for determining the interfaces of large protein-protein complexes. *Nat Struct Biol* 7:220–223
- Tamm LK, Abildgaard F, Arora A, Blad H, Bushweller JH (2003) Structure, dynamics and function of the outer membrane protein A (OmpA) and influenza hemagglutinin fusion domain in detergent micelles by solution NMR. *FEBS Lett* 555:139–143
- Tjandra N, Bax A (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* 278:1111–1114
- Tjandra N, Omichinski JG, Gronenborn AM, Clore GM, Bax A (1997) Use of dipolar ^1H - ^{15}N and ^1H - ^{13}C couplings in the structure determination of magnetically oriented macromolecules in solution. *Nat Struct Biol* 4:732–738

- Tugarinov V, Kay LE (2003a) Ile, Leu, and Val methyl assignments of the 723-residue malate synthase G using a new labeling strategy and novel NMR methods. *J Am Chem Soc* 125:13868–13878
- Tugarinov V, Kay LE (2003b) Quantitative NMR studies of high molecular weight proteins: application to domain orientation and ligand binding in the 723-residue enzyme malate synthase G. *J Mol Biol* 327:1121–1133
- Tugarinov V, Kay LE (2004a) An isotope labeling strategy for methyl TROSY spectroscopy. *J Biomol NMR* 28:165–172
- Tugarinov V, Kay LE (2004b) Stereospecific NMR assignments of prochiral methyls, rotameric states and dynamics of valine residues in malate synthase G. *J Am Chem Soc* 126:9827–9836
- Tugarinov V, Muhandiram R, Ayed A, Kay LE (2002) Four-dimensional NMR spectroscopy of a 723-residue protein: chemical shift assignments and secondary structure of malate synthase G. *J Am Chem Soc* 124:10025–10035
- Tugarinov V, Hwang PM, Ollerenshaw JE, Kay LE (2003) Cross-correlated relaxation enhanced ^1H - ^{13}C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. *J Am Chem Soc* 125:10420–10428
- Tugarinov V, Hwang PM, Kay LE (2004) Nuclear magnetic resonance spectroscopy of high-molecular-weight proteins. *Annu Rev Biochem* 73:107–146
- Tycko R, Blanco FJ, Ishii Y (2000) Alignment of biopolymers in strained gels: a new way to create detectable dipole-dipole couplings in high-resolution biomolecular NMR. *J Am Chem Soc* 122:9340–9341
- Veglia G, Opella SJ (2000) Lanthanide ion binding to adventitious sites aligns membrane proteins in micelles for solution NMR spectroscopy. *J Am Chem Soc* 122:11733–11734
- Venters RA, Thompson R, Cavanagh J (2002) Current approaches for the study of large proteins by NMR. *J Mol Struct* 602:275–292
- Vinogradova O, Sönnichsen F, Sanders CR (1998) On choosing a detergent for solution NMR studies of membrane proteins. *J Biomol NMR* 11:381–386
- Vogt J, Schulz GE (1999) The structure of the outer membrane protein OmpX from *Escherichia coli* reveals possible mechanisms of virulence. *Struct Fold Des* 7:1301–1309
- Wagner G, Braun W, Havel TF, Schaumann T, Go N, Wüthrich K (1987) Protein structures in solution by nuclear magnetic resonance and distance geometry. The polypeptide fold of the basic pancreatic trypsin inhibitor determined using two different algorithms, DISGEO and DISMAN. *J Mol Biol* 196:611–639
- Wang YX, Jacob J, Cordier F, Wingfield P, Stahl SJ, Lee-Huang S, Torchia D, Grzesiek S, Bax A (1999) Measurement of $^3\text{h}J_{\text{NC}}$ connectivities across hydrogen bonds in a 30 kDa protein. *J Biomol NMR* 14:181–184
- Weigelt J, van Dongen M, Uppenberg J, Schultz J, Wikstrom M (2002a) Site-selective screening by NMR spectroscopy with labeled amino acid pairs. *J Am Chem Soc* 124:2446–2447
- Weigelt J, Wikstrom M, Schultz J, van Dongen MJ (2002b) Site-selective labeling strategies for screening by NMR. *Comb Chem High Throughput Screen* 5:623–630
- Wider G (1998) Technical aspects of NMR spectroscopy with biological macromolecules and studies of hydration in solution. *Prog NMR Spectrosc* 32:193–275
- Wider G (2000) Structure determination of biological macromolecules in solution using nuclear magnetic resonance spectroscopy. *BioTechniques* 29:1278–1280
- Wider G (2003) NMR structure of the micelle-bound polypeptide hormone glucagon. *Magn Reson Chem Spec Issue* S56–S63

- Wider G, Wüthrich K (1999) NMR spectroscopy of large molecules and multimolecular assemblies in solution. *Curr Opin Struct Biol* 9:594–601
- Williamson MP, Havel TF, Wüthrich K (1985) Solution conformation of proteinase inhibitor IIA from bull seminal plasma by ^1H nuclear magnetic resonance and distance geometry. *J Mol Biol* 182:295–315
- Wimley WC (2002) Toward genomic identification of beta-barrel membrane proteins: composition and architecture of known structures. *Protein Sci* 11:301–312
- Woestenenk EA, Hammarstrom M, Hard T, Berglund H (2003) Screening methods to determine biophysical properties of proteins in structural genomics. *Anal Biochem* 318:71–79
- Wüthrich K (1986) *NMR of proteins and nucleic acids*. Wiley, New York
- Wüthrich K (1998) The second decade – into the third millennium. *Nat Struct Biol* 5:492–495
- Wüthrich K, Wider G (2003) Transverse relaxation-optimized NMR spectroscopy with biomacromolecular structures in solution. *Magn Reson Chem* 41:S80–S88
- Xia YL, Sze KH, Li N, Shaw PC, Zhu G (2002) Protein dynamics measurements by 3D HNCO based NMR experiments. *Spectrosc Int J* 16:1–13
- Xu R, Ayers B, Cowburn D, Muir TW (1999) Chemical ligation of folded recombinant proteins: segmental isotopic labeling of domains for NMR studies. *Proc Natl Acad Sci USA* 96:388–393
- Yabuki T, Kigawa T, Dohmae N, Takio K, Terada T, Ito Y, Laue ED, Cooper JA, Kainosho M, Yokoyama S (1998) Dual amino acid-selective and site-directed stable-isotope labeling of the human c-Ha-Ras protein by cell-free synthesis. *J Biomol NMR* 11:295–306
- Yamazaki T, Lee W, Arrowsmith CH, Muhandiram DR, Kay LE (1994) A suite of triple resonance NMR experiments for the backbone assignment of ^{15}N , ^{13}C , ^2H labeled proteins with high sensitivity. *J Am Chem Soc* 116:11655–11666
- Yamazaki T, Otomo T, Oda N, Kyogoku Y, Uegaki K, Ito N, Ishino Y, Nakamura H (1998) Segmental isotope labeling for protein NMR using peptide splicing. *J Am Chem Soc* 120:5591–5592
- Yan XZ, Kong XM, Xia YL, Sze KH, Zhu G (2000) Determination of internucleotide $^1J_{\text{HN}}$ couplings by the modified 2D J_{NN} -correlated [^{15}N , ^1H] TROSY. *J Magn Reson* 147:357–360
- Yang DW, Kay LE (1999a) Improved (HN)- ^1H -detected triple resonance TROSY-based experiments. *J Biomol NMR* 13:3–10
- Yang DW, Kay LE (1999b) TROSY triple-resonance four-dimensional NMR spectroscopy of a 46 ns tumbling protein. *J Am Chem Soc* 121:2571–2575
- Yang DW, Venters RA, Mueller GA, Choy WY, Kay LE (1999) TROSY-based HNCO pulse sequences for the measurement of $^1\text{H}_\text{N}$ - ^{15}N , ^{15}N - ^{13}CO , $^1\text{H}_\text{N}$ - ^{13}CO , ^{13}CO - $^{13}\text{C}^\alpha$ and $^1\text{H}_\text{N}$ - $^{13}\text{C}^\alpha$ dipolar couplings in ^{15}N , ^{13}C , ^2H -labeled proteins. *J Biomol NMR* 14:333–343
- Yokoyama S (2003) Protein expression systems for structural genomics and proteomics. *Curr Opin Chem Biol* 7:39–43
- Zhu G, Kong XM, Sze KH (1999) Gradient and sensitivity enhancement of 2D TROSY with water flip-back, 3D NOESY-TROSY and TOCSY-TROSY experiments. *J Biomol NMR* 13:77–81
- Zhu G, Xia YL, Nicholson LK, Sze KH (2000) Protein dynamics measurements by TROSY-based NMR experiments. *J Magn Reson* 143:423–426

Emerging Techniques in Fast Multidimensional NMR

RAY FREEMAN, ERIKS KUPČE

6.1 Introduction

The advent of two-dimensional Fourier transform methods triggered a veritable revolution in the practice of NMR spectroscopy (Jeener 1971; Aue et al. 1976). It presented the chemist with a clear pictorial representation of interactions between different chemical sites, displaying them in the form of “correlation peaks” on a contour map in two frequency dimensions. For the first time the structural chemist could use NMR spectra directly, without the recourse to tedious assignment techniques, such as double resonance. The correlations between interacting chemical sites were plain to see on the new charts, almost as if touched by a magic wand. Chemistry journals were soon full of two-dimensional spectra.

It was not long before the two-dimensional concept was extended to additional frequency dimensions. By concatenating two or more “evolution” intervals during which the nuclear spins moved according to different rules, and performing a Fourier transformation with respect to each evolution period in turn, the new technique generates a multidimensional spectrum. The problem of display is solved by recording appropriate plane sections through the frequency-domain data matrix. The availability of biochemical samples isotopically labelled in carbon-13 and nitrogen-15 has initiated a veritable explosion of applications of multidimensional NMR to the study of proteins, aided by the improved sensitivity and frequency dispersion of modern NMR spectrometers.

However there is a fly in this particular ointment – the experiments tend to be of long duration, often several days at a time. The problem is that each and every evolution dimension has to be sampled systematically and independently. This puts an unwelcome brake on projects of really high dimensionality, or on experiments to monitor the time-dependence of the interactions. NMR spectroscopists are forced to make compromises with respect to resolution, so that fewer increments are used in the evolution dimensions, sometimes extending the time-domain data artificially by means of “linear prediction” algorithms. The situation is clearly unsatisfactory; multidimensional NMR ties up an expensive spectrometer for long periods and reduces the throughput of samples.

Desperate measures are called for. This review outlines two radical new approaches to fast multidimensional NMR. But first we must accept a fundamental limitation on all experiments carried out in a very short time – sensitivity inevitably suffers because less energy is transferred from the spin system to the spectrometer receiver. Fortunately, in many high-field studies, particularly if a cryogenic receiver

coil is used, the signal-to-noise ratio is not a limiting factor. Speed is far more important. Two radically different approaches to the speed question are described in the following.

6.2 Spatially Encoded Single-Scan Multidimensional NMR

An ingenious solution to the speed problem (Frydman et al. 2002, 2003) conflates the operations of evolution and signal detection into a single scan lasting less than a second. Consider the simple case of a two-dimensional spectrum. The idea is to break the effective sample volume into a set of independent “packets” of spins by selective excitation in an applied magnetic field gradient, exploiting an art well known in magnetic resonance imaging. In the simple case, the cylindrical sample would be divided into a set of N parallel slices stacked along the axis of the cylinder, but in principle field gradients could be applied in all three spatial dimensions. Typically $N=50$ distinct slices might be used. Quite intense gradients are required to ensure that the frequency dispersion of the slices is much greater than the NMR spectral width under investigation, so that spectra from adjacent slices do not overlap. We can think of this as creating a set of N “pigeon holes” – separate regions within the sample that can be used for storing different items of NMR information.

6.2.1 Evolution

Frequency-selective radiofrequency pulses (each at a different frequency) are used to excite the nuclear spins in the various slices. The slices are excited in sequence, so that spins evolve for a different time in each slice– those in the top slice having the longest evolution time and so on, in a linear progression down through the sample (Fig. 6.1). Whereas the conventional experiment has to employ a fresh scan for each and every evolution increment, the single-scan method simply puts information from each evolution step into a different pigeon hole. We may liken this to a form of parallel processing rather than the more protracted serial processing used in the traditional mode of operation. The number N is a compromise between the desire to obtain an adequate number of samples during evolution and the restriction imposed by the need to fit N slices into a sample of finite effective length.

At this point the applied field gradient is reversed for an equal duration, thus refocusing the effect of the first gradient pulse. Spin packets that found themselves initially in a positive local field are now situated in an equal negative local field, so the two effects cancel at the end of the bipolar gradient pair. This compensation is important because the entire sequence is to be repeated many times, and cumulative effects must be avoided. The distribution of the slices in space remains unaffected (just as if a horizontal rack of pigeon holes had been relabelled right-to-left rather than left-to-right, without disturbing any of the contents.) Note that spins in a given slice continue to precess at their characteristic NMR frequencies in the main applied

magnetic field, essentially unaffected by the bipolar gradient pair. This precession is “spectroscopic” rather than “spatial”.

After the bipolar gradient pulses, the spins in the various slices have evolved for different time intervals, accumulating different precession phases, ϕ_k . Their magnetizations can be represented as a helix of vectors in the rotating reference frame (Fig. 6.2). No macroscopic NMR signal is detectable because the various magnetization components of the ensemble mutually interfere. Next a mixing pulse is applied, affecting spins in all the slices equally. Mixing may take many possible forms but suppose, for the purposes of illustration, that we are performing a simple correlation spectroscopy (COSY) experiment (Aue et al. 1976; Bax and Freeman 1981) designed

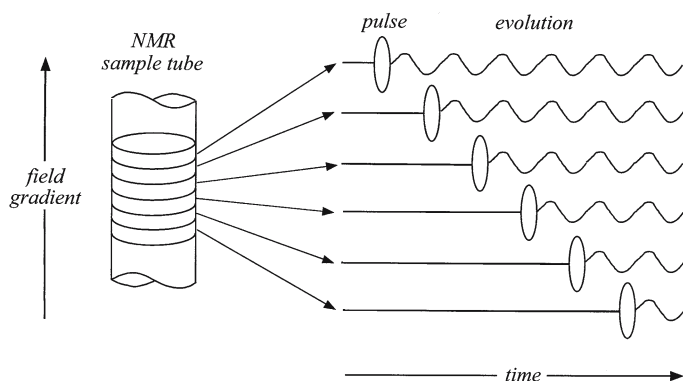


Fig. 6.1. The NMR sample is divided into several slices by selective excitation in an applied magnetic field gradient aligned along the tube axis. The slices are excited in succession – the top slice having the longest evolution time while the bottom slice has the shortest. After the field gradient is extinguished, the spins in different slices “remember” how long they precessed during the evolution interval

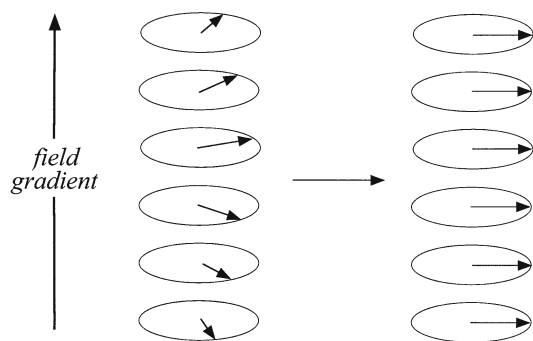


Fig. 6.2. At the end of the evolution stage, the effects of the bipolar field gradient are cancelled, but precession in the main field of the magnet has dispersed the magnetization vectors into a helix. During the subsequent detection interval an applied magnetic field gradient gradually brings these vectors into alignment to form a spin echo

to correlate chemical sites connected by a resolvable spin–spin coupling. A non-selective (“hard”) radiofrequency pulse is employed, causing coherence to be transferred between interacting sites. The affected spins are now poised to precess at the frequency of the new site, but they still “remember” their spatial locations because they are labelled according to their evolution history (ϕ_k).

6.2.2 Detection

Another field gradient pulse is now applied in a sense to counteract the NMR precession that occurred during evolution. The helix of vectors is gradually unwound, bringing spins from all the slices into phase coincidence at a particular instant to form a spin echo (Fig. 6.2). This relies on the fact that the accumulated phase (ϕ_k) of spins in any given slice is proportional to the evolution time associated with that slice, and hence to the distance along the tube axis (Fig. 6.1). This is a most unusual kind of spin echo: a *macroscopic* field gradient has been used to refocus the effect of a *spectroscopic* precession. The fact that all the slices contribute to the intensity of the echo is crucial. *No signal strength is lost as a consequence of the slice-selection process*, otherwise the sensitivity of the single-scan method would be diluted by a factor N , and this would be unacceptable. If the sensitivity of the single-scan method is limited in practice, this is not through loss of signal but rather through acceptance of additional noise (see later).

So far we have considered the problem as if only a single chemical site were involved. In any practical case there are several different chemical sites and their accumulated phase handicaps are proportional to their respective chemical shifts. Each chemical site has its magnetization vectors wound into a helix of a different pitch. For example if there are three different chemical shifts Ω_A , Ω_B and Ω_C , the corresponding vectors are brought to a focus at three different times (Fig. 6.3). Consequently there is a separate spin echo for every distinguishable chemical site, and they appear at times proportional to their characteristic chemical shifts. We call this sequence of spin echoes an “echo spectrum”. Although it *resembles* the expected NMR spectrum in the F_1 frequency dimension, is a very different animal indeed. For one thing, it displays intensity as a function of *time* not *frequency*. In complete contrast to the conventional experiment, no Fourier transformation stage is involved.

Once this echo spectrum has been obtained, the field gradient is reversed, bringing the magnetization vectors back again into focus one by one, but in the reverse order. This doubling of the evolution period is unavoidable and has some practical repercussions as far as noise is concerned. To improve sensitivity, this second echo spectrum may be recorded and added to the first one after reversing its sense and correcting some trivial phase shifts. However, the main reason for using a bipolar gradient pair is to avoid cumulative effects, because this detection cycle is repeated many times in order to extract the information from the F_2 dimension.

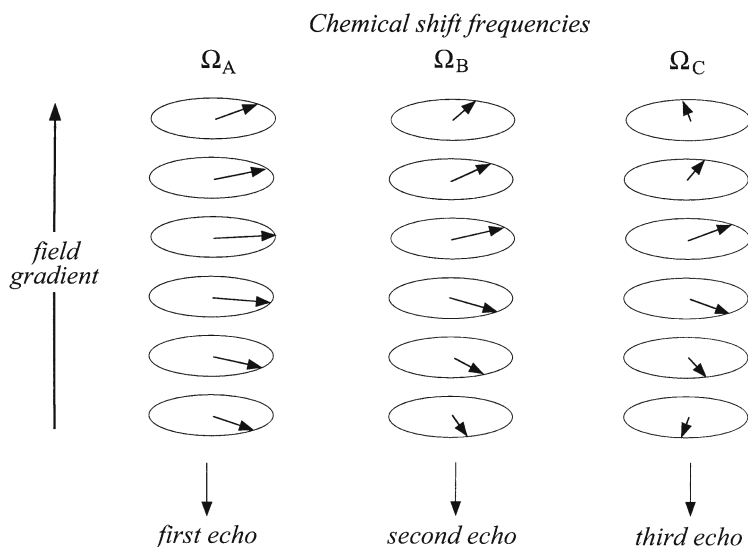


Fig. 6.3. If there are three different chemical shifts Ω_A , Ω_B and Ω_C , the corresponding vectors are dispersed into helices of different pitches. Allowed to precess in the applied magnetic field gradient, these helices unwind to form spin echoes at different times. This creates an “echo spectrum” that has the same general form as the conventional NMR spectrum

6.2.3 Resolution

Good resolution in the evolution dimension (F_1) of a *conventional* two-dimensional experiment imposes a minimum acceptable value for the final evolution time $(t_1)_{\max}$. If $(t_1)_{\max}$ is too short, the observed resonances are broadened or, worse still, are distorted by sinc-function oscillations. Typically $(t_1)_{\max}$ would be set somewhere in the range 10–100 ms. It is not possible to work with much shorter times. Sometimes the time-domain data are artificially extended by linear prediction, but this becomes unreliable beyond a factor of about 2, or for noisy data.

Resolution in the “echo spectrum” is not governed by the conventional rules of Fourier transformation; as we have seen, no transformation is performed in this dimension. One of the crucial factors that allows the single-scan experiment to be completed so quickly is that the time allowed for evolution in the presence of the field gradient is very short, typically 500 μs , 1 or 2 orders of magnitude shorter than in conventional spectroscopy. The spread of precession phases due to chemical shifts is thus quite limited. In principle, the more intense the applied field gradient during detection, the narrower the spin echo, because magnetization vectors from different slices come into phase quickly and then get out of phase equally rapidly, so the echoes should be quite narrow. In practice, there are other factors that limit resolution in the F_1 dimension. Only a small number N of pigeon holes are employed to store the various evolution increments, so there is no advantage to be gained by taking

more than N samples across the entire echo spectrum. (We shall see later that sensitivity considerations also require N to be small.) Consequently, the F_1 dimension is poorly digitized in comparison with the equivalent conventional NMR spectrum, which would typically use between 10^3 and 10^4 samples.

This lack of resolution would seem to be the inevitable price to be paid for the speed advantage of the single-scan technique. In fact the practical limit on the number of slices N can be lifted by changing the excitation scheme from the sequence of selective radiofrequency pulses (described earlier) to an adiabatic frequency sweep in the presence of the applied field gradient (Pelupessy 2003). In this manner the sampling can be much finer because the excitation is continuous. Although this is a more effective scheme and could improve resolution in the F_1 dimension, finer sampling sacrifices sensitivity, as described later.

6.2.4 The F_2 Dimension

Correlation peaks in a COSY spectrum build up gradually, at a rate determined by the inverse of the relevant spin–spin coupling constant, as antiphase components created by the mixing pulse evolve into detectable magnetization. Consequently the first echo spectrum described earlier shows very little effect of the mixing pulse. But when the detection cycle is repeated at increasing delays after the mixing pulse, coherence is gradually transferred between interacting chemical sites, causing frequency jumps within the F_1 spectra, and a consequent intensity modulation. Sampling in this time dimension (t_2) is more comprehensive; typically 256 repeated gradient pairs would be applied in order to achieve acceptable digital resolution. Fourier transformation of the intensity of each echo peak as a function of t_2 generates the COSY spectrum, displaying correlation peaks that serve to map the spin–spin interactions. The entire sequence can be completed in less than a second, much faster than the comparable COSY experiment performed by conventional methodology.

6.2.5 Signal-to-Noise Ratio

As mentioned earlier, any measurement completed in a short time suffers a degradation in signal-to-noise ratio compared with a conventional experiment where data is accumulated over a much longer period. This loss is inevitable and applies to all fast multidimensional measurements. It is tempting to argue that a technique which temporarily stores signals from different evolution increments in N different slices must dilute the signal strength N times. This is *not* the case, because each echo collects signal intensity from all the spins in the effective sample volume, not merely from a single slice. This is a consequence of the unconventional way in which the F_1 spectrum is generated – as a string of spin echoes, each one derived from signals from all N slices.

However there are additional sources of noise in the single-scan method that inevitably arise because the acquisition operation is so fast. There are two equiva-

lent ways to explain this loss of sensitivity. The first attributes the additional noise to the significantly increased receiver bandwidth (“filter bandwidth”). Bandwidth is determined by the inverse of the receiver dwell time, which is short, typically only 2 μ s. Consequently the band-pass filter in the receiver has to accept a very wide range of frequencies, typically 500 kHz. This band contains far more noise than a conventional proton NMR spectrum which would span only about 5 kHz. Noise is proportional to the square root of the bandwidth, so this amounts to a tenfold loss in sensitivity.

The second (equivalent) explanation is based on considerations of receiver duty cycle, the ratio of the times for which the receiver is active and inactive. Suppose there are N data points in each “echo spectrum” in the F_1 dimension, with another N points collected when the gradient is reversed (for practical simplicity this “reflected” echo spectrum is sometimes discarded). Each of these $2N$ points from the F_1 dimension must be examined as a function of t_2 and the result Fourier transformed to give the F_2 spectrum. Only one out of a total of $2N$ evolution samples is used in any given t_2 trace, so the receiver duty cycle is effectively $1/(2N)$. The signal-to-noise ratio is determined by the square root of the duty cycle, so for a typical case where $2N=100$, the sensitivity is degraded tenfold. These numbers are quoted merely for guidance; the practical parameters in any particular experiment may well be different. For example, if the “reflected” echo spectra are included, the effective receiver duty cycle would be doubled and, as a consequence, some of the lost sensitivity retrieved.

Both analyses predict a roughly tenfold loss in sensitivity. Here lies the irony – dividing the sample into N slices does *not* reduce the signal strength (as many would have imagined) but it *does* increase the noise by the smaller factors $\sqrt{2N}$ or \sqrt{N} . This is why resolution in the F_1 dimension has to be a compromise. It is N that limits the attainable resolution, but if it is too large, the sensitivity suffers.

Any transients induced by the intense field gradient pulses may also contribute noiselike artifacts, although this problem could be mitigated by technological improvements in the equipment used for generating gradients. Although in principle a single-scan measurement benefits from the fact that the initial state has the full equilibrium Boltzmann spin populations without the need to introduce a relaxation delay, this advantage is sacrificed when multiscan averaging is required. Furthermore, some applications in biochemistry require a presaturation period for water suppression, slowing down the process appreciably.

Consequently the lack of sensitivity is probably the limiting factor in the application of the single-scan method to samples where concentration is inherently limited, such as proteins. At the present time, two-dimensional spectra of proteins would probably require unrealistically high concentrations, unless further sensitivity enhancement schemes can be incorporated. Nevertheless, the enormous increase in speed of the single-scan technique naturally suggests some exciting new applications, particularly for time-dependent studies. It has recently been used for monitoring two-dimensional spectra from samples in continuous-flow schemes such as liquid chromatography (Shapira et al. 2004).

6.2.6

Application of Single-Scan Two-Dimensional Spectroscopy

The simple case of the proton total correlation spectroscopy (TOCSY) spectrum of 2-ethylindanone (5% by volume) illustrates the application of single-scan two-dimensional NMR. The mixing time was 60 ms and the duration of signal acquisition was 96.6 ms. There were 151 complex data points in the F_1 dimension (the echo spectrum) and 160 (extended by linear prediction to 256) in the F_2 dimension. The sweep widths were 1,600 Hz in both dimensions. Figure 6.4 shows the two-dimensional spectrum in the range 0.6–3.4 ppm, with necessarily poorer resolution in the F_1 dimension. The conventional one-dimensional spectrum (which incidentally required a significantly longer measurement time) is set out along the top margin, indicating the assignment. This fast technique should prove invaluable for screening experiments on large sets of related molecules, where the conventional spectra in just one dimension are not sufficiently specific.

6.3

Projection–Reconstruction of Multidimensional Spectra

Systematic, step-by-step examination of all the coordinates in a three-dimensional matrix is not necessarily the best approach. The time-domain signals of a traditional three-dimensional NMR experiment form such a matrix, made up of two successive evolution periods t_1 and t_2 followed by a detection stage t_3 . If there are N increments in t_1 and M increments in t_2 this requires NM repeated scans, where neither N nor M can be significantly reduced without compromising resolution. Clearly any scheme for speeding up such experiments must aim to gather the data in a more economical fashion.

The human eye sees two stereoscopic images and the brain processes these to form a three-dimensional impression of our surroundings. Not all the information in the scene is directly available – certain objects in the foreground may obscure more distant features – but these difficulties are soon resolved by adopting a second viewpoint. So, if we think of a three-dimensional NMR spectrum as an object, we should be able to reconstruct it from a suitable set of plane projections. This idea has been widely exploited in imaging technology, for example in X-ray tomography and (to a lesser extent) in magnetic resonance imaging. The NMR spectrum, although complex, presents a much more favourable case than an anatomical sample, where the density function is continuous in all three dimensions. The three-dimensional NMR spectrum is made up of discrete resonances, sparsely distributed in space, with a rather small number of close encounters between resonances. This suggests that just a few different “views” of the spectrum should provide sufficient information to define the entire three-dimensional structure. Indeed such spectra are sometimes portrayed in scientific journals as just two stereoscopic images.

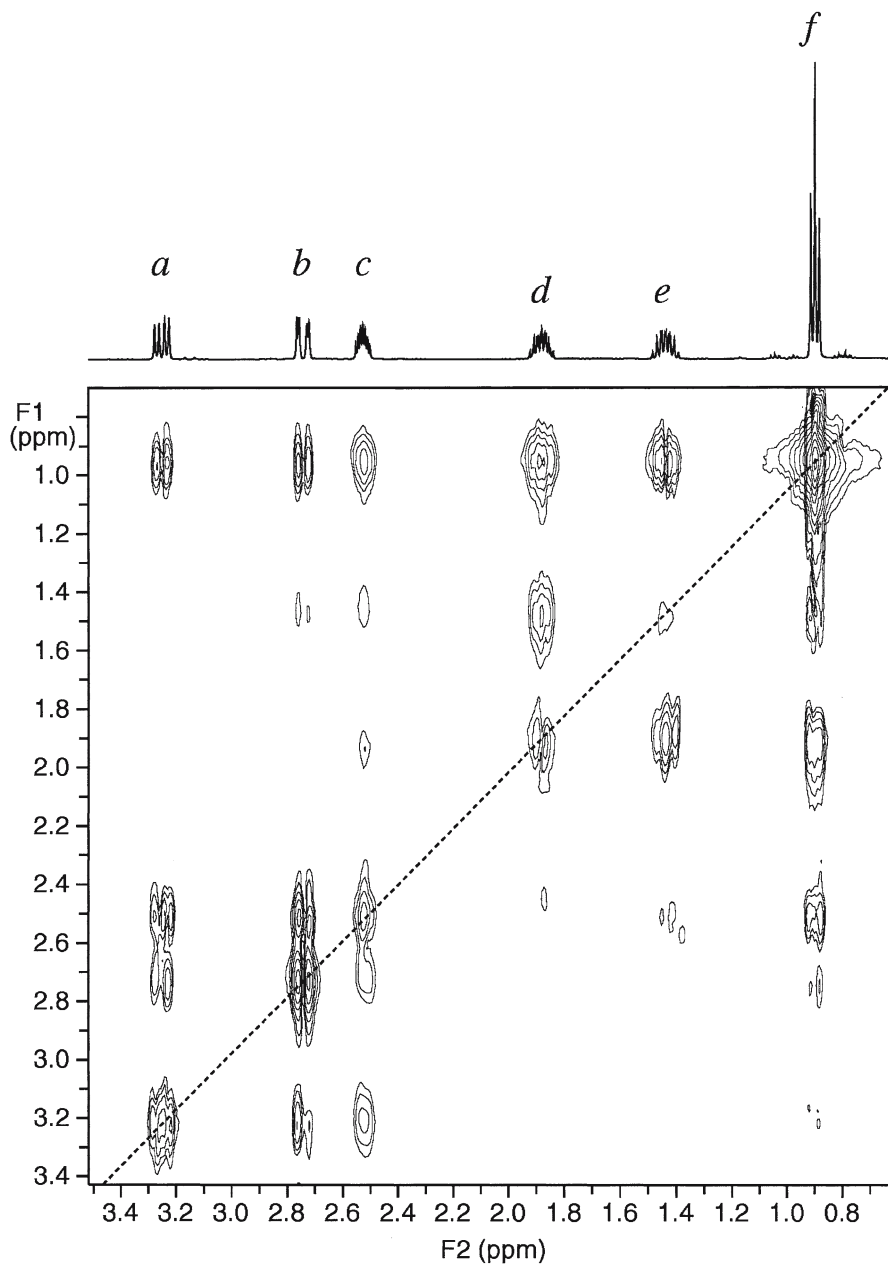


Fig. 6.4. The spatially encoded single-scan proton total correlation spectrum of a 5% solution of 2-ethylindanone, showing the aliphatic region. The measurement was completed in less than 157 ms. Provisional assignment: *a, b* the non-equivalent methylene at C3; *c* the single proton at C2; *d, e* the non-equivalent methylene at C2; *f* the methyl group

6.3.1 Projection of NMR Spectra

It turns out to be a simple matter to obtain projections of the spectrum in any given direction. There is a well-established Fourier transform theorem which states that a section through the origin of the time-domain signal making an angle α in the t_1t_2 plane transforms into a projection of the frequency-domain spectrum F_1F_2 onto a plane inclined at the same angle α (Fig. 6.5c). If this seems too arcane, imagine the simpler case where the second evolution period is not used, that is to say $t_2=0$. Time-domain signals are acquired in the t_1t_3 plane, and after Fourier transformation (Fig. 6.5a) we obtain a projection on the F_1F_3 plane ($\alpha=0$). Alternatively, if $t_1=0$, then time-domain signals are collected in the t_2t_3 plane and after Fourier transformation (Fig. 6.5b) they generate a projection onto the F_2F_3 plane ($\alpha=90^\circ$). In the general case, joint incrementation of t_1 and t_2 (with increments Δt_1 and Δt_2) simply swings the projection plane between these two extremes, through a tilt angle α , defined by $\tan\alpha=\Delta t_2/\Delta t_1$ (Fig. 6.5c).

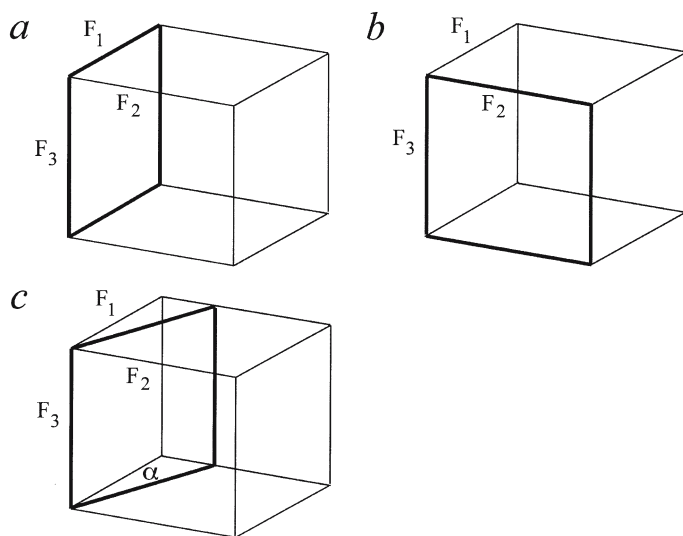


Fig. 6.5. **a** In a three-dimensional experiment, Fourier transformation of signals gathered as a function of a single evolution parameter t_1 (with $t_2=0$) generates a projection onto the F_1F_3 plane (*bold lines*). **b** When t_2 is the sole evolution parameter (with $t_1=0$) the corresponding projection is onto the F_2F_3 plane. **c** When both t_1 and t_2 are varied jointly with increments defined by $\tan\alpha=\Delta t_2/\Delta t_1$ the projection plane falls between these two extremes, subtending an angle α with respect to the F_1F_3 plane

In practice, Fourier transformation of the time-domain signals generates *two* tilted projections, inclined at $\pm\alpha$ with respect to the F_1F_3 plane. This arises because a signal evolving at a chemical shift Ω_A during t_1 modulates the signal at a chemical shift Ω_B in the second evolution period t_2 . Since it is essential to define the sense of precession in both evolution dimensions, the signals are acquired in quadrature (sine and cosine components). Four scans generate four time-domain responses:

$$S_1 = \cos(\Omega_A t_1) \cos(\Omega_B t_2) \quad (6.1)$$

$$S_2 = \sin(\Omega_A t_1) \sin(\Omega_B t_2) \quad (6.2)$$

$$S_3 = \sin(\Omega_A t_1) \cos(\Omega_B t_2) \quad (6.3)$$

$$S_4 = \cos(\Omega_A t_1) \sin(\Omega_B t_2) \quad (6.4)$$

The evolution times are incremented at rates given by

$$t_1 = t \cos\alpha \quad (6.5)$$

$$t_2 = t \sin\alpha. \quad (6.6)$$

Then the responses are combined in pairs:

$$S_1 - S_2 = \cos[(\Omega_A \cos\alpha + \Omega_B \sin\alpha)t] \quad (6.7)$$

$$S_1 + S_2 = \cos[(\Omega_A \cos\alpha - \Omega_B \sin\alpha)t] \quad (6.8)$$

$$S_3 + S_4 = \sin[(\Omega_A \cos\alpha + \Omega_B \sin\alpha)t] \quad (6.9)$$

$$S_3 - S_4 = \sin[(\Omega_A \cos\alpha - \Omega_B \sin\alpha)t] \quad (6.10)$$

Fourier transformation as a function of the common time parameter t generates the sum and difference frequencies $\Omega_A \cos\alpha \pm \Omega_B \sin\alpha$, scaled according to the tilt angle. These expressions differ only in the sign of α , indicating that the projections occur in pairs inclined at $\pm\alpha$ with respect to the F_1F_3 plane.

Each projection provides an independent view of the spectrum, and if any ambiguities remain, a further pair of projections, taken at a different tilt angle, should serve to resolve them. In this manner a tedious step-by-step exploration of all the elements of a three-dimensional matrix (the conventional method), is replaced by two (or possibly four) two-dimensional acquisitions. This offers an order of magnitude improvement in speed (Kupče and Freeman 2003). The idea is readily extended to higher-dimensional NMR spectra where the speed factor increases by an order of magnitude for each additional evolution dimension (Kupče and Freeman 2004b). In situations where a two-dimensional spectrum is marred by frequency overlap, a single tilt operation will often resolve the problem (Freeman and Kupče 2003) without the need to perform the complete three-dimensional analogue.

For biochemists principally interested in automated assignment programs, the projections can be used to derive chemical shifts and correlations directly, without

reconstructing the three-dimensional spectrum. The “mixed” frequencies $f(+\alpha)$ and $f(-\alpha)$ observed in the pair of projections tilted at $+\alpha$ and $-\alpha$ are used to compute the corresponding “pure” frequencies f_1 and f_2 in the orthogonal planes F_{13} and F_{23} .

$$f_1 = [f(+\alpha) + f(-\alpha)] / 2 \cos \alpha \quad (6.11)$$

$$f_2 = [f(+\alpha) - f(-\alpha)] / 2 \sin \alpha \quad (6.12)$$

This would be particularly useful in cases of overlap in the orthogonal planes. For those who prefer to examine actual three-dimensional spectra, a reconstruction program can be used. Reconstructed spectra have the advantage that they bring to light any instrumental artifacts that might pass unnoticed in a purely numerical analysis.

6.3.2 Reconstruction from Projections

This is a classic “inverse” problem. We can go directly from a three-dimensional image to any chosen projection, but the reverse process can only be accomplished indirectly. One solution is to start with a three-dimensional matrix consisting of arbitrary intensities and then readjust these values to be consistent with the plane projections, using an iterative algorithm. Two other approaches have proved to be valuable for the NMR situation. Both involve the concept of “back-projection”.

The first method is useful in situations of good sensitivity. The starting point is the two orthogonal projection planes F_1F_3 and F_2F_3 illustrated in Fig. 6.5, corresponding to $t_2=0$ and $t_1=0$ respectively. These two-dimensional versions of the experiment are often used for the initial setting-up procedure as they give strong signals. A three-dimensional array is created by scaling the intensities in the F_1F_3 plane according to the intensities in the F_2F_3 plane. For simplicity we can consider the F_1F_2 planes one at a time, recorded for a given value of F_3 . Some responses in this plane correspond to genuine correlations, others do not – there is no discrimination. This “provisional” array (Fig. 6.6, array b) suggests that every response on the F_1 axis is correlated with every response on the F_2 axis. One could say that the frequency coordinates are correct but the intensities remain to be determined; some intensities will later turn out to be zero.

These ambiguities are resolved by recording a pair of tilted projections as described earlier. Genuine correlation peaks must be consistent with this new information. A simple way to impose this additional constraint involves back-projection and processing according to the “lower-value algorithm” (Baumann et al. 1981; McIntyre et al. 1990). One of the tilted projections is extended into a set of parallel ridges running across the F_1F_2 plane. Let us call this the “back-projection” array (Fig. 6.6, array a). The intensity at each location in the provisional array (Fig. 6.6, array b) is compared with the corresponding intensity in the back-projection array and the lower value retained. Peaks in the former that do not lie under the ridges of the latter are eliminated, leaving fewer potential correlation peaks (Fig. 6.6, array c). A further iteration involving back-projection of the second tilted projection should resolve

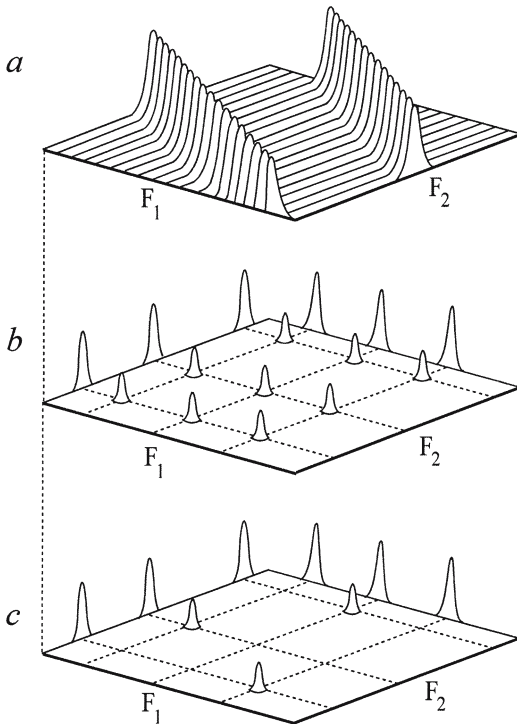


Fig. 6.6. If the projections onto the F_1 and F_2 axes comprise three peaks each, multiplication creates a “provisional array” (b) with nine responses. The lower-value algorithm is employed to identify the genuine correlation peaks. A tilted projection is back-projected (a) to create parallel ridges running across the F_1F_2 plane. The intensities at corresponding frequency coordinates in a and b are compared and the lower value retained. The genuine correlation peaks (c) lie directly underneath the ridges

any remaining ambiguities in intensity. If not, a further pair of tilted projections is employed.

An inherent disadvantage of the lower-value algorithm is that the sensitivity is determined solely by the signal-to-noise ratio in the two orthogonal projections; the tilted projections act merely in an editing role, making no contribution to the overall signal intensities. As a consequence this method is best restricted to samples with a good signal-to-noise ratio. Where sensitivity is marginal, an alternative processing scheme is preferred, employing a somewhat larger number of different projections. It is based on the inverse Radon transform (Deans 1992; Kupče and Freeman 2004a), a method widely used in other fields of science to create an image from a set of plane projections (often spaced at regular intervals around a circle). By back-projection and summing this information, a set of crisscrossing ridges is generated. Genuine correlation peaks correspond to points where *all* the ridges cross, but there remain weaker spurious peaks where fewer ridges intersect. The more tilted projections collected, the more intense the true peaks in comparison with the false ones. Eventually these undesirable artifacts sink below the noise floor. Because this is an additive procedure, each new projection improves the signal-to-noise ratio, in contrast to the lower-value algorithm. Hybrids of the two processing schemes can strike a balance between the advantages and drawbacks of the two approaches.

6.3.3

Application of Projection–Reconstruction NMR

The principle of three-dimensional projection–reconstruction can be illustrated by reference to the 900-MHz constant-time HNCO spectrum of the 187-residue protein HasA, isotopically enriched in nitrogen-15 and carbon-13. The sample was a 1.7 mM aqueous solution containing 10% heavy water. The carbon–proton plane was the first to be recorded, using a two-dimensional analogue of the three-dimensional experiment with t_1 set to zero. The orthogonal nitrogen–proton plane was “borrowed” from the corresponding two-dimensional heteronuclear single-quantum correlation (HSQC) experiment which has inherently better sensitivity. Each plane required 20 min of data gathering. The three-dimensional HNCO information was acquired in the form of 100 complex data points with t_1 and t_2 incremented jointly. Projections were recorded onto planes tilted at $\pm 30^\circ$ and $\pm 60^\circ$ with respect to the F_1F_3 plane, requiring a total of 60 min of data accumulation. Reconstruction employed the lower-value algorithm in this case. The total elapsed time for data gathering was 1 h and 40 min, to be compared with an estimated time for the traditional experiment of roughly 17 h, even with diminished digital resolution. A typical carbon–nitrogen plane from the reconstruction is shown in Fig. 6.7, where nine correlations are clearly visible.

6.3.4

Related Methods

Projection–reconstruction was not the first technique to exploit the simultaneous incrementation of two evolution parameters. Once it was realized that it was not necessary to monitor *all* the elements of a three-dimensional time-domain matrix, several authors (Ding and Gronenborn 2002; Kim and Szyperski 2003; Kozminski and Zhukov 2003) employed joint exploration of t_1 and t_2 to derive the essential correlation information. This technique has come to be known as G-matrix Fourier transform (GFT) NMR (Kim and Szyperski 2003). However, there is a practical complication – GFT spectra consist of sums and differences of chemical shifts, rather than the actual shifts themselves. This arises from the way the data are collected. As demonstrated earlier, Fourier transformation of the time-domain signal obtained when t_1 and t_2 are linked together generates projections onto two planes, $T(+)$ and $T(-)$, tilted at $\pm\alpha$ with respect to the F_1F_3 plane (Fig. 6.8). But instead of recording these tilted planes directly, GFT-NMR gathers these two data sets as if they were further projected onto the F_1F_3 plane. In this projection a typical response at coordinates (f_1, f_2) would have apparent frequencies given by

$$(f_1\cos\alpha + f_2\sin\alpha)\cos\alpha \text{ and } (f_1\cos\alpha - f_2\sin\alpha)\cos\alpha.$$

Since the GFT experiment increments t_1 and t_2 at the same rate, $\alpha=45^\circ$ and the signals appear at the sum and difference frequencies $(f_1+f_2)/2$ and $(f_1-f_2)/2$ rather than at the true chemical shifts f_1 and f_2 . As a result, there are many more resonance lines than in the conventional spectrum, and the intensities are halved for every additional splitting. The subsequent analysis entails the identification of the subspectra

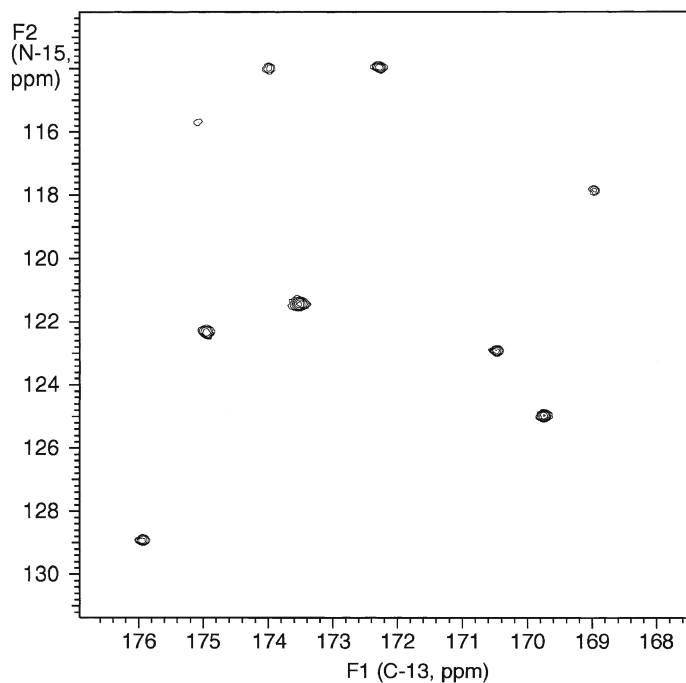


Fig. 6.7. A typical carbon–nitrogen plane (taken at a proton frequency of 8.61 ppm) extracted from the reconstructed three-dimensional 900-MHz constant-time HNCO spectrum of a 1.7 mM aqueous solution of the HasA protein. Total accumulation time 1 h 40 min. This offers an order of magnitude improvement in speed over the traditional mode, with the bonus of better digital resolution

from each chemical site, peak-picking, the measurement of the resonance frequencies, and extraction of the actual chemical shifts.

The projection–reconstruction technique differs from GFT-NMR in two key respects. First, it emphasizes the importance of incrementing the evolution times t_1 and t_2 at *different* rates, thus projecting the three-dimensional spectrum in any desired direction. There is a numerical program for choosing the appropriate tilt angle (Kupče and Freeman 2004b). Secondly, the philosophy behind the data processing is quite different, culminating in a result where chemical shift intermodulation is a distinct advantage in projection–reconstruction NMR, rather than a complicating factor. Both methods offer an impressive improvement in speed, particularly for spectra of higher dimensions.

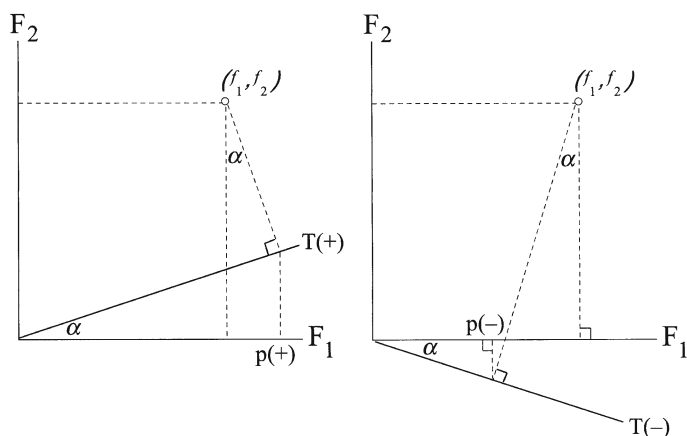


Fig. 6.8. Joint evolution of t_1 and t_2 in a three-dimensional experiment projects a typical response at (f_1, f_2) onto two tilted planes $T(+)$ and $T(-)$ subtending an angle α with respect to the F_1F_3 plane (For simplicity the F_3 dimension is not shown.) When signals are detected in the conventional manner (involving a further projection onto the F_1F_3 plane) the new F_1 ordinates are $p(+)= (f_1\cos\alpha + f_2\sin\alpha)\cos\alpha$ and $p(-)= (f_1\cos\alpha - f_2\sin\alpha)\cos\alpha$. In the special case when t_1 and t_2 are incremented at the same rate, the tilt angle α is 45° and the sum and difference frequencies $(f_1+f_2)/2$ and $(f_1-f_2)/2$ are detected

6.4 Conclusions

Spatially encoded single-scan multidimensional NMR (Frydman et al. 2002, 2003) offers a very fast mode of data acquisition, providing a complete two-dimensional spectrum in a fraction of a second. In its present form the technique is limited by poor inherent sensitivity, largely as a result of higher-than-normal levels of instrumental noise, but the incorporation of sensitivity-enhancement schemes could alleviate this difficulty. Compared with conventional three-dimensional spectroscopy, projection-reconstruction NMR (Kupče and Freeman 2003, 2004a, b) delivers a speed advantage of an order of magnitude, with a further order of magnitude for every additional evolution dimension. It also offers an intuitive and transparent interpretation of related schemes that employ joint incrementation of the evolution parameters. These emerging multidimensional techniques show clear promise for extending the scope of biomolecular NMR spectroscopy, particularly in the study of proteins.

Acknowledgements. We thank Lucio Frydman of the Weizmann Institute, Rehovot, for illuminating discussions, and Anne Lacroisey and Muriel Delepierre of the Institut Pasteur, Paris, for the protein sample HasA.

References

- Aue WP, Bartholdi E, Ernst RR (1976) Two-dimensional spectroscopy. Application to nuclear magnetic resonance. *J Chem Phys* 64:2229
- Baumann R, Wider G, Ernst RR, Wüthrich K (1981) Improvement of 2D NOE and 2D correlated spectra by symmetrization. *J Magn Reson* 44:402
- Bax A, Freeman R (1981) Investigation of complex networks of spin-spin coupling by two-dimensional NMR. *J Magn Reson* 44:542
- Deans SR (1992) *The Radon transform and some of its applications*. Wiley, New York
- Ding K, Gronenborn A (2002) Novel 2D triple-resonance NMR experiments for sequential resonance assignments of proteins. *J Magn Reson* 156:262–268
- Freeman R, Kupče E (2003) New methods for fast multidimensional NMR. *J Biomol NMR* 27:101–113
- Frydman L, Scherf T, Lupulescu A (2002) The acquisition of multidimensional NMR spectra within a single scan. *Proc Natl Acad Sci USA* 99:15859–15862
- Frydman L, Lupulescu A, Sherf T (2003) Principles and features of single-scan two-dimensional NMR spectroscopy. *J Am Chem Soc* 125:9204–9217
- Jeener J (1971) Ampère International Summer School, Basko Polje, Yugoslavia
- Kim S, Szyperski T (2003) GFT NMR, a new approach to rapidly obtain precise high-dimensional NMR spectral information. *J Am Chem Soc* 125:1385–1393
- Kozminski W, Zhukov I (2003) Multiple quadrature detection in reduced dimensionality experiments. *J Biomol NMR* 26:157–166
- Kupče E, Freeman R (2003) Reconstruction of the three-dimensional NMR spectrum of a protein from a set of plane projections. *J Biomol NMR* 27:383–387
- Kupče E, Freeman R (2004a) The Radon transform: a new scheme for fast multidimensional NMR. *Concepts Magn Reson A* 22:4–11
- Kupče E, Freeman R (2004b) Projection-reconstruction techniques for speeding up multidimensional NMR spectroscopy. *J Am Chem Soc* 126:6429–6440
- McIntyre L, Wu X-L, Freeman R (1990) Fine structure of cross-peaks in truncated COSY experiments. *J Magn Reson* 87:194–201
- Pelupessy P (2003) Adiabatic single-scan two-dimensional NMR spectroscopy. *J Am Chem Soc* 125:12345–12350
- Shapira B, Karton A, Aronzon D, Frydman L (2004) Real-time 2D NMR identification of analytes undergoing continuous chromatographic separation. *J Am Chem Soc* 126:1262–1265

Principles and Application of Projected Multidimensional NMR Spectroscopy – G-matrix Fourier Transform NMR

THOMAS SZYPERSKI

7.1 Introduction

Higher-dimensional Fourier transform (FT) NMR (Ernst et al. 1987; Cavanagh et al. 1996) is a key analytical technique in natural sciences but suffers from two major drawbacks. First, the minimal measurement time of an N -dimensional (ND) FT NMR experiment, which is constrained by the need to sample $N-1$ indirect dimensions, may exceed by far the measurement time required to achieve workable signal-to-noise (S/N) ratios. In this “sampling limited data collection regime” (Szyperski et al. 2002) valuable NMR spectrometer time is used to sample the indirect dimensions and not for achieving sufficient “signal averaging.” Second, the comparably low digital resolution in the indirect dimensions limits the accuracy of the measurement of NMR parameters such as chemical shifts. These drawbacks have fostered the development of G-matrix FT (GFT) NMR as a technique for rapid sampling of multidimensional NMR data (Kim and Szyperski 2003). In GFT NMR, the conventional sampling of an ND time domain subspace is replaced by phase-sensitive joint sampling of all N chemical shift evolution periods spanning this subspace. The ND FT is replaced by a “G-matrix” transformation for editing chemical shift multiplets arising from the joint sampling, followed by a 1D FT. Since the chemical shifts are encoded in the chemical shifts in a redundant manner, greatly increased accuracy for the measurement of the shifts can be obtained. As a result, the GFT NMR data acquisition scheme affords highly precise 4D, 5D or higher-dimensional NMR spectral information within a few hours or less, and enables one to accurately adjust measurement times to sensitivity requirements. This makes GFT NMR particularly attractive for high-throughput structure determination in structural genomics (Montelione et al. 2000), and opens new perspectives to study dynamic processes such as slow protein folding and systems exhibiting a high degree of chemical shift degeneracy.

7.2 Background

GFT NMR is rooted in two-spin coherence spectroscopy (Szyperski et al. 1993a) and its generalization, an approach that was coined reduced-dimensionality (RD) NMR spectroscopy (Szyperski et al. 1993b). The concept of RD NMR is simple: instead of encoding N chemical shifts in an ND FT NMR experiment, one of the chemical shifts (“projected chemical shift”) is encoded into an in-phase splitting in

a “projected” ($N-1$)D spectrum (Fig. 7.1). Hence, the ND spectral information is obtained by sampling only $N-1$ dimensions, yielding significantly increased data acquisition speed. A key challenge, however, which is associated with implementing RD NMR is to retain upon projection all information of the parent experiment.

7.2.1 RD and Accordion NMR Spectroscopy

The joint sampling of two evolution periods is a feature that RD NMR has in common with “accordion” NMR spectroscopy (Bodenhausen and Ernst 1981). However, the principles behind accordion spectroscopy solely enable one to co-increment a “real” observable (i.e., one that has no phase properties) with an indirect shift evolution period. In its original implementation, accordion spectroscopy was based on the joint sampling of a chemical shift evolution period with a mixing period of exchange spectroscopy. Subsequently, accordion spectroscopy has also been used for measuring other real observables, such nuclear spin relaxation times (Kay and Prestegard 1988; Mandel and Palmer 1994), scalar coupling constants (Tolman and Prestegard 1996) and total correlation spectroscopy (TOCSY) mixing (Kontaxis and Keeler 1995). In contrast, RD NMR focuses on the joint sampling of a chemical shift with a second chemical shift, which itself evidently represents a complex observable. In other words, both jointly sampled observables in RD NMR possess phase properties. As will be outlined in the following, this feature has far reaching consequences for the design and implementation of RD (and thus GFT) NMR experiments, suggesting that RD and accordion NMR represent distinct concepts.

7.2.2 RD NMR Spectroscopy

7.2.2.1 Joint Sampling of Two Shift Evolution Periods

In RD NMR, two chemical shift evolution periods are jointly sampled by co-incrementing the corresponding delays in the radiofrequency pulse scheme. One of

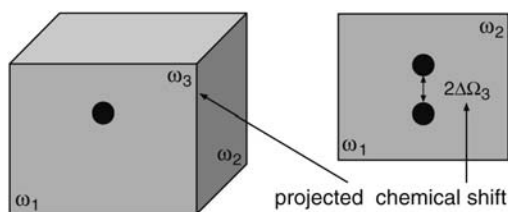


Fig. 7.1. Principle of reduced-dimensionality (RD) NMR exemplified with a projection of a 3D spectrum into a 2D plane (Szyperski et al. 1993a, b). The third chemical shift, Ω_3 , is encoded into in-phase splitting and is named the “projected shift”

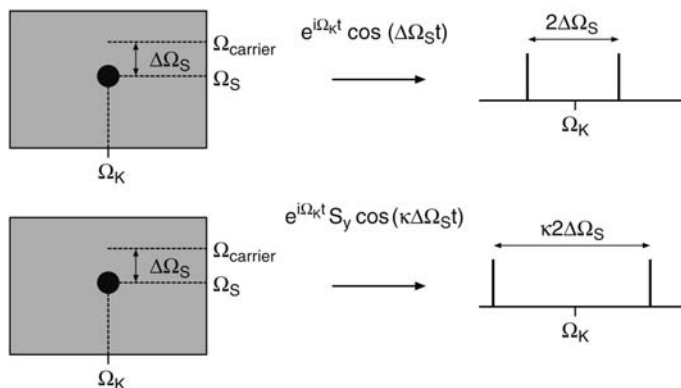


Fig. 7.2. Scaling of chemical shift evolution periods in RD NMR (Szyperski et al. 1993b, 1994). Assuming (1) that the shifts of spins K and S are jointly sampled and (2) that the shift of spin K is detected in quadrature, we can tune the in-phase splitting arising from the shift evolution of spin S by scaling the increment for sampling the chemical shift of S. On the *left*, a 2D subspace of an ND FT NMR experiment is sketched, and on the *right* the peak pattern detected in $(N-1)$ D RD NMR is shown. *Top* no scaling, *bottom* scaling with factor κ

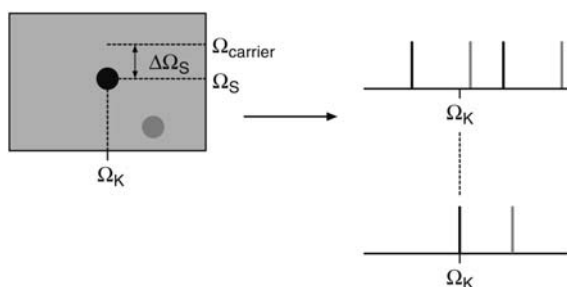


Fig. 7.3. Extension of Fig. 7.2 to demonstrate the impact of central peak detection in RD NMR (Szyperski et al. 1995, 1996). The identification of peak pairs is hampered by chemical shift degeneracy of the other conventionally sampled shifts. Two peaks in the plane give rise to two peak pairs. If the intensities of all lines are identical, it is a priori not clear which peaks form the two pairs. This ambiguity is resolved when peaks *defining* the centers of the peak pairs are detected, as is shown at the *lower right*

the chemical shifts (spin K in Fig. 7.2) is detected in quadrature (Ernst et al. 1987), while a second chemical shift (spin S in Fig. 7.2) provides a cosine modulation of the transfer amplitude. This leads to an in-phase splitting encoding the shift of spin S (Fig. 7.2, top). Since the frequency separation from the carrier position is measured, one cannot distinguish signals located upfield or downfield from the carrier, that is, the projected chemical shift (Fig. 7.1) is a priori not detected in a phase-sensitive manner. Hence, the carrier position for *spin* S must be set at the edge of the spectral range in order to avoid ambiguities.

As was suggested in Szyperski et al. (1993b), the increment of the chemical shift evolution of the projected shift S can be scaled relative to the one of spin K (Szyperski et al. 1994). The corresponding scaling factor (κ in Fig. 7.2) represents an important RD (and GFT) NMR acquisition parameter. Furthermore, detection of peaks defining the center of the peak pairs resolves remaining ambiguities (Fig. 7.3) (Szyperski et al. 1995, 1996). These “central peaks” ensure that the information of the parent ND experiment is retained upon projection in the $(N-1)$ D RD NMR experiment.

7.2.2.2

Time-Proportional Phase Incrementation in RD NMR

To avoid signal losses due to off-resonance effects of radiofrequency pulses, the carrier position associated with the projected shift (spin S in Fig. 7.2) can be moved to the edge of the spectral range by using time-proportional phase incrementation (TPPI) (Marion and Wüthrich 1983) in RD NMR (Szyperski et al. 1995; Brutscher et al. 1995a) (Fig. 7.4). In combination with scaling of the projected chemical shift evolution period (Fig. 7.2), employment of TPPI allows one to efficiently place central peaks and peaks forming the peak pair (Fig. 7.3) into separate spectral regions (Szyperski et al. 1995). As a result, spectral congestion arising from the projection is avoided.

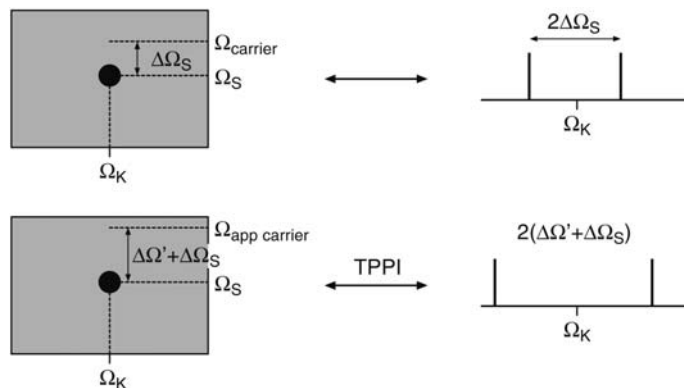


Fig. 7.4. Extension of Fig. 7.2 to demonstrate the employment of time-proportional phase incrementation (TPPI) (Marion and Wüthrich 1983) on the radiofrequency pulse exciting spin S . This moves the apparent carrier position and thus allows one to place upfield and downfield components (and central peaks) into distinct spectral regions to avoid overlap due to the projection (Szyperski et al. 1995)

7.2.2.3 Editing Peak Pairs in RD NMR

Brutscher et al. (1995a) proposed a variant of RD NMR spectroscopy in which the projected shifts are likewise detected in quadrature. This is achieved by recording a second data set in which the projected chemical shift gives rise to a sine modulation of the transfer amplitude. In this data set, the projected shift is encoded in an antiphase splitting, and combination with the cosine RD NMR data set allows one to edit the two peaks forming the peak pair into two subspectra.

7.2.2.4 Peak Pattern in RD NMR Spectra

It was recognized early on that the symmetry properties of RD NMR spectra manifested in detection of peak pairs greatly supports peak identification (Szyperski et al. 1993b). Moseley et al. 2004 have very recently developed a generalized approach for automated peak picking of RD NMR spectra. Moreover, central peak positions can be used to symmetrize RD NMR spectra. This facilitates peak pair identification and increases S/N ratios (Szyperski et al. 1995; Brutscher et al. 1995b).

7.2.2.5 Double RD NMR

The principle of RD NMR can be applied twice, yielding a modulation of the transfer amplitude with the cosine of two chemical shifts (Löhr and Rüterjans 1995). Such double RD NMR experiments enable a twofold reduction in dimensionality. Recently, this idea was applied by Ding and Gronenborn (2002) to design experiments where the transfer amplitude is cosine-modulated twice in the same dimension.

7.2.2.6 Use of Heteronuclear Magnetization

RD NMR experiments were the first NMR experiments designed to simultaneously recruit both proton and heteronuclear magnetization for signal detection (Szyperski et al. 1996). This strategy enables one to take advantage of the entire spin-1/2 polarization in $^{13}\text{C}/^{15}\text{N}$ double-labeled proteins. Specifically, RD NMR experiments have been designed to take advantage of ^{13}C steady-state magnetization to detect central peaks, while the corresponding peak pairs are derived from ^1H magnetization (Szyperski et al. 1996, 1998, 2002; Xia et al. 2002).

7.2.2.7

Application of RD NMR in Structural Genomics

A package of NMR parameter sets for employing RD NMR spectroscopy for complete protein resonance assignment (Szyperski et al. 2002) has been developed. This RDpack comprises pulse sequences, parameter sets, scripts and macros for efficient de novo implementation of RD NMR experiments as well as rapid adjustment of parameter sets when using Varian INOVA spectrometers (Fig. 7.5). The RDpack is freely available for academic users from the author of this article and contains 11 RD NMR experiments. 3D $H^{\alpha\beta}C^{\alpha\beta}(CO)NHN$, 3D $HACA(CO)NHN$ and 3D $HC(C-TOCSY-CO)NHN$ sequentially correlate proton and carbon shifts of residue $i-1$ with the amide proton and nitrogen shifts of residue i , 3D $H^{\alpha\beta}C^{\alpha\beta}NHN$, 3D $HNNCAHA$ and 3D $H^{\alpha\beta}C^{\alpha\beta}(CO)HA$ provide complementary intraresidue connectivities, and 3D $HNN<CO,CA>$ affords both sequential and intraresidue connectivities. Aliphatic side chains are assigned by use of 3D $HCCH$ correlation spectroscopy (COSY) and TOCSY, while aromatic spin system identification relies on 2D $HBCB(CGCD)HD$ and 2D 1H -TOCSY HCH COSY. All parameter sets offer flags to conveniently select (1) central peak acquisition from ^{13}C steady-state magnetization (Szyperski et al. 1996), (2) transverse relaxation optimized spectroscopy (Pervushin et al. 1997) type data acquisition, and (3) 2H decoupling for data collection with deuterated proteins (Gardner and Kay 1998). The experiments of the RDpack have been employed to solve the structures of ten target proteins of the Northeast Structural Genomics Consortium (<http://www.nesgc.org>).

7.3

GFT NMR

GFT NMR spectroscopy represents a generalization of RD NMR and uses modules developed for RD NMR. GFT NMR is based on (1) projecting multiple chemical shifts, (2) (time-domain) editing of the components of the chemical shift multiplets arising from the multiple projection, and (3) multiple “bottom-up” central peak detection. The resulting formalism represents a flexible, generally applicable NMR data acquisition scheme. Provided that $m=K+1$ chemical shift evolution periods of an ND experiment are jointly sampled in a single indirect “GFT dimension”, $p=2^m-1$ different $(N-K)D$ spectra represent the GFT NMR experiment containing the information of the parent ND experiment. Hence, such a set of p spectra was named an $(N,N-K)D$ GFT NMR experiment (Kim and Szyperski 2003) (Fig. 7.6).

7.3.1

Theory of GFT NMR

In GFT NMR, the chemical shift evolution periods spanning a given multidimensional subspace of an FT NMR experiment are *jointly* sampled. As outlined earlier, combining two dimensions is accomplished in RD NMR, where the chemical shift of a “projected dimension” is encoded in a cosine modulation yielding in-phase

Complete Protein Resonance Assignment

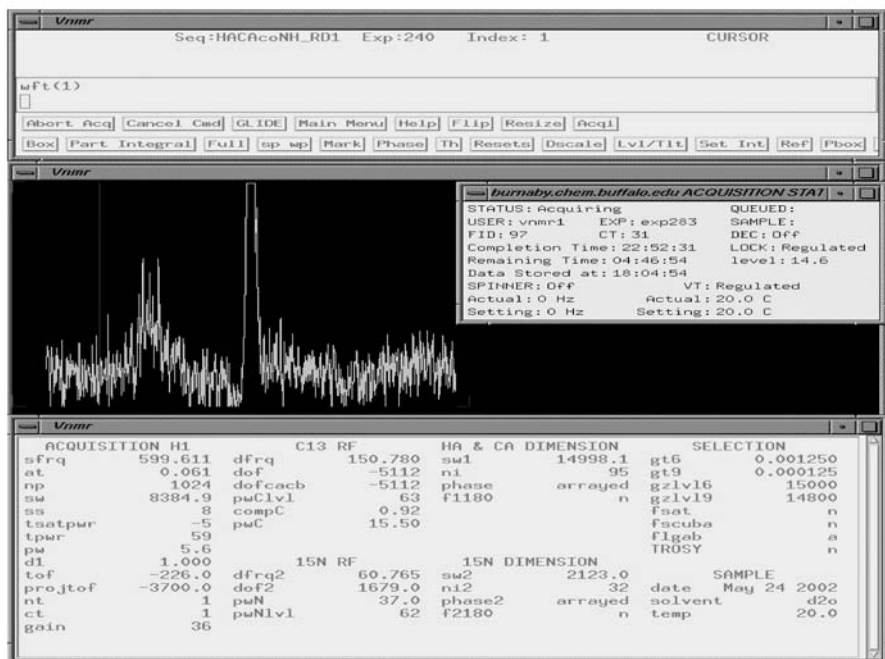
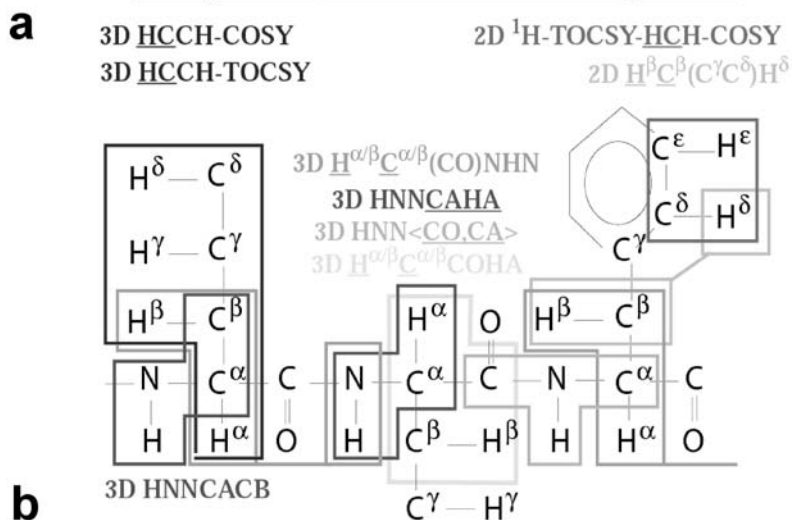


Fig. 7.5. **a** RD NMR experiments for complete protein resonance assignment (Szyperki et al. 2002). **b** Varian VNMR interface displaying the RD 3D ${}^1\text{H}$ ACACO(NH)NHN parameter set taken from the RDpack (see text)

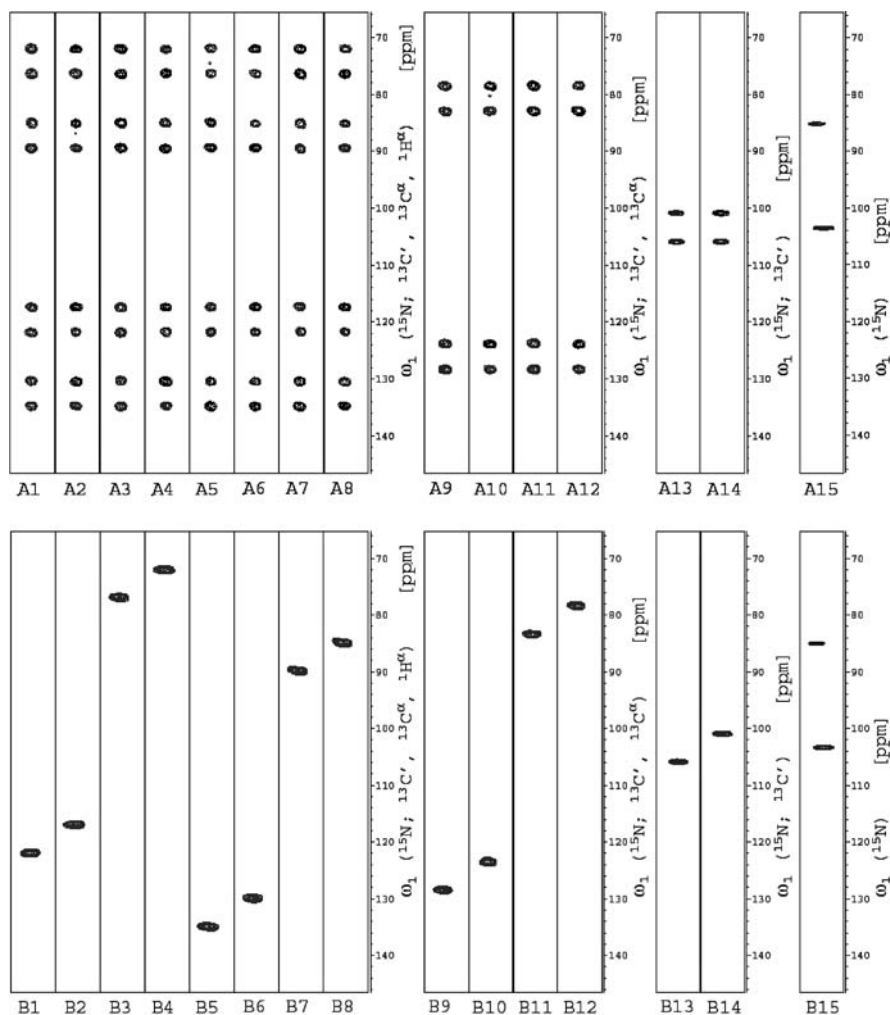


Fig. 7.6. $\omega_1[(^{15}\text{N};^{13}\text{C};^{13}\text{C}^{\alpha}, ^1\text{H}^{\alpha})]$, $\omega_2(^1\text{H}^{\alpha})$, $[\omega_1(^{15}\text{N};^{13}\text{C};^{13}\text{C}^{\alpha})]$, $\omega_2(^1\text{H}^{\alpha})$, $[\omega_1(^{15}\text{N};^{13}\text{C})]$, $\omega_2(^1\text{H}^{\alpha})$, and $[\omega_1(^{15}\text{N})]$, $\omega_2(^1\text{H}^{\alpha})$ strips taken from a (5,2)D HACACONHN G-matrix Fourier transform (GFT) NMR experiment (Kim and Szyperski 2003). The signals were detected on an amide proton. Positive and negative contour levels are shown in red and blue, respectively. At the top, the chemical shift multiplets arising from the joint sampling of chemical shifts are shown. At the bottom, spectra containing the individual edited chemical shift multiplet components obtained after G-matrix transformation are displayed. (From Kim and Szyperski 2003 and reproduced with permission of the American Chemical Society)

peak doublets, and the sine-modulated counterpart exhibiting antiphase splittings enables “frequency discrimination.” GFT NMR generalizes these earlier developments: the dimensionality N of an FT NMR spectrum is adjusted to a “target di-

mensionality” N_t , by jointly sampling $K+1$ chemical shifts $\Omega_0, \Omega_1, \dots, \Omega_K$ ($K=N-N_t$). Assuming that Ω_0 is detected in quadrature and that the setting of the phases ϕ_j of the radiofrequency pulses exciting the spins of dimension j ($j=1\dots K$) ensures cosine modulation, the transfer amplitude of the N_t D experiment is proportional to .

$$e^{i\Omega_0 t} \cdot \prod_{j=1}^K \cos(\Omega_j t) \quad (7.1)$$

The resulting peak centered about the position of Ω_0 contains 2^K components and is designated a “chemical shift multiplet.” A shift of ϕ_j by 90° yields $\sin(\Omega_j t)$ modulation, and $2^K N_t$ D spectra have to be recorded if all phases ϕ_k are systematically varied between 0 and 90° . In turn, linear combination of these 2^K spectra allows editing of the chemical shift multiplet components. For brevity, we define $c_j = \cos(\Omega_j t)$, $s_j = \sin(\Omega_j t)$ and , so that $e^{ij} = c_j + is_j$. With $K=1$, one obtains for the time evolution of the two shift multiplet components encoding the sum and the difference of Ω_0 and Ω_1 .

$$\begin{bmatrix} e^{i1} \\ e^{-i1} \end{bmatrix} \otimes e^{i0} = \begin{bmatrix} e^{i0+i1} \\ e^{i0-i1} \end{bmatrix} = \begin{bmatrix} 1 & i \\ 1 & -1 \end{bmatrix} \otimes \begin{bmatrix} 1 & i \end{bmatrix} \cdot \begin{bmatrix} c_1 \\ s_1 \end{bmatrix} \otimes \begin{bmatrix} c_0 \\ s_0 \end{bmatrix}.$$

In general, one has that

$$\begin{bmatrix} e^{ik} \\ e^{-ik} \end{bmatrix} \otimes \dots \otimes \begin{bmatrix} e^{i1} \\ e^{-i1} \end{bmatrix} \otimes e^{i0} = \begin{bmatrix} e^{i0+i1} \\ e^{i0-i1} \end{bmatrix} =$$

$$\begin{bmatrix} 1 & i \\ 1 & -1 \end{bmatrix} \otimes \dots \otimes \begin{bmatrix} 1 & i \\ 1 & -1 \end{bmatrix} \otimes \begin{bmatrix} 1 & i \end{bmatrix} \cdot \begin{bmatrix} c_k \\ s_k \end{bmatrix} \otimes \dots \otimes \begin{bmatrix} c_1 \\ s_1 \end{bmatrix} \otimes \begin{bmatrix} c_0 \\ s_0 \end{bmatrix} \Leftrightarrow$$

$$\hat{T}_c(K) = \hat{G}_c(K) \cdot \hat{S}(K).$$

$\hat{S}(K)$ represents the 2^{K+1} -dimensional real vector of the 2^{K+1} trigonometric modulations which are encoded in the real, S_{jr} , and imaginary, S_{ji} , parts of the $2^K N_t$ D spectra containing the chemical shift multiplets. $\hat{G}_c(K)$ denotes a complex matrix, and $T_c(K)$ represents the 2^K -dimensional complex vector comprising the shift evolutions of the multiplet components which are encoded in the resulting 2^K edited NMR data sets. The 2^K spectra of $T_c(K)$ are designated “basic spectra” which yield after Fourier transformation 2^K peaks at $\Omega_0 \pm \Omega_1 \pm \dots \pm \Omega_K$ encoding the desired $K+1$ chemical shifts. If two multiplets exhibit degenerate chemical shifts in the other N_t-1 dimensions, acquisition of peaks defining the centers of the chemical shift splittings (“central peaks”) can ensure unambiguous identification of the peaks forming a given shift multiplet. Moreover, degeneracy may occur among the shifts encoded in splittings. Hence, the information of the ND experiment, which resolves degeneracy in up to $N-1$ dimensions, is made available only if central peaks are detected at all frequencies $\Omega_0 \pm \Omega_1 \pm \dots \pm \Omega_{K-1}$, $\Omega_0 \pm \Omega_1 \pm \dots \pm \Omega_{K-2}$, ..., $\Omega_0 \pm \Omega_1$, and Ω_0 as described for the basic spectra. This yields a total of $p=2^{K+1}-1$ N_t D spectra, including 2^K basic spectra and 2^K-1 central peak spectra. These p data sets constitute an “(N, N_t)D GFT NMR experiment,” and one data set denotes central peaks arising from omission of m chemical shifts to be of m th order (Fig. 7.6).

7.3.2 Sensitivity of GFT NMR

The joint sampling of $K+1$ *non-constant-time* chemical shift evolution periods leads to transfer amplitudes being attenuated by

$$\exp \left[-\sum_{j=0}^K R_{2,j} \cdot t \right] \quad (7.2)$$

where $R_{2,j}$ represents the transverse relaxation rate constant of the j th dimension. However, higher-dimensional heteronuclear FT NMR spectra are quite often recorded with *constant-time* frequency labeling (Cavangh et al. 1996) and/or with $t_{\max} = 1/R_{2,j}$ for all j . Then, the linewidth in the GFT dimension is entirely determined by t_{\max} , and the 2^K lines of the chemical shift multiplet exhibit the same width as the corresponding single peak in ND FT NMR along each of the dimensions (Kim and Szyperski 2003). Each splitting arising from a jointly sampled shift reduces the S/N ratio of a peak detected in the parent FT NMR by a factor of 2. However, a factor of $\sqrt{2}$ is gained because frequency discrimination for the projected chemical shift is not associated with a FT, that is, sine- and cosine-modulated peak patterns contribute equally to the signal height detected for each component of the chemical shift multiplets. Thus, the S/N ratio of each of the 2^K components in the basic spectra is reduced by $1/\sqrt{2^K}$ when compared with the corresponding single peak in the parent FT NMR spectrum. This intrinsic loss of sensitivity is, at least partly, compensated by the fact that unambiguous identification of the GFT NMR peak pattern can be achieved at lower S/N ratios (Sect. 7.3.3). Furthermore, after symmetrization about central peaks, the S/N ratio of each of the components of the shift multiplet is comparable to the corresponding peak in the parent FT NMR spectrum (Fig. 7.7).

7.3.3 Precision of Measurements in GFT NMR

Chemical shifts are encoded in the shift multiplets in a redundant manner (Fig. 7.6), and the shifts are calculated from a system of equations that is overdetermined (Kim and Szyperski 2003, 2004; Atreya and Szyperski 2004). As a result, the accuracy of chemical shift measurements in constant-time GFT NMR experiments scales with $1/\sqrt{n}$, where n is the number of linear combinations contributing to the determination of a shift (assuming identical maximal evolution times). The increased accuracy of the measurement can be assessed by comparing the shifts measured in intra- and inter-residue GFT data (Kim and Szyperski 2004), or by comparison with conventional FT NMR spectra (Atreya and Szyperski 2004). A more detailed reasoning shall be provided in the following.

Depending on the particular protocol that is employed to extract chemical shifts from frequency-domain data, the precision of the shift measurement may depend on both the resonance linewidth at half height, $\Delta\nu_{1/2}$, and the S/N ratio. For example, if a Lorentzian function is fitted to the frequency-domain data, the shifts are obtained as a fitting parameter. Straightforward simulations show that the standard deviation of the shift measurement scales, in a very good approximation, linearly

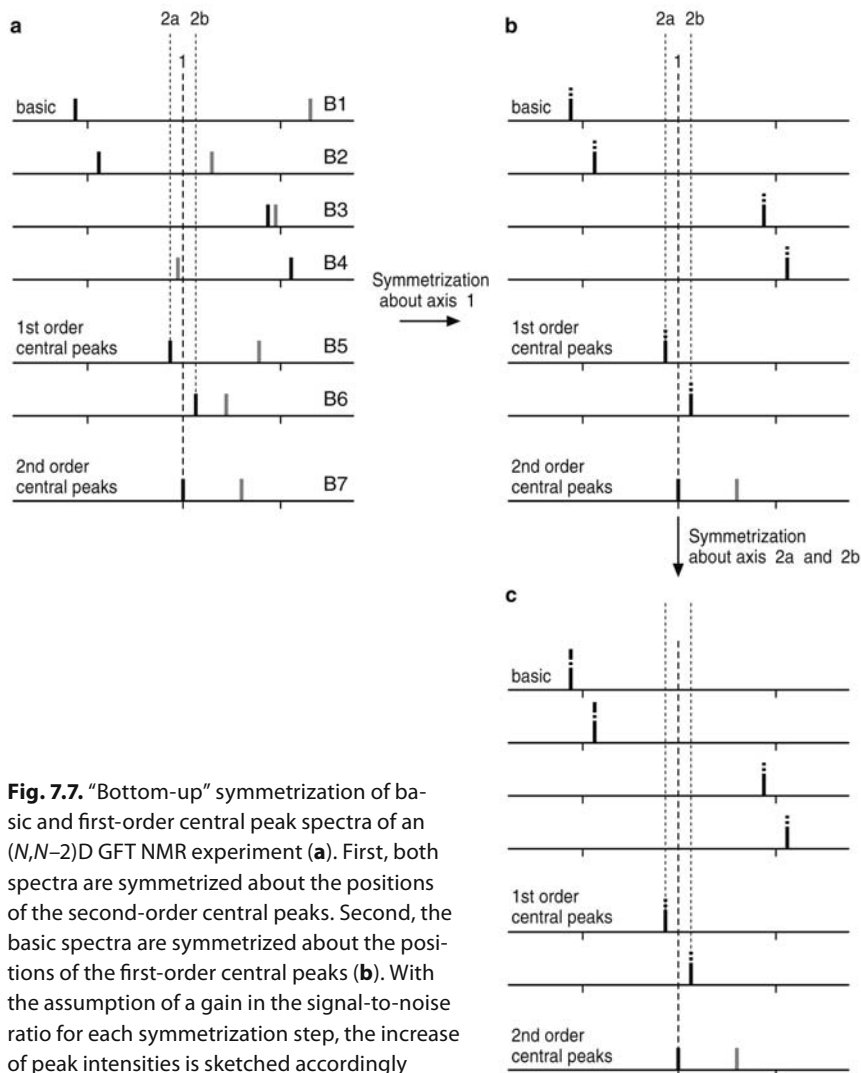


Fig. 7.7. "Bottom-up" symmetrization of basic and first-order central peak spectra of an $(N, N-2)$ D GFT NMR experiment (a). First, both spectra are symmetrized about the positions of the second-order central peaks. Second, the basic spectra are symmetrized about the positions of the first-order central peaks (b). With the assumption of a gain in the signal-to-noise ratio for each symmetrization step, the increase of peak intensities is sketched accordingly

with the inverse of the S/N ratio for a S/N ratio greater than 2. Assuming identical measurement times, the 2^K peaks in the basic spectra of a GFT NMR experiment, which correspond to a single peak in the parent experiment, exhibit S/N ratios that are reduced by $1/\sqrt{2^K}$ compared with the parent peak (Sect. 7.3.2). Consequently, the standard deviation for each of the linear combinations of shifts measured in GFT NMR is increased by about $1/\sqrt{2^K}$ when compared with the shift measurement in the parent experiment. On the other hand, a reduction of the deviation by $\sqrt{2^k}$ is gained in constant-time GFT NMR spectroscopy owing to the overdetermined system of 2^K equations yielding the desired shift. Hence, in the ideal case of having NMR data

affected solely by Gaussian white noise and having high digital resolution allowing one to fit a (known) line shape function, the precision of the single-quantum shift measurements obtained from the set of 2^K basic GFT spectra or from the conventional NMR spectrum could be quite comparable. However, caveats for realizing accurate measurements of shifts in conventional FT NMR by line-shape fitting include (1) poor digitization of indirect dimensions, (2) the presence of non-Gaussian noise (Grage and Akke 2003) and spectral artefacts, (3) line-shape variations arising from linear prediction of time-domain data for resolution enhancement (Cavanagh et al. 1996), (4) smaller phase errors arising from off-resonance effects of radiofrequency pulses and (5) peak overlap. Hence, if one could obtain some increase in *precision* through line fitting this would not necessarily warrant increased *accuracy* for the shift measurement. Consequently, such spectra are most often analyzed by identifying the peak maxima (“peak picking”), that is, above a threshold ensuring secure peak identification their intensities and line shapes are not considered.

Hence, approaches such as GFT NMR are valuable for accurate measurement of chemical shifts registered in indirect dimensions of multidimensional spectroscopy. In order to assess errors of shift measurements and their propagation in GFT NMR, a statistical model was thus adopted in which the error for the identification of the peak maximum is associated with a Gaussian distribution and the line width at half height, $\Delta\nu_{1/2}$, represents the corresponding 99.7% confidence interval. Such a Gaussian distribution model (Kim and Szyperki 2003) has been chosen in order to consider the other sources for erroneous peak picking described earlier. Hence, this model focuses on the line widths as the primary determinants for the accuracy, and is conservative in the sense that it is only requested that the peak maxima are identified with 99.7% confidence within $\Delta\nu_{1/2}$. Importantly, spectral analysis is often impeded by spectral artefacts. Thus, it is of advantage that peaks in the $2^{K+1}-1$ spectra exhibit symmetry properties resulting from the complete sampling of linear combinations of shifts. The likelihood that a chemical shift multiplet is occurring by chance as a result of noise fluctuations and/or spectral artifacts is negligible for $K>1$. Hence, the symmetry properties allow one to identify peaks with intensities that are closer to the noise level when compared with the analysis of conventional data.

7.3.4

Survey: GFT NMR and Comparison with FT NMR

The GFT NMR acquisition scheme is generally applicable and allows one to accurately adjust measurement times to sensitivity requirements. Moreover, it enables one, for the first time, to obtain 5D or higher-dimensional NMR spectral information with high precision and in a short time. Data acquisition can be further accelerated by longitudinal proton relaxation optimization (Pervushin et al. 2002; Atreya and Szyperki 2004) and nonlinear sampling combined with maximum entropy reconstruction (Hoch and Stern 1996). Concomitantly, GFT NMR data sizes are tremendously reduced when compared with FT NMR data sets, and data processing is rapid because Fourier transformation (and linear prediction) of a multidimensional subspace is replaced by fast **G**-matrix transformation and a 1D FT. As was described for RD NMR, central peaks can be derived from ^{13}C steady-state magnetization and

are obtained “for free” while acquiring data for the basic spectra derived from proton magnetization (Szyperski et al. 1996). Implementation of GFT NMR requires no additional hardware, and no a priori information is required on chemical shifts. GFT NMR data analysis is robust (Atreya and Szyperski 2004), and combination with transverse relaxation optimization (Pervushin et al. 1997) enables one to study large systems. These advantages have made GFT NMR a most valuable approach for rapid sampling of NMR data (Constans 2003; Service 2003).

7.3.5 Application of GFT NMR

7.3.5.1 Protein Resonance Assignment

By November 2004, eight NMR structures of target proteins of the Northeast Structural Genomics Consortium had been solved in our laboratory (molecular mass between 9 and 16 kDa, average 13 kDa), demonstrating high efficiency of GFT-NMR-based structure determination in high throughput (Liu et al. 2004). Automated resonance assignment (Szyperski et al. 1998; Moseley et al. 2001) for high-throughput structure determination in structural genomics (Montelione et al. 2000) strongly benefits from use of GFT NMR. First, peak lists of GFT NMR spectra may be used to establish sequential connectivities. High robustness arises from the fact that the extended set of connectivities (Atreya and Szyperski 2004; Kim and Szyperski 2004) corresponding to the matching of several *linear combinations* of shifts is redundant. This is of particular interest since automated resonance assignment protocols are sensitive to the lack of even a smaller number of connectivities. Alternatively, the GFT NMR peak lists can be used to calculate ND peak lists containing accurate single-quantum shifts. Subsequently, sequential connectivities are established on the basis of matching of the thus obtained chemical shifts. Owing to increased accuracy, correspondingly reduced matching tolerances (defined as the chemical shift difference between two shift values below which these are considered to be identical) can be employed (Atreya and Szyperski 2004; Kim and Szyperski 2004).

7.3.5.2 G-matrix Transformation for Measurement of Residual Dipolar Couplings

Residual dipolar couplings (RDCs) provide valuable information for NMR-based structural biology (Tjandra and Bax 1997; Prestegard 1998). Recently, methods for measuring multiple types of RDCs in a single experiment were devised which rely on using a G-matrix transformation (Ding and Gronenborn 2003; Atreya and Szyperski, unpublished results; Hoshino and Otting 2004). In these approaches, the cosine and sine modulations with chemical shifts are replaced by corresponding modulations with $\cos(\pi Dt)$ and $\sin(\pi Dt)$, where D represents the RDC.

7.3.6 GFT projection-reconstruction

GFT NMR experiments recorded with scaling of the jointly sampled chemical shifts (Fig. 7.2) (Kim and Szyperski 2003) have been used to reconstruct parent ND experiments to facilitate visual inspection (Coggins et al. 2003; Kupče and Freeman 2003, 2004). Such reconstruction efforts may be of value in cases where manual analysis of the reconstructed spectrum is easier than the analysis of the projected GFT NMR subspectra and/or the corresponding peak lists. A drawback of the approach is related to artefacts introduced by additional processing associated with the reconstruction.

Acknowledgements. Support of the National Institutes of Health (GM62413-01) and the National Science Foundation (MCB 0075773 and 0416899) is acknowledged.

References

- Atreya HS, Szyperski T (2004) G-matrix Fourier transform NMR spectroscopy for complete protein resonance assignment. *Proc Natl Acad Sci USA* 101:9642–9647
- Bodenhausen G, Ernst RR (1981) The accordion experiment, a simple approach to 3-dimensional NMR-spectroscopy. *J Magn Reson* 45:367–373
- Brutscher B, Cordier F, Simorre JP, Caffrey MS, Marion D (1995a) High-resolution 3D HNCOCA experiment applied to a 28-kDa paramagnetic protein. *J Biomol NMR* 5:202–206
- Brutscher B, Morelle N, Cordier F, Marion D (1995b) Determination of an initial set of NOE-derived distance constraints for the structure determination of N-15/C-13-labeled proteins. *J Magn Reson* B109:238–242
- Cavanagh C, Fairbrother WJ, Palmer AG, Skelton NJ (1996) *Protein NMR spectroscopy*. Academic, San Diego
- Coggins BE, Venters RA, Zhou P (2003) Generalized reconstruction of *n*-D NMR spectra from multiple projections: application to the 5-D HACACONH spectrum of protein G B1 domain. *J Am Chem Soc* 126:1000–1001
- Constans A (2003) NMR hits the big time. *Scientist* 17:45–48
- Ding K, Gronenborn AM (2002) Novel 2D triple-resonance NMR experiments for sequential resonance assignments of proteins. *J Magn Reson* 156:262–268
- Ding K, Gronenborn AM (2003) Simultaneous and accurate determination of one-bond ^{15}N - $^{13}\text{C}'$ and two-bond $^1\text{H}^{\text{N}}$ - $^{13}\text{C}'$ dipolar couplings. *J Am Chem Soc* 125:11504–11505
- Ernst RR, Bodenhausen G, Wokaun A (1987) *Principles of nuclear magnetic resonance in one and two dimensions*. Oxford University Press, London
- Gardner K, Kay LE (1998) The use of ^2H , ^{13}C , ^{15}N multidimensional NMR to study the structure and dynamics of proteins. *Annu Rev Biophys Biomol Struct* 27:357–406
- Grage H, Akke M (2003) A statistical analysis of NMR spectrometer noise. *J Magn Reson* 162:176–188
- Hoch JC, Stern AS (1996) *NMR data processing*. Wiley-Liss, New York

- Hoshino M, Otting G (2004) Sensitivity-enhanced double-TROSY experiment for simultaneous measurement of one-bond (15)N-(1)H, (15)N-(13)C' and two-bond (1)H-(13)C' couplings. *J Magn Reson* 171:270–276
- Kay LE, Prestegard JH (1988) Spin-lattice relaxation of coupled spins from 2D Accordion spectroscopy. *J Magn Reson* 77:599–605
- Kim S, Szyperski T (2003) GFT NMR, a new approach to rapidly obtain precise high-dimensional spectral information. *J Am Chem Soc* 125:1385–1393
- Kim S, Szyperski T (2004) GFT NMR experiment for polypeptide backbone and $^{13}\text{C}^{\beta}$ chemical shift assignment. *J Biomol NMR* 28:117–130
- Kontaxis G, Keeler J (1995) The accordion approach for “Tailored” TOCSY. *J Magn Reson* 115:35–41
- Kupče E, Freeman R (2003) Projection-reconstruction of three-dimensional NMR spectra. *J Am Chem Soc* 125:13958–13959
- Kupče E, Freeman R (2004) Fast reconstruction of four-dimensional NMR spectra from plane projections. *J Biomol NMR* 28:391–395
- Liu G, Shen Y, Atreya HS, Parish D, Shao Y, Sukumaran D, Xiao R, Yee A, Lemak A, Bhattacharya A, Acton TA, Arrowsmith CH, Montelione GT, Szyperski T (2005) NMR data collection and analysis protocol for high-throughput protein structure determination. *Proc Natl Acad Sci USA* 102:10487–10492
- Löhr F, Rüterjans H (1995) A new experiment for the sequential assignment of backbone resonances in proteins. *J Biomol NMR* 6:189–197
- Mandel AM, Palmer AG (1994) Measurement of relaxation-rate constants using constant-time accordion NMR spectroscopy. *J Magn Reson* 110:62–72
- Marion D, Wüthrich K (1983) Application of phase sensitive two-dimensional correlated spectroscopy (COSY) for measurements of ^1H - ^1H spin-spin coupling-constants in proteins. *Biochem Biophys Res Commun* 113:967–974
- Montelione GT, Zheng D, Huang Y, Gunsalus C, Szyperski T (2000) Protein NMR spectroscopy for structural genomics. *Nat Struct Biol* 7:982–984
- Moseley HNB, Monleon D, Montelione GT (2001) Automatic determination of protein backbone resonance assignments from triple resonance nuclear magnetic resonance data. *Methods Enzymol* 339:91–108
- Moseley HNB, Riaz N, Aramini JM, Szyperski T, Montelione GT (2004) A generalized approach to automated NMR peak list editing: application to reduced dimensionality triple resonance spectra. *J Magn Reson* 170:263–277
- Pervushin K, Riek R, Wider G, Wüthrich K (1997) Attenuated T-2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. *Proc Natl Acad Sci USA* 94:12366–12371
- Pervushin K, Vogeli B, Eletsky A (2002) Longitudinal ^1H relaxation optimization in TROSY NMR spectroscopy. *J Am Chem Soc* 124:12898–12902
- Prestegard JH (1998) New techniques in structural NMR–anisotropic interactions. *Nat Struct Biol* 5:517–522
- Service RF (2003) Propelled by recent advances, NMR moves into the fast lane. *Science* 299:503
- Szyperski T, Wider G, Bushweller JH, Wüthrich K (1993a) 3D ^{13}C - ^{15}N heteronuclear two-spin coherence spectroscopy for polypeptide backbone assignments in ^{13}C - ^{15}N -double labeled proteins. *J Biomol NMR* 3:127–132

- Szyperski T, Wider G, Bushweller JH, Wüthrich K (1993b) Reduced dimensionality in triple resonance NMR experiments. *J Am Chem Soc* 115:9307–9308
- Szyperski T, Pellecchia M, Wüthrich K (1994) 3D $^1\text{H}^{\alpha/\beta}\text{C}^{\alpha/\beta}(\text{CO})\text{NHN}$, a Projected 4D NMR experiment for the sequential correlation of polypeptide $^1\text{H}^{\alpha/\beta}$, $^{13}\text{C}^{\alpha/\beta}$ and backbone ^{15}N and $^1\text{H}^{\text{N}}$ chemical shifts. *J Magn Reson* B105:188–191
- Szyperski T, Braun D, Fernandez C, Bartels C, Wüthrich K (1995) A novel reduced-dimensionality triple resonance experiment for efficient polypeptide backbone assignment, 3D COHNNCA . *J Magn Reson* B108:197–203
- Szyperski T, Braun D, Banecki B, Wüthrich K (1996) Useful information from axial peak magnetization in projected NMR experiments. *J Am Chem Soc* 118:8147–8148
- Szyperski T, Banecki B, Braun D, Glaser RW (1998) Sequential assignment of medium-sized $^{15}\text{N}/^{13}\text{C}$ -labeled proteins with projected 4D triple resonance NMR experiments. *J Biomol NMR* 11:387–405
- Szyperski T, Yeh DC, Sukumaran DK, Moseley HNB, Montelione GT (2002) Reduced-dimensionality NMR spectroscopy for high-throughput protein resonance assignment. *Proc Natl Acad Sci USA* 99:8009–8014
- Tjandra N, Bax A (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium. *Science* 278:1111–1114
- Tolman JR, Prestegard JH (1996) Measurement of amide ^{15}N - ^1H one-bond couplings in proteins using accordion heteronuclear shift correlation experiments. *J Magn Reson* B112:269–274
- Xia Y, Arrowsmith C, Szyperski T (2002) Novel projected 4D triple resonance experiments for polypeptide chemical shift assignment. *J Biomol NMR* 24:41–51

Protein Nanomechanics – as Studied by AFM Single-Molecule Force Spectroscopy

MARIANO CARRIÓN-VÁZQUEZ, ANDRÉS F. OBERHAUSER, HÉCTOR DÍEZ,
RUBÉN HERVÁS, JAVIER OROZ, JESÚS FERNÁNDEZ, DAVID MARTÍNEZ-MARTÍN

I am inspired by the biological phenomena in which chemical forces are used in repetitious fashion to produce all kinds of weird effects (one of which is the author).

*Richard Feynman,
1959 talk at the CalTech (Feynman 1960)*

8.1 Introduction

Proteins are in charge of virtually every biological process that occurs in modern cells. In isolation or in complexes they often perform their functions through structural changes in their intra- or intermolecular bonding. Understanding the inner workings of these biological machines remains one of the major frontiers in biology. Many of these “bionanomachines” use mechanical forces in different cell processes ranging from replication, transcription, translation, and protein degradation to organelle transport, cell adhesion, and cell crawling. Several manipulation techniques can now be employed to investigate these mechanical processes at the single-molecule level. This review will start by introducing both protein machines and single-molecule manipulation techniques (Sect. 8.2).

Atomic force microscopy (AFM. Also used through the text as an acronym for atomic force microscope) in its single-molecule force spectroscopy mode (for simplicity, referred to as SMFS here; see Sect. 8.3) is one of the nanomanipulation techniques most extensively used for the study of the mechanical properties of proteins, particularly to examine their response to stretching (i.e., molecular elasticity and mechanical stability) and to unbinding. Several reviews have recently been dedicated to this topic, focusing on the analysis of intramolecular (Carrion-Vazquez et al. 2000; Isralewitz et al. 2001; Best et al. 2003a; Zhuang and Rief 2003; Bustamante et al. 2004; Rounsevell et al. 2004; Samori et al. 2005) or intermolecular interactions (Hinterdorfer 2002; Weisel et al. 2003), or both (Zlatanova et al. 2000; Merkel 2001).

The aim of this review is to provide the nonspecialist with a quick guide so that she/he can become familiar with the central themes of this new and exciting field. The main concepts, ideas, achievements, biological implications, and limitations will be discussed. Particular emphasis will be placed on the analysis of the intramolecular interactions of proteins (given that they have been more extensively characterized) and we will also highlight the advantages of combining this technique with protein engineering and molecular dynamics in order to elucidate the molecular determinants underlying the mechanical stability of proteins.

Comprehensive descriptions that provide the experimental details behind this new methodology can be found elsewhere (Carrion-Vazquez et al. 2000; Rounsevell et al. 2004). For an in-depth discussion on the role of mechanical forces in biochemistry, as well as for a review of the closely related field of motor proteins, see the works of Howard (2001) and Bustamante et al. (2004).

8.2

Biological Machines, Single-Molecule Manipulation and Protein Mechanics

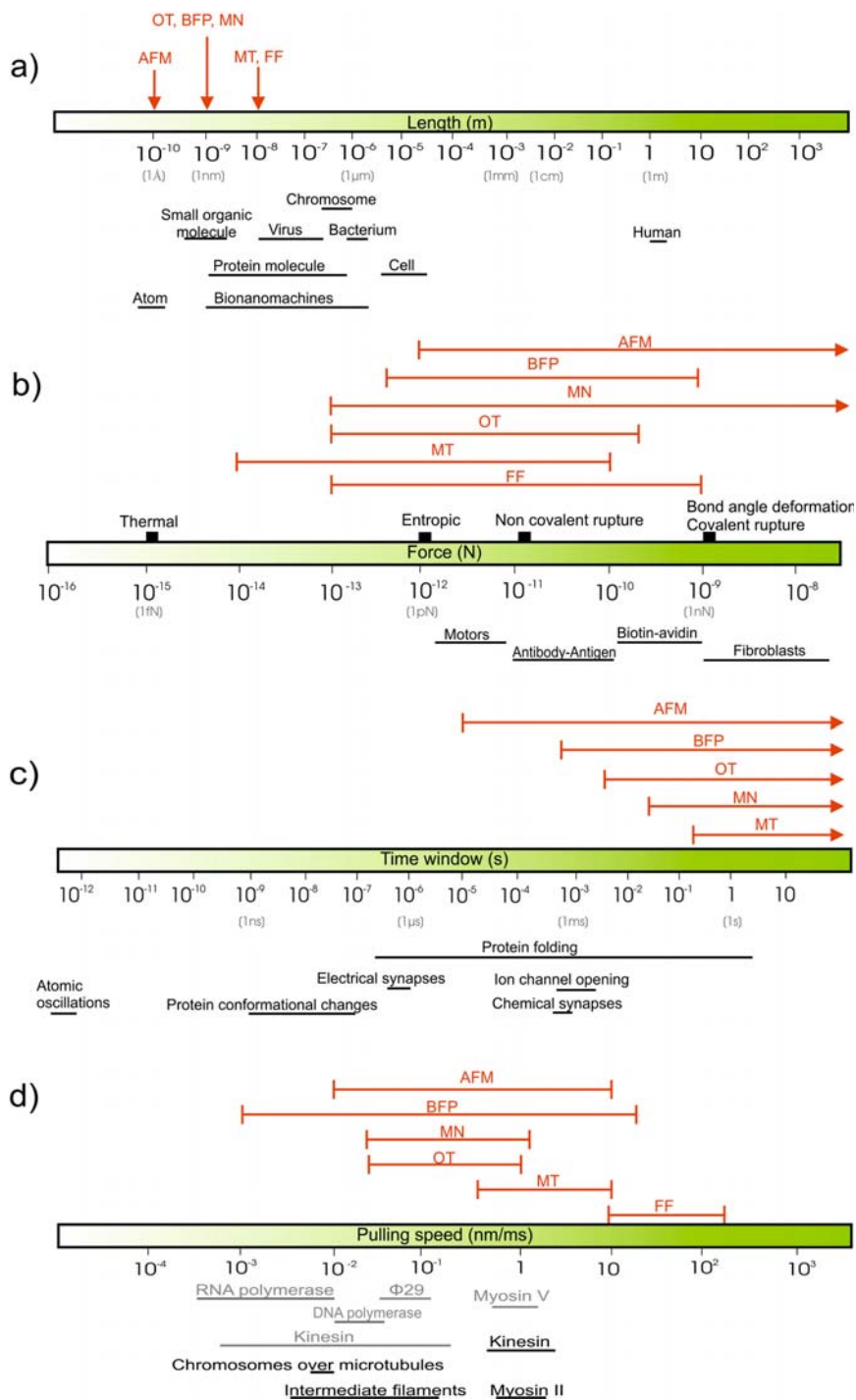
8.2.1

Protein Machines

Under the present paradigm of modern biochemistry, the cell can be regarded as a factory-like system. It is crowded with a variety of specialized molecular machines of nanometer size (Fig. 8.1a), most of which are proteins either as single polypeptides or as complex assemblies of protein “parts” (Alberts 1998). Very often, their functions involve the conversion of chemical energy (stored or supplied) into mechanical work through changes in their conformation. Thus, from the cell nucleus to the extracellular matrix there are “bionanomachines” involved in mechanical processes as diverse as replication, transcription, translation, protein folding, protein and nucleic acid unfolding, protein degradation, nucleic acid and protein translocation, organelle transport, cell adhesion, membrane fusion, or cell crawling (Howard 2001; Bustamante et al. 2004).

Although the “machine” metaphor is a useful first approximation to describe these cellular structures and there are striking parallels between artificial and natural machines, it does not hold for two important aspects. Indeed, these are issues that prevent us from learning about the behavior of natural machines by analogy to artificial ones. First, since bionanomachines are products of evolution, rather than being “designed” by engineers, their structure has been shaped by natural selection instead of human planning. Second, in contrast to macroscopic machines (i.e., conventional machines, excluding the artificial nanomachines) that are mainly affected by gravity and inertia, bionanomachines are mainly dominated by the so-called thermal “forces”, owing to their smaller mass (strictly speaking, in the current view of phys-

Fig. 8.1. Single-molecule manipulation (SMM): relevant parameters and scales. **a** Spatial resolution of SMM techniques (*red*) and size of representative biological structures (*black*). **b** Force ranges of SMM techniques (*red*), physicochemical forces at the molecular level (*black squares*), and force ranges of typical biological machines and interactions (*black lines*). **c** Temporal resolution of SMM techniques (*red*) and the time duration of characteristic biological processes (*black*). **d** Pulling speed range of SMM techniques (*red*) and representative speed range of protein machines measured *in vitro* (*grey*) and *in vivo* (*black*). *AFM* atomic force microscopy, *OT* optical tweezers, *BFP* biomembrane force probe, *MN* microneedles, *MT* magnetic tweezers, *FF* flow field



ics there are only four fundamental forces in nature from which the other “forces” derive: strong nuclear, weak nuclear, electromagnetic, and gravitational). Since thermal forces are ubiquitous and random in nature (Fig. 8.1b), at room temperature, bionanomachines in their aqueous environment experience a constant “bombardment” by the numerous water molecules that surround them. These forces are in the femtonewton (10^{-15} N) range and result in what is called Brownian motion (a phenomenon originally described by the botanist Robert Brown, in 1827, and later explained in physical terms by Albert Einstein, in 1905). This unfamiliar environment confers on them the particular properties that dominate their behavior. It is through this thermal agitation of their molecular structure that proteins reach the high-energy transition states that are essential in biochemical reactions (thermal energy is $k_B T = 4.1$ pN nm = 0.6 kcal mol $^{-1}$, at room temperature; k_B is the Boltzmann constant and T is the absolute temperature). To understand life inside the cell, it is essential to know how these protein machines move their parts and change shape in response to the mechanical and thermal forces present in their nanoenvironment. The energies involved in protein conformational changes (the “signals” in our experiments) are just above thermal energy levels (the “noise”), typically ranging from $1 k_B T$ (thermal energy) to $25 k_B T$ (ATP hydrolysis), such that the structures are stable enough to prevail at physiological temperatures. Owing to these natural fluctuations, the behavior of these machines is stochastic and typically very “noisy.” In contrast, artificial machines use much higher energy values than thermal energy to work rapidly, accurately, and deterministically.

Like the other major biological macromolecules (DNA, RNA, and polysaccharides), proteins can be thought of as polymers (more specifically, they are heteropolymers or copolymers), i.e., articulated chains of atoms, that interact in a few, well-defined ways. It would be necessary to apply quantum mechanics to provide a quantitative description of the properties of the atoms involved, but fortunately most of the basic properties of these biomolecules can be understood qualitatively using a small set of simplified rules such as entropic elasticity, steric repulsion, van der Waals forces, electrostatic interactions, and hydrogen bonding. The central concept here is that two types of bonds exist: covalent bonds, connecting monomeric units into a primary structure; and noncovalent interactions, which stabilize their tertiary structure and permit structural (conformational) changes upon binding to other biomolecules. Proteins are synthesized as flexible chains that typically fold spontaneously into a unique and compact 3D structure, the so-called native conformation. Thus, through the course of evolution random molecular fluctuations have been harnessed into self-organized protein structures. These protein “folds” are impressive examples of self-organization (a key feature of biological systems), which is also responsible for the spontaneous assembly into the functional complexes we call bionanomachines.

8.2.2 Range of Relevant Forces in Biology

The native conformation of a protein represents a minimum of its free energy. Nevertheless the stability of proteins is very marginal as their free energies of un-

folding range from 5 to 15 kcal mol⁻¹ (8–25 $k_B T$), a minute amount of energy that results from subtracting the two large quantities that represent the energy of the native and unfolded states (Fersht 1999). The native conformation is mainly stabilized by weak localized interactions (electrostatic interactions, van der Waals forces, hydrogen bonds, and the hydrophobic effect). These are the same forces that stabilize intermolecular bonds in protein pairs. Therefore, it is not surprising that, as we will see later (Sect. 8.3.3), the process of forced unfolding of proteins has been modeled in close analogy with that of the force dissociation of bonded pairs.

What then is the magnitude of the biologically relevant forces that affect proteins? Given that changes in protein conformation are measured in the angstrom to nanometer range and the energies involved range from 1 to 25 $k_B T$, the relevant biological forces are expected to be in the piconewton range. The ranges of the diverse biological forces known are summarized in Fig. 8.1b. Because proteins are subject to the ubiquitous thermal forces, the number of possible configurations (entropy) is at its maximum when a protein forms a random coil or is denatured. This entropy becomes progressively reduced with the formation of secondary and tertiary structures. Stretching these structures in the low-force regime, to overcome “entropic forces,” has been achieved experimentally and requires the application of forces on the order of a few piconewtons (Sect. 8.2.5). Several molecular motors such as myosin, kinesin, and RNA/DNA polymerases also generate forces in the same range, from a few piconewtons to tens of piconewtons. Indeed, the strongest molecular motors known to date are the portal protein of the $\Phi 29$ bacteriophage (Smith et al. 2001) and the pilus motor of bacteria (Maier et al. 2002), which are capable of generating forces of up to approximately 60 and 110 pN, respectively. The next force magnitude group, the “enthalpic force” regime, includes the forces needed to unfold folded domains of proteins (intramolecular interactions) as well as those required to overcome specific intermolecular interactions such as ligand/receptor or antigen/antibody. These forces are typically in the range 100–300 pN, when measured at a high loading rate (this parameter is discussed in Sect. 8.3.3). However, since the typical loading rates *in vivo* may be much lower in some cases, the corresponding forces may also be lower as shall be seen later (Sect. 8.3.3). Finally, the forces needed to break apart covalent bonds are almost 2 orders of magnitude larger, in the range of a few nanonewtons (reviewed by Zlatanova and Leuba 2003). The kinetic range covered by the different nanomanipulation techniques and the typical velocities of some biological processes are shown in Fig. 8.1d.

8.2.3

Mechanical Proteins and Mechanical Nanomachines

Some proteins in the cell have a defined “mechanical function” related to their molecular elasticity and/or plasticity. These proteins can be classified into two groups: the so-called elastomeric proteins, which present entropic elasticity (e.g., elastin, PEVK, and N2B regions of titin) (Sect. 8.3.3), and the “enthalpic proteins,” which show also an enthalpic component in their elasticity. The “mechanical stability,” i.e., resistance to unfolding in response to an applied mechanical force, of this latter group is a parameter of physiological importance (Sect. 8.3.3). These

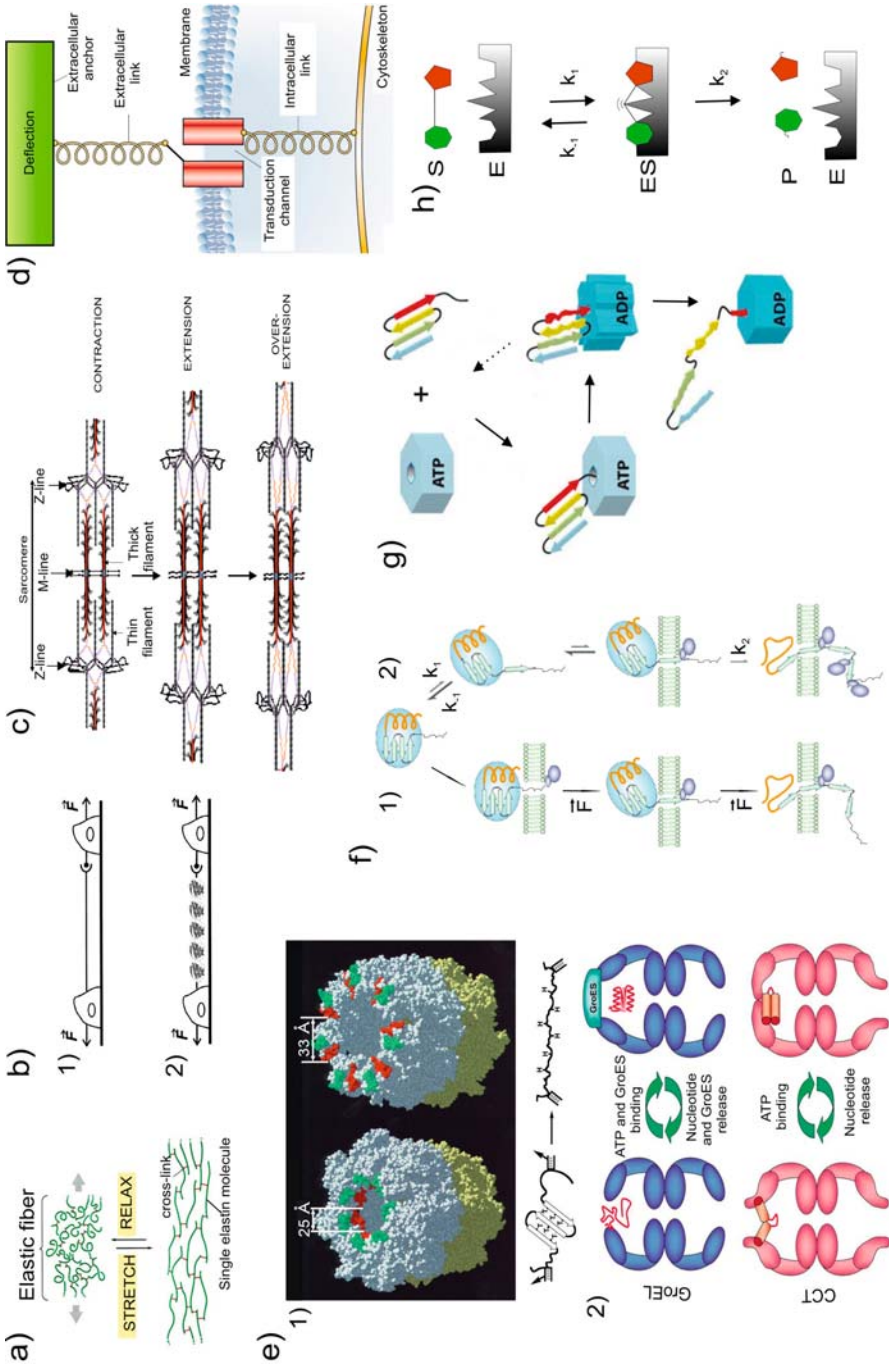


Fig. 8.2. Protein mechanics in vivo. **a** The extracellular matrix protein elastin is an entropic elastomer. After pulling, elastin molecules stretch owing to their entropic elasticity (adapted, with permission, from Alberts et al. 2002, copyright 2002 Taylor & Francis Group). **b** Adhesion protein mechanics: modular cell adhesion receptors (2) may function as shock absorbers by increasing the range and lifetime of adhesion bonds, whereas nonextensible receptors (1) may display higher bond rupture forces (see text for details). The “protective” unfolding of sacrificial modules (weaker than the adhesive bond) of a modular protein results in an increase in the total work needed for the unbinding process. This mechanical shock-absorber effect increases the range and lifetime of bonds and offers an explanation for why adhesion proteins (whose function is often related to mechanical resistance) are typically modular (adapted from Fisher et al. 2000). **c** Changes in the elastic protein titin during sarcomere contraction, extension, and overextension. The regions from titin are represented by different colors: fibronectin and fibronectin/immunoglobulin (Ig) domains anchored to the thick filament (red), PEVK and N2B unique sequences (orange), and Ig domains (proximal and distal to the Z disk; violet); the last two regions are the extensible part of titin (modified, with permission, from Tskhovrebova and Trinick 2003, copyright 2003 Macmillan Publishers Ltd.). **d** Mechanosensitive ion channels. These types of ion channels are present, for instance, in the cilia of the renal epithelial cells and in the stereocilia of the hair cells in the auditory and vestibular systems. The bending of the cilia opens mechanically gated transduction channels allowing ion influx. The key proteins involved in the mechanics of these channels are still under debate (modified from Gillespie and Walker 2001, with permission, copyright 2001 Macmillan Publishers Ltd.). **e** Folding–unfolding cell machines: chaperonins. 1 In the bacterial chaperonin GroEL, prior to the folding process, conformational changes increase the relative distance between the substrate binding sites, suggesting a “racklike” model for forced protein unfolding (after Shtilerman et al. 1999, with permission, copyright 1999 American Association for the Advancement of Science). 2 Comparison between protein folding mediated by GroEL–GroES and CCT chaperonins. GroEL–GroES is assumed to perform a “free folding” of protein substrates, while CCT may induce conformational changes that mechanically force the substrate protein to fold in the correct way (after Valpuesta et al. 2002, with permission, copyright 2002 Elsevier Science). **f** Mitochondrial protein transport. 1 “Pulling model”: mitochondrial chaperone mtHsp70 involved in protein translocation functions as a molecular motor that actively pulls to unfold the protein. 2 This chaperone may alternatively act as a Brownian motor taking advantage of spontaneous unfolding events to translocate the protein substrate into the mitochondrion (“trapping model”). It has recently been proposed that these two models could also coexist, each one being just an extreme case of the same process (reviewed by Oster and Wang 2003) (modified from Neupert and Brunner 2002, with permission, copyright 2002 Macmillan Publishers Ltd.). **g** Compartmental proteases: the AAA+ (hexameric ring) ATPase from the proteasome and other related proteases unfold proteins, in an ATP-dependent manner, prior to their translocation to the catalytic chamber for degradation. Force attempt failures would result in the release of the substrate (dotted arrow) (modified from Sauer et al. 2004, with permission, copyright 2004 Elsevier Science). **h** The Haldane–Pauling model (“strain and distortion”) for enzyme-mediated catalysis postulates that mechanical tension would break the links between molecules

proteins are located within particular cellular machines such as the cytoskeleton (sarcomeric and cortical), the adhesion machinery located at the cell–cell interface (cell–cell and cell–substrate adhesion proteins), or the mechanosensitive ion channels. Examples of these proteins can be seen in Fig. 8.2a–d.

The enthalpic group of mechanical proteins seems to have evolved to resist force such that mechanical unfolding is critical for their function. However, mechanical unfolding in a cell may not be limited to proteins with a mechanical function. There is increasing evidence to indicate that many “unfoldases” may operate mechanically, i.e., these proteins would actively unfold (or fold) their protein substrates by exerting mechanical forces on them using chemical energy from the hydrolysis of ATP (reviewed by Valpuesta et al. 2002; Matouschek 2003; Prakash and Matouschek 2004; Sauer et al. 2004). These unfoldases include molecular chaperones, compartmental proteases such as proteasomes, and the protein translocases that constitute the protein import machinery of the mitochondria, chloroplasts, and endoplasmic reticulum (examples of these machines can be seen in Fig. 8.2e–g). If this model is confirmed, most proteins in the cell would be mechanically unfolded at some point or another during their lifespan. This means that the mechanical properties of many proteins may be critically important. In addition to being generated by these mechanochemical enzymes, mechanical forces may also underlie the activity of other enzymes, according to the hypothesis of the tension-induced catalysis proposed by Haldane–Pauling (reviewed by Bustamante et al. 2004; Fig. 8.2h). This hypothesis, which remains to be demonstrated, postulates that enzyme catalysis may work by inducing mechanical tension on the enzyme–substrate complex.

The nanomachinery involved in all these processes and the mechanochemistry behind them are still not well understood. Although current technology still does not allow us to directly measure the forces involved *in vivo* (except for the case of extracellular proteins), much can be learned *in vitro* by the study of well-defined model systems, while we await the exciting developments to come. In the meantime, a first generation of nanomanipulation techniques allows us to apply mechanical forces to single protein molecules and single interacting proteins and to analyze their responses.

8.2.4 Single-Molecule Techniques

Technical innovation has always been central to scientific progress. Classical biochemical techniques deal with large ensembles of molecules in the range of the Avogadro’s number ($N_A \sim 10^{23}$ molecules per mol) and can therefore only provide average parameters from them. In contrast, single-molecule techniques allow the direct study of individual molecules, eliminating the “population noise” and the need for the difficult-to-accomplish synchronization of reactions (a must when analyzing the dynamics of multistep processes in the ensemble). Using these techniques, we can obtain a histogram of the actual distribution of values for an experimental parameter (i.e., the probability distribution function), instead of the population average. One clear advantage of this approach is that it enables us to directly probe the existence of rare intermediates. Thus, single-molecule biophysics permits the dynamics

of molecular processes to be analyzed in real time with an unprecedented resolution (reviewed in Leuba and Zlatanova 2001). A pioneer among these techniques was patch-clamping, which enabled the electrical activity of a single ion channel to be recorded directly (Neher and Sakmann 1976).

8.2.5 Single-Molecule Manipulation Techniques

With the recent advent of the so-called single-molecule manipulation techniques, we are now able to manipulate single proteins and study their mechanical properties as well as their function (i.e., the dynamics of protein structure, which typically involves structural changes in their intramolecular – “conformation” – or intermolecular – “interactions” – bonding). The techniques for single-molecule manipulation include mechanical force transducers such as AFM, biomembrane force probe (also dubbed dynamic force spectroscopy, another form of SMFS), microneedles and optical fibers, as well as external field manipulators like optical tweezers, magnetic tweezers, and the flow-field apparatus. These are complementary techniques that cover overlapping ranges of force and that also differ in their spatial resolution and dynamic range (time window) (Fig. 8.1). Three of these techniques have been used extensively to characterize the mechanical properties of intra- and intermolecular interactions of single proteins: optical tweezers (Ashkin 1970), biomembrane force probe (Evans et al. 1995), and AFM (Binnig et al. 1986). While the first two are based on soft springs, the last one is based on stiff springs.

The biomembrane force probe technique has essentially been used to measure the bond strength of intermolecular interactions in protein–biomolecule pairs (Evans and Ritchie 1999). Optical tweezers have been widely used to study the mechanics of motor proteins (reviewed by Howard 2001, Oster and Wang 2003 and Bustamante et al. 2004). Finally, AFM has been applied mostly to characterize the mechanical resistance of both individual polypeptides (intramolecular interactions) and protein–biomolecule bonds (intermolecular interactions) (Bustamante et al. 2001). With the exception of AFM (several companies now sell apparatuses designed to carry out this technique) most single-molecule manipulation instruments are currently custom-built, they are not commercially available and, therefore, remain inaccessible to the nonspecialist.

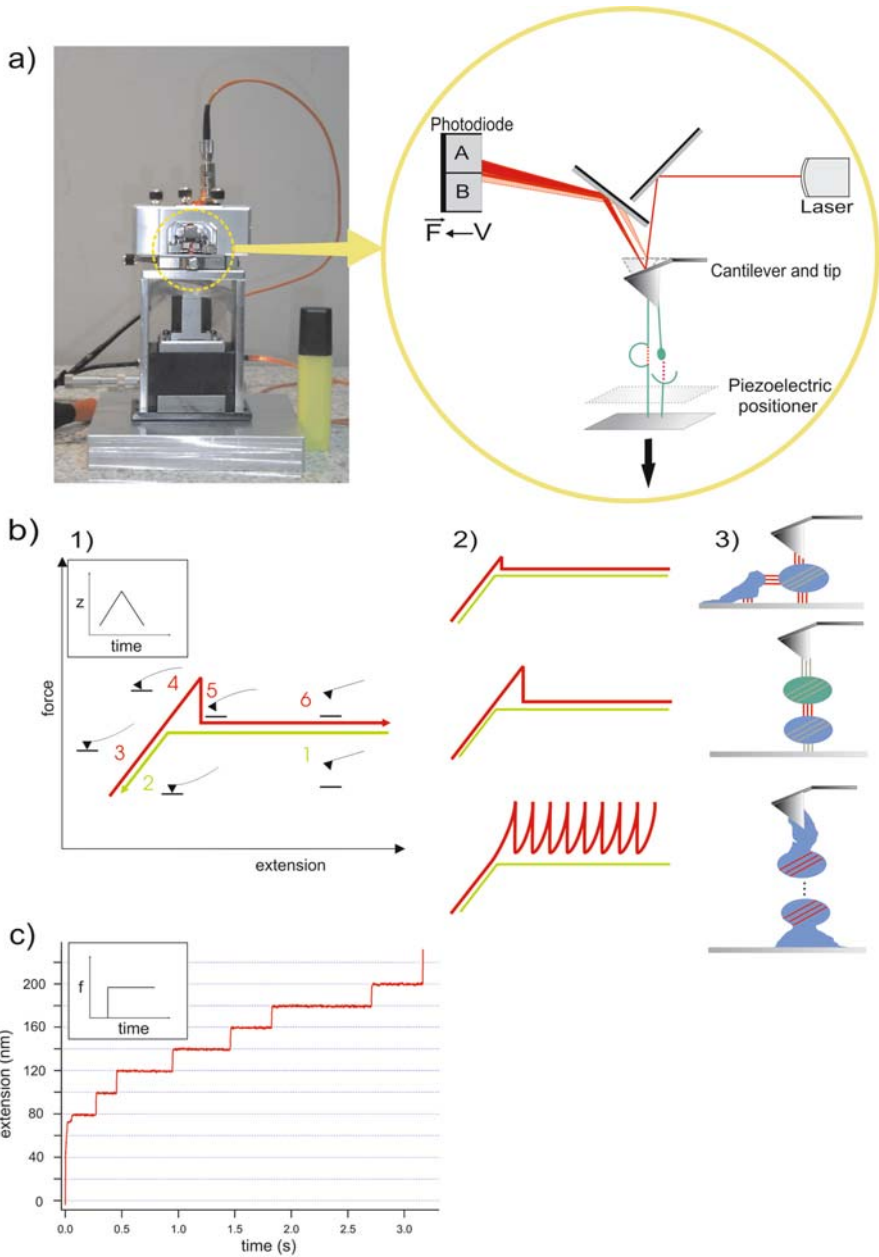
Compared to all the other nanomanipulation techniques, the remarkable advantages of AFM are the distance resolution and the precision of positioning, while its bottleneck is the limited resolution in the low-force regime. Furthermore, AFM is particularly suitable for single-molecule studies (whose goal is to address a single molecule or a single molecular pair), since the curvature radius of an AFM tip (see section 8.3.1.1) is much smaller than that of optical or magnetic beads and, therefore, the number of individual molecules (or molecule pairs) that it can hold is orders of magnitude smaller.

8.3 SMFS of Proteins: Physical Principles and Methodology

The AFM, invented in 1986 by Binning and coworkers, is a relatively simple apparatus that allows the measurement of the mechanical properties of materials at atomic resolution. It is essentially a mechanical force transducer based on a flexible sensor (the cantilever) that ends in a sharp tip, which directly contacts the sample. The forces are measured on the basis of the deflection of the cantilever, and distances are calculated from the travel of a high-resolution positioner. It works like a miniature record player or phonograph that connects our macroscopic world with the atomic world. This microscope (“nanoscope” may be a more appropriate word to describe it from a biologist’s point of view) has been most widely used in its “imaging” configuration to mechanically map the topography of surfaces at atomic resolution. The principle of this technique is more akin to touch than to vision. By using a 3D positioner, a sharp probe (the “eye” of this microscope) combs the surface of the sample recording the height at each point, yielding a topographic map of the surface (Bustamante et al. 1997). The resolution gets only truly “atomic” in the *z*-axis for hard materials, whereas for soft biological material the typical lateral resolution in this mode is around 1 nm.

The AFM is not just a high-resolution imaging tool, it can also be used to probe and manipulate atoms and molecules. Subsequently, the so-called force spectroscopy

Fig. 8.3. Single-molecule force spectroscopy (SMFS): physical principle and the modes of operation. **a** Schematic diagram of the physical principle of typical AFM pulling experiments. The *line* depicts the laser light path before (*dashed line; pale red*) and after (*solid lines; red*) pulling on the protein system. The protein system connects mechanically the tip and the substrate, which is in turn attached to a piezoelectric positioner. In this setup the movement of the positioner along the *z*-axis results in bending of the cantilever along the same axis. This bending is tracked by changes in the reflected angle of a laser beam bounced off the cantilever, which in turn is detected by a split photodiode as a voltage difference between the two channels and is converted into force using Hooke’s law (Eq. 8.1). SMFS is used to measure intra- and intermolecular interactions (*red dots*), which are represented as a protein and a protein–biomolecule pair, respectively (*green*). **b** 1 Typical force curve diagram in length-clamp SMFS showing different snapshots of the movement of the cantilever and tip as the positioner performs an approach–retraction cycle. When the substrate is not in contact with the tip (1); when in contact with the tip, which bends the cantilever (2); when it is withdrawn from the tip, which bends the cantilever the other direction (3); adhering to the tip, originating a force peak (4); on “jumping off contact” from the tip (5); when is not in contact with the tip (6). 2 Different schematic diagrams of force–extension showing recordings of nonspecific interactions, specific interactions, and a typical sawtooth pattern obtained by stretching of a multidomain protein molecule. 3 Cartoon representing the different interactions depicted in 2. *Top*: nonspecific interactions between the protein and the tip, the substrate, or another protein are detected (*red bars*), but not intramolecular interactions (*light brown bars*). *Center*: a specific protein–protein interaction is detected (*red bars*). *Bottom*: intramolecular interactions in a multidomain protein are detected (*red bars*). Proteins or domains denatured by adsorption (to the substrate



or the tip) are represented by amorphous symbols (modified from Zlatanova and Leuba 2003, with permission, copyright 2003 Elsevier Science). **c** Force-clamp mode of SMFS showing the typical staircase extension–time recording. This particular example shows a polyubiquitin protein (N–C linked) being stretched at a constant force (110 pN) (Schlierf et al. 2004)

copy or force-measuring configuration of the AFM was designed exclusively to record force–extension curves and, accordingly, it is based on a single piezoelectric positioner (z -axis; Fig. 8.3). These “force spectrometers” or “pullers” are typically more compact and rugged instruments (in which thermal drift and mechanical vibrations can be minimized) and they are used to measure the nanomechanical properties of immobilized materials such as biomolecules just by stretching them (Burnham and Colton 1989). They have been used in this way to mechanically unfold biopolymers and to determine the unbinding forces for various biomolecular pairs. Typically, this mode of SMFS is capable of measuring forces with a sensitivity of tens of piconewtons (Fig. 8.1b) and changes in length with nanometer resolution (Fig. 8.1a; Sects. 8.3.1, 8.6.1). The timescale resolution of this technique is in the submillisecond range (Fig. 8.1c). The resolution of this instrument allows the analysis of single-molecule unfolding events as well as single rupture events. In this way, the intermolecular binding forces of complementary DNA strands (Lee et al. 1994) and those of receptor–ligand systems (Florin et al. 1994) can be measured as can the intramolecular forces that maintain the conformation of polysaccharide rings (Rief et al. 1997a; Marszalek et al. 1998, 1999a) and the fold of the different domains of a modular protein (Rief et al. 1997b; Fig. 8.3b).

This section will mainly focus on the use of this technique to analyze intramolecular interactions in proteins since the systems involved for these studies are relatively simpler than those used to characterize intermolecular interactions.

8.3.1 Mechanical Force Transduction by AFM

As shown in Fig. 8.3, the AFM transduces the forces exerted on its flexible cantilever by measuring the angular deviations of a laser beam bounced back off it. Through this optical method, deviations are converted into voltage differences by a split photodiode detector. However, the raw data provided by an AFM (photodiode voltage, ΔV , and positioner extension, Δz_p) have to be transformed in order to obtain the final force–distance plot (Sect. 8.3.1.2). The cantilever of an AFM behaves as a Hookean spring. Following Hooke’s law, the force is calculated as the product of the cantilever deflection (in the z -axis) and its spring constant:

$$F = -k\Delta z_c, \quad (8.1)$$

where k is the spring constant of the cantilever and Δz_c is its displacement along the z -axis (Figs. 8.3a, 8.4a).

8.3.1.1 Cantilever and Tip Choices

In these experiments, proteins typically attach nonspecifically to the cantilever tip, which can pick out single molecules even from a dense layer of proteins. Although unsharpened cantilever tips are very much larger than the size of a typical protein (these tips have a radius of curvature of approximately 50 nm while

the major dimension of a typical protein is approximately 5 nm), they are preferred to sharpened ones. The reason for this is because the larger tips have a better chance of finding single-molecule events, which as we will see later, is a process of trial and error. The spring constant of the cantilever, k , is typically in the range of 10–100 pN nm⁻¹. For a typical cantilever of 60 pN nm⁻¹ the thermal noise of the force measurements is calculated to be approximately 16 pN (the root-mean-square force fluctuation; Bustamante et al. 2000). In SMFS of proteins the use of a stiff cantilever simplifies the analysis by ensuring that the macromolecule is the dominant elastic component in the whole system. As we will see later, this is the main assumption in the Monte Carlo simulations used to obtain the kinetic parameters of the unfolding process.

8.3.1.2

Calibration of the AFM

The mechanical resistance (restoring force) of a protein bond probed with this sensor is obtained using Eq. 8.1. However, the photodiode voltage (ΔV) and positioner extension (Δz_p) should be transformed to get the final force–distance plot. In order to do that, prior to an SMFS experiment two constants have to be determined: the ratio of photodiode output voltage to cantilever deflection ($\Delta V/\Delta z_p$) to obtain the Δz_c values, and the spring constant of the cantilever (k) (Fig. 8.4a).

Determination of the Optical Lever Sensitivity

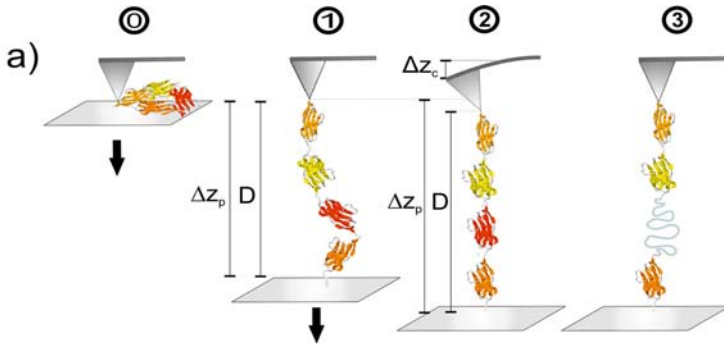
The split photodiode detector converts the incident light in each half of the split photodiode into voltage and then outputs the voltage difference. The ratio between the photodiode output voltage and the displacement of the piezoelectric positioner in the z direction ($\Delta V/\Delta z_p$) is determined by performing an approach–retraction cycle and measuring the slope of ΔV vs. Δz_p when the tip is in contact with a hard “substrate” (also called sample stage, typically made of glass or mica) at high applied forces (high compliance region), assuming that there is no deformation of the substrate. In these conditions the displacement of the cantilever, Δz_c , equals that of the piezoactuator, Δz_p . The value of this slope is often called optical lever sensitivity or sensitivity for short, C (in units of volts per nanometer), which depends on the dimensions and the shape of the laser spot on the photodiode and on the refractive index of the medium used. This proportionality factor allows the raw photodiode data, ΔV , to be transformed into the cantilever deflection, Δz_c :

$$\Delta z_c = \Delta z_p = \Delta V / C. \quad (8.2)$$

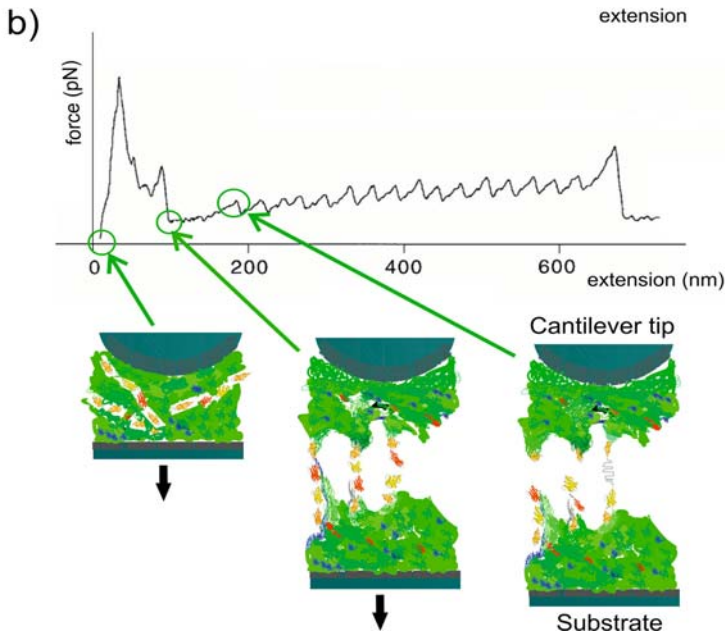
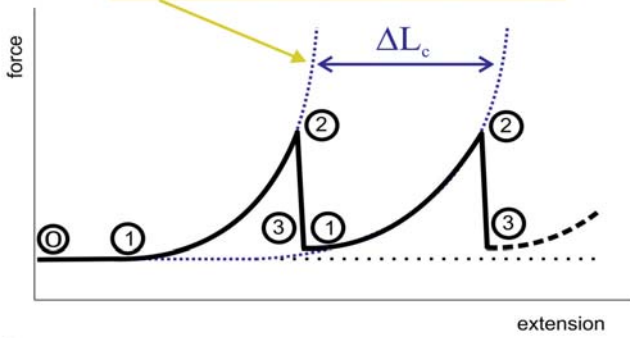
Then, from Eq. 8.1 the force value can be calculated as long as the spring constant of the cantilever is known (see next section).

Determination of the Spring Constant of the Cantilever

The most commonly used method for determining the spring constant of a cantilever is the thermal method, which models the cantilever as a damped simple har-



WLC
$$F = \frac{kT}{p} \left[\frac{1}{4(x-L_c)^2} + \frac{x}{L_c} - \frac{1}{4} \right]$$



monic oscillator fluctuating in response to thermal noise. The mechanical properties of the cantilever are related to the frequency and amplitude of these oscillations. Hence, its spring constant can be calculated using the so-called equipartition theorem:

$$\langle \Delta x^2 \rangle = k_B T / k, \quad (8.3)$$

where the first member of the equality represents the mean-square displacement noise. This calculation has a typical error of $\pm 20\%$ (Hutter and Bechhoffer 1993; Florin et al. 1995). It must be noted also that the position of the spot on the back of the cantilever has an influence on the determination of its spring constant and that a method has been developed to correct this effect (Proksch et al. 2004). An alternative method to the thermal noise one is based on the shift in the resonance frequency of the cantilever after the addition of a small mass. This method is more accurate but it requires specialized equipment (Cleveland et al. 1993).

8.3.1.3

Obtaining the Final Force–Distance Curve

Finally, using the standard SMFS mode (Sect. 8.3.3) the raw data recorded in a typical force curve correspond to the displacement of the positioner along the z direction, Δz_p . This raw curve should *sensu stricto* be called a force–extension or

←

Fig. 8.4. Mechanical unfolding of a modular protein. **a** Cartoon summarizing the interpretation of the sawtooth patterns observed in force–extension curves. The extension of the piezoelectric positioner (Δz_p) results in bending of the cantilever (Δz_c) and in an elongation of the trapped stretch of the protein molecule, which mechanically links the two elements. As the distance between the substrate and the tip of the cantilever (D) increases, the protein elongates and owing to the reduction of its entropy (each of their modules tends to maximize its disorder), it generates a restoring force that bends the cantilever (states 1–2). This entropic elasticity of proteins can be described by the wormlike chain (WLC) model of polymer elasticity, where the two adjustable parameters, p (persistence length) and ΔL_c (contour length), can be estimated after fitting the force recording to this model (*dotted blue lines*; see Sect. 8.3.3.1 for an explanation of these parameters). When the domain with the lowest mechanical stability unfolds (state 3), the contour length of the protein increases, relaxing the force on the cantilever to near zero. Repetition of this cycle will produce a sawtooth pattern. The last force peak of the recording typically reflects the detachment of the molecule from either the tip or the substrate. **b** Cartoon showing the hypothetical series of events behind a sawtooth pattern. In each cycle, the AFM tip is pressed against a monolayer of proteins and then withdrawn. Randomly, a single modular protein forms the only connection between the two surfaces in some recordings. The *colors* represent different conformations of the same modular protein (including denatured or aggregated), with yellow–red color range representing the native fold. Note that the proximal region of the recording (here spanned around 100 nm) is dominated by nonspecific interactions (modified from Carrion-Vazquez et al. 2000; Rief et al. 1997b; with permission, copyright 1997 American Association for the Advancement of Science)

force–displacement curve. However, we are interested in plotting the force vs. the real distance between the tip and the sample, D . These two distances differ because of the cantilever deflection and are related through

$$D = \Delta z_p - \Delta z_c, \quad (8.4)$$

again assuming no sample deformation. Therefore, by combining Eqs. 8.2 and 8.4, we can obtain the final plot of the force vs. real tip–sample distance (D), yielding what should be called more properly “force–distance” curves (Sect. 8.3.3.1). However, as these curves are typically referred to in the literature as force–extension, *sensu lato*, we will adopt this convention in this review to refer to force–distance curves.

8.3.2

Establishing a “Mechanical Circuit”: Protein Attachment and Functionalization

The analysis of protein intramolecular interactions by SMFS typically relies on the attachment of the protein by adsorption to two points: on the site of the “substrate” (typically a glass coverslip and often gold-coated) and on the tip side (Figs. 8.3, 8.4). An aqueous solution of the protein in an appropriate buffer (normally at a concentration above 1 μM) is deposited on a glass coverslip that is mounted on top of the piezoelectric positioner. After a few minutes protein molecules will have been randomly attached to the surface by physisorption (i.e., nonspecific adsorption via “physical forces”) forming a layer that is 20–50-nm thick. However, when adsorbed onto a surface, the protein molecules tend to denature to some extent (Buijs et al. 1996; Hlady and Buijs 1998). To minimize this problem, two terminal cysteine residues can be added to the end of the protein (usually the C-terminus) by genetic engineering for them to covalently attach to a gold-coated coverslip (Rief et al. 1997b). Still, although this “less-destructive” method offers the protein an alternative means of attaching specifically, it does not rule out the simultaneous occurrence of adsorption.

The use of physisorption to immobilize the system studied is perfectly adequate for protein unfolding studies, since during the whole experiment there is a linkage between the substrate and the force probe. However, this is not an ideal method to immobilize proteins for single-bond dissociation analysis, since it does not allow us to distinguish true bond failure from the desorption of biomolecules (from either the substrate or the force probe). For this reason different protocols of “functionalization” have been developed for the analysis of intermolecular interactions using SMFS (Sect. 8.5; reviewed in Zlatanova et al. 2000; Weisel et al. 2003).

8.3.3

Mechanical Unfolding of Proteins

Proteins can be unfolded by applying a mechanical force using AFM in its SMFS mode. This imposes a specific reaction coordinate on the unfolding process (i.e., the end-to-end distance). These experiments are typically performed under nonequilib-

rium conditions. Depending on the variable being “clamped” (i.e., maintained constant), there are two basic modes currently being used: length-clamp (which yields a force–extension curve) and force-clamp (which yields an extension–time curve).

8.3.3.1 Length-Clamp (Standard) SMFS

In a typical SMFS experiment, a solution containing the sample protein is deposited on top of a glass coverslip (often coated with a thin gold layer of less than 30 nm), which is mounted on the piezoelectric positioner. The tip of the cantilever is then brought into contact with the protein layer, adsorbing some protein molecules which establish a mechanical circuit between the cantilever tip and the positioner (Fig. 8.4). In the standard mode (i.e., length-clamp) the positioner then retracts at a constant speed (typically around 0.5 nm ms^{-1}) to a set length while the cantilever reports the force of resistance opposing the extension of whatever has been trapped between the cantilever tip and the substrate. Often, during this random pickup, a segment of a single protein molecule (up to its full length) remains as the only link between the two components, thereby creating a mechanical circuit in which the mechanical properties of a single molecule can be measured. When this happens, the first source of resistance to extension in a protein molecule comes typically from the entropic forces. This entropic elasticity is a general property of polymers. It results from the tendency of a chain to form a coil in order to maximize the conformational freedom (entropy) of its constituent monomers under the drive of thermal fluctuations. Stretching a polymeric molecule reduces its entropy, producing a restoring force that results in the bending of the cantilever. In contrast with the linear relationship that governs the bending of the cantilever (Hooke’s law, Eq. 8.1), the entropic elasticity of a protein follows a nonlinear relationship that can be formally described by the so-called worm-like chain (WLC) model of polymer elasticity (Fig. 8.4a; Bustamante et al. 1994; Marko and Siggia 1995). This model predicts the entropic restoring force, F , of the polypeptide at any given extension from two descriptive parameters: L_c , the contour length (total length of the unfolded polypeptide), and p , the persistence length (the smallest rigid component of the chain).

Further extension of the protein may cause its fold to unravel, typically in an all-or-none fashion, which can be described by a two-state model as we will see later. This exposes previously “force-hidden” amino acid residues to force, which results in an increase in the end-to-end length of the protein trapped between the tip and the positioner and relaxes the force acting on the cantilever to near zero. For a modular protein a further extension repeats this cycle for each of the remaining folded modules, giving rise to a characteristic sawtooth pattern in the force–extension curve (Fig. 8.4b). Note that during the process of stretching, once the first module unfolds, the resulting polymer is no longer homogeneous as the modular protein is now connected in series with a long denatured polypeptide chain, in which the monomeric unit is an amino acid residue. As can be expected, in the resulting “force-spectrum” curve the force peaks are sorted according to their strength, such as the “weakest” modules unfold first. In these experiments the last peak corresponds to the rupture of the mechanical circuit (typically the detachment of the protein either from the

tip or from the substrate), which indicates the end of the experiment. As a result of this process the contour length of the molecule increases with each unfolding event, reflecting, at each point, the length of the “force-bearing” amino acid regions. The distance between two consecutive peaks represents the “force-hidden” amino acid region of the protein, ΔL_c (Fig. 8.4a).

Polyprotein Strategy: Protein Repeats as Single-Molecule Fingerprints, Linkers, and Amplifiers

As we have seen (Sect. 8.3.2), proteins tend to denature to some extent when adsorbed onto a surface, and for this reason long modular proteins are very good systems for SMFS studies. Such proteins are made up of several modules and, thus, two of these can readily be partially or totally “sacrificed” for the required attachments of the molecule to the cantilever tip and to the substrate. Furthermore, the “proximal region” to the substrate of a force–extension curve is often the main source of non-specific interactions and spans typically the first 30–70 nm (Fig. 8.4b). Therefore, the forces involved in extending short proteins or a typical module (even if they were not denatured by the adsorption required for the attachments) may be masked by the “noise” caused by such nonspecific interactions. Thus, if we also consider that modular proteins tend to have a mechanical function, it is not surprising that the earliest SMFS experiments (and most of the SMFS studies published to date) focused on multidomain mechanical proteins (native or recombinant) such as titin, tenascin, spectrin, and fibronectin (Oberhauser et al. 1998, 2000, 2002; Rief et al. 1997b, 1999). However, modular proteins are composed of heterogeneous populations of modules (with different sizes, sequences, structures, and stabilities); therefore it is not possible to identify which force peaks correspond to which modules in the resulting sawtooth pattern.

Thus, although these studies provide a simple way of first approaching the study of the global mechanical properties of a modular protein, it seems clear that pure-periodic proteins (i.e., perfectly repetitive), the so-called polyproteins, would be an ideal choice system to allow the study of the mechanical properties of a single protein domain (or even a nonmodular protein). However, there are only a few naturally occurring polyproteins that are composed of identical repeats (at the amino acid level) without having any spacers between them, such as polyubiquitins and the yeast α -factor. Indeed, in most cases the repeats are not perfect or are linked by spacer sequences of varying lengths (Finley and Varshavsky 1985; Wiborg et al. 1985 and references therein). The advent of methods to synthesize homomeric polyproteins (i.e., head-to-tail covalently linked tandem repeats of a single domain or protein; referred to in this review as polyproteins, *sensu lato*) opened the way to the mechanical analysis of single modules and nonmodular proteins (Carrion-Vazquez et al. 1999a, 2000). Since then, several additional strategies have been established for the artificial synthesis of polyproteins for SMFS experiments. All the strategies developed so far fall into two basic categories:

1. Methods based on the construction of repeats at the DNA level through genetic engineering (Carrion-Vazquez et al. 1999a). In order to permit directional head-to-tail cloning, two strategies were originally used: one using a single nonpalindromic (asymmetrical) restriction endonuclease recognition site (*AvaI*), which

adds a Leu-Gly linker; the other uses palindromic (symmetrical) restriction sites and two restriction enzymes with compatible cohesive ends (*Bam*HI/*Bgl*II and *Kpn*I), which add an Arg-Ser linker. Sequencing is typically carried out at the monomer level using this method, before concatemerization, although the cloned concatemer could in principle be sequenced either directly by deletion sequencing or after its digestion to monomers (and processing an “oversampled” number of clones). A “cassette” strategy has recently been established based on the use of multiple restriction sites. Although this method generates a polyprotein with different amino acid linkers, the sequencing of the concatemer is easier and it permits the selective introduction of proteins or domains at specific positions (Steward et al. 2002).

2. A method based on the synthesis of repeats directly at the protein level, by disulfide-bonding cysteine residues (Yang et al. 2000). Using this method, directionality (i.e., head-to-tail links) in the assembling of monomers is only possible if the synthesis is performed in the solid state.

Polyproteins provide multiple advantages for SMFS studies (Carrion-Vazquez et al. 2000; Fig. 8.5). Their monotonous nature for the first time permitted single-module analysis and the experimental demonstration of the stochastic nature of the process of mechanical unfolding. These systems are also an ideal choice for force spectroscopy experiments because their periodical design provides a way to sort out nonspecific interactions and to unequivocally identify single-molecule events (i.e., single-molecule “fingerprinting”). Furthermore, some repetitions of the polyprotein can have “sacrificial” roles: they can be used for protein attachment (through partial or complete denaturation), they can act as polymer spacers that physically separate most of the repeats from the proximal region of the substrate, and they can also take up the pressure denaturation effects that may likely result from the high forces applied to attach the polyprotein to the cantilever tip (while in the nanonewton range, these pressures are estimated to range up to several gigapascals). The use of polyproteins also permits some spatial features to be amplified, which has enabled the detection of small amino acid insertions, conformational intermediates that are spatially very close to the native state, as well as aberrant folding events. In the case of unstructured proteins, which present featureless force–extension curves (e.g., PEVK or N2B titin domains, see later), the construction of a chimeric polyprotein, which includes a marker domain (e.g., I27 titin domain; see later) fused to the protein under study, has enabled single-molecule events to be unequivocally fingerprinted (Li et al. 2001, 2002). In the case of “structured proteins” the use of heteromeric polyproteins (instead of homomeric polyproteins) allows the inclusion of a convenient internal standard (e.g., I27 module), covalently built in the molecule, which may also permit the “solubilization” of proteins that are insoluble as homomeric polyproteins (as shown for the I27 module; Forman et al. 2005 and references therein). Lastly, because they are perfect repetitions of the system under study, polyproteins have the added bonus of significantly increasing the rate of data acquisition in SMFS experiments. An alternative strategy to the use of homomeric and heteromeric polyproteins in SMFS is “to sandwich” the protein under study between two reference

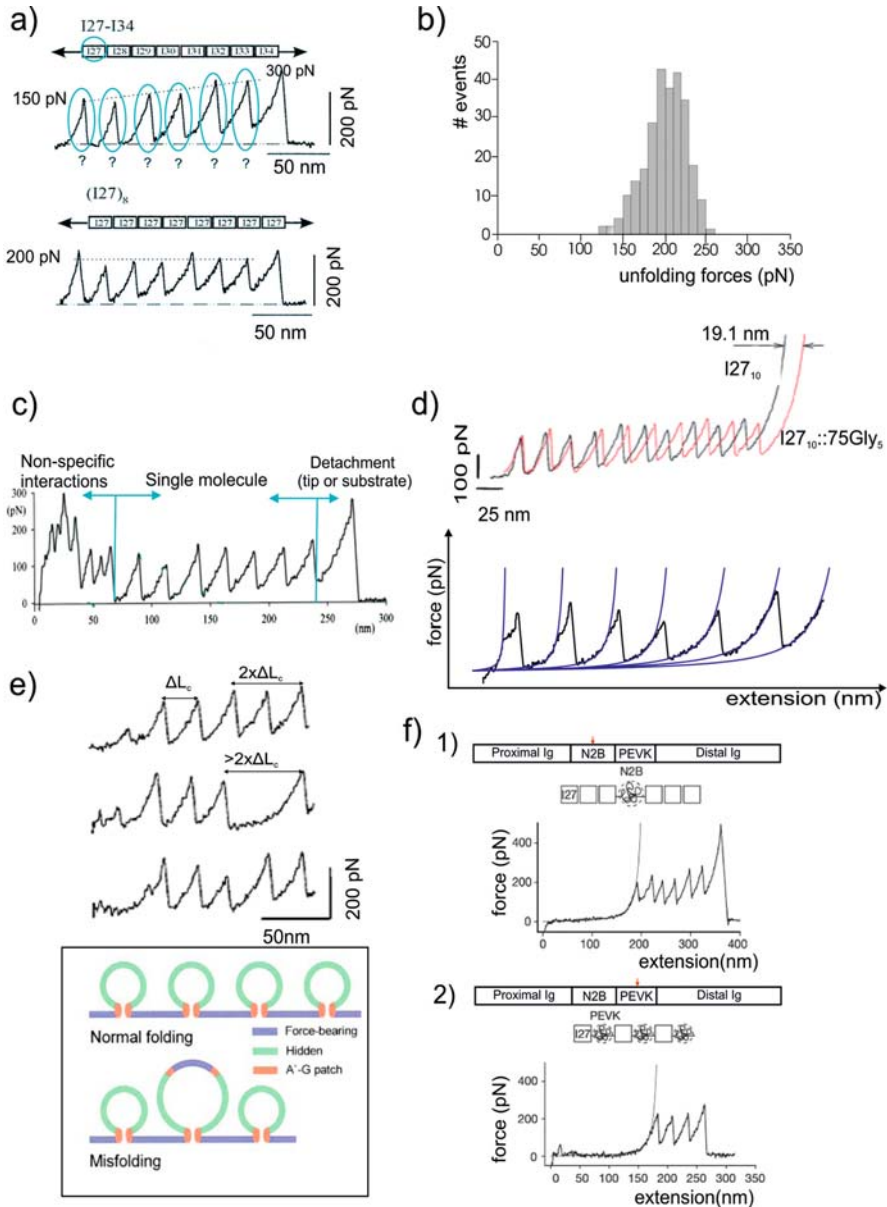


Fig. 8.5. The use of polyproteins in SMFS. **a** Mechanical analysis of single protein modules. The force–extension curve of a modular protein (recombinant I27-I34 Ig region of I band of titin) typically shows a steady rise in the force peaks, indicating the hierarchical unfolding of its modules. However, none of the peaks in the spectrum can be specifically attributed to any module. On the other hand, stretching an I27 polyprotein produces a sawtooth pattern in which the force fluctuates around an average value. This permits an individual module to be characterized (modified from Li et al. 2000b). All the sawtooth pattern recordings in the

polyproteins made out of either the same module or two different modules (see Table 8.1 for examples and references).

Nevertheless, in spite of all the advantages of using polyproteins in SMFS studies, the technique in its current form is based on a trial-and-error approach in which both the attachment points of the polyprotein molecules and the number of modules trapped in each trial are random.

←

remaining panels of the figure were obtained using I27 polyproteins. **b** Demonstration of the stochastic nature of mechanical unfolding. Forced unfolding experiments on polyproteins generate a distribution of forces rather than a specific force value, reflecting the dominant role of thermal forces at the single-molecule level (modified from Carrion-Vazquez et al. 1999b). **c** Polyproteins also act as spacer arms, increasing the length of the protein beyond the region of the nonspecific interactions between the cantilever tip and the polyprotein layer (which typically spans about 30–70 nm). The pseudoperiodicity of the sawtooth pattern identifies single domain unfolding (modified from Carrion-Vazquez et al. 2000). **d** Amplification of mechanical properties. *Top*: when compared with the wild type (*black trace*), a five glycine residue insertion in a mutant polyprotein (*red trace*) elongates the contour length of the molecule with each unfolding event (1.91 nm per module) such that it can be readily detected from the third or fourth force peak of the sawtooth pattern (after Carrion-Vazquez et al. 1999b). *Bottom*: an unfolding intermediate originating from the rupture of a patch of hydrogen bonds between the A and B β -strands of I27 (Fig. 8.9) is detected as a deviation (“hump”) of the WLC (*blue lines*). In the polyprotein analyzed in each experiment, the first peak reflects the rupture of the AB patch of all the modules trapped in this stretch (after Fisher et al. 2000). **e** Detection of misfolding events. During repeated unfolding–refolding cycles, some sawtooth patterns (less than 2%) displayed missing unfolding force peaks (“skips”), which were interpreted as originating from the rupture of “superfolds” formed during the refolding process. This interpretation was based on the observation that the increment in the contour length was longer than the sum of the lengths of two consecutive peaks in the unfolding pattern, and the difference accounted for the stretching of the linkers that separate two consecutive modules (from Fisher et al. 2000; Oberhauser et al. 1999; with permission, copyright 1999 Macmillan Publishers Ltd.). **f** Single-molecule fingerprinting with polyproteins. The mechanical properties of “unstructured” (i.e., without tertiary structure) proteins or domains can be done using polyproteins as an internal marker for single-molecule identification. 1 The “sandwich” method: Force–extension curve of a protein chimera containing the N2B region from human cardiac titin flanked on either side by three I27 domains (I27₃-N2B-I27₃). 2 “Alternating copolymer” method: force–extension curve of a protein chimera containing three PEVK domains from human cardiac titin, alternated with three Ig I27 domains (I27-PEVK)₃. In both curves, a nonlinear least-squares fit (i.e., a Levenberg–Marquardt curve fit) of the WLC equation (*thin line*) to the force–extension curve prior to the first I27 unfolding event was used to measure the contour length, L_c , and persistence length, p , both of PEVK as well as of N2B (after Li et al. 2002)

Information Extracted from SMFS Data Basic Parameters in SMFS

The Increase in Contour length (ΔL_c) Reflects the Number of Force-Hidden Amino Acid Residues in a Fold

The force–extension curves of a homomeric polyprotein analyzed by SMFS experiments in the length-clamp mode reveal a series of equally spaced force peaks as the tip is pulled away from the coverslip (substrate). Each peak typically corresponds to the unfolding of a single domain (if no intermediates are present in the process). The last peak typically corresponds to the rupture of either the tip–protein or the substrate–protein attachments, which relaxes the force to zero (Fig. 8.5c). As mentioned earlier, the rising part of each peak can be modeled by means of the WLC equation of polymer elasticity, while the stretching of a whole polyprotein can be modeled as a series of polymer stretching events with a constant stepwise increase of the polypeptide length between two consecutive unfolding events (increase in contour length, ΔL_c ; Fig. 8.4a). Thus, homomeric polyproteins typically result in equally spaced force peaks for which ΔL_c is the same between any two consecutive unfolding events. This parameter gives an indication of the mechanical structure of the domain, since it is related to the number of amino acid residues “trapped” between the bonds that form the major resistance points to mechanical unfolding, i.e., the length of the force-hidden region of the domain (Carrion-Vazquez et al. 1999b).

Unfolding Force (F_u), a Parameter for Mechanical Stability

Force is a vector quantity, so the mechanical stability of a protein is expected to depend not only on the magnitude but also on the direction and point of application of the force (Sect. 8.4.4). Furthermore, it has a time-dependence as discussed later).

Forced unfolding of a protein in SMFS is the convoluted result of two competing forces: thermal and mechanical. Therefore, if we disregard the effects of systematic and random errors (e.g., those from the calibration of the spring constant of the cantilever), the frequency histogram of the unfolding events for a polyprotein should reveal the actual contribution of thermal forces to each unfolding event, i.e., the distribution of thermal energy available to each module prior to unfolding (Fig. 8.5b). It is predicted that these histograms are skewed towards the lower forces (Evans and Ritchie 1997) as observed in large data sets (Carrion-Vazquez et al. 1999a). Furthermore, the shape of the distribution of unfolding forces at any particular speed carries information regarding the unfolding energy surface, from which the kinetic parameters of the process can be extracted (Carrion-Vazquez et al. 1999a; Evans and Ritchie 1999). However, because the experimental errors broaden the force histogram, it is difficult to deconvolve them from the inherent thermal spread. In practice, the method of choice to obtain the kinetic parameters of the process is based not only on the force histogram but also on the dependence of the unfolding force on the pulling speed, which determines the time window of the experiment.

Protein unfolding is a stochastic (probabilistic) process, i.e., each time an experiment is repeated, even in the same (macroscopic) conditions, a different result will be obtained. As previously mentioned, the mechanical stability of a protein fold can be defined as its resistance to unfolding in response to an applied mechanical force. However, because of the probabilistic nature of the process we cannot strictly

speaking refer to a specific “unfolding force” (i.e., a “critical force”) but rather to a “most probable” unfolding force, at a given pulling speed (see later). By changing the pulling speed we can vary the time exposure of the protein to the mechanical and thermal forces. Furthermore, any protein domain will eventually unfold under the slightest force, or even in the absence of force, if one waits long enough for the thermal forces to do all the work.

The Effect of a Mechanical Force on the Thermodynamics and Kinetics of the Protein Unfolding Reaction

Considerations from Thermodynamics and Statistical Mechanics: Equilibrium vs. Nonequilibrium Unfolding

Classical thermodynamics (equilibrium thermodynamics or “thermodynamics” for short) deals with quasistatic processes, which are idealized, “infinitely slow” processes that can be approximated in practice by performing them “very slowly.” Thus, thermodynamics is not concerned with the rate at which a process takes place (i.e., kinetics). Time-dependent thermodynamic processes are studied by nonequilibrium thermodynamics. In quasistatic processes the system often goes through a sequence of states that are infinitesimally close to equilibrium, in which case the process is typically reversible. Thus, technically, “quasistatic” and “reversible” are different terms, not synonyms as they are sometimes treated in the literature. Whereas reversible processes are almost always quasistatic (only two exceptions are known at present: superfluidity and superconductivity), the converse is not always true. Similarly, nonequilibrium processes are usually irreversible. In single-molecule mechanics, the overlapping of the force–extension curves for pulling and relaxation is normally taken, in a first approximation, as a hallmark of thermodynamic reversibility for the conditions of the experiment (although, strictly speaking, curve overlapping is only a necessary but not a sufficient condition). On the other hand, the existence of different curves, i.e., hysteresis, is a sensitive indicator of the irreversibility for the process. In contrast to thermodynamics, which deals with systems composed of large molecular populations from a macroscopic point of view (i.e., “large scale”), statistical mechanics approaches these systems from a microscopic perspective (meaning atomic scale in the physicist’s jargon), which in our case typically means single molecules.

Studies on the mechanical unfolding of RNA have shown that the breaking of the secondary structure elements of RNA occurs close to thermodynamic equilibrium (this also being a reversible process), within the millisecond to second timescales of the experiment (i.e., the molecule follows the same force–extension curve upon stretching as relaxation). In contrast, breaking of the more complex tertiary structure is often a nonequilibrium process (and is also irreversible on the timescale of the experiment), in which interactions equilibrate over a slower timescale and, therefore, the pulling–relaxation cycle presents hysteresis. Because the distance to the transition state is often shorter in tertiary interactions than in secondary structures, they tend to be brittle, breaking at high forces and after small deformations. In contrast, secondary interactions are compliant, breaking at low forces and after large deformations (Bustamante et al. 2004).

Unlike RNA, in which secondary and tertiary interactions are independent and additive, the unfolding of most proteins seems to be highly cooperative, with the stability of secondary structures depending on their tertiary context. At the typical pulling speeds used in SMFS experiments (approximately 0.5 nm ms^{-1}) there are only a few protein structures that show coincident stretching and relaxation curves (i.e., at thermodynamic equilibrium). These proteins are true elastic elements acting as reversible springs that store/restore elastic energy with 100% efficiency and no heat dissipation. They include elastomeric proteins (reviewed by Tatham and Shewry 2000) like elastin (Fig. 8.6a, panel 2; Urry et al. 2002), the PEVK and N2B domains of cardiac titin (Li et al. 2001, 2002), and the EH domain of myomesin (Schoenauer et al. 2005). The myosin II tail is also an elastic protein structure though under experimental pulling conditions the pulling/relaxation curves show some hysteresis because some energy is dissipated as heat, i.e., the pulling process is close to equilibrium (pale blue area in Fig. 8.6a, panel 3, see also Fig. 8.8, Table 8.1; Schwaiger et al. 2002).



mission, copyright 2002 The Royal Society, UK). 3 Force–extension curve of an elastic protein structure: myosin II tail (*insert*). This structure behaves like a truly elastic protein showing little hysteresis when relaxed, which reflects the small amount of energy dissipated between extension and retraction (after Zhuang and Rief 2003, with permission, copyright 2003 Elsevier Science). 4 Force–extension curve of an entropic–enthalpic modular protein structure showing a force spectrum of the Ig region of titin. The folding–unfolding of titin of the Ig region is a nonequilibrium process (note the marked hysteresis), making this protein a perfect shock absorber that dissipates as heat part of the energy put into the system by the mechanical work exerted upon stretching the molecule. *Insert*: structure of the titin I27 Ig module (from Zhuang and Rief 2003, with permission, copyright 2003 Elsevier Science). **b** Effect of the pulling speed on the unfolding force of the titin I27 module. The unfolding force rises with increasing pulling speeds (from Carrion-Vazquez et al. 1999a), indicating that mechanical unfolding is a kinetic process. This dependence allows the kinetic parameters of the process to be extracted. **c** 1 The stretching of a single Ig domain modeled as a two-state kinetics system, folded–unfolded. Under an applied force, an Ig domain unravels causing an increase in the end-to-end length, Δx (after Carrion-Vazquez et al. 1999a). 2 The effect of a mechanical force on the free-energy diagram of a protein that unfolds following a two-state model (f folded, u unfolded). The *dashed blue curve* represents the process in the absence of an applied force. An applied force (*black dashed line*) tilts the energy diagram of the process, decreasing the barrier to the transition state, $\ddagger [\Delta G^\ddagger(F) < \Delta G^{0\ddagger}]$ (*red line*), which increases the rate of the forward reaction. The application of force also lowers the energy of the unfolded state relative to that of the folded state $[\Delta G(F) < 0]$, which populates the unfolded state. The mechanical reaction coordinate is x . With an applied force, the positions of the free-energy minima (x_f and x_u) and the maximum (x^\ddagger) shift such that Δx_u becomes shorter ($\Delta x_u, red < \Delta x_u, blue$) and Δx_f longer ($\Delta x_f, red > \Delta x_f, blue$). Local curvature of the free-energy surface dictates these relative shifts in position. Note that Δx_u and Δx_f here refer to the unfolding distances to the transition state in a two-state kinetics model, while Δx in panel 1 refers to the increase in the end-to-end distance of the molecule (modified from Bustamante et al. 2004, with permission, copyright 2004 Annual Reviews)

Table 8.1. Mechanical properties of proteins and single modules obtained by single-molecule force spectroscopy (SMFS) compared with other physical and chemical parameters. We survey here only wild-type proteins and modules (mutant variants of some of them can be found in the literature). These are classified as nonmechanical (*regular font*) and mechanical (*bold font*). Polyubiquitins are considered here as putative mechanical proteins since their role as a mechanical handle for proteasomal traction has not been established yet

Protein	Function	Conformational (and mechanical) topologies at "breakpoint"	Single-molecule identification [system]	Unfolding force (pN)	Pulling speed (nm ms ⁻¹)	ΔL_c (nm)	p (nm) ^a
Dihydrofolate reductase (DHFR)	Conversion of dihydrofolate (DHF) to tetrahydrofolate (THF) in the presence of NADPH	Unknown breakpoint (mixed-type: 8-stranded β -sheet and several α -helices)	Homomeric polyprotein [(DHFR) ₆] Heteromeric polyprotein [(I27-DHFR) ₄]	27 (without MTX) 74–82 ^b (with MTX) 83 ^b (with DFH) 98 (with NADPH) 83–89 (with MTX-NADPH)	0.4	67 ΔL_c (N→U)	–
			[(ddFLN1-3)-DHFR-(ddFLN4-5)]	~60 (without ligand and also with NADPH, MTX, and MTX-NADPH)		22±3 ΔL_c (N→I)	–
T4 lysozyme	Peptidoglycan hydrolysis	Unknown breakpoint ("all- α " protein)	Homomeric polyprotein [(T4-lysozyme) _n] ^d	64±16 (44±13 in 1M GuHCl)	1	32	0.65±0.25
Barnase	Ribonuclease	5-stranded antiparallel β -sheet and 2 α -helices (disruption of hydrophobic core)	Heteromeric polyprotein [(I27-barnase) ₃ -(I27) ₂]	70	0.3	Peaks variable in position	0.35
C2A domain synaptotag min I	Calcium sensor for vesicle fusion	(AG) antiparallel β -sheet(zipper)	Homomeric polyprotein [(C2A) ₉]	60	0.6	38	0.4
Calmodulin	Calcium transducer involved in several processes	Unknown breakpoint (antiparallel α -strands and 2 antiparallel β -sheet hairpins)	– (homomeric polyprotein [(CaM) ₄] but no force peaks observed)	<20	0.6	–	0.32

Table 8.1. Continued

$k_u^0(F=0)$ (s^{-1})	Δx_u^\ddagger (nm)	Mechanical kinetic stability ($kcal\ mol^{-1}$)	$k_u^0([D]=0)$ (s^{-1})	Chemical thermodynamic stability ($kcal\ mol^{-1}$)	T_m ($^{\circ}C$)	Refolding	References
$k_u^0(-MTX)$ $=25k_u^0(+MTX)$	0.25 (taken from I27)	-	-	Without ligands 2.4 ^c , with folate and NADP ⁺ 3.5 ^c	~53 (<i>E. coli</i>)	-	Appleman et al. (1988); Tsay et al. (1990); Wallace and Matthews (2002); Ainavarapu et al. (2005)
-	-	-	-	-	-	-	Junker et al. (2005)
-	0.81 ^e	-	10 ⁻⁴	18.2 (at 12 $^{\circ}C$ and pH 5.7)	Monomeric 49.5 Polymeric 40 (apparent)	Occurs within 1 s of relaxation with an efficiency close to 100%	Chen et al. (1992); Yang et al. (2000)
6.7 $\times 10^{-2}$	0.35	-	3.37 $\pm 0.69 \times 10^{-5}$ (isolated barnase) 2.34 $\pm 0.74 \times 10^{-5}$ (I27-barnase-I27)	10.2 ± 0.4 (isolated barnase) 9.8 ± 1.2 (I27-barnase-I27)	-	-	Best et al. (2001); Best and Clarke (2002)
-	-	-	-	-	55 74 (+Ca ²⁺)	-	Shao et al. (1996); Lu and Schulten (1999); Carrion-Vazquez et al. (2000)
-	-	-	-	-	55 >90 (+Ca ²⁺)	-	Brzeska et al. (1983); Carrion-Vazquez et al. (2000)

Table 8.1. Continued

Protein	Function	Conformational (and mechanical) topologies at "breakpoint"	Single-molecule identification [system]	Unfolding force (pN)	Pulling speed (nm ms ⁻¹)	ΔL_c (nm)	p (nm) ^a
GFP	Conversion of blue chemiluminescence of aequorin into green fluorescent light (by energy transfer)	Unknown breakpoint (β -barrel)	"Sandwich fingerprinting" with modular proteins[(I27-I30)-GFP-(I31-I34)][[ddFLN1-3)-GFP-(ddFLN4-5)]	104±40 (intermediate GFPD α at 35)	0.3	76.6±0.3 3.2±1 (GFPD α) 6.8±0.6 (GFPD α D β)	0.5
B1 domain protein L	Cell-wall protein of <i>Peptostreptococcus magnus</i> involved in evasion from immune system	(I–V) parallel β -sheet (shear)	Homomeric polyprotein [(protein L) ₅] carrying a Y47W mutation in each module	136±1	0.4	18.6±0.1	0.4
E2Lip3 N-C	Lipoyl domain of dihydrolipoyl acetyl transferase of PDH	Antiparallel β -sheet (zipper)	Heteromeric polyprotein [(I27) ₄ -E2lip3] and "sandwich fingerprinting" with homomeric polyprotein [(I27) ₂ -E2lip3-(I27) ₂]	<20	0.7	–	–
N-K41		Antiparallel β -sheet (?) (shear) (?)	Heteromeric polyprotein [(I27) ₄ -E2lip3]	177±3	0.7	10.9±0.2	0.4
Polyubiquitins N-C	Precursor of ubiquitin	(I–V) parallel β -sheet (shear)	Natural homomeric polyprotein [(N-C) ₉]	203±35	0.4 (0.25–0.41)	24±5 (20.3±0.9 measured at 100–200 pN in F-clamp) ^f	0.4
K48-C	Signaling for proteasomal degradation	(III–V) antiparallel β -sheet (shear)	Natural homomeric polyprotein [(K48-C) ₂₋₇]	85±20	0.3 (0.28–0.31)	7.8±2.8	0.4
α -Elastin [synthetic penta-peptides (GVGVP) ₂₅₁ and (GVGIP) ₂₆₀]	Source of elasticity in vertebrate tissues (i.e., arteries, vocal cords, skin and certain cartilages)	Undetected breakpoint (β -spiral)	–	–	?	–	0.4 at lower forces and 0.6 at higher forces for (GVGVP) ₂₅₁ 0.7 with 20% error for (GVGIP) ₂₆₀

Table 8.1. Continued

$k_{off}^0(F=0)$ (s^{-1})	Δx_{off}^\ddagger (nm)	Mechanical kinetic stability (kcal mol $^{-1}$)	$k_{off}^0([D]=0)$ (s^{-1})	Chemical thermodynamic stability (kcal mol $^{-1}$)	T_m ($^{\circ}C$)	Refolding	References
-	0.28±0.03 (from FPDα) 0.55±0.1 (from GFPD αΔβ)	<22 (native GFP) 14 (GFPΔα) 12–15 (GFPΔαΔβ)	-	-	-	Refolding time of α-helix (in GFPΔα) less than 1 ms	Dietz and Rief (2004)
5±3×10 $^{-2}$	0.22±0.02	-	2±0.2×10 $^{-2}$ ~4×10 $^{-4}$	4.6±0.2	-	-	Scalley et al. (1997) Brockwell et al. (2005)
-	-	-	-	-	-	-	Brockwell et al. (2003)
7.6×10 $^{-3}$	0.29	-	-	-	-	-	
4×10 $^{-4}$ or 0.015–0.0375 (F-clamp)	0.25 or 0.14–0.17 (F-clamp)	-	4×10 $^{-4}$	6.7	-	A continuum of intermediates during mechanical refolding is reported. Duration of refolding is strongly force dependent	Khorasanizadeh et al. (1993); Carrion-Vazquez et al. (2003); Fernandez and Li (2004); Schlierf et al. (2004)
4×10 $^{-4}$	0.63	-	4×10 $^{-4}$	6.7	-		
-	-	-	-	-	-	-	Tatham and Shewry (2000); Urry et al. (2002)

Table 8.1. Continued

Protein	Function	Conformational (and mechanical) topologies at "breakpoint"	Single-molecule identification [system]	Unfolding force (pN)	Pulling speed (nm ms ⁻¹)	ΔL_c (nm)	p (nm) ^a
PEVK titin	Entropic spring in titin (I band)	Undetected breakpoint (PP11-like helix \leftrightarrow unfolded)	Heteromeric polyproteins [(I27-PEVK) ₃]	–	0.4–0.6	–	0.3–2.3 (average 0.91)
			[(exon 120,161,184 PEVK-I27) ₃]	–	0.4	–	0.89±0.42 (exon 120) 0.92±0.38 (exon 161) 0.98±0.4 (exon 184)
N2B cardiac titin	Entropic spring in cardiac titin (I band)	Undetected breakpoint	"Sandwich fingerprinting" with homomeric polyprotein [(I27) ₃ -N2B-(I27) ₃]	–	0.4–0.6	–	0.66
Myosin	Muscle contraction motor	Unknown breakpoint (2 parallel and left-handed α -helical coiled coils)	Heterogeneous modularity (rabbit skeletal myosin, myosin II from <i>Dictyostelium discoideum</i> and myosin II tail from <i>D. discoideum</i>)	20–25 (31±6 in 0.6M acetate)	0.04–0.13	–	25±10 (p1) 0.4±0.1 (p2)
Titin^h	Source of passive elasticity in muscle (I band)	Several domains: see I1, I27, I28, and fnIII (I band/A band)	Heterogeneous modularity: native + recombinant	Hierarchical 150–300	0.5–1	25–30	0.4
			Proximal Ig region [(I4-I11)] (two-state)	150–200 (weak hierarchy)	–	–	0.66
			Distal Ig region [(I27-I34)] (intermediate, three-state)	150–330 (strong hierarchy) (–)	–	~25 (–)	0.66 (–)
			[(I60-I65)] ⁱ	~185	0.5	31.9	0.56±0.02
			[(I91-I98)] ⁱ	~225	0.5	28.7	0.58±0.02
SIs-kettin	Muscle elasticity in insects' flight muscles	Ig domains	Native fragment	Hierarchical 120–250	0.5	28.5	0.4
Projectin	Muscle elasticity in insects' flight muscles	Ig and fnIII domains	Heterogeneous modularity: native + recombinant (2Ig and 3fnIII)	Hierarchical: 150–250 (ascribed to Igs); 50–150 (ascribed to fnIIIs)	0.5	28.3	0.38

Table 8.1. Continued

$k_{\text{off}}^0(F=0)$ (s^{-1})	$\Delta x_{\text{off}}^{\ddagger}$ (nm)	Mechanical kinetic stability (kcal mol^{-1})	$k_{\text{off}}^0([D]=0)$ (s^{-1})	Chemical thermodynamic stability (kcal mol^{-1})	T_m ($^{\circ}\text{C}$)	Refolding	References
-	-	-	-	-	-	-	Li et al. (2001); Linke et al. (2002); Sarkar et al. (2005)
-	-	-	-	-	-	-	
-	-	-	-	-	-	-	Li et al. (2002)
-	-	2.3 ± 0.7^9	-	-	-	Reversible transition (true elastic protein with a small hysteresis) on a timescale of less than 1 s	Schwaiger et al. (2002)
3×10^{-5}	0.3 ± 0.1	-	-	-	-	The folding rate for distal domains is higher than the folding rate for proximal domains (1.2 s^{-1} vs. 0.33 s^{-1}) (intermediate of distal Ig unfolding domains 10^2 s^{-1})	Labeit and Kolmerer (1995); Rief et al. (1997b, 1998b); Freiburg et al. (2000); Li et al. (2002); Watanabe et al. (2002)
3.3×10^{-3}	0.25	-	-	-	-		
8×10^{-5} (1.0×10^{-2})	0.25 (0.33)	- (-)	- (-)	- (-)	- (-)		
1.0×10^{-5} 6.1×10^{-5}	0.35 0.25	- -	- -	- -	- -		
8×10^{-3}	0.17	-	-	-	-	-	Bullard et al. (2006)
2×10^{-3} (ascribed to Igs) 1.7×10^{-2} (ascribed to fnIIIs)	0.15 (ascribed to Igs) 0.1 (ascribed to fnIIIs)	-	-	-	-	2 rates for native protein: 15 and 0.2 s^{-1}	Bullard et al. (2006)

Table 8.1. *Continued*

Protein	Function	Conformational (and mechanical) topologies at "breakpoint"	Single-molecule identification [system]	Unfolding force (pN)	Pulling speed (nm ms ⁻¹)	ΔL_c (nm)	p (nm) ^a
I1 titin	Structural element in titin	2 breakpoints: (AB) antiparallel β -sheet (shear); (A'G) parallel β -sheet (shear)	Heteromeric polyprotein [(I27-I1) ₄]	127±18	0.4–0.6	30.0±0.62 A small percentage of peaks of 21.7±1.28 (disulfide-bonded modules)	?
I27 titin	Structural element in titin	2 breakpoints: (A'G) parallel β -sheet (shear) (AB) antiparallel β -sheet (shear)	Homomeric polyprotein [(I27) ₈], [(I27) ₁₂] Heteromeric polyprotein [(I27-I28) ₄]	204±26 (Intermediate at 108±19)	0.6 (0.3–0.5)	28.1±0.17 (22 at ~150 pN in F-clamp)	0.39±0.07
I28 titin	Structural element in titin	2 breakpoints: (A'G) parallel β -sheet(?) ^j (shear) (?) ^j (AB) antiparallel β -sheet(?) ^j (shear) (?) ^j	Homomeric polyprotein [(I28) ₈] Heteromeric polyprotein [(I27-I28) ₄]	257±27 (in [(I28) ₈]); 300 (in [(I27-I28) ₄]) (Intermediate at 151±16)	0.6 (0.3–0.5)	?	–
TifnIII titin	Structural elements in titin	2 breakpoints: (AB) antiparallel β -sheet (?) ^j (shear) (?) ^j ; (AG) parallel β -sheet (?) ^j (shear) (?) ^j	Heterogeneous modularity: [ConFn (I band titin)] [AFn (A band titin)]	200 180	0.5 0.5	31 31.5	0.4 0.4

Table 8.1. Continued

$k_{off}^0(F=0)$ (s^{-1})	Δx_{off}^\ddagger (nm)	Mechanical kinetic stability (kcal mol $^{-1}$)	$k_{off}^0([D]=0)$ (s^{-1})	Chemical thermodynamic stability (kcal mol $^{-1}$)	T_m ($^{\circ}C$)	Refolding	References
5.0×10^{-3}	0.35	–	–	–	–	–	Gao et al. (2002); Li and Fernandez (2003)
1×10^{-6} (from native I27) 3.3×10^{-4} (in [(I27)I2]) (from intermediate) or $4. \times 10^{-4}$ F-clamp in [(I27) ₁₂])	0.59 (from native I27) 0.25 (from intermediate)	22.2	$4.91 \pm 0.53 \times 10^{-4}$ (isolated I27) $6.0 \pm 3.0 \times 10^{-4}$ (in 27-I28 dimer) $3.67 \pm 0.96 \times 10^{-4}$ (I27-barnase-I27)	7.6 ± 0.1 (isolated I27) 8.1 ± 0.1 (in I27-I28 dimer) 8.6 ± 0.5 (I27-barnase-I27) 7.2 ± 0.3 (in [(I27) ₃])	72.6	Refolding rate following force-induced extension ($1.2 s^{-1}$) and following chemical denaturation ($32 s^{-1}$)	Politou et al. (1995); Lu et al. (1998); Carrion-Vazquez et al. (1999a); Marszalek et al. (1999b); Klimov and Thirumalai (2000); Li et al. (2000b); Paci and Karplus (2000); Oberhauser et al. (2001); Best et al. (2001); Williams et al. (2003)
2.8×10^{-5} (in [(I28)8]) 2.5×10^{-6} (in [(I27-I28) ₄]) or 1.9×10^{-6} (F-clamp in [(I28) ₈])	0.25	23.7	$1.6 \pm 0.3 \times 10^{-5}$ (isolated I28) $4.0 \pm 0.9 \times 10^{-6}$; (in I27-I28 dimer) $2.1 \pm 0.4 \times 10^{-5}$ (isolated I28e) ^k	3.0 ± 0.1 (isolated I28) 3.5 ± 0.1 (in I27-I28 dimer) 3.2 ± 0.2 (isolated I28e) ^k	54.1	Refolding rate following force-induced extension ($0.0025 s^{-1}$) and following chemical denaturation ($30 s^{-1}$)	Politou et al. (1995); Lu et al. (1998); Marszalek et al. (1999b); Li et al. (2000b); Oberhauser et al. (2001); Scott et al. (2002)
–	0.3 ± 0.1	–	–	–	–	–	Rief et al. (1998b)
–	0.3 ± 0.1	–	–	–	–	–	

Table 8.1. Continued

Protein	Function	Conformational (and mechanical) topologies at "breakpoint"	Single-molecule identification [system]	Unfolding force (pN)	Pulling speed (nm ms ⁻¹)	ΔL_c (nm)	p (nm) ^a
FNfnIII fibronectin	Cell–substrate adhesion protein involved in tissue elasticity, cell migration, matrix assembly and other several process	1,7,8,9,10,12, 13,14 FNfnIII. 2 breakpoints: (AB) antiparallel β -sheet (shear) ^l (AG) parallel β -sheet (shear)	Heterogeneous modularity: native fibronectin Hetero/homomeric polyproteins: [(I27- ¹ FNfnIII) ₄] (¹ FNfnIII ² FNfnIII) ₆ [(¹⁰ FNfnIII) ₈] (¹² FNfnIII ¹³ FNfnIII) ₅ [(I27- ¹³ FNfnIII) ₈]	Hierarchical 80–200	0.6	28.5	–
			220 (¹ FNfnIII)	0.6	–	–	
			220 (² FNfnIII)	0.6	–	–	
			74±20 (¹⁰ FNfnIII)	0.6	–	–	
			124±18 (¹² FNfnIII)	0.6	–	–	
			81.7±14 (¹³ FNfnIII)	0.6	–	–	
90±20 (N→I)	0.4	12±2	0.4				
50±20 (I→U)	0.4	19±2	0.4				
100±20 (N→U apparent)	0.4	31±2	0.4				
TNfnIII tenascin	Cell–substrate adhesion protein involved in tissue development, organogenesis, and other several process	Two breakpoints: (AB) antiparallel β -sheet (?) ^l (shear) (?) ^l ; (AG) parallel β -sheet (?) ^l (shear) (?) ^l	Heterogeneous modularity: native human tenascin-C	137±12	0.2–0.6	24.8±2.3	0.42±0.22
			[TNfnALL]	138±50	0.3–0.5	24.7±1.2	0.42±0.22
			[TNfnA-D]	138±50	0.3–0.5	24.7±1.2	–
			[TenFn]	113	0.5	28.6	0.4
³TNfnIII tenascin	Structural element in tenascin	2 breakpoints: (AB) antiparallel β -sheet (shear) ^l ; (AG) parallel β -sheet (shear)	Homomeric polyprotein (³ TNfnIII) ₈ ^k	~100 (MD, simulations postulate a putative intermediate not detected by AFM)	1	–	0.4
Myomesin	Putative component of the M-bridges of the thick filament lattice of sarcomere	Ig domains, fnIII domains, and unstructured EH-segment	Homomeric polyprotein: (My6) ₈ (My10) ₉	190±41	1	26.3±1.3	–
				203±50	1	27.9±1.5	–
			Heteromeric polyprotein: (My6-EH) ₄	–	0.5	–	~0.3

Table 8.1. Continued

$k_{off}^0(F=0)$ (s^{-1})	Δx_u^\ddagger (nm)	Mechanical kinetic stability (kcal mol $^{-1}$)	$k_{off}^0([D]=0)$ (s^{-1})	Chemical thermodynamic stability (kcal mol $^{-1}$)	T_m ($^{\circ}C$)	Refolding	References
–	–	–	–	–	–	FNfnIII modules refold with similar rate constants (refolding rate at zero force for $^{10}FNfnIII$ is $0.9 s^{-1}$ and $0.8 s^{-1}$ for $^{12,13}FNfnIII$) in contrast to solution experiments ($^{10}FNfnIII$ refolds much faster than 9FNfnIII)	Litvinovich and Ingham (1995); Plaxco et al. (1997); Kramer et al. (1999); Lu and Schulten (1999); Paci and Karplus (1999); Cota and Clarke (2000); Craig et al. (2001, 2004); Gao et al. (2002, 2003); Oberhauser et al. (2002); Li et al. (2005)
4.0×10^{-3} (1FNfnIII)	0.17 (1FNfnIII)	20.5 (1FNfnIII)	–	–	121 (3FNfnIII)		
4.0×10^{-3} (2FNfnIII)	0.17 (2FNfnIII)	20.5 (2FNfnIII)	–	–	111 (6FNfnIII)		
2.0×10^{-2} ($^{10}FNfnIII$) ^m	0.38 ($^{10}FNfnIII$) ^m	22.2 ($^{10}FNfnIII$)	2.3×10^{-4} ($^{10}FNfnIII$)	–	102 ($^{10}FNfnIII$)		
–	–	–	–	–	95 (7FNfnIII)		
$^{12}FNfnIII$	$^{12}FNfnIII$	$^{12}FNfnIII$	–	–	78 (1FNfnIII)		
2.2×10^{-2} ($^{13}FNfnIII$)	0.34 ($^{13}FNfnIII$)	19.5 ($^{13}FNfnIII$)	–	–	57 ($^{2,4,5,8,9}FNfnIII$)		
0.02	0.438	–	–	–	48 (1FNfnIII)		
0.5	0.45	–	–	–	–		
–	–	–	–	–	–		
–	0.3	–	–	–	–	30% of the recovery occurs at a fast rate ($42 s^{-1}$), whereas 57% occurs at a slower rate ($0.5 s^{-1}$)	Clarke et al. (1997); Oberhauser et al. (1998); Rief et al. (1998b); Paci and Karplus (2000); Craig et al. (2004)
–	0.3	–	–	–	–		
–	0.3	–	–	–	–		
–	0.55 ± 0.1	–	–	–	–		
1.1×10^{-4}	0.48	–	$4.6 \pm 1.1 \times 10^{-4}$ (isolated 3TNfnIII)	~ 2.8 (isolated 3TNfnIII , pH 7)	~ 44 (isolated 3TNfnIII , pH 7)	The rate constant for 3TNfnIII (from bulk solution methods) is $3 s^{-1}$	Ng et al. (2005)
–	–	–	–	–	–	Refolding rates, My6 similar to fn domains of fibronectin; My10 similar to lg domains of titin	Schoenauer et al. (2005)
–	–	–	–	–	–		
–	–	–	–	–	–		

Table 8.1. Continued

Protein	Function	Conformational (and mechanical) topologies at "breakpoint"	Single-molecule identification [system]	Unfolding force (pN)	Pulling speed (nm ms ⁻¹)	ΔL_c (nm)	p (nm) ^a
Spectrin ⁿ	Actin-binding cytoskeletal protein involved in cortical membrane mechanics	3 antiparallel and left-handed α -helical coiled coils (disruption of tertiary arrangement)	Heterogeneous modularity: Native spectrin	25–35	0.3	31.7±0.3	0.8 (at <50 pN) 0.4 (at 50–300 pN)
			α -Spectrin ₁₃₋₁₈	2 events at 55	0.3	2 events at 15 and 26	–
			Homomeric polyprotein [(R16) ₄]	2 events at 60 and 80	3	2 events at 15.5±3.5 and 31±7	0.59±0.22
Filamin from <i>Dictyostelium discoideum</i> A from human (native)	Actin-binding cytoskeletal protein involved in cellular mechanical responses	2 breakpoints: parallel β -sheet (?) ^j ; antiparallel β -sheet (?) ^j (shear) (?) ^j	Heterogeneous modularity: native filamin <i>D. d.</i>	60–90	0.25–0.35	31.5±0.1	0.9 (at <30 pN) 0.5 (at >30 pN)
			"Sandwich fingerprinting" with modular proteins [(I27-I30) ⁻⁴ ddFLN-(I31-I34)] [(I27-I30) ⁻² ddFLN-(I31-I34)]	(62.9±3 for ⁴ ddFLN N→I and 52.5±3 for ⁴ ddFLN I→U)	0.25–0.35	(⁴ ddFLN presents an intermediate at 14–15)	0.9 (at <30 pN) 0.5 (at >30 pN)
			Heterogeneous modularity: native filamin A	Hierarchical 50–220 (histogram with peaks around 80, 130, and 180)	0.37	30±5	0.33
Polycystin-1	Receptor linking the extracellular matrix to actin cytoskeleton (via focal adhesions)	PKD domains	Recombinant hetero-modularity: AF20-Fc	~50–250	0.5–0.7	27.8±3.3	0.4
			QIF-4 (PKD ₁₃₋₁₆)	~50–250 (mean, 152±59)	0.5–0.7	–	–
			PKD ₆₋₁₀	102±50	0.5–0.7	–	–
	Putative mechanosensor	"Sandwich fingerprinting" with homomeric polyprotein [(I27) ₃ -PKD _{2,3,4} -(I27) ₃]	~180 (I27 similar to PKDd)	1	–	–	

Table 8.1. *Continued*

Protein	Function	Conformational (and mechanical) topologies at "breakpoint"	Single-molecule identification [system]	Unfolding force (pN)	Pulling speed (nm ms ⁻¹)	ΔL_c (nm)	p (nm) ^a
PKDd1 polycystin-1	Structural element in polycystin-1	Non-native A-B loop-G interaction (not experimentally tested)	Heteromeric polyprotein: [(I27-PKDd1) ₃ -I27]	~200	1	~28	–
			(I27-PKDd1) ₃	~160	0.5–0.7	–	–
			Heterogeneous modularity: recombinant QIF-2 (containing PKDd1)	157±44	0.5–0.7	~28	–

[D] concentration of denaturing agent

^aEstimated for the unfolded state

^bNote that the dissociation constant for human DHFR-DHF ($K_d \sim 580$ nM, Tsay et al. 1990) and DHFR-MTX ($K_d < 10$ nM, Appleman et al. 1988) does not correlate with the mechanical unfolding force (around 80 pN in both cases; substrate DHF binds to DHFR at the same site as MTX).

^cValues obtained from human DHFR, in urea denaturing conditions

^dPolyprotein obtained using chemical linkage of protein monomers. Here, replacing residues at the intermolecular contact points with cysteine residues (T21C and K124C) generates a polyprotein with a linking geometry different to the typical N–C recombinant polyprotein, which offers an alternative pulling direction (Yang et al. 2000).

^eThis value was calculated indirectly (Yang et al. 2000).

^fIn force-clamp experiments, Schlierf et al. (2004) observed that in 5% of the cases the protein unfolds through one or more intermediate states.

^gIn kilocalories per mol nanometer

^hModules I4, I5 (proximal I band), I32, and I34 (distal I band) have also been studied by SMFS, but they are not included here owing to the limited amount of data available from them. The average force values found for these modules were 171, 155, 298 and 281 pN (0.4–0.6 nm ms⁻¹), respectively (Li et al. 2002).

Table 8.1. *Continued*

$k_u^0(F=0)$ (s^{-1})	Δx_u^\ddagger (nm)	Mechanical kinetic stability (kcal mol $^{-1}$)	$k_u^0([D]=0)$ (s^{-1})	Chemical thermodynamic stability (kcal mol $^{-1}$)	T_m ($^{\circ}C$)	Refolding	References
-	-	-	-	-	-	-	-
2.3×10^{-4}	0.27	-	5.6×10^{-3} (isolated PKDd1)	1-2 (isolated PKDd1)	-	-	Forman et al. (2005); Qian et al. (2005)
-	-	-	-	-	-	-	-

ⁱTitin domains are named following the original domain numbering (Labeit and Kolmerer 1995). Here, I54-59 and I27-34 correspond to I65-70 and I91-98, respectively in the nomenclature from Freiburg et al. (2000).

^jThese are putative conformational and mechanical topologies at the breakpoint since no MD data are available.

^kThese constructs include a two amino acid extension at the N- and C-termini for I28 and a two amino acid extension at the C-terminus in the case of ³TNfnIII.

^lMD simulations of forced unfolding show that the breaking of the AB patch is followed by partial solvation and alignment of the structure.

^mLi et al. (2005) assumed that these values correspond to the intermediate unfolding reaction ($N \rightarrow I$ instead of $N \rightarrow U$).

ⁿFor α -actinin (Rief et al. 1999) and dystrophin (Bhasin et al. 2005), members of the spectrin superfamily, the values of F_u and ΔL_c were found to be similar.

^oData from the PKDd1 domain could not easily be distinguished from those of the I27 used as an internal marker in the heteromeric polyprotein, although some recordings suggest that it has a slightly lower stability than the I27.

However, most single-molecule unfolding experiments in proteins are done under nonequilibrium conditions, where the molecule is pulled or relaxed at a faster rate than its spontaneous rate of equilibration (i.e., its molecular relaxation time). As a consequence, a marked hysteresis is apparent and at the same time the transition forces become speed-dependent. The design of modular proteins such as the elastic immunoglobulin (Ig) regions of titin (the molecular spring of the sarcomere) allows these molecules to act as perfect shock absorbers that dissipate most of the elastic energy as heat (pale blue area in Fig. 8.6a, panel 4; Rief et al. 1997b). Although the unfolding process is not quasistatic in these conditions (and is also irreversible), it may still be perfectly relevant to processes *in vivo* where the pulling speeds are considerably high (Fig. 8.1d; Howard 2001).

At equilibrium it is possible to directly determine, from a single molecule, the change in Gibbs free energy (ΔG , which as a state function depends only on the initial and final stages of the process and can be obtained by integrating a reversible force–extension curve), the equilibrium constant, the reaction kinetics, and their dependence on force (Bustamante et al. 2004). However, if the mechanical unfolding of a molecule occurs under nonequilibrium conditions (i.e., the extension and relaxation curves do not overlap), the efficiency of the process is less than 100%. This indicates that not all the mechanical work carried out during the unfolding reaction is converted into a change in the free energy of the molecule. It has recently been demonstrated that it is possible to recover the free energy of unfolding RNA molecules not only from near-equilibrium conditions (Liphardt et al. 2002) but also in far-from-equilibrium conditions (Collin et al. 2005). The latter method is a promising tool to extract the equilibrium free energy of protein unfolding, as most protein unfolding reactions occur far from equilibrium.

Dependence of the Force (F_u) on the Loading Rate (r)

The force at which an unfolding event occurs depends on the loading rate used during extension of the protein:

$$r = dF/dt = kv, \quad (8.5)$$

where v is the pulling speed. As we have seen, different pulling speeds simply reflect different time windows during which the mechanical force is acting (i.e., time exposure). Thus, in forced unfolding experiments a protein domain unfolds driven by two competing forces: thermal and mechanical. Since the former has less time to act at high pulling speeds, the unfolding force and hence the mechanical stability depend on the rate of extension. The faster the pulling speed, the higher the force required to cause the unfolding of the domain. The speed-dependence of a typical protein system can be seen in Fig. 8.6b (i.e., titin I27 domain; Sect. 8.4.1).

The range of pulling speeds of a typical AFM is limited to 3 orders of magnitude, between approximately 0.01 and 10 nm ms⁻¹ (Fig. 8.1d). The lower limit of the pulling speed is due to the large mechanical loop of a standard AFM apparatus, which makes thermal drift the limiting factor. At very high pulling speeds the viscous drag of the solvent and the cantilever response are the factors that introduce errors. Nevertheless, as we will see in Sect. 8.6.1 the use of a short, low-noise cantilever can reduce this effect (Viani et al. 1999).

In cases where the physiological range of speeds are lower than the AFM ones (Fig. 8.1d), can we extrapolate the behavior of the system to the lower range? Put in other words, is this dependence linear? Recent data from both the direct measurement of intermolecular interactions using the biomembrane force probe (Sect. 8.6.2; Merkel et al. 1999) and the analysis of protein mutations on a titin Ig domain using the AFM (Sect. 8.4.1; Williams et al. 2003) suggest that this may not be the case.

Kinetic Parameters of Mechanical Unfolding: Spontaneous Rate Constant (k_u^0) and Unfolding Distance (Δx_u)

While the kinetics of a reaction is pathway-dependent, thermodynamic properties such as the free energy of unfolding only depend on the initial and final stages of the process. Thus, the question arises as whether the “mechanical stability” of a protein is a thermodynamic or a kinetic property. If mechanical stability were a property of the equilibrium it should depend only on the state variable that describes the system, i.e., the end-to-end extension of the molecule. The dependence of the force on the pulling speed reveals that mechanical unfolding is a kinetic process and it holds the key to extracting kinetic information from these experiments.

What then is the effect of a mechanical force on the kinetics of the protein unfolding reaction? The effect of force on the lifetime of a bond was first addressed for intermolecular bonds by Bell (1978); Evans and Ritchie then provided a more detailed interpretation in 1997. This formulation is directly applicable to the case of intramolecular bonds since the forces stabilizing both types of interactions are the same (Sect. 8.2.2). Still, there is an important difference between both processes as forced unfolding of a protein can be reversed and, therefore, refolding cannot be neglected as it is rebinding, in the case of force unbinding.

From basic classical mechanics, these authors developed a phenomenological procedure to obtain intrinsic rate constants from pulling experiments. Based on conventional transition state theory, the mechanical unfolding reaction is described by a simple two-state kinetics model in which the protein adopts only a native (folded) or a denatured (unfolded) state (Fig. 8.6c, panel 1), with a single high-energy transition (or activation) state (\ddagger). In this model the process is dominated by the thermal fluctuations of the protein and it is characterized by the relative height of the activation energy barrier (ΔG^\ddagger) and the distance to the barrier, i.e., the position of the transition state along the mechanical reaction coordinate, also dubbed the width of the folding potential (Δx_u ; Fig. 8.6c, panel 2). As with any denaturant, proteins unfold under force because the free-energy barrier (transition state) to unfolding is lowered, in comparison to the native state. The force-dependent rate constants (k_u , unfolding rate, and k_f , folding rate) are given in this treatment by

$$k_u(F) = A e^{-(\Delta G_u^\ddagger - F\Delta x_u)/k_B T} \quad (8.6)$$

and

$$k_f(F) = A e^{-(\Delta G_f^\ddagger + F\Delta x_f)/k_B T} \quad (8.7)$$

where A is the frequency factor and Δx_u and Δx_f are the distances along the reaction coordinate (end-to-end length) from the folded state to the transition state and from the unfolded state to the transition state, respectively; k_B is the Boltzmann constant and T the absolute temperature. Thus, the application of a force F_u on a protein increases its unfolding rate (by reducing the activation energy of unfolding by $F_u \Delta x_u$) but decreases its refolding rate. A force applied at an angle θ to the reaction coordinate x adds a mechanical potential $-(F \cos \theta)x$, which tilts the reaction landscape and lowers the barrier (Fig. 8.6c, panel 2). The rate constant of the unfolding process therefore depends exponentially on the force applied and the strain induced (deformation). This constant reflects the stochastic transition from the folded to unfolded states and thus it determines the time required to cross the barrier.

We can calculate a probability density for unfolding, which predicts the most likely force of unfolding in terms of the spontaneous unfolding rate constant as follows:

$$F_u = (k_B T / \Delta x_u) \ln(r \Delta x_u / k_u^0 k_B T). \quad (8.8)$$

Using an analytical solution of this type, we should in principle be able to calculate the kinetic parameters for the process, k_u^0 (spontaneous rate of unfolding) and Δx_u (width of the activation energy barrier, i.e., the distance on the reaction coordinate over which the force must be applied to reach the transition state). However, although pulling at a fixed loading rate (i.e., constant force, Eq. 8.5) is experimentally feasible (see later), most protein unfolding experiments have been carried out at a constant length. In these experiments the loading rate varies as the molecule is stretched because the stiffness of the molecule depends on the applied force; hence, this equation cannot be used directly. Instead, k_u^0 and Δx_u are typically determined using Monte Carlo simulations that mimic the stochastic nature of the thermally driven unfolding of a protein molecule stretched at a constant rate (Oberhauser et al. 1998; Rief et al. 1998a; Carrion-Vazquez et al. 1999a). This procedure models the extensibility of the protein with an entropic elasticity, such as that described by the WLC, and a force-dependent all-or-none (two-state model) unfolding of its individual modules. The unfolding of the modules is described by Eyring rate theory with a simple two-state Markovian model and a force-dependent rate constant. In this description the probability of observing unfolding of any module at each step is

$$P = N k_u \Delta t \quad (8.9)$$

where N is the number of folded modules, k_u is the unfolding rate constant, and Δt is the polling interval (Rief et al. 1998a). The simulation is performed for a number of pulling speeds using different combinations of k_u^0 and Δx_u until they replicate the observed speed dependence (Fig. 8.6b; Carrion-Vazquez et al. 1999a). To make this method more accurate, the values of k_u^0 and Δx_u must also fit the force histogram simultaneously (Fig. 8.5b). An alternative method has recently been proposed that compares mutant and wild-type data to calculate these parameters (Best et al. 2002).

The mechanical kinetics parameters (i.e., k_u^0 and $\Delta\chi_u$) offer a useful description of the energy landscape of a protein under force that, although simplistic, allows us to compare the mechanical strength between different proteins (Sect. 8.4).

Finally, it has recently been suggested that the Eyring kinetic model may not be the most appropriate theory to describe this process (Howard 2001; Bustamante et al. 2004). In the Eyring model, the transition state is considered similar to the initial state but, in principle, this is only applicable when covalent bonds are made or broken. In contrast, Kramers rate theory may better apply to the typical protein conformational changes, in which a large number of bonds are made or broken (reviewed by Bieri and Kiefhaber 2000).

Protein Engineering Analysis

Mutational analysis is a high-resolution tool that permits minimal perturbations of the system to be introduced in order to interrogate the process under study with a view to revealing molecular details of the underlying mechanism. It has provided a unique way of studying the molecular basis of how proteins respond to mechanical force. Several types of mutant proteins of the titin I27 domain (Sect. 8.4.1) have been used in SMFS studies. Loop insertion using glycine residues was used to show the existence of a “mechanical clamp” in I27 (a model system in these pulling experiments, Sect. 8.4.1) and demonstrated the amino acid resolution of the technique (Carrion-Vazquez et al. 1999b). Proline mutagenesis was used to show the existence of a mechanical folding intermediate (Marszalek et al. 1999b) and to alter the mechanical stability of the domain (Li et al. 2000a). This approach takes advantage of the fact that proline (in addition to being a breaker of secondary structure) is an imino acid and, as such, cannot form backbone hydrogen bonds. Conservative substitution and deletion mutagenesis have been used to demonstrate the existence of the I27 folding intermediate (Fowler et al. 2002). Furthermore, analysis of the so-called mechanical Φ value holds the promise for a more detailed examination of the structure of the transition state in a forced unfolding reaction (Best et al. 2002). This is the mechanical equivalent to the well-established Φ -value analysis (Fersht et al. 1992), which is used to examine the conformational effect of a mutation based on the relative changes in the free energy of the native, transition, and denatured states ($\Delta\Delta G_{N-\ddagger}$, $\Delta\Delta G_{U-\ddagger}$). The Φ value is obtained from the changes in thermodynamic stability measured from the shifts in the equilibrium denaturation curves. The unfolding Φ value, defined as $\Phi_u = \Delta\Delta G_{N-\ddagger} / \Delta\Delta G_{N-D}$, is determined by comparing the effect of the mutation on the transition state and measures the amount of native structure that is present around the mutated residue in the transition state. $\Delta\Delta G_{N-D}$ is determined using equilibrium denaturation experiments. This analysis can only be applied when $\Delta\chi_u$ remains the same for wild-type and mutant proteins.

These methods have allowed us to establish that mechanical unfolding of titin I27 is a three-state process in which the transition state is very similar to the native state (Sect. 8.4.4).

8.3.3.2 Force-Clamp SMFS

As we have seen in the previous section, mechanical unfolding experiments are typically performed at a constant length (not at a constant force), where it is not possible to directly calculate the unfolding rate constants as a function of the force. However, it is now possible to stretch the molecule at a constant force and directly obtain the kinetic parameters of the process in a model-independent manner (Oberhauser et al. 2001). These force-clamp experiments have so far been performed on a few polyproteins: titin I27, titin I28 (Oberhauser et al. 2001), and ubiquitin (Fernandez and Li 2004; Schlierf et al. 2004). In this mode, instead of a sawtooth pattern of well-separated unfolding peaks in a force–extension curve, the unfolding of the individual modules of the polyprotein is here monitored as a stepwise increase in the length of the polyprotein as a function of time, at a given force, giving rise to a characteristic “staircase” pattern (Fig. 8.3c). It must be noted that, whereas in the length-clamp mode the fingerprint for assessing protein unfolding is based on both the magnitude of the unfolding force (F_u) and the increment in the contour length (ΔL_c), in this mode the step height is the only parameter available to do this identification.

Thus, the extension of a polyprotein in force-clamp presents the unfolding events in a very different way (i.e., an extension–time curve), complementing the constant-length mode and resolving many of the uncertainties inherent in the analysis of the data from a sawtooth pattern. In addition to simplifying the kinetic analysis, as it directly measures the force dependence of the unfolding probability of proteins, the force-clamp mode holds the promise of capturing new unfolding intermediates (or exhibiting them more clearly), as they may survive longer under these conditions. Still, this mode is limited by the tip stability (drift) and, particularly, by the speed of the force-feedback system (at present approximately 3 ms).

8.3.3.3 Computer Simulations of Mechanical Unfolding

SMFS experiments can measure the forces required for the mechanical unfolding of protein molecules and can resolve the changes in length with single amino acid resolution (Carrion-Vazquez et al. 1999b). However, at present they cannot provide a detailed structural resolution of the process. To this end computer simulations have proven to be very important for the atomic analysis of this process. Simulations allow us to follow how mechanical forces change the structure of the proteins under stress at the atomic level. These simulations are easy to implement since the denaturing agent (force) and the reaction coordinate (N–C distance) are simple to simulate. Since these techniques also deal with single molecules they are especially suitable for direct comparison with SMFS results, which constitutes one of the main appeals of the SMFS field. Two main groups of methods have been used to this end: all-atom molecular dynamics (with and without explicit water solvent) and on- and off-lattice simulations (Israelewitz et al. 2001).

Simulation techniques have much higher time resolution than the single-molecule techniques (which can only capture slow conformational motions). They also use shorter computational times in the range of nanoseconds or tenths of nanoseconds, which is achieved by increasing the pulling speed in the case of constant-velocity simulations (or by increasing the force in the case of constant-force simulations). In most of the cases, however, a direct comparison is not possible because these techniques simulate the unfolding process over a very short period of time (picosecond to nanosecond range), whereas SMFS experimental data are obtained over a much longer period of time (millisecond to second range). The simulations are generally performed at pulling speeds many orders of magnitude faster than the experiment (eight or more, typically 10^6 – 10^7 vs. 0.1 – 1 nm ms⁻¹). As a consequence of this time reduction, much irreversible work is done to stretch the molecule, which typically results in force peaks 1 order of magnitude higher than those observed in the AFM experiments. Hence, it is not clear whether they should precisely model the real process in these short periods of time and they may not fully reproduce the experimental conditions.

In spite of these difficulties, the synergy between experiment and simulation is very powerful. It is expected that this will be central to the future development of the field in obtaining the necessary high-resolution picture of the mechanical unfolding of proteins. A good example of this will be shown in Sect. 8.4.1.

8.3.4 Mechanical Refolding of Proteins

More than 40 years ago, Anfinsen et al. (1961) demonstrated that the amino acid sequence is the primary determinant of the 3D structure of a folded protein. Although a lot of effort has since been directed at elucidating the rules that govern the protein folding reaction, the folding code remains elusive. In his famous paradox, Levinthal (1969) pointed out that a random-coil polypeptide chain would require an astronomical amount of time to reach the native state if it were to simply explore all the possible conformations at random. Nevertheless, this process takes typically less than 1 s *in vitro*. This consideration triggered the search for partially folded intermediates that might guide the process of folding.

The main obstacle to solve this “folding problem” is the high degree of cooperativity of the unfolding transition, which at equilibrium leads to the accumulation of large populations of native and unfolded forms only. Partially folded forms along the kinetic pathway (kinetic intermediates) are rarely stable enough to be detected, having lifetimes in the millisecond range. As a result, their structural characterization is difficult. In this section we will review how SMFS may contribute to help solving this important problem.

As mentioned, pulling experiments occur far from equilibrium. Unfolding rates along the force–unfolding pathway can be estimated, although refolding rates along the same pathway cannot be typically measured because the high forces applied often prevent the folding of proteins. Still, it is possible to measure refolding rates following relaxation of a polyprotein, although these measurements are still difficult to make (Carrion-Vazquez et al. 1999a).

These experiments use the length-clamp mode of SMFS as follows. Once a stretch of a single protein molecule has been trapped between the tip and the substrate it can be held there by setting the retraction distance at less than that of the unfolded polyprotein. This prevents its detachment from the tip or the substrate, and allows repeated cycles of extension and relaxation using a length-clamp protocol of double pulse (Carrion-Vazquez et al. 1999a).

Although these measurements are performed in the absence of force (i.e., unloaded), very low residual forces may still be acting, which may drastically slow refolding. Hence, there is no guarantee that this pathway is the same as the “solution” pathway, especially considering that both pathways (mechanical and chemical) have been found to be different in the case of the unfolding of the titin I27 domain (Sect. 8.4.1.).

Force-clamp SMFS has recently enabled a protein molecule to be stretched under a constant force and its refolding followed under a low defined force (Fernandez and Li 2004). In these experiments, a polyubiquitin chain was first unfolded at a high, constant force to check the number of modules trapped in the stretch. The force was then relaxed to lower values so that refolding could be monitored by measuring the changes in the end-to-end length of the protein (Fig. 8.7). After the polyprotein chain had reached an appropriate, constant length (which takes a variable amount of time), the force was increased again to verify that folding had occurred. Thus, the new unfolding events indicated the occurrence of folding during the relaxation of the protein to its original length.

It has recently been argued, on the basis of simulations, that the reason why individual protein folding events were not fully resolvable in these experiments may be due to the large fluctuations to which the unfolded modules are subject (i.e., they may be acting as soft entropic spring “linkers”), which could be masking the decrease in contour length produced by the unfolding of each module (Best and Hummer 2005). However, these simulations failed to reproduce the experimental observation that all the domains fold in a cooperative way towards the end of the folding trajectories (Brujic and Fernandez 2005).

Mechanical refolding has also been reported recently for filamin (a cytoskeletal actin-binding protein), revealing the existence of discrete intermediates during the process (Schwaiger et al. 2005). In this study a mechanical double-jump method was used, the equivalent to the classical double-jump stopped-flow method.

It must be noted that in all these SMFS refolding studies, there is some concern that the folding dynamics of proteins may be significantly altered by surface immobilization to both the tip and the substrate.

8.4 Intramolecular Interactions

SMFS has been used to study the multiple intra- and intermolecular forces that determine the structure and dynamics and, therefore, the function of protein molecules. In this section we will review the studies on mechanical unfolding and folding. We also include here proteins that occur naturally in complexes and whose mechanical properties have been studied as a whole. Subsequently, we will devote the

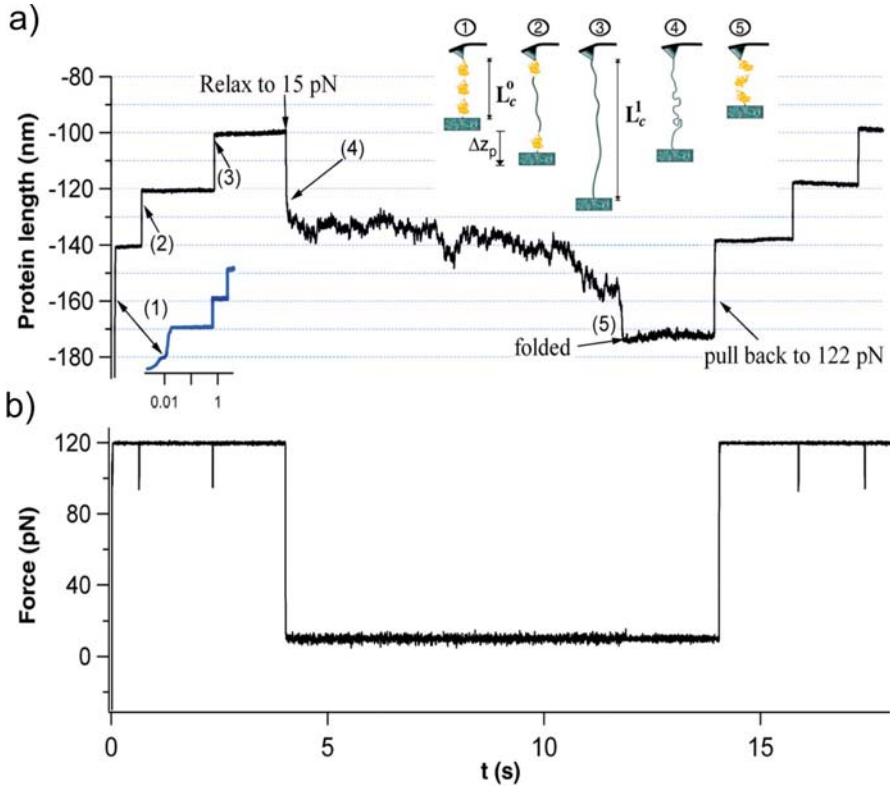


Fig. 8.7. Mechanical refolding of polyubiquitin monitored by force-clamp SMFS. **a** Protein length vs. time. After unfolding a polyprotein stretch at 122 pN (first staircase pattern), the force is lowered to 15 pN and the changes in the polyprotein folds are observed. Afterwards, the force is increased to unfold the polyprotein stretch to check that the protein was completely refolded. **b** Force–time relationship. Note the fluctuation of the force as the feedback system adjusts the force back to the set value. This follows and marks the unfolding of each module (after Fernandez and Li 2004, with permission, copyright 2004 American Association for the Advancement of Science)

following section to studies that specifically probe intermolecular forces, and which typically use different methodologies.

Both mechanical and nonmechanical proteins have been studied by SMFS. Figure 8.8 shows their modular and atomic structures. We have summarized the findings of these mechanical analyses in Table 8.1, which includes the values of the WLC adjusting parameters used to fit the family of experimental curves (p and ΔL_c) as well as representative values of mechanical parameters such as F_u (at a particular pulling speed), k_u^0 , and Δx_u .

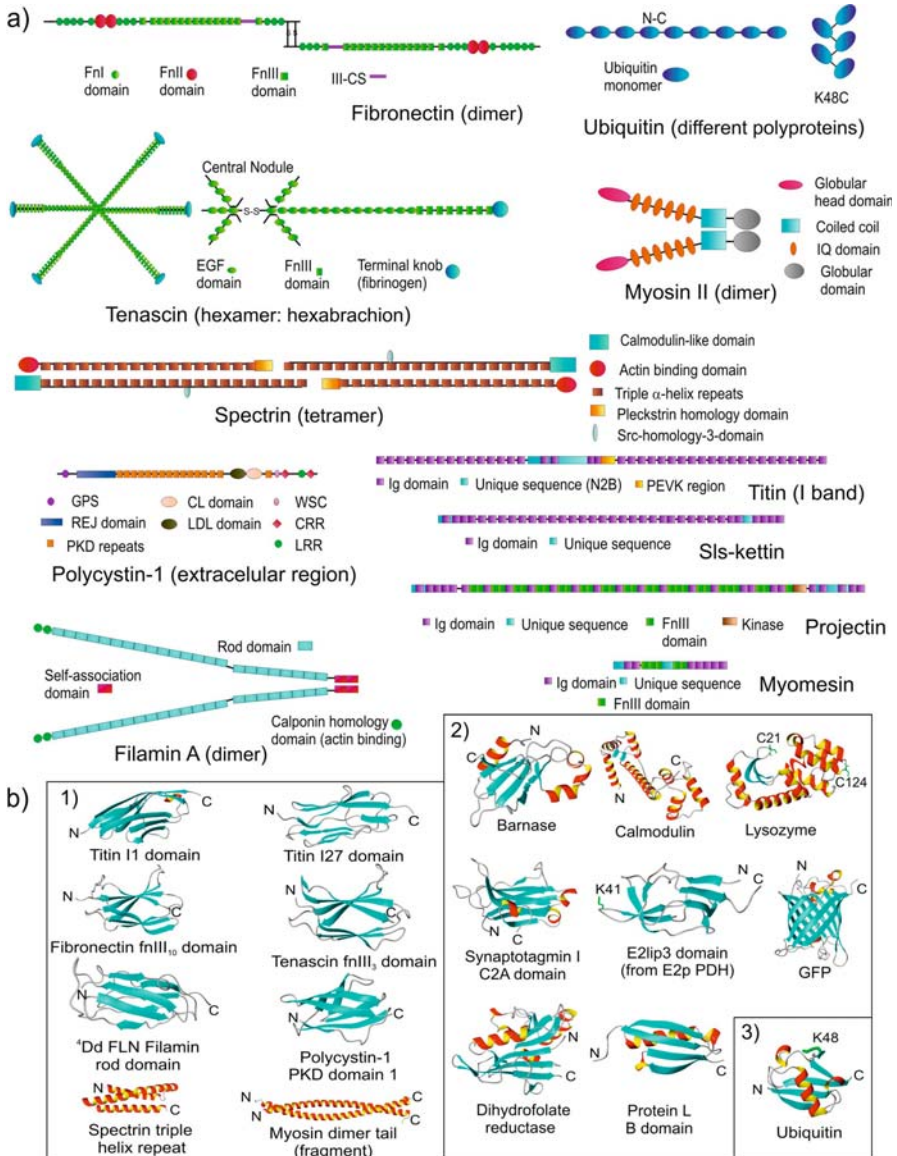


Fig. 8.8. Structure of proteins studied by SMFS. **a** Native proteins: modular structure. **b** Atomic structures of proteins and protein modules (displayed using the MOLMOL program, Koradi et al. 1996). The application point of the force is indicated by either N or C when the protein was pulled from its termini, or by the name and number of the specific amino acid residue in the one-letter code. **1** Modules from mechanical proteins. **2** Modules from nonmechanical proteins. T4 lysozyme polyprotein was synthesized by forming disulfide bonds between two exogenous cysteine residues. **3** We have classified ubiquitin in a distinct category since its mechanical role as a handle for the proteasome unfoldase is still hypothetical

8.4.1 Model System: I27 Module, an Immunoglobulin Domain from Titin

One of the model systems most commonly used to study mechanical unfolding/refolding is the I27 module from titin, a gigantic multimodular protein responsible for the so-called passive elasticity of muscle (Sect. 8.4.4, Figs. 8.2c, 8.8a, reviewed by Tskhovrebova and Trinick 2003). The I27 module from human cardiac titin is an Ig domain from the elastic (I-band) region of the protein that was originally chosen for several reasons. Its tertiary structure was known (Fig. 8.8b), its stretching had been simulated by molecular dynamics, and it was found to be thermodynamically very stable (Table 8.1; Carrion-Vazquez et al. 1999a and references therein). This module, 89 amino acids long, is a β -sandwich fold composed of seven β -strands (AA', B, C, D, E, F, and G) that folds into two face-to-face β -sheets through backbone hydrogen bonding and hydrophobic core interactions (Improta et al. 1996). With the exception of the parallel pair A'G, all adjacent β -strands are antiparallel (Fig. 8.10a, left).

Simulations of I27 stretching by molecular dynamics identified its force-bearing structural components as the β -strands A and G (proximal to the N- and C-termini of the domain, respectively), which are connected to β -strands B and A', respectively, through two patches of backbone hydrogen bonds (Figs. 8.9, 8.10a, left; Lu et al. 1998; Klimov and Thirumalai 2000; Lu and Schulten 2000; Paci and Karplus 2000; Li and Makarov 2003). These two patches of hydrogen bonds are true "structural" barriers with a different mechanical resistance: a low force barrier located between β -strands A and B, which involves two hydrogen bonds between the β -strands (outside the β -strands there is an additional, weaker hydrogen bond), and a high-force one between β -strands A' and G involving six hydrogen bonds. Both patches of hydrogen bonds are perpendicular to the direction of the force vector (a "shear" mechanical topology; Figs. 8.9, 8.10a, left). The remaining hydrogen bonds in the structure are parallel to the force vector (a "zipper" mechanical topology; Figs. 8.9, 8.10a, right) and like the hydrophobic core offer low resistance to extension.

The predictions of this model were found to be remarkably consistent with the experimental data from SMFS using polyproteins. In the length-clamp mode they revealed a series of force peaks of approximately 200 pN (at 0.6 nm ms^{-1}), each one corresponding to the unfolding of a domain. When the overall process was modeled by the WLC equation of polymer elasticity (as a series of polymer stretching events), a stepwise increase of the polymer length (ΔL_c) of 28.1 nm was found, which corresponds to the length of the force-hidden region located after the A'G patch. The hypothetical role of the A'G patch as a mechanical clamp was tested by loop insertion (Fig. 8.5d) and proline mutagenesis (Carrion-Vazquez et al. 1999b; Li et al. 2000a). In the force-clamp mode the extension-time curve revealed approximately 22-nm steps (at a constant force of 180 pN; Fig. 8.3c), corresponding to the expected increase in length due to the unfolding of a single domain. It should be noted that the 28.1-nm increase in contour length calculated from the length-clamp mode using the WLC model corresponds to the length of an unfolded module stretched at infinitely high force. At lower forces partial coiling of the amino acid chain, spontaneously driven by polymer entropic elasticity, would reduce this length (Oberhauser et al. 2001).

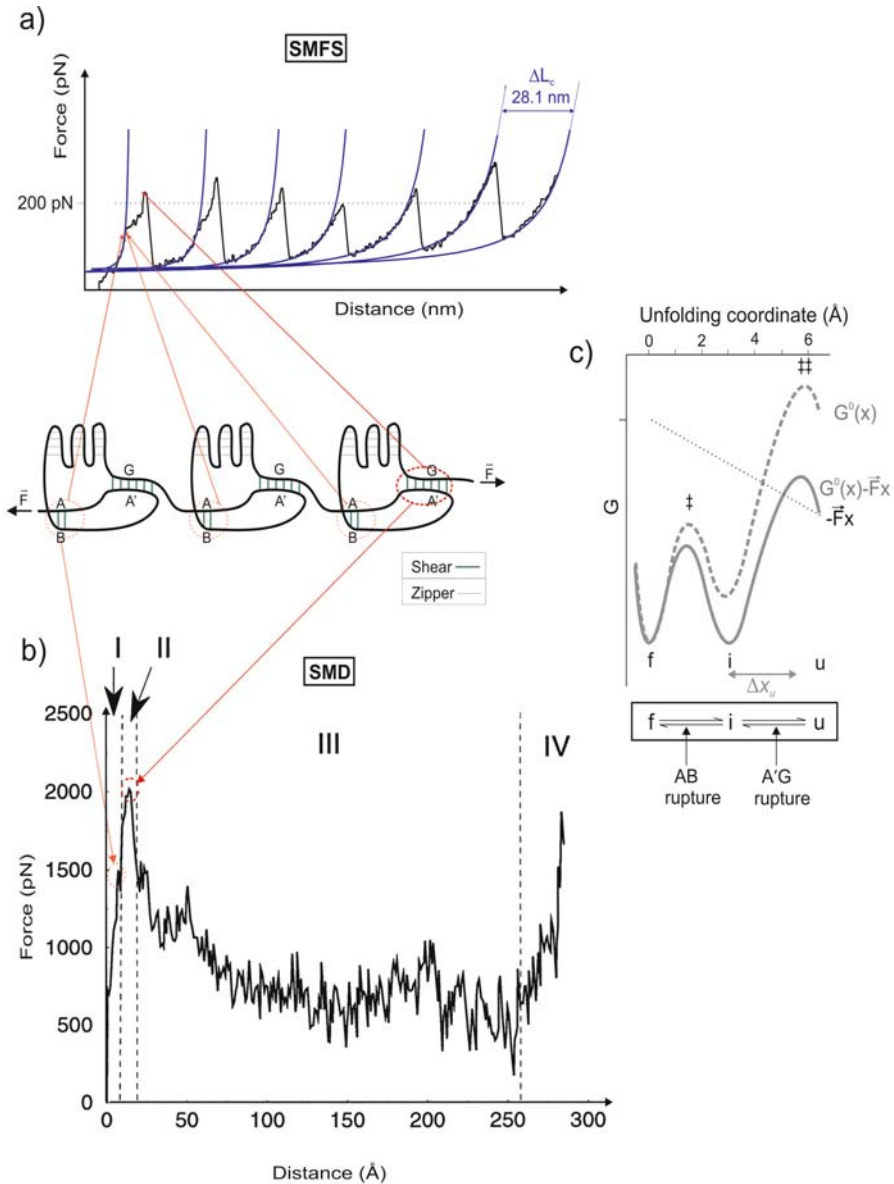


Fig. 8.9. Mechanical architecture of the I27 module. Schematic representation of an I27 polypeptide. Zipper bonds in the structure are represented by *dashed grey lines* and shear patches (AB, A'G) by *solid green lines*. **a** Force spectrum measured by AFM (from Fisher et al. 2000). **b** Force spectrum predicted by steered molecular dynamics (SMD) simulations. The main stability determinants of this module are a minor mechanical barrier (A-B patch of hydrogen bonds in the polypeptide backbone) and a major one (A'G patch of hydrogen bonds in the backbone). The major resistance barrier corresponds to the force peaks in the force–extension

The stretching of I27 polyproteins also confirmed another prediction, of a higher resolution, made by the simulations: the existence of a weaker mechanical barrier at the AB patch. Thus, the corresponding force–extension recordings revealed a fine mechanical detail consisting of a small deviation from the pure entropic behavior described by the WLC model, which appeared as a “hump” of decreasing intensity on each of the saw teeth (Fig. 8.9). This deviation was interpreted as evidence for the existence of an unfolding intermediate involving the rupture of the AB patch. These humps are found at lower force than the unfolding peaks (approximately 100 pN, at 0.3–0.5 nm ms⁻¹) and extend each module by approximately 7 Å. Each hump represents the simultaneous rupture of the hydrogen bonds from the AB patches of all the modules that remain folded in the polyprotein stretch that has been trapped in the experiment. The deviation is more evident in the first peak, it decreases progressively with module unfolding, and it completely disappears when all the modules of the trapped fragment have been unraveled. Hence, it appears that once a module unfolds and relaxes the force on the system, the remaining domains of the polyprotein fragment quickly refold their AB patches into their native conformation. The existence of this unfolding intermediate in the wild-type I27 module was later confirmed by stretching mutant proteins (Marszalek et al. 1999b; Fowler et al. 2002). Moreover, it was suggested that it may play a role in the elasticity of titin beyond the normal physiological range of extensions (i.e., at high forces; Sect. 8.4.4)

In spite of the potential pitfalls of these computer simulations, they have provided a wealth of data that have deepened our understanding of how proteins respond to a mechanical force. Thus, simulations rightly predicted the experimental behavior of the I27 domain under a mechanical force. Considering, as we have seen, that simulations are often performed at pulling speeds many orders of magnitude faster than those used in experiments, in principle it is not clear that they should rightly reproduce the true process. However, the close agreement between the experiments

←
sion spectrum (a). The minor barrier is detected as a deviation of the WLC (“hump”), which is more apparent in the first peak of the spectrum since it collects the contribution of all the events from the stretched modules (*pink arrows*). This intermediate refolds after each unfolding event, as the force relaxes. These results are correlated with SMD simulations, where a low force peak and a high one can be observed when stretching a single domain (b). Although the extension data are similar (28.1 nm in AFM vs. 25–30 nm in SMD), the force data are different. Thus, the AFM mean peak value is about 200 pN, whereas the main force peak in the SMD simulation is about 2,000 pN. This discrepancy can be attributed to the different timescale of the experiments; simulations are much shorter processes (1 ns to tenths of 1 ns) than experiments (milliseconds to seconds) and thus they cannot fully reproduce the experimental conditions (from Lu et al. 1998, with permission, copyright 1998 The Biophysical Society). c Free-energy diagram for I27 unfolding. The process is modeled as a three-state system (folded–intermediate–unfolded). The *solid grey curve* represents the free energy under a force. The *dashed grey curve* represents the free energy with no applied force. The application of force lowers the energy barriers of both transition states (‡, ††) and the energy of the different states. This increases the rate of the forward reaction (from Williams et al. 2003, with permission, copyright 2003 Macmillan Publishers Ltd.)

and simulations for the I27 domain is intriguing and does raise many important questions regarding how proteins respond to a mechanical force.

A unified model for the forced unfolding of I27 has recently been proposed, based on a three-state system (Fig. 8.9, panel c) with three different unfolding pathways at three different force ranges: at high forces an intermediate state is rapidly reached and “module unfolding” occurs from this state, at lower forces unfolding occurs directly from the native state, and at forces below approximately 43 pN it follows a “solution” pathway (Li et al. 2002; Fowler et al. 2002; Williams et al. 2003).

8.4.2

Other Protein Folds: from “Mechanical” and “Nonmechanical” Proteins

Ever since the first SMFS study on titin (Rief et al. 1997b) several proteins and protein modules with and without “mechanical” functions have been analyzed. Their atomic structures are shown in Fig. 8.8, while Table 8.1 summarizes the main findings, as well as the details of the methods used. We have listed the criteria for single-molecule identification, their mechanical properties and the kinetic parameters extracted, as well as comparable parameters obtained by other techniques. The kinetic parameters of the forced unfolding process, k_u^0 (thermal unfolding rate) and Δx_u (width of the folding potential), are simple, yet useful indicators of the energy landscape of a protein. They permit rough comparisons of the mechanical strength of different proteins to be made.

The protein structures analyzed by SMFS include (protein classification based on the SCOP database, <http://scop.mrc-lmb.cam.ac.uk>; Murzin et al. 1995) (1) the “all- β ” protein E2Lip3, (2) several all- β structures of the β -sandwich type including Ig-like domains (from titin, fibronectin, tenascin, filamin, polycystin-1, Sls-kettin, and projectin) and a C2 domain-like module (from synaptotagmin I), (3) a protein of the β -barrel type (green fluorescent protein, GFP), (4) several “ $\alpha+\beta$ ” proteins (T4 lysozyme, barnase, dihydrofolate reductase, ubiquitin, protein L), (5) several “all- α ” structures (calmodulin, spectrin, myosin II tail), and (6) several unstructured polypeptides (elastin, PEVK, N2B, EH). With the exception of the disulfide-bonded T4 lysozyme, E2lip3, and ubiquitin (Sect. 8.4.4) all the proteins studied to date have been pulled apart from their N–C termini only. This pulling geometry is functionally relevant for most naturally occurring modular proteins subject to mechanical force such as titin, tenascin, fibronectin, or spectrin.

Overall, this comparison shows that the proteins studied so far have widely different mechanical stability, measured by the most probable unfolding force (F_u). Interestingly, “mechanical” proteins tend to be mechanically more stable than “non-mechanical” ones (Table 8.1). On the basis of their structure, β -strand proteins tend to be mechanically more stable than α -helical ones. The β -strand proteins with a shear mechanical topology were found to be mechanically more stable than those with a zipper one (Sect. 8.4.4, Fig. 8.10a). The Ig-like β -sandwich fold seems to be a platform that can tolerate higher mechanical stress. Inside this class, the Ig (forces ranging from 150 to 330 pN, at 0.6 nm ms^{-1}) and PKD domains (to which polycystin-1 belongs; 50–200 pN, at 1 nm ms^{-1}) appear to be mechanically more stable than

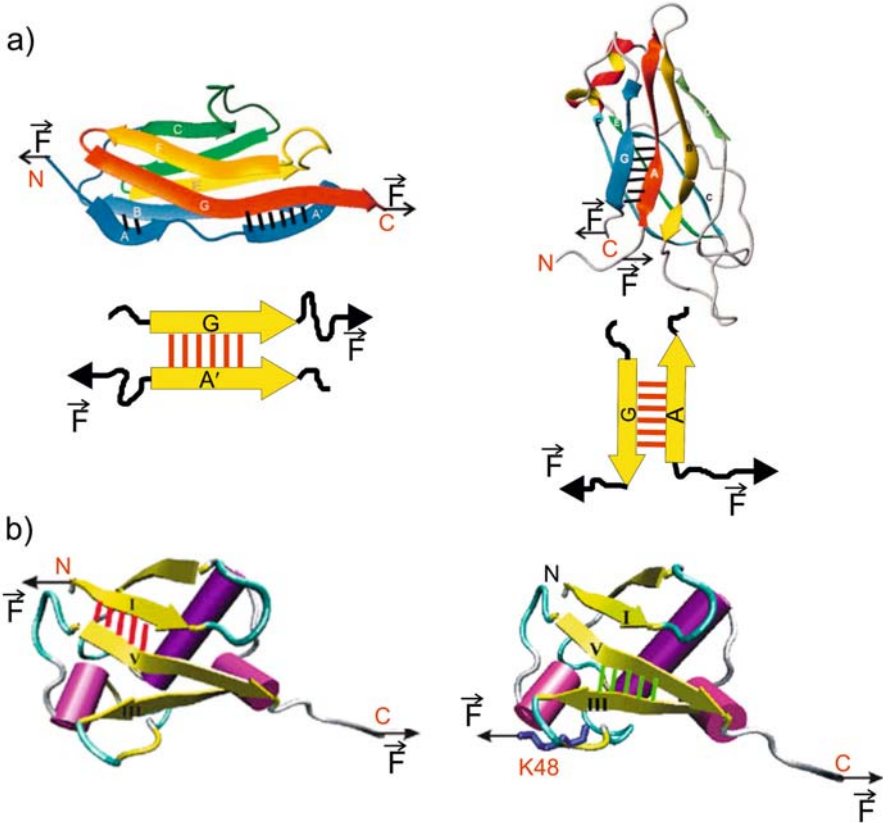


Fig. 8.10. Directional effects of a mechanical force in proteins. The directional effects derive from the fact that force is a vector quantity. **a** Shear vs. zipper configurations. The mechanical stability determinants of two Ig-like modules (I27 and C2A from synaptotagmin I) are patches of backbone hydrogen bonds that can either be orthogonal (shear configuration, *left*; the A'G patch of the I27 titin module) or parallel to the direction of the applied force (zipper configuration, *right*; AG patch of the C2A module). The zipper configuration leads to a lower mechanical stability presumably because the critical bonds are broken one by one (serial bonds), in contrast with the all-or-none manner of breakage required in the shear configuration (parallel bonds) (from Carrion-Vazquez et al. 2000). **b** The mechanical stability of ubiquitin is linkage-dependent. The force needed to unfold a ubiquitin domain is larger when the force is applied to the N–C termini (*left*) than if the force is applied between the K48 residue and the C-terminus (*right*). The application points of the force are indicated in red. N and C represent the N- and C-termini and K48 is lysine residue 48 (after Carrion-Vazquez et al. 2003)

the fibronectin type III (fnIII; 75–220 pN, at 0.6 nm ms⁻¹) or the E-set domains (to which filamin belongs; 50–220 pN, at 0.37 nm ms⁻¹).

In the lower regime of forces that SMFS can detect we can find both elastomeric (e.g., elastin, PEVK, N2B, EH) and “nonmechanical” proteins (e.g., calmodulin). For high forces, the correlation between mechanical stability and mechanical func-

tion holds relatively well but it is less true below approximately 100 pN (Table 8.1). This may reflect the possibility that for some proteins with no known mechanical function their relatively high mechanical stability (e.g., protein L, GFP) may just be an epiphenomenon, unrelated to the biological function of the protein. Indeed, a number of proteins from the survey may never be mechanically pulled in the cell or if they are the pulling geometry in situ may be different to that of the SFMS experiments. Alternatively, for some proteins it may be a neutral trait, not directly selected by evolution (i.e., a remnant by-product of their evolutionary history). For instance, the mechanical stability of the A-band fnIII domains of titin (bound to the myosin thick filament of the sarcomere) is only slightly inferior to those of the I band (180 vs. 200 pN). Hence, in the case of the A-band domains (nonelastic) it seems that this trait was not evolutionarily selected but rather it may be an epiphenomenon originating through possible constraints found during the evolution of this elastic protein.

Mechanical stability seems to be roughly predicted by the unloaded unfolding rate constant (Table 8.1). Furthermore, recent interesting findings suggest that mechanical stability can be modulated by ligand binding. As shown in Table 8.1, two independent reports (Ainavarapu et al. 2005; Junker et al. 2005) indicate that the binding of several ligands to dihydrofolate reductase increases its mechanical stability. However, as the findings of these two groups are somewhat different (a direct increase of the “mechanical stability” vs. an increase in the lifetime of an unfolding intermediate), the molecular origin of this interesting stabilization effects remains to be investigated.

8.4.3

Supramolecular Mechanical Properties of Protein Complexes

Most protein machines are made of complexes of noncovalently bound polypeptides. Characterization of the mechanical properties of these multimolecular structures is also an important challenge for “single-molecule” force spectroscopy. Supramolecular complexes with an integral structure and behavior, such as the long coiled-coil structure of the myosin II tail (Fig. 8.8b), have also been characterized using this technique (Table 8.1; Schwaiger et al. 2002). The myosins are a large family of motor proteins that move along actin filaments as they hydrolyze ATP. Myosin II is a hexamer formed by four light chains and two heavy chains that have a globular N-terminal motor and a tail that dimerizes into a 150 nm long parallel coiled coil (Fig. 8.8). The typical force–extension curve of a myosin tail presents a characteristic molecular fingerprint. This suggests a mechanism for myosin elasticity in which the folded coiled coil initially extends by entropic elasticity up to 20 pN, and then at 20–25 pN, short coiled-coil segments start to unfold, leading to a very long force plateau (which marks a structural transition of the tail). At the end of the process the completely unfolded coiled coils additionally extend as the force rises rapidly. During the force plateau the coiled coil extends to about 2.5 times its original length. Furthermore, the force–extension profile of relaxation almost exactly follows the trace observed during pulling, indicating that the myosin coiled coil is a true elastic protein structure, able to refold against forces of up to 30 pN in much less than 1 s.

The comparison of the unfolding/refolding cycles shows that the transition presents little hysteresis. Thus, very little energy is dissipated in the complete cycle as the process is basically in equilibrium. Therefore, evolution seems to have selected this simple coiled-coil fold to confer nearly pure entropic elasticity to the tail of the myosin II machine (Fig. 8.6a, panel 3).

Other attempts have been made to measure supramolecular mechanics in a few other protein systems, but the results are often unclear and must be treated with caution. Future experimentation in such complex systems should ensure that all the appropriate controls are included in order to rule out the spurious interactions present in the system (intermolecular interactions and single-molecule events).

8.4.4 Achievements

8.4.4.1 Single-Molecule Mechanics vs. Bulk Biochemical Assays

Until recently, the only way to measure the stability of a protein was to change its physical (temperature or pressure) or chemical environment (using concentrated solutions of guanidinium chloride or urea, acid or alkaline pH) and monitor the loss of protein conformation by spectroscopic techniques in order to determine the change in the Gibbs free energy. Most of these folding studies have been performed using chemical denaturants acting on untethered proteins (i.e., in solution). However, a considerable number of proteins from the cytoskeleton and the extracellular matrix, as well as unfoldase substrates from chaperone, translocation, and degradation machines, are likely to be subjected to mechanical forces and are tethered. Thus, at least for those proteins, SMFS experiments may more closely mimic the real physiological conditions in which they function in the cell, i.e., typically immobilized and subjected to a mechanical force (in some cases acting as a denaturing agent) that imposes on them a “natural” reaction coordinate (i.e., the length of the protein in the direction of the force to which it is subjected). Protein length is also a well-defined reaction coordinate with a clear physical meaning, in contrast to the less well physically defined kinetic “*m* values” used in chemical folding experiments (defined as the derivative of the natural logarithm of the folding, or unfolding, rate constant with respect to the denaturant concentration, they measure the sensitivity of the rate of the process to denaturant concentration and are generally interpreted as a measure of the change in solvent exposure of the lateral chains of amino acid residues).

Thermodynamic comparisons of the change in free energy (a function of state) between both methods should in principle give identical results, provided that the entropic contribution of tethering the ends of the molecule is properly corrected. On the other hand, we should consider that the kinetics of the reaction depends on the pathway and that SMFS unfolding experiments impose a reaction coordinate to the molecule that is different from that of the bulk (chemical) experiments. Hence, the kinetic parameters of unfolding obtained by both methods may differ. Thus, force acts along a single dimension in specific regions of the protein, typically the N- and

C-termini, while conventional denaturants have a more global effect. Interestingly, in this survey as a whole there is poor correlation between mechanical and chemical unfolding kinetic rates (a measure of kinetic stability) (Table 8.1). In contrast, a remarkable agreement has been found between the unfolding rate constants of the distal Ig modules of titin (Table 8.1, titin modules I27 and I28; Carrion-Vazquez et al. 1999a; Li et al. 2000b). This is noteworthy and raises interesting questions on the mechanical design of these particular domains. The unfolding pathway of I27 has been analyzed in detail comparing chemical and mechanical methods and using protein engineering based on the Φ -value (Fowler et al. 2002) and hydrophobic core destabilization analyses (Brockwell et al. 2002; Fowler et al. 2002; Best et al. 2003b; Williams et al. 2003). In this way it was confirmed that the two unfolding pathways are different. It was concluded that the I27 mechanically unfolds through an intermediate which is not populated in the chemical pathway, and that the transition state presents a different structure in both pathways. However, the A'G region is the only region in the I27 structure that is critical in both pathways, being responsible for kinetic stability in both cases. This may explain why the two unfolding rates are so close in this module.

The relationship between mechanical and chemical refolding for the two proteins that have been studied so far by SMFS, i.e., ubiquitin and filamin, has not yet been explored (Fernandez and Li 2004; Schwaiger et al. 2005). Still, both refolding pathways may differ considerably, based on the entropic cost of tethering (Carrion-Vazquez et al. 1999a) and on the fact that the application of force imposes a different reaction coordinate from that of the chemical experiments.

8.4.4.2 Detection of Rare Misfolding Events

During refolding experiments “skip” events have been observed. Analysis of their contour lengths suggests that these are due to the formation of a “superfold” upon refolding, which includes two consecutive domains and the linker region between them (Fig. 8.5e). This misfolded “superdomain” unfolds at forces similar to those of a single domain and refolds back to two normal domains. SMFS of the titin poly-I27 and the fnIII domains of tenascin have demonstrated the power of these single-molecule studies in detecting rare events, which occur in a small number of molecules within the population (2–4%). In this way, the efficiency (i.e., fidelity) of refolding after mechanical unfolding can be estimated (Oberhauser et al. 1999). Such unusual unfolding events could not be observed using ensemble techniques and they open the door to investigate the means of reversing undesirable misfolding processes that occur in the cell in a number of pathological conditions.

8.4.4.3

Directional Aspects of the Force: Shear vs. Zipper Configurations and “Achilles Heels”

Since force is a vector quantity, the mechanical stability of a protein will depend not only on the magnitude but also on the direction and application point of the force. Thus, a force will couple to a reaction whenever it has a component in the direction of the reaction pathway.

We must first consider the geometrical errors present in SMFS measurements. So far, we have assumed for simplicity an orthogonal pulling geometry, i.e., the polyprotein is pulled at a 90° angle with respect to the substrate plane (Fig. 8.4a). To be more realistic, most of the time the polyprotein would be pulled at a slight angle. However, it has been demonstrated that the range of the geometrical errors introduced in this way is negligible (Carrion-Vazquez et al. 1999b).

Protein domains from muscle (titin), cell adhesion (tenascin, fibronectin), cytoskeletal (filamin), and surface receptor usually belong to the Ig-like β -sandwich family of proteins (Fig. 8.8a). These are likely to have somewhat similar mechanical topologies at the breakpoint. This superfamily of Ig folds, which includes Ig (titin, projectin, Sls-kettin, myomesin), fnIII (tenascin, fibronectin, titin, projectin, myomesin), E-set (filamin), and PKD (polycystin-1) types, consists of seven stranded β -sandwich structures in which the N- and C-terminal strands are parallel to each other and point 180° in different directions (Fig. 8.10a, left). They seem to have evolved to withstand forces when connected in series based on a shear mechanical topology of the hydrogen bonds at the breakpoint (i.e., the force vector is orthogonal to the bonds, which are arranged in parallel in the mechanical circuit). This arrangement may provide these domains with a considerable resistance to mechanical stretching (Table 8.1, Fig. 8.10a, left). Interestingly, protein L, a protein with no known mechanical function but with a shear mechanical topology at the breakpoint, also has relatively high mechanical stability (Table 8.1, Fig. 8.8b).

On the other hand, the C2 domain, found for instance in the secretory protein synaptotagmin I, is a β -sandwich domain composed of eight antiparallel strands, with the N- and C-terminal strands pointing in the same direction, and a zipper mechanical topology at the breakpoint, i.e., the force vector is parallel to the bonds which are mechanically arranged in series (Fig. 8.10a, right). As can be seen in Table 8.1, the two proteins with a zipper configuration at the breakpoint studied so far, i.e., C2 domain (Carrion-Vazquez et al. 2000) and E2Lip3 domain (Brockwell et al. 2003), both have low mechanical stability when pulled in the N–C direction (see later).

The different mechanical resistance between shear (concerted rupture) and zipper (sequential rupture or “peeling”) topologies was predicted on the basis of molecular dynamics simulations of proteins (Rohs et al. 1999). This concept was experimentally demonstrated for DNA oligomers, choosing sequences where the only difference between the two configurations was the orientation of the tails (the sequence of the DNA was the same, only the topology was different). These two configurations were found to be thermodynamically (energetically) and kinetically equivalent when probed with traditional bulk methods (i.e., they have the same binding energy as well as the same thermal on and off rates). However, upon forced dissociation, the

complex with the shear geometry has an approximately 3-fold higher mechanical stability and a 15-fold greater unbinding probability than that with the zipper geometry (Albrecht et al. 2003 and references therein).

Recently, it has been possible to experimentally pull proteins from different points of application such that the results of two different pulling geometries can be compared (regular N–C pulling vs. pulling from the terminus and an internal residue; Fig. 8.10b). This is possible by taking advantage of the existence of naturally occurring covalent linkages for internal amino acid residues in a couple of proteins (i.e., ubiquitin and E2lip3). These experiments show that the pulling direction affects dramatically the mechanical stability of a protein so that the two pulling geometries present very different mechanical stabilities (Table 8.1; Brockwell et al. 2003; Carrion-Vazquez et al. 2003). This raises the possibility that evolution may have selected specific mechanisms in unfoldases to take advantage of the “Achilles heels” of their substrate proteins (e.g., the internal ubiquitination of protein substrates for proteasomal degradation may be one of those), as this would have the advantage of saving energy to the cell.

These directional effects on the mechanical stability of a protein are observed because a mechanical force acts locally along a single dimension, instead of globally. They illustrate the anisotropy of the mechanical unfolding landscape and certainly could not be found by chemical denaturation.

8.4.4.4

Molecular Determinants of the Mechanical Stability of Proteins

SMFS experiments have shown that (1) the mechanical stability, given by the most probable unfolding force (F_u), is not correlated with thermodynamic stability (ΔG) or with melting temperature ($T_m = \Delta G / \Delta S$; Table 8.1, reviewed by Carrion-Vazquez et al. 2000), (2) chemical and mechanical unfolding follow different pathways (Best et al. 2001; Fowler et al. 2002) and have different unfolding barriers (Best et al. 2001; Brockwell et al. 2002), (3) the mechanical stability and the mechanical unfolding pathway depend on the pulling geometry and the application points of the force such that proteins present mechanically weak spots (Brockwell et al. 2003; Carrion-Vazquez et al. 2003), and (4) mechanical stability can be modulated by ligand binding (Ainavarapu et al. 2005).

SMFS experiments do not provide direct structural information on the dynamics of unfolding but they do provide many parameters that are sensitive to the structure of the protein under study (Table 8.1; Li et al. 2000a). We have seen that proteins show a broad range of responses to mechanical stress, which cannot easily be rationalized in terms of predictors of mechanical resistance (Table 8.1). However, to date only a few proteins have been analyzed, each with a varying degree of detail (19 different ones in this survey; Table 8.1). Thus, although the molecular basis underlying the mechanical resistance of proteins is still unclear, we have seen that several determinants can be identified through these studies: amino acid sequence, mechanical topology, unloaded unfolding rate constant, and pulling geometry.

The force equation for the unfolding process (Eq. 8) predicts that the mechanical stability (F_u) depends on the unfolding distance, Δx_u (i.e., the location of the transi-

tion state along the mechanical reaction coordinate), the height of the barrier (ΔG_u^\ddagger which depends on k_u^0), and the loading rate (r). The effect of a force is stronger the further from the native state it acts (Bustamante et al. 2004). For instance, to overcome a $12\text{-}k_B T$ barrier (7.2 kcal mol^{-1}) would require an unfolding force of 200 pN for a protein that has an unfolding distance of 0.25 nm, while for a different protein, a similar barrier would require only 100 pN if it were located at an unfolding distance of 0.50 nm (at the same loading rate).

The molecular structure of a protein poses constraints on the location of the transition state in mechanical unfolding. It has been suggested that tertiary interactions have shorter distances to their transition states than secondary structures, and they tend to be brittle than secondary interactions (i.e., they break at high forces and after small deformations), which are more compliant (breaking at low forces and after large deformations). Furthermore, tertiary interactions may require more time to equilibrate than secondary ones and, therefore, they often present hysteresis in the pulling–relaxation cycle (reviewed by Bustamante et al. 2004). Mechanical proteins tend to be assortments of modules, often mostly of the same type (e.g., Ig, fibronectin), although they frequently display distinct mechanical stabilities. Low forces would stretch the more compliant domains (with longer distances to their transition states, e.g., ¹⁰FNfnIII domain; Table 8.1) while the brittle domains (with shorter distances to their transition states, e.g., I27 domain) would maintain their structure at higher applied forces, requiring extreme forces to unfold them.

Most proteins show a high degree of connectivity and as a result their unfolding seems to be highly cooperative with the stability of secondary structures depending on their tertiary context and often presenting no intermediates. Still, owing to the local action of the applied force, their mechanical stability tends to be related to highly localized molecular structures near the mechanical “breakpoint” rather than to the global structure. For instance, the mechanical stability of titin I27 has been found to depend exclusively on a specific, highly localized clamp of six backbone hydrogen bonds, the A'G patch (reviewed by Carrion-Vazquez et al. 2000), and the associated side-chain packing (i.e., hydrophobic) interactions between the A' and G strands. The hydrophobic core of this structure plays no role in resisting force (Best et al. 2003b). As we have seen (Sect. 8.4.2, Table 8.1) and as predicted by simulations (Lu and Schulten 1999), β -strand proteins tend to unfold at higher unfolding forces and to be brittle than α -helical ones. Furthermore, β -strand proteins with a shear mechanical topology are mechanically more stable than zipper β -strand proteins. This seems to derive from two effects: the higher energy barrier necessary for breaking many bonds in parallel (vs. breaking individual bonds in series) and the greater stiffness of this link (compared with the stiffness of a single bond or several bonds in series). As a result, the distance to the transition state is shorter owing to a deep well in the potential energy (Bustamante et al. 2004).

In contrast to the findings on the I27 domain, recent experiments combining SMFS, molecular dynamics simulations, and Φ -value analysis have revealed that the barrier to mechanical unfolding in the TNfnIII-3 domain involves only a few hydrogen bonds but significant side-chain packing (i.e., hydrophobic) interactions of the amino acid residues from both the peripheral β -strands and the core (Ng et al. 2005). Thus, while it remains to be explored whether other fnIII domains share this mechanism, it seems clear now that the highly localized hydrogen-bonding and

associated packing interactions (as found in I27) may not be the dominant determinant of mechanical stability for other groups of protein structures.

8.4.4.5

Muscle Elasticity, a Macroscopic Biological Property, Reduced to the Single-Molecule Level

The sarcomere is the basic contractile unit of muscle. Early studies showed that this relatively simple contraction/extension machine (i.e., it works in a single dimension) is also highly elastic (Huxley and Hanson 1954). The pioneering work by Wang et al. (1979) and Maruyama et al. (1981) demonstrated that the so-called «passive elasticity» of muscle (i.e., generation of restoring forces that resist stretch independently of ATP) is mainly mediated by titin, a giant protein (more than 3 MDa, the longest polypeptide known to date) that spans half a sarcomere (approximately 1 μm ; from the Z disk to the M line) and acts as a molecular spring (Fig. 8.2c). Passive elasticity plays an important role in muscle function since, typically, a muscle actively contracts against the elastic strain of a passively elongating muscle. This property ensures that the sarcomere recovers its initial dimensions on muscle relaxation (Tskhovrebova and Trinick 2003).

One of the remarkable feats of SMFS has been the reconstruction of the passive elasticity of intact myofibrils on the basis of a simple scaling up from the mechanical properties of single molecules of titin (Li et al. 2002). The mechanics of single titin molecules was reconstituted in turn from the mechanical properties of representative elements of its elastic region (i.e., N2, PEVK, and the proximal and distal Ig regions relative to the N-terminus of the protein), showing that titin behaves very differently from a Hookean spring (e.g., the cantilever of an AFM; Fig. 8.6a, panel 1). It is essentially an entropic spring in which force and extension do not follow a linear relationship.

Through this reductionist approach it was possible to explain muscle passive elasticity, a macroscopic property, from the additive mechanical properties of a single sarcomeric protein at the single-molecule level. According to this model, within the physiological range of sarcomere extension (i.e., low forces below 4 pN) unfolding would rarely happen; thus most of the elasticity of titin in the physiological range would result from the entropic elasticity of straightening the Ig domains in the I band, and of extending its unique sequences (PEVK and N2B). Mechanical unfolding of the Ig modules would basically function as a buffer of both length and additional entropic elasticity. This “shock-absorber” effect would only occur at a high force, i.e., in nonphysiological conditions, to prevent damage of the sarcomere (i.e., it is essentially a safety mechanism). Therefore, in response to axial tension, titin behaves as a multistage spring that adjusts both its length and apparent stiffness by virtue of its particular modular design.

The example of titin elasticity shows how single-molecule experiments can be used to elucidate, at a more fundamental level, the physiological function of a protein through the biophysical dissection of a complex hierarchical system.

8.5

Intermolecular Interactions: Findings and Limitations

Although this review focuses largely on the analysis of intramolecular interactions of proteins by SMFS, for comparison we also wish to briefly outline the main findings in the field of protein mechanical unbinding. Recent and more comprehensive reviews on the subject can be found elsewhere (Zlatanova et al. 2000; Hinterdorfer 2002; Weisel et al. 2003).

Mechanical forces, both external (shear flow) and internal (cytoskeletal rearrangements, motor proteins), control many cellular functions that are mediated by specific interactions between receptor and ligand molecules, such as rolling, motility, or adhesion. Proteins can be unbound from their partners by applying a mechanical force with an AFM. The unbinding force values obtained in these experiments can be related to the affinity values of the pairs at the single-molecule level. During the last decade, SMFS has been used to measure the unbinding force of a number of interactions between proteins and other biomolecules. Similar to mechanical unfolding, mechanical unbinding experiments are performed in the length-clamp mode under nonequilibrium conditions, and a reaction coordinate (the end-to-end distance) is imposed on the unbinding process.

Although the subject of study (i.e., the intermolecular bonds) seems in principle to be simpler and more directional than the intramolecular interactions, the setup here is somehow more complex as it involves two molecules, each with its own intramolecular interactions (Fig. 8.3b). In these systems sorting out intramolecular interactions is a difficult task that must be resolved in order to be able to extract clear-cut conclusions.

In this field it is acknowledged that the main difficulty of these experiments is finding a solid criterion to identify single unbinding events. Unlike the case of stretching a modular protein in unfolding experiments for which there is a straightforward single-molecule fingerprinting, to date there is no standard strategy to assure that the forces observed correspond to the rupture of the interaction between a single pair of molecules. Since in this field there is no single accepted standard capable of formally proving, in a sufficient way, that the observations correspond to single unbinding events, it is therefore assumed that a reasonable degree of confidence in the results can only be derived from a cumulative body of evidence of diverse origin (Zhu et al. 2002; Weisel et al. 2003). Perhaps the strongest argument for the occurrence of single bonds is the quantal behavior in measurements of a low number of molecular interactions, i.e., multiple force peaks with a periodic distribution (Florin et al. 1994). However, it must be stressed that this is simply a necessary but not a sufficient condition, as the quantal unit may represent just the smallest detectable binding unit, not the elementary unit.

In comparison with the unfolding experiments, these experiments present some additional differences. Firstly, the interacting molecules need to be firmly linked to the tip and coverslip surfaces to avoid detachment during the pulling process. The commonest methods involve covalent links and strong noncovalent interactions. Covalent links are obtained by using bifunctional agents or spacer linkers that impart higher mobility on the molecules, but if they are too long they can alter the force dynamics by increasing the elasticity or even changing the structure of the system

(Weisel et al. 2003). Noncovalent interactions are easier to use for functionalization, but are typically avoided for measuring strong interaction forces since they are weaker than covalent links. The most widely employed one is the biotin–avidin system, which resists forces up to about 160 pN (257 pN, in the case of streptavidin) at typical pulling speeds (Sect. 8.5.1). Secondly, the ligand and receptor have to be in the appropriate conformation and direction so they can interact properly (this can be determined by prior AFM imaging). In this sense, spacer linkers provide enough flexibility to the molecules so they can adopt the correct position. Finally, since adhesion between biomolecules always presents nonspecific interactions that may mask the specific ones, one must always make sure that nonspecific interactions are blocked or alternatively ensure that the appropriate controls are included. Typical controls consist of force measurements done after blocking the receptor attached to the coverslip (with free ligand), in the absence of receptors, or using “inert,” unrelated proteins.

In the first studies in this field, the unbinding force was related to the activation enthalpy, ΔH . Thus, comparisons of bond strengths of different analogs of biotin bound to (strept)avidin showed that the physical force required to break a bond was related to the activation enthalpy and not to the overall free energy of the bond (Sect. 8.5.1; Moy et al. 1994). However, at present the most widely accepted description of the unbinding process is based on the same force equation we have seen for mechanical unfolding. This is derived from the formulations of Bell (1978) and Evans and Ritchie (1997), originally developed for the case of intermolecular bonds (Eq. 8.8). In this equation, the unbinding force (F) depends on analog factors, loading rate (r), unbinding distance or width of potential (x_β), kinetic dissociation rate constant (k_{off}), and on the thermal energy ($k_B T$). Similarly, most of the later studies in this field do not use a single loading rate (to obtain a single value of unbinding force) but rather explore a full range of loading rates and forces to determinate the values of x_β and k_{off} (including its prediction at zero force). Interestingly, the properties of the interaction usually change at different loading ranges, indicating the complex nature of the energy landscape of receptor–ligand bonds (Merkel et al. 1999). Such experiments have allowed the rate constants, affinity, and structural data of the binding site to be estimated.

8.5.1

Model System: Biotin–(Strept)avidin

Ground-breaking SMFS experiments measured the interaction force of the biotin–(strept)avidin system (Florin et al. 1994; Moy et al. 1994; Table 8.2), which has become the main model system in this field. The methodology used in these first studies is essentially the same, with only small variations, as the one still used nowadays in most experiments. In this technique, the AFM substrate is functionalized with one interaction partner and is brought into contact with the AFM tip, which is functionalized with the other member of the interaction system. In this way, the interactions that occur can be examined by pulling back the positioner to separate the interacting molecules, thereby measuring the force of the interaction. A typical force–extension recording obtained in this manner is shown in Fig. 8.3b. The un-

binding force measured for the biotin–avidin pair was found to be approximately 160 pN, while that of the biotin–streptavidin pair was approximately 257 pN (details of the experiment and controls are listed in Table 8.2).

8.5.2 Other Protein–Biomolecule Pairs

Since the pioneering studies on biotin–(strept)avidin, mechanical unbinding of a number of protein–biomolecule pairs has been reported (both macromolecules and small ligands, see references in Table 8.2). These include interactions between proteins (adhesion proteins, cytoskeletal and motor proteins, GroEL chaperonin, and denatured proteins), receptor–ligand pairs, antigen–Ig pairs, protein–DNA, and protein–lipid. The mechanical properties of a selection of these protein–biomolecule pairs are presented in Table 8.2. Studies on the interaction between proteins, antigen–Ig pairs, and receptor–ligand pairs are the most extended works (Zlatanova et al. 2000; Hinterdorfer 2002; Weisel et al. 2003), while interactions between DNA and proteins have received less attention.

As we have seen, a mechanical force causes bond rupture and, intuitively larger forces just increase bond failure rates. This is the characteristic behavior of “slip bonds.” However, the so-called catch bonds would strengthen under force and their existence has recently been demonstrated experimentally using SMFS in the P-selectin adhesion bond. This finding might be physiologically relevant since the biphasic transition between catch and slip bonds provides an interesting mechanical switch for the regulation of leukocyte rolling on selectins, which initially increases and then decreases as the wall shear stress increases (Table 8.2; Marshall et al. 2003). Catch bonds also seem to regulate flow-dependent bacterial adhesion.

Recently, AFM has been used to study the energetics of extracting single membrane proteins from their lipidic environment. For example, several studies have analyzed the anchoring forces of bacteriorhodopsin to its native “purple membrane” (a well-characterized seven-transmembrane α -helical protein from *Halobacterium salinarum*) and the mechanical stability of an α -helix in a hydrophobic environment (reviewed by Frederix et al. 2003; Janovjak et al. 2006). In these experiments, which use a strategy that departs considerably from the classical receptor–ligand methods, single-molecule identification was achieved by AFM imaging (before and after the mechanical pulling), while desorption from the mica substrate was controlled by repeating the experiment on a double layered membrane, which produced similar results. The contour length of the force peaks was compared with the expected lengths of the loop and transmembrane regions of the protein. In this way, the α -helices were found to detach from the membrane two by two (except for the A α -helix) in the expected order, with forces ranging from 100 to 200 pN (at 0.04 nm ms^{-1}). The forces needed to simultaneously extract and unfold the A α -helix were found to be similar to those needed to either unfold other α -helices or dissociate them from the membrane once already unfolded (87, 99, and 105 pN respectively, at 0.04 nm ms^{-1}). It should be noted that the 99 pN value represents the force needed to unfold (and partially extract) an α -helix in a hydrophobic environment, not in an aqueous solution (Frederix et al. 2003 and references therein). Hence, the bacteriorhodopsin

system is a good example of the complexity of this type of experiment and it offers a unique opportunity to get a handle on the elusive structure of membrane proteins. We have not included this information in Table 8.1 as there is not conclusive evidence attributing these force values exclusively to the unfolding of an α -helix in the membrane. Indeed, we cannot rule out the contribution of unbinding interactions with lipids or other proteins in the membrane.

Recently, Kienberger et al. (2005) genetically engineered the bacteriorhodopsin system in order to estimate the force range of intermolecular vs. intramolecular interactions in an antibody–antigen system. They found that the intermolecular antibody/antigen unbinding force was significantly lower (126 vs. 204 pN at the same loading rate) than the force required to mechanically extract and unfold the helix pairs from the membrane.

8.6 Limitations and Perspectives

8.6.1 Technical Limitations

Using state-of-the-art instrumentation, force resolution in SMFS is limited only by the thermal fluctuations detected by the force sensor. Thus, the cantilever is a critical element in SMFS because its mechanical properties dictate the force sensitivity, response times, and spatial resolution. According to the so-called fluctuation–dissipation theorem, a reduction in the size of the force sensor should improve the signal-to-noise ratio (Gittes and Schmidt 1998). The signal-to-noise ratio of force measurements is independent of the stiffness of the cantilever but can be improved by reducing its dimensions, as has been verified experimentally. As a result, smaller but still soft cantilevers allow for higher sensitivity, faster response times, and a high spatial resolution (Viani et al. 1999). This new generation of cantilevers is a promising development permitting resolutions of 7–10 pN (e.g., Schwaiger et al. 2002). However, they are currently expensive, as they have to be custom-made using micromachining techniques. They also require a modified SMFS head, in order to avoid optical interference from some laser light spilling over the cantilever and reflecting off the substrate. At present, although SMFS experiments can readily resolve single amino acid length changes (Carrion-Vazquez et al. 1999b), they cannot provide detailed structural information yet. With the further development of smaller cantilevers, the secondary structure of proteins and single hydrogen bonds may be resolved in the future, although some theoretical considerations have been raised against this possibility (Thompson and Siggia 1995). If we extrapolate to the size limit, the cantilever should be replaced by a single elastic molecule to increase the precision of this method. A first step in this direction has been the development of an all-or-none force sensor based on a short DNA duplex (Albrecht et al. 2003). In the future we may also see the development of molecular force transducers with much better force resolution and capable of being used for *in vivo* applications (see later).

Table 8.2. Mechanical properties of a selection of protein-biomolecule pairs studied by SMFS

Interaction pair	F_u (pN)	v_r (nm ms^{-1})	X_{fp} (nm)	k_{off} (s^{-1})	Thermo-dynamic data	Attachment (T: tip; S: surface)	Single unbinding event criterion	Controls and experiment design	Remarkable aspects of the work	References
Avidin-biotin	160 ± 20	–	0.93^a	–	$\Delta H = 21.5 \text{ kcal mol}^{-1}$; $\Delta G = -20.4 \text{ kcal mol}^{-1}$	T: receptor via biotin-labeled BSA; S: agarose beads with ligands	Quantal behavior	Blocking with avidin, elastic agarose beads, different ligands/receptor, thermodynamical data, changes in pH	Pioneer work on interaction studies	Florin et al. (1994); Moy et al. (1994) ^b
HAS-antiHSA*	240 ± 48	0.2 ($r=54 \text{ nN s}^{-1}$)	6	6.7×10^{-4}	$K_D = 10 \text{ nM}$	T: Ab via PEG; S: Ag attached to mica via PEG	Spacer linkers	Spacer linker, blocking with free HSA, PEG spacer to control unspecific interactions	Spacer linkers, k_{off} directly measured	Hinterdorfer et al. (1996) ^c
Biotinylated BSA-IgG antibody*	60 ± 20	–	–	–	–	T and S to Ag and Ab via DSU	Quantal behavior	Blocking with free streptavidin and biotin, changes in the pH, BSA instead of Ag, (-) control antibody, switching tip/surface	One of the first Ag-Ab AFM studies	Dammer et al. (1996)
Ferritin-antiferritin IgG2a*	49 ± 10	–	–	–	–	T: covalently to ferritin; S: Ab via glutaraldehyde	Quantal behavior	Control antibody, polypeptopic antigen, blocking with ferritin	Using a polypeptopic antigen	Allen et al. (1997) ^d
Fluorescein-Ab scFv fragments*	50 ± 4	1	–	–	$\Delta G = -12.14 \text{ kcal mol}^{-1}$; $K_D = 0.75 \text{ nM}$	T: fluorescein via PEG; S: Ab bound to gold by Cys	Spacer linker, AFM imaging	Spacer linker, AFM imaging, blocking fluorescein, mutant scFv, thermodynamics data	AFM imaging, mutants	Ros et al. (1998) ^e
Proteoglycan of <i>Micrococcia proliferans</i> [†]	40 ± 15	–	–	–	–	Covalently bound to T and S	Quantal behavior	Changes in Ca^{2+} , Mg^{2+} as control, blocking with Ab, AFM and EM imaging	Pioneer AFM study of interaction forces	Dammer et al. (1995)
$\alpha_4\beta_3$ -GRGDSP peptide [†]	42 ± 4	1	–	–	–	T: peptides via PEG; S: $\alpha_4\beta_3$ expressing osteoclasts	Quantal behavior	Spacer linker, control peptides, control cells, changes in pH, blocking with peptides	First study using intact cells	Lehenkari and Horton (1999)

Table 8.2. Continued

Interaction pair	F_u (pN)	v_r (nm ms^{-1})	X_{fp} (nm)	k_{off} (s^{-1})	Thermodynamic data	Attachment (T: tip; S: surface)	Single unbinding event criterion	Controls and experiment design	Remarkable aspects of the work	References
VE-cadherin (homophilic interaction) [†]	40	0.8	0.59	1.8	$K_D = 1.15$ nM	T: VE-cadherin via PEG; S: VE-cadherin covalently via PEG	Unimodal Gaussian fit	Spacer linker, no calcium, blocking with Ab, changes in speed	Lateral interaction, multiple speeds	Baumgartner et al. (2000)
LFA-1-ICAM-1 [†]	20–320	0.1–15 ($r = 20$ –50,000 pN s^{-1})	0.018–0.21	4–57	–	T: T cells via Con A; S: soluble ICAM-1 adsorbed to mica surface	Unimodal Gaussian fit	Blocking with antibodies, changes in Mg^{2+} ; LFA-1 varieties	Multiple r ; two loading regimes	Zhang et al. (2002) ^f
P-selectin-PSGL-1 [†]	5–50	0.25	0.46	0.22	–	T: PSGL-1 via Abs; S: lipid bilayers with P-selectin	Comparison between monomeric and dimeric data	Monomeric/dimeric PSGL-1, Ab as ligand, blocking with Ab, changes in free Ca^{2+} , flow chamber	Catch bonds vs. slip bonds	Marshall et al. (2003)
GroEL-denatured pepsin [#]	48±20 (ATP); 44±20 (no ATP)	0.12	2.5 ±3.4 (ATP); 9.2±5.9 (no ATP)	–	–	T: pepsin via Sulfo-LC-SPDP; S: GroEL adsorbed to mica surface	"Compression-free" unimodal Gaussian fit	AFM imaging, blocking with pepsin, compression-free method, changes in ATP	Compression-free method	Sekiguchi et al. (2002, 2003)
Bacterio-rhodopsin-native purple membrane [§]	100–200	0.04	–	–	–	Nonspecific attachment (S: native purple membrane adsorbed to mica)	AFM imaging, comparison with unfolded protein length	AFM imaging, comparison of expected length with length of helices, changes in pH, retinal configuration changes	Integral membrane protein, combination of AFM imaging and force spectroscopy	Müller et al. (2002) ^h
Helix-pair ED	167±20	–	3.2 ^g	1.0×10 ^{-2g}	–	–	–	–	–	–
Helix-pair BC	99±16	–	8.6 ^g	3.4×10 ^{-5g}	–	–	–	–	–	–

Table 8.2. Continued

Interaction pair	F_u (pN)	v_r (nm ms^{-1})	χ_B (nm)	k_{off} (s^{-1})	Thermo-dynamic data	Attachment (T: tip; S: surface)	Single unbinding event criterion	Controls and experiment design	Remarkable aspects of the work	References
ExpG–target DNA ζ^y	50 –165	0.003– 6.08	7.5±1	$(1.2\pm 1)\times 10^{-3}$	–	T: DNA via PEG; S: ExpG attached to silanized mica	Unimodal Gaussian fit	Functionalized AFM tip and clean surface, blocking with free DNA, comparison with EMSA	First protein–DNA interaction AFM study, various r	Bartels et al. (2003)
LexA–target DNA ζ^y	Variable (depending on r)					T: DNA via CMA; Reducing S: TF via CMA	Reducing success rate	Spacer linker, fluorescence assays, CMA to control unspecific interactions, changes in spacer	Changes in r , changes in spacer length	Kühner et al. (2004)
LexA– recA		0.3–3	0.54±0.1	0.045±0.02	$K_D=2\text{ nM}$					
LexA–yebG		1–3	0.49±0.05	0.13±0.08	$K_D=20\text{ nM}$					

Interaction pairs are classified as avidin–biotin, antigen (Ag)–antibody (Ab) (*), adhesion proteins (t), cytoplasmatic proteins (#), protein–lipids (S), and protein–DNA (Ψ).

^aCalculated as $-\Delta H/F_u$

^bRelated articles, Izrailev et al. (1997) (steered molecular dynamics simulation of biotin–avidin), Wong et al. (1998), Merkel et al. (1999)

^cRelated article, Wielert-Badt et al. (2002)

^dRelated article, Allen et al. (1999) (discussion on the influence of epitope availability in AFM studies of the Ag–Ab interaction).

^eRelated article, Schwesinger et al. (2000) (the correlation between unfolding force and k_{off} is shown here studying antibody mutants and using different loading rates).

^fThis methodology has been also applied to study the interactions between fibronectin and $\alpha_5\beta_1$ integrin (Li et al. 2003) and P-selectin and its ligand (Hanley et al. 2003).

^gData from Janovjak et al. (2004)

^hRelated articles, Oesterhelt et al. (2000) (first published report on the subject), Janovjak et al. (2003) (temperature dependence of the process), Janovjak et al. (2004) (energy landscape)

The future use of SMFS in proteomics will require new developments to enhance the speed of analysis and sample throughput, so that parallel analysis of samples can be efficiently performed. In addition to the development of small cantilevers, there are other steps in this direction: the “micro SMFS” (based on a multibridged cantilever; Chu et al. 2000), the so-called tip arrays (Minne et al. 1998), and the “Millipede” device (Vettiger et al. 2002). Besides increasing the throughput, multiple cantilevers have the advantage over single cantilevers that reference cantilevers can be used to effectively remove thermal drifts, mechanical disturbances coming from noise (such as vibrations), and false positives.

Current technology does not permit SMFS to directly measure the forces involved in unfolding/folding inside cells, although it has already been used to measure the mechanical properties of the extracellular portions of interacting proteins located on the plasma membrane of living cells (Benoit and Gaub 2002). The ultimate goal of protein mechanics is to fully access the machinery of living cells. Several avenues are being explored to achieve this, including the use of magnetic beads or fluorophores. The recent development of a hybrid AFM/total internal reflection fluorescence microscopy apparatus (Sarkar et al. 2004) and the characterization of the mechanical properties of GFP (Dietz and Rief 2004) are the first steps towards the development of a molecular force sensor based on fluorescence that could be fused to the protein of interest by genetic engineering. Many alternative approaches are also being explored at present, as the biosensor field is a very active area of research nowadays.

Two additional constraints limit SMFS potential: the lack of knowledge of the initial state of the adsorbed sample and the fact that there is no real choice of the specific system (molecules or other entities to be analyzed). As we have mentioned, SMFS is typically a “blind” fishing expedition where single-molecule pickup is random. We cannot choose generally what we get in each event and the options are many since the immobilized/adsorbed protein sample may consist of a heterogeneous mixture of species including native, partially denatured, completely denatured, and misfolded single molecules, as well as aggregates and native supramolecular complexes. Therefore, AFM imaging of the sample prior to SMFS would be of paramount importance to “have a look” before and after performing the pulling experiment. In this way, specific molecules or other entities could be chosen and pulled at specific points. Moreover, this would also enable specific immobilization systems to be developed, to better preserve protein conformation (as mentioned, physisorption can often cause partial or total denaturation of the protein). However, with the exception of a few amenable systems such as bacteriorhodopsin, which forms rigid 2D crystals in the purple membrane (Sect. 8.5.2), AFM imaging/SMFS analyses in parallel (this method is also known as “simultaneous,” *sensu lato*) are rarely seen (Meadows et al. 2003). The main reason for this is that proteins are soft materials that are “hard to image” using classical imaging modes (i.e., contact and “tapping”). In this sense, the development of the so-called jumping mode, as it is milder on the sample, has generated some expectation for this particular application (de Pablo et al. 1998). Furthermore, a compromise in the experimental conditions must be established, since the requirements of each of the two AFM configurations are typically different: relatively thick layers and mild attachment to substrate (typically gold) in

SMFS vs. monolayers and strong attachment to substrate (usually mica) in AFM imaging.

It was the advent of polyprotein production that moved the field of intramolecular protein mechanics forwards by providing unequivocal single-molecule fingerprinting. Development of a similar tool (of universal applicability to proteins) to unmistakably identify single unbinding events would be desirable to provide a further impulse to the field of intermolecular protein mechanics. Still, since the methods available for constructing polyproteins are very laborious, a means to simplify this process would also be welcomed. Thus, the lengthy construction of the concatemer, the main bottleneck in the most widely used of those methods (i.e., the recombinant polyprotein), should be shortened. Implementing a single-step method to translate the concatemer *in vitro* would also be useful (expression *in vivo*, in *Escherichia coli*, remains the norm) as only small amounts are typically needed to complete the mechanical analysis of a polyprotein (a few tens of micrograms of protein). This method would give us more control of the expression conditions, which may help to overcome possible problems related to toxicity and solubility of the protein. Recent improvements of *in vitro* expression systems, both in terms of yield and in the length of the protein that can be produced, offer some promise for the future simplification of this step.

8.6.2

Methodological Limitations

New developments are required not only in SMFS instrumentation but also in the data analysis department. Recent efforts in this direction have led to the proposal of a detailed protocol for data analysis that includes a set of criteria for choosing the specific experimental conditions, and that aids in the selection of recordings in SMFS experiments (reviewed by Rounsevell et al. 2004).

Mechanical unbinding experiments on receptor–ligand systems performed with the biomembrane force probe have provided compelling evidence that the energy landscape is not homogeneous. Indeed, the force varies in a nonlinear fashion with the pulling speed over a wide range of speeds. This variation indicates that extrapolating to low pulling speeds, which may be physiologically significant in some cases, may not adequately describe the response of the system in these conditions (Merkel et al. 1999). SMFS is an extremely sensitive technique that can resolve differences in the contour length of a protein in the subnanometer range and forces above a couple of tens of piconewtons. However, the range of pulling speeds of a typical SMFS apparatus is limited to a range of 3 orders of magnitude because of current experimental limitations (thermal drift and instrumental stability), and these speeds may be significantly faster than a number of physiological pulling speeds. In order to cover the range of low loading rates, a new method has recently been proposed that extends the dynamic range by employing a model based on protein engineering, i.e., analyzing the behavior of conservative mutations that shift the unfolding rate of the wild type (Sect.8.3.3). In the case of I27 this approach has uncovered a complex energy landscape that illustrates the possible dangers of directly extrapolating from experimental to physiological conditions (Sect. 8.4.1, Fig. 8.9, panel c; Williams et al.

2003). Alternatively, it would be better to directly use complementary nanomanipulation techniques with a lower speed range (Sect. 8.2.5). However, this may require the development of a method to adapt the use of polypeptides to techniques such as optical tweezers (e.g., using either long polypeptides or a long spacer or “handle”) or the biomembrane force probe.

In these experiments, the limited spatial and temporal resolution implies that only the maximum of the force profile can typically be accessed (e.g., by measuring a rupture or unfolding force). Thus, most details of the reaction pathways currently remain unresolved. A theory has been developed to overcome this limitation, without the need to resort to an increase in technical resolution. This theory enables force profiles to be reconstructed from force spectra obtained through the measurements obtained at varying loading rates, and by combining the complementary information obtained with stiff (AFM) and soft springs (e.g., optical tweezers and biomembrane force probe; Heymann and Grubmüller 2000). There is therefore a powerful synergy between experimental data and modeling/simulation and this is expected to be central for the future development of the field.

The available data are both incomplete and somewhat heterogeneous owing to the different experimental conditions used (e.g., pulling speeds). In the future, and similar to what has been recently proposed in the field of chemical protein folding (Maxwell et al. 2005), it would be desirable to establish a standard set of experimental conditions to perform SMFS such that data from the protein database would be strictly comparable (e.g., pulling speed, calibration method, buffer, physiological pH, room temperature).

8.6.3 SMFS and Protein Mechanics *in Vivo*

How well do SMFS experiments reflect protein mechanics *in vivo*? For the case of extension machines like the sarcomere, SMFS seems to adequately mimic the pulling geometry since proteins are pulled apart from both ends of the polypeptide chain. This may also be the case for other cytoskeletal machineries, the adhesion machinery, and some mechanosensitive ion channels. Furthermore, it has been suggested that some chaperonins may pull, in a similar way, their protein substrates apart prior to their refolding (Shtilerman et al. 1999). However, the case of other unfoldase machines is not so clear-cut. The accepted model for the case of protein translocases (from the mitochondrion, chloroplast, and endoplasmic reticulum) and compartmental proteases such as the proteasome is also a mechanical one. Nevertheless, rather than a linear pulling geometry with two attachment points, the evidence here favors a different geometry that involves a single attachment point from which the pulling would be done by threading the protein towards the entrance of a narrow channel present in these nanomachines (Sect. 8.2.3, Fig. 8.2). This model is mainly based on the fact that the susceptibility of substrate proteins to be unfolded by these nanomachines (*in vivo*) correlates more closely with the mechanical stability obtained by mechanical unfolding (using SMFS) than with thermodynamic or kinetic stability (measured *in vitro* by bulk chemical or heat denaturation). In the case of compartmental proteases, the AAA+ ATPase motor involved in the pulling

process seems to unfold the structure adjacent to the degradation tag by trapping local unfolding fluctuations. Global unfolding then occurs immediately, driven by the cooperativity of the protein unfolding process (Matouschek 2003; Prakash and Matouschek 2004; Sauer et al. 2004). In this “local stability” model, like in SMFS findings, the structure and pulling geometry at the attachment point (i.e., local stability) are more important than their global counterparts. This highlights the importance of the existence of Achilles heels in proteins. As a result, mechanical unfoldases may have evolved specific pulling mechanisms to take advantage of the presence of weak spots in their protein substrates in order to unfold them more economically. For instance, since polyubiquitins may act as handles for mechanical pulling (Carrion-Vazquez et al. 2003; Pickart and Cohen 2004), it is tempting to suggest that proteasomal substrates may have placed lysine residues for ubiquitination at specific locations in order to minimize the forces required for their unfolding. It is also conceivable that the cell’s metabolism may on occasion generate some protein structures (misfolded proteins or aggregates) with very high mechanical stability that may jam the proteasomal motor, leading to pathological situations. It therefore seems critical in the future to develop experimental systems to mechanically analyze unfoldase protein substrates in a way that closely mimics the physiological configuration.

In the case of mechanical proteins, the mechanical design of their structures probably reflects precise solutions to adapt their function to specific mechanical challenges (e.g., mechanical elasticity or resistance) organisms have encountered through evolution. In fact, as a result of eons of evolution, most modern bionanomachines are probably perfect or near-perfect solutions to the variety of challenges organisms have encountered (some of which must certainly have been mechanical). These bionanomachines are truly amazing devices from the engineering point of view. For instance, the ATP synthase achieves near 100% efficiency (Wang and Oster 1998), while artificial machines do not surpass 40%. Evolution therefore remains the best engineer and a constant source of ideas for the development of biologically inspired materials.

8.7 Conclusions

Biological nanomachines carry out many processes in which mechanical forces are used to trigger conformational transitions. AFM, in its SMFS mode, combined with nanoscale physics, computational techniques, and protein engineering, has established a new field of investigation (i.e., protein mechanics), whose aim is to directly study the mechanical properties of intramolecular (folding/unfolding) and intermolecular (binding/unbinding) protein interactions.

Single-molecule mechanical studies have already produced many remarkable new insights into protein folding and function although only a few protein structures have been investigated so far and the available data are both incomplete and heterogeneous owing to the lack of a standard for the experimental conditions. Also, most of these studies have been necessarily phenomenological, i.e., limited to the characterization of some of the main mechanical properties, but lacking a detailed mechanistic description of the process.

The first findings have revealed that proteins show a broad range of responses to mechanical stress. It has been found that mechanical proteins tend to be mechanically more stable than nonmechanical ones. β -strand proteins are usually mechanically more stable than α -helical ones, and β -strand proteins with a shear mechanical topology were found to be mechanically more stable than zipper β -strand proteins. Furthermore, patches of hydrogen bonds in the protein backbone have often been identified as the breakpoint of mechanical resistance.

The identification of the force-bearing components of proteins has begun to yield information regarding the molecular determinants underlying their mechanical behavior. To date, SMFS experiments have demonstrated that the mechanical stability of a protein is not correlated with its thermodynamic stability and that it can be modulated by ligands. It has also been revealed that chemical and mechanical unfolding follow different pathways and have different unfolding barriers. Moreover, it has been established that the mechanical stability and the mechanical unfolding pathway depend on the pulling geometry and the application points of the force. Although the molecular basis of the mechanical resistance of proteins remains in most cases unclear, several determinants have been identified so far: amino acid sequence, topology, unloaded unfolding rate constant, and pulling geometry.

In less than a decade since the first SMFS experiments were carried out on proteins, this technique has identified a variety of components involved in a number of different mechanical functions in the protein machinery. Elastic proteins and elastic protein components (elastin, titin PEVK, and N2B) seem to function mainly as pure entropic springs (with entropic elasticity). In contrast, mechanical modular proteins (like fibronectin, tenascin, or titin Ig domains) seem to rely on module unfolding as a “mechanical buffer” (enthalpic elasticity), in this way behaving as perfect mechanical shock absorbers. Thus, the mostly modular cell adhesion proteins have their elastic elements (modules) connected in series through intermolecular bonds (Fig. 8.2b). When these proteins are subjected to mechanical tension, often present at the cell–cell interface, it has been proposed that this organization helps to maintain the interactions of these adhesive molecules over larger extensions, and increases their lifetime when compared with that of a rigid element (Oberhauser et al. 1998; Evans and Ritchie 1999).

Single-molecule mechanical techniques are still in their infancy, but once this first generation of techniques evolves, we can expect them to provide us with more and more fundamental information on the structure and function of proteins, and to become an indispensable tool in understanding how these molecules fold and work. With the newfound capacity to manipulate and look at the “secret life” of single molecules, we should be prepared for many surprises from the mechanochemistry of proteins. Indeed, we are entering a new and exciting time in biology which, in combination with the knowledge generated in this proteomic era, is likely to help us understand how proteins work in real time.

Acknowledgements and Notes added in Proof. We are in debt to J. Clarke, H. Li, F. Ritort, and S. García-Manyes for their critical reading of the manuscript. We thank J.M. Fernandez, J. Clarke, and F. Ritort for kindly sending preprints of their current works. We also thank the members of the M.C.-V. laboratory for helpful discussions and comments on the manuscript. This work was supported by grants from the former Spanish Ministry of Science and Technology (BIO2003-08004), the Consejería de Educación of the Madrid Community (GR/SAL/0836/2004) and the institutional RED CIEN (G03/06) to M.C.-V. and from the National Institutes of Health (DK067443) and the John Sealy Memorial Endowment Fund for Biomedical Research to A.F.O. We apologize to many of our colleagues for omitting or only indirectly citing their work. Owing to the limitations of space, some important contributions to the main points of our review have not been cited and some review articles, rather than the primary sources, have sometimes been preferred as references. While this chapter was in proof several papers have been published reporting important facts for this field. Thus the surprising Hookean behaviour of ankyrin (Lee et al. 2006) as well as the mechanical refolding of ankyrin repeats (Lee et al. 2006) and RNase H (Cecconi et al. 2005); the later study has been done with optical tweezers and found RNase H to be a three-state folding protein; also a recent publication has reported a higher mechanical stability for DHFR (around 130 pN), although at a higher loading rate than previously used (Wilcox et al. 2005); this study has also brought additional evidence supporting the dependence of mitochondrial unfolding on the local structure adjacent to the targeting sequence, similar to the findings from AFM unfolding. Finally, another study suggests that the mechanism for mitochondrial unfolding may be somehow different from those involved in AFM and chemical techniques (Sato et al. 2005).

References

- Ainavarapu SRK, Li L, Badilla CL, Fernandez JM (2005) Ligand binding modulates the mechanical stability of dihydrofolate reductase. *Biophys J* 89:3337–3344
- Alberts B (1998) The cell as a collection of protein machines: preparing the next generation of molecular biologists. *Cell* 92:291–294
- Alberts B, Johnson A, Lewis J, Raff M, Roberts K, Walter P (2002) The molecular biology of the cell, 4th edn. Garland Science, New York, p 1103
- Albrecht C, Blank K, Lalic-Multhaler M, Hirler S, Mai T, Gilbert I, Schiffmann S, Bayer T, Clausen-Schaumann H, Gaub HE (2003) DNA: A programmable force sensor. *Science* 301:367–370
- Allen S, Chen X, Davies J, Davies MC, Dawkes AC, Edwards JC, Roberts CJ, Sefton J, Tendler SJ, Williams PM (1997) Detection of antigen-antibody binding events with the atomic force microscope. *Biochemistry* 36:7457–7463
- Allen S, Davies J, Davies MC, Dawkes AC, Roberts CJ, Tendler SJ, Williams PM (1999) The influence of epitope availability on atomic-force microscope studies of antigen-antibody interactions. *Biochem J* 341:173–178
- Anfinsen CB, Haber E, Sela M, White FH (1961) The kinetics of formation of native ribonuclease during oxidation of the reduced polypeptide chain. *Proc Natl Acad Sci USA* 47:1309–1314

- Appleman JR, Prendergast N, Delcamp TJ, Freisheim JH, Blakley RL (1988) Kinetics of the formation and isomerization of methotrexate complexes of recombinant human dihydrofolate reductase. *J Biol Chem* 263:10304–10313
- Ashkin A (1970) Acceleration and trapping of particles by radiation pressure. *Phys Rev Lett* 24:156–159
- Bartels FW, Baumgarth B, Anselmetti D, Ros R, Becker A (2003) Specific binding of the regulatory protein ExpG to promoter regions of the galactoglucan biosynthesis gene cluster of *Sinorhizobium meliloti* – a combined molecular biology and force spectroscopy investigation. *J Struct Biol* 143:145–152
- Baumgartner W, Hinterdorfer P, Ness W, Raab A, Vestweber D, Schindler H, Drenckhahn D (2000) Cadherin interaction probed by atomic force microscopy. *Proc Natl Acad Sci USA* 97:4005–4010
- Bell GI (1978) Models for the specific adhesion of cells to cells. *Science* 200:618–627
- Benoit M, Gaub HE (2002) Measuring cell adhesion forces with the atomic force microscope at the molecular level. *Cells Tissues Organs* 172:174–189
- Best RB, Clarke J (2002) What can atomic force microscopy tell us about protein folding? *Chem Commun* 7:183–192
- Best RB, Hummer G (2005) Comment on “Force-clamp spectroscopy monitors the folding trajectory of a single protein”. *Science* 308:498
- Best RB, Li B, Steward A, Daggett V, Clarke J (2001) Can non-mechanical proteins withstand force? Stretching barnase by atomic force microscopy and molecular dynamics simulation. *Biophys J* 81:2344–2356
- Best RB, Fowler SB, Toca-Herrera JL, Clarke J (2002) A simple method for probing the mechanical unfolding pathway of proteins in detail. *Proc Natl Acad Sci USA* 99:12143–12148
- Best RB, Brockwell DJ, Toca-Herrera JL, Blake AW, Smith DA, Radford SE, Clarke J (2003a) Force mode atomic force microscopy as a tool for protein folding studies. *Anal Chim Acta* 479:87–105
- Best RB, Fowler SB, Toca-Herrera JL, Steward A, Paci E, Clarke J (2003b) Mechanical unfolding of a titin Ig domain: structure of transition state revealed by combining atomic force microscopy, protein engineering and molecular dynamics simulations. *J Mol Biol* 330:867–877
- Bhasin N, Law R, Liao G, Safer D, Ellmer J, Discher BM, Sweeney HL, Discher DE (2005) Molecular extensibility of mini-dystrophins and a dystrophin rod construct. *J Mol Biol* 352:795–806
- Bieri O, Kiefhaber T (2000) Kinetic models in protein folding. In: Pain R (ed) *Frontiers in molecular biology: mechanisms of protein folding*, 2nd edn. Oxford University Press, Oxford, pp 34–64
- Binnig G, Quate CF, Gerber C (1986) Atomic force microscope. *Phys Rev Lett* 56:930–933
- Brockwell DJ, Beddard GS, Clarkson J, Zinober RC, Blake AW, Trinick J, Olmsted PD, Smith DA, Radford SE (2002) The effect of core destabilization on the mechanical resistance of I27. *Biophys J* 83:458–472
- Brockwell DJ, Paci E, Zinober RC, Beddard GS, Olmsted PD, Smith DA, Perham RN, Radford SE (2003) Pulling geometry defines the mechanical resistance of a beta-sheet protein. *Nat Struct Biol* 10:731–737
- Brockwell DJ, Beddard GS, Paci E, West DK, Olmsted PD, Smith DA, Radford SE (2005) Mechanically unfolding the small topologically simple protein L. *Biophys J* 89:506–519
- Brujic J, Fernandez JM (2005) Response to Comment on “Force-clamp spectroscopy monitors the folding trajectory of a single protein”. *Science* 308:498c

- Brzeska H, Venyaminov SV, Grabarek Z, Drabikowski W (1983) Comparative studies on thermostability of calmodulin, skeletal muscle troponin C and their triptic fragments. *FEBS Lett* 153:169–173
- Buijs J, Norde W, Lichtenbelt JWT (1996) Changes in the secondary structure of adsorbed IgG and F(ab')₂ studied by FTIR spectroscopy. *Langmuir* 12:1605–1613
- Bullard B, Garcia T, Benes V, Leake MC, Linke WA, Oberhauser AF (2006) The molecular elasticity of the insect flight muscle proteins projectin and kettin. *Proc Natl Acad Sci USA* 103:4451–4456
- Burnham NA, Colton RJ (1989) Measuring the nanomechanical properties and surface forces of materials using an atomic force microscope. *J Sci Technol A* 7:2906–2913
- Bustamante C, Marko JF, Siggia ED, Smith S (1994) Entropic elasticity of lambda-phage DNA. *Science* 265:1599–1600.
- Bustamante C, Rivetti C, Keller DJ (1997) Scanning force microscopy under aqueous solutions. *Curr Opin Str Biol* 7:709–716
- Bustamante C, Macosko JC, Wuite GJ (2000) Grabbing the cat by the tail: manipulating molecules one by one. *Nat Rev Mol Cell Biol* 1:130–136
- Bustamante C, Keller D, Oster G (2001) The physics of molecular motors. *Acc Chem Res* 34:412–420
- Bustamante C, Chemla YR, Forde NR, Izhaky D (2004) Mechanical processes in biochemistry. *Annu Rev Biochem* 73:705–748
- Carrion-Vazquez M, Oberhauser AF, Fowler SB, Marszalek PE, Broedel SE, Clarke J, Fernandez JM (1999a) Mechanical and chemical unfolding of a single protein: a comparison. *Proc Natl Acad Sci USA* 96:3694–3699
- Carrion-Vazquez M, Marszalek PE, Oberhauser AF, Fernandez JM (1999b) Atomic force microscopy captures length phenotypes in single proteins. *Proc Natl Acad Sci USA* 96:11288–11292
- Carrion-Vazquez M, Oberhauser AF, Fisher TE, Marszalek PE, Li H, Fernandez JM (2000) Mechanical design of proteins studied by single-molecule force spectroscopy and protein engineering. *Prog Biophys Mol Biol* 74:63–91
- Carrion-Vazquez M, Li H, Lu H, Marszalek PE, Oberhauser AF, Fernandez JM (2003) The mechanical stability of ubiquitin is linkage dependent. *Nat Struct Biol* 10:738–743
- Cecconi C, Shank EA, Bustamante C, Marqusee S (2005) Direct observation of the three-state folding of a single protein molecule. *Science* 309:2057–2060
- Chen BL, Baase WA, Nicholson H, Schellmann JA (1992) Folding kinetics of T4 lysozyme and nine mutants at 12 degrees. *Biochemistry* 31:1464–1476
- Chu J, Wang Z, Macda R, Kataoka K, Itoh T, Suga T (2000) Novel multibrIDGE-structures piezo-electric microdevice for scanning force microscopy. *J Vac Sci Technol B* 18:3604–3607
- Clarke J, Hamill SJ, Johnson CM (1997) Folding and stability of a fibronectin type III domain of human tenascin. *J Mol Biol* 270:771–778
- Cleveland JP, Manne S, Bocek D, Hansma PK (1993) A nondestructive method for determining the spring constant of cantilevers for scanning force microscopy. *Rev Sci Instrum* 64:403–405
- Collin D, Ritort F, Jarzynski C, Smith SB, Tinoco I Jr, Bustamante C (2005) Verification of the Crooks fluctuation theorem and recovery of RNA folding free energies. *Nature* 437:231–234
- Cota E, Clarke J (2000) Folding of β -sandwich proteins: three-state transition of a fibronectin type III module. *Protein Sci* 9:112–120

- Craig D, Krammer A, Schulten K, Vogel V (2001) Comparison of the early stages of forced unfolding for fibronectin type III modules. *Proc Natl Acad Sci USA* 98:5590–5595
- Craig D, Gao M, Schulten K, Vogel V (2004) Tuning the mechanical stability of fibronectin type III modules through sequence variations. *Structure* 12:21–30
- Dammer U, Popescu O, Wagner P, Anselmetti D, Güntherodt HJ, Misevic GN (1995) Binding strength between cell adhesion proteoglycans measured by atomic force microscopy. *Science* 267:1173–1175
- Dammer U, Hegner M, Anselmetti D, Wagner P, Dreier M, Huber W, Güntherodt HJ (1996) Specific antigen/antibody interactions measured by force microscopy. *Biophys J* 70:2437–2441
- De Pablo PJ, Colchero J, Gomez-Herrero J, Baro AM (1998) Jumping mode scanning force microscopy. *Appl Phys Lett* 73:3300–3302
- DeSilva TM, Harper SL, Kotula L, Hensley P, Curtis PJ, Otvos L Jr, Speicher DW (1997) Physical properties of a single-motif erythrocyte spectrin peptide: a highly stable independently folding unit. *Biochemistry* 36:3991–3997
- Dietz H, Rief M (2004) Exploring the energy landscape of GFP by single-molecule mechanical experiments. *Proc Natl Acad Sci USA* 101:16192–16197
- Evans E, Ritchie K (1997) Dynamic strength of molecular adhesion bonds. *Biophys J* 72:1541–1555
- Evans E, Ritchie K (1999) Strength of a weak bond connecting flexible polymer chains. *Biophys J* 76:2439–2447
- Evans E, Ritchie K, Merkel R (1995) Sensitive force technique to probe molecular adhesion and structural linkages at biological interfaces. *Biophys J* 68:2580–2587
- Fernandez JM, Li H (2004) Force-clamp spectroscopy monitors the folding trajectory of a single protein. *Science* 303:1674–1678
- Fersht AR (1999) *Structure and mechanism in protein science: a guide to enzyme catalysis and protein folding*, 2nd edn. Freeman, New York
- Fersht AR, Matouschek A, Serrano L (1992) The folding of an enzyme. I. Theory of protein engineering analysis of stability and pathway of protein folding. *J Mol Biol* 224:771–782
- Feynman RP (1960) There's plenty of room at the bottom. An invitation to enter a new field of physics. *Eng Sci* 23:22–36
- Finley D, Varshavsky A (1985) The ubiquitin system: functions and mechanisms. *Trends Biochem Sci* 10:343–346
- Fisher TE, Carrion-Vazquez M, Oberhauser AF, Li H, Marszalek PE, Fernandez JM (2000) Single molecular force spectroscopy of modular proteins in the nervous system. *Neuron* 27:435–446
- Florin EL, Moy VT, Gaub HE (1994) Adhesion forces between individual ligand-receptor pairs. *Science* 264:415–417
- Florin EL, Rief M, Lehmann H, Ludwig M, Dornmair C, Moy VT, Gaub HE (1995) Sensing specific molecular interactions with the atomic force microscope. *Biosens Bioelectron* 10:895–901
- Forman JR, Qamar S, Paci E, Sandford RN, Clarke J (2005) The remarkable strength of polycystin-1 supports a direct role in mechanotransduction. *J Mol Biol* 349:861–871
- Fowler SB, Best RB, Toca-Herrera JL, Rutherford TJ, Steward A, Paci E, Karplus M, Clarke J (2002) Mechanical unfolding of a titin Ig domain: structure of unfolding intermediate revealed by combining AFM, molecular dynamics simulations, NMR and protein engineering. *J Mol Biol* 322:841–849

- Frederix PL, Akiyama T, Staufer U, Gerber Ch, Fotiadis D, Müller DJ, Engel A (2003) Atomic force bio-analytics. *Curr Opin Chem Biol* 7:641–647
- Freiburg A, Trombitas K, Hell W, Cazorla O, Fougères F, Centner T, Kolmerer B, Witt C, Beckmann JS, Gregorio CC, Granzier H, Labeit S (2000) Series of exon-skipping events in the elastic spring region of titin as the structural basis for myofibrillar elastic diversity. *Circ Res* 86:1114–1121
- Furuike S, Ito T, Yamazaki M (2001) Mechanical unfolding of single filamin A (ABP-280) molecules detected by atomic force microscopy. *FEBS Lett* 498:72–75
- Gao M, Craig D, Vogel V, Schulten K (2002) Identifying unfolding intermediates of FN-III 10 by steered molecular dynamics. *J Mol Biol* 323:939–950
- Gao M, Craig D, Lequin O, Campbell ID, Vogel V, Schulten K (2003) Structure and functional significance of mechanically unfolded fibronectin type III intermediates. *Proc Natl Acad Sci USA* 100:14784–14789
- Gillespie PG, Walker RG (2001) Molecular basis of mechanosensory transduction. *Nature* 413:194–202
- Gittes F, Schmidt CF (1998) Thermal noise limitations on micromechanical experiments. *Eur Biophys J* 27:75–81
- Hanley W, McCarty O, Jadhav S, Tseng Y, Wirtz D, Konstantopoulos K (2003) Single molecule characterization of P-selectin/ligand binding. *J Biol Chem* 278:10556–10561
- Heymann B, Grubmüller H (2000) Dynamic force spectroscopy of molecular adhesion bonds. *Phys Rev Lett* 84:6126–6129
- Hinterdorfer P (2002) Molecular recognition studies using the atomic force microscope. *Methods Cell Biol* 68:115–139
- Hinterdorfer P, Baumgartner W, Gruber HJ, Schilcher K, Schindler H (1996) Detection and localization of individual antibody-antigen recognition events by atomic force microscopy. *Proc Natl Acad Sci USA* 93:3477–3481
- Hlady V, Buijs J (1998) Local and global optical spectroscopic probes of adsorbed proteins. In: Malmsten M (ed) *Biopolymers at interfaces*. Dekker, New York, pp 181–220
- Howard J (2001) *Mechanics of motor proteins and the cytoskeleton*. Sinauer Associates, Sunderland, MA
- Hutter JL, Bechhoffer J (1993) Calibration of atomic-force microscope tips. *Rev Sci Instrum* 64:1868–1873
- Huxley H, Hanson J (1954) Changes in the cross-striations of muscle during contraction and stretch and their structural interpretation. *Nature* 173:973–976
- Improta S, Politou AS, Pastore A (1996) Immunoglobulin-like modules from titin I-band: extensible components of muscle elasticity. *Structure* 4:323–337
- Isralewitz B, Gao M, Schulten K (2001) Steered molecular dynamics and mechanical functions of proteins. *Curr Opin Struct Biol* 11:224–230
- Izrailev S, Stepaniants S, Balsera M, Oono Y, Schulten K (1997) Molecular dynamics study of unbinding of the avidin-biotin complex. *Biophys J* 72:1568–1581
- Janovjak H, Kessler M, Oesterhelt D, Gaub H, Müller DJ (2003) Unfolding pathways of native bacteriorhodopsin depend on temperature. *EMBO J* 22:5220–5229
- Janovjak H, Struckmeier J, Hubain M, Kedrov A, Kessler M, Müller DJ (2004) Probing the energy landscape of the membrane protein bacteriorhodopsin. *Structure* 12:871–879
- Janovjak H, Kedrov A, Cisneros DA, Sapra KT, Struckmeier J, Müller DJ (2006) Imaging and detecting molecular interactions of single transmembrane proteins. *Neurobiol Aging* 27:546–561

- Junker JP, Hell K, Schlierf M, Newpert W, Rief M (2005) Influence of substrate binding on the mechanical stability of mouse dihydrofolate reductase. *Biophys J* 89:L46-48
- Khorasanizadeh S, Peters ID, Butt TR, Roder H (1993) Folding and stability of a tryptophan-containing mutant of ubiquitin. *Biochemistry* 32:7054-7063
- Kienberger F, Kada G, Mueller H, Hinterdorfer P (2005) Single molecule studies of antibody-antigen interaction strength versus intra-molecular antigen stability. *J Mol Biol* 347:597-606
- Klimov DK, Thirumalai D (2000) Native topology determines force-induced unfolding pathways in globular proteins. *Proc Natl Acad Sci USA* 97:7524-7529
- Koradi R, Billeter M, Wuthrich K (1996) MOLMOL: a program for display and analysis of macromolecular structures. *J Mol Graph* 14:51-55
- Krammer A, Lu H, Isralewitz B, Schulten K, Vogel V (1999) Forced unfolding of the fibronectin type III module reveals a tensile molecular recognition switch. *Proc Natl Acad Sci USA* 96:1351-1356
- Kühner F, Costa LT, Bisch PM, Thalhammer S, Heckl WM, Gaub HE (2004) LexA-DNA bond strength by single molecule force spectroscopy. *Biophys J* 87:2683-2690
- Labeit S, Kolmerer B (1995) Titins: giant proteins in charge of muscle ultrastructure and elasticity. *Science* 270:293-296
- Lee G, Abdi K, Jiang Y, Michaely P, Bennett V, Marszalek PE (2006), Nanospring behaviour of anyrin repeats. *Nature* 440:246-249
- Lee GU, Chrisley LA, Colton RJ (1994) Direct measurement of the forces between complementary strands of DNA. *Science* 266:771-773
- Lehenkari PP, Horton MA (1999) Single integrin molecule adhesion forces in intact cells measured by atomic force microscopy. *Biochem Biophys Res Commun* 259:645-650
- Lenne PF, Raae AJ, Altmann SM, Saraste M, Horber JK (2000) States and transitions during forced unfolding of a single spectrin repeat. *FEBS Lett* 476:124-128
- Levinthal C (1969) How to fold graciously. In: DeBrunner P, Tsigris JCM, Münck E (eds) Mössbauer spectroscopy in biological systems. Proceedings of a meeting held at Allerton House, Monticello, IL. University of Illinois Press, Champaign, IL, p 22
- Leuba S, Zlatanova J (eds) (2001) *Biology at the single-molecule level*. Pergamon, Amsterdam
- Li F, Redick SD, Erickson HP, Moy VT (2003) Force measurements of the $\alpha 5 \beta 1$ integrin-fibronectin interaction. *Biophys J* 84:1252-1262
- Li H, Fernandez JM (2003) Mechanical design of the first proximal Ig domain of human cardiac titin revealed by single molecule force spectroscopy. *J Mol Biol* 334:75-86
- Li H, Carrion-Vazquez M, Oberhauser AF, Marszalek PE, Fernandez JM (2000a) Point mutations alter the mechanical stability of immunoglobulin modules. *Nat Struct Biol* 7:1117-1120
- Li H, Oberhauser AF, Fowler SB, Clarke J, Fernandez JM (2000b) Atomic force microscopy reveals the mechanical design of a modular protein. *Proc Natl Acad Sci USA* 97:6527-6531
- Li H, Oberhauser AF, Redick SD, Carrion-Vazquez M, Erickson HP, Fernandez JM (2001) Multiple conformations of PEVK proteins detected by single-molecule techniques. *Proc Natl Acad Sci USA* 98:10682-10686
- Li H, Linke WA, Oberhauser AF, Carrion-Vazquez M, Kerkvliet JG, Lu H, Marszalek PE, Fernandez JM (2002) Reverse engineering of the giant muscle protein titin. *Nature* 418:998-1002
- Li L, Huang HH, Badilla CL, Fernandez JM (2005) Mechanical unfolding intermediates observed by single-molecule force spectroscopy in a fibronectin type III module. *J Mol Biol* 345:817-826

- Li PC, Makarov DE (2003) Theoretical studies of the mechanical unfolding of the muscle protein titin: bridging the time-scale gap between simulation and experiment. *J Chem Phys* 119:9260–9268
- Linke WA, Kulke M, Li HB, Fujita-Becker S, Neagoe C, Manstein DJ, Gautel M, Fernandez JM (2002) PEVK domain of titin: an entropic spring with actin-binding properties. *J Struct Biol* 137:194–205
- Liphardt J, Dumont S, Smith SB, Tinoco IJr, Bustamante C (2002) Equilibrium information from nonequilibrium measurements in an experimental test of Jarzynski's equality. *Science* 296:1832–1853
- Litvinovich SV, Ingham KC (1995) Interactions between type III domains in the 110kDa cell-binding fragment of fibronectin. *J Mol Biol* 248:611–626
- Lu H, Schulten K (1999) Steered molecular dynamics simulations of force-induced protein domain unfolding. *Proteins* 35:453–463
- Lu H, Schulten K (2000) The key event in force-induced unfolding of titin's immunoglobulin domains. *Biophys J* 79:51–65
- Lu H, Isralewitz B, Krammer A, Vogel V, Schulten K (1998) Unfolding of titin immunoglobulin domains by steered molecular dynamics simulation. *Biophys J* 75:662–671
- Maier B, Potter L, So M, Long CD, Seifert HS, Sheetz MP (2002) Single pilus motor forces exceed 100 pN. *Proc Natl Acad Sci USA* 99:16012–16017
- Marko JF, Siggia ED (1995) Stretching DNA. *Macromolecules* 28:8759–8770
- Marshall BT, Long M, Piper JW, Yago T, McEver RP, Zhu C (2003) Direct observation of catch bonds involving cell-adhesion molecules. *Nature* 423:190–193
- Marszalek PE, Oberhauser AF, Pang YP, Fernandez JM (1998) Polysaccharide elasticity governed by chair-boat transitions of the glucopyranose ring. *Nature* 396:661–664
- Marszalek PE, Pang YP, Li H, El Yazal, J, Oberhauser AF, Fernandez JM (1999a) Atomic levers control pyranose ring conformations. *Proc Natl Acad Sci USA* 96:77894–77898
- Marszalek PE, Lu H, Li H, Carrion-Vazquez M, Oberhauser AF, Schulten K, Fernandez JM (1999b) Mechanical unfolding intermediates in titin modules. *Nature* 402:100–103
- Maruyama K, Kimura S, Ohashi K, Kuwano Y (1981) Connectin, an elastic protein of muscle. Identification of "titin" with connectin. *J Biochem (Tokyo)* 89:701–709
- Matouschek A (2003) Protein unfolding-an important process in vivo? *Curr Opin Struct Biol* 13:98–109
- Maxwell KL, Wildes D, Zarrine-Afsar A, de los Rios MA, Brown AG et al (2005) Protein folding: defining a "standard" set of experimental conditions and a preliminary kinetic data set of two-state proteins. *Protein Sci* 14:602–616
- Meadows PY, Bemis JE, Walker GC (2003) Single-molecule force spectroscopy of isolated and aggregated fibronectin proteins on negatively charged surfaces in aqueous liquids. *Langmuir* 19:9566–9572
- Merkel R (2001) Force spectroscopy on single passive biomolecules and single biomolecular bonds. *Phys Rep* 346:344–385
- Merkel R, Nassoy P, Leung A, Ritchie K, Evans E (1999) Energy landscapes of receptor-ligand bonds explored with dynamic force spectroscopy. *Nature* 397:50–53
- Minne SC, Yeraliogly G, Manalis SR, Adams JD, Zesch J, Atalar A, Quate CF (1998) Automated parallel high-speed atomic force microscopy. *Appl Phys Lett* 72:2340–2302
- Moy VT, Florin EL, Gaub HE (1994) Intermolecular forces and energies between ligands and receptors. *Science* 266:257–259

- Müller DJ, Kessler M, Oesterhelt F, Möller C, Oesterhelt D, Gaub H (2002) Stability of bacteriorhodopsin alpha-helices and loops analyzed by single-molecule force spectroscopy. *Biophys J* 8:3578–3588
- Murzin AG, Brenner SE, Hubbard T, Clothia C (1995) SCOP: a structural classification of proteins database for the investigation of sequences and structures. *J Mol Biol* 247:536–540
- Neher E, Sakmann B (1976) Single-channel current recorded from membrane of denervated frog muscle fibers. *Nature* 260:799–802
- Neupert W, Brunner M (2002) The protein import motor of mitochondria. *Nat Rev Mol Cell Biol* 3:555–565
- Ng SP, Rounsevell RWS, Steward A, Geierhaas CD, Williams PM, Paci E, Clarke J (2005) Mechanical unfolding of TNfn3: the unfolding pathway of a fnIII domain probed by protein engineering, AFM and MD simulation. *J Mol Biol* 350:776–789
- Oberdorfer Y, Fuchs H, Janshoff A (2000) Conformational analysis of native fibronectin by means of force spectroscopy. *Langmuir* 16:9955–9958
- Oberhauser AF, Marszalek PE, Erickson HP, Fernandez JM (1998) The molecular elasticity of tenascin, an extracellular matrix protein. *Nature* 393:181–185
- Oberhauser AF, Marszalek PE, Carrion-Vazquez M, Fernandez JM (1999) Single protein misfolding events captured by atomic force microscopy. *Nat Struct Biol* 6:1025–1028
- Oberhauser AF, Hansma PK, Carrion-Vazquez M, Fernandez JM (2001) Stepwise unfolding of titin under force-clamp atomic force microscopy. *Proc Natl Acad Sci USA* 98:468–472
- Oberhauser AF, Badilla-Fernandez C, Carrion-Vazquez M, Fernandez JM (2002) The mechanical hierarchies of fibronectin observed with single-molecule AFM. *J Mol Biol* 319:433–447
- Oesterhelt F, Oesterhelt D, Pfeiffer M, Engel A, Gaub HE, Müller DJ (2000) Unfolding pathways of individual bacteriorhodopsins. *Science* 288:143–146
- Oster G, Wang H (2003) Rotary protein motors. *Trends Cell Biol* 13:114–121
- Paci E, Karplus M (1999) Forced unfolding of fibronectin type 3 modules: an analysis by biased molecular dynamics simulations. *J Mol Biol* 288:441–459
- Paci E, Karplus M (2000) Unfolding proteins by external forces and temperature: the importance of topology and energetics. *Proc Natl Acad Sci USA* 97:6521–6526
- Pantazatos DP, MacDonald RI (1997) Site-directed mutagenesis of either the highly conserved Trp-22 or the moderately conserved Trp-95 to a large, hydrophobic residue reduces the thermodynamic stability of a spectrin repeating unit. *J Biol Chem* 272:21052–21059
- Pickart CM, Cohen RE (2004) Proteasomes and their kin: proteases in the machine age. *Nat Rev Mol Cell Biol* 5:177–187
- Plaxco KW, Spitzfaden C, Campbell ID, Dobson CM (1997) A comparison of the folding kinetics and thermodynamics of two homologous fibronectin type III modules. *J Mol Biol* 270:763–770
- Politou AS, Thomas DJ, Pastore A (1995) The folding and stability of titin immunoglobulin-like modules, with implications for the mechanism of elasticity. *Biophys J* 69:2601–2610
- Prakash S, Matouschek A (2004) Protein unfolding in the cell. *Trends Biochem Sci* 29:593–600
- Proksch R, Schäffer TE, Cleveland JP, Callahan RC, Viani MB (2004) Finite optical spot size and position corrections in thermal spring constant calibration. *Nanotechnology* 15:1344–1350
- Qian F, Wei W, Germino G, Oberhauser AF (2005) The nanomechanics of polycystin-1 ectodomain. *J Biol Chem* 280:40723–40730
- Rief M, Oesterhelt F, Heymann B, Gaub HE (1997a) Single molecule force spectroscopy on polysaccharides by atomic force microscopy. *Science* 275:1295–1297

- Rief M, Gautel M, Oesterhelt F, Fernandez JM, Gaub HE (1997b) Reversible unfolding of individual titin immunoglobulin domains by AFM. *Science* 276:1109–1112
- Rief M, Fernandez JM, Gaub HE (1998a) Elastically coupled two-level systems as a model for biopolymer extensibility. *Phys Rev Lett* 81:4764–4767
- Rief M, Gautel M, Schemmel A, Gaub HE (1998b) The mechanical stability of immunoglobulin and fibronectin III domains in the muscle protein titin measured by atomic force microscopy. *Biophys J* 75:3008–3014
- Rief M, Pascual J, Saraste M, Gaub HE (1999) Single molecule force spectroscopy of spectrin repeats: low unfolding forces in helix bundles. *J Mol Biol* 286:553–561
- Rohs R, Etchebest C, Lavery R (1999) Unraveling proteins: a molecular mechanics study. *Biophys J* 76:2760–2768
- Ros R, Schwesinger F, Anselmetti D, Kubon M, Schäfer R, Plückthun A, Tiefenauer L (1998) Antigen binding forces of individually addressed single-chain Fv antibody molecules. *Proc Natl Acad Sci USA* 95:7402–7405
- Rounsevell R, Forman JR, Clarke J (2004) Atomic force microscopy: mechanical unfolding of proteins. *Methods* 34:100–111
- Samori B, Zuccheri G, Baschieri R (2005) Protein unfolding and refolding under force: methodologies for nanomechanics. *Chemphyschem* 6:29–34
- Sarkar A, Robertson RB, Fernandez JM (2004) Simultaneous atomic force microscope and fluorescence measurements of protein unfolding using a calibrated evanescent wave. *Proc Natl Acad Sci USA* 101:12882–12886
- Sarkar A, Caamano S, Fernandez JM (2005) The elasticity of individual titin PEVK exons measured by single molecule atomic force microscopy. *J Biol Chem* 280:6261–6264
- Sato T, Esaki M, Fernandez JM, Endo T (2005) Comparison of the protein-unfolding pathways between mitochondrial protein import and atomic-force microscopy measurements. *Proc Natl Acad Sci USA* 102:17999–18004
- Sauer RT, Bolon DN, Burton BM, Burton RE, Flynn JM, Grant RA, Hersch GL, Joshi SA, Kenniston JA, Levchenko I, Neher SB, Oakes E., Siddiqui SM, Wah DA, Baker TA (2004) Sculpting the proteome with AAA(+) proteases and disassembly machines. *Cell* 119:9–18
- Scalley ML, Yi Q, Gu H, McCormack A, Yates JR, Baker D (1997) Kinetics of folding of the IgG binding domain of Peptostreptococcal protein L. *Biochemistry* 36:3373–3382
- Schlierf M, Li H, Fernandez JM (2004) The unfolding kinetics of ubiquitin captured with single-molecule force-clamp techniques. *Proc Natl Acad Sci USA* 101:7299–7304
- Schoenauer R, Bertocini P, Machaidze G, Aebi U, Perriard JC, Hegner M, Agarkova (2005). Myomesin is a molecular spring with adaptable elasticity. *J Mol Biol* 349:367–379
- Schwaiger I, Kardinal A, Schleicher M, Noegel AA, Rief M (2004) A mechanical unfolding intermediate in an actin-crosslinking protein. *Nat Struct Mol Biol* 11:81–85
- Schwaiger I., Sattler C, Hostetter DR, Rief M (2002) The myosin coiled-coil is a truly elastic protein structure. *Nat Mater* 1:232–235
- Schwaiger I, Schleicher M, Noegel AA, Rief M (2005) The folding pathway of a fast-folding immunoglobulin domain revealed by single-molecule mechanical experiments. *EMBO Rep* 6:1–6
- Schwesinger F, Ros R, Strunz T, Anselmetti D, Güntherodt H J, Honegger A, Jermutus L, Tiefenauer L, Pluckthun A (2000) Unbinding forces of single antibody-antigen complexes correlate with their thermal dissociation rates. *Proc Natl Acad Sci USA* 97:9972–9977
- Scott KA, Steward A, Fowler SB, Clarke J (2002) Titin; a multidomain protein that behaves as the sum of its parts. *J Mol Biol* 315:819–829

- Sekiguchi H, Arakawa H, Okajima T, Ikai A (2002) Non-destructive force measurement in liquid using atomic force microscope. *Appl Surf Sci* 188:489–492
- Sekiguchi H, Arakawa H, Taguchi H, Ito T, Kokawa R, Ikai A (2003) Specific interaction between GroEL and denatured protein measured by compression-free force spectroscopy. *Biophys J* 85:484–490
- Shao X, Davletov BA, Sutton RB, Südhof TC, Rizo J (1996) Bipartite Ca²⁺-binding motif in C2 domains of synaptotagmin and protein kinase C. *Science* 273:248–251
- Shtilerman M, Lorimer GH, Englander SW (1999) Chaperonin function: folding by forced unfolding. *Science* 284:822–825
- Smith DE, Tans SJ, Smith SB, Grimes S, Anderson DL, Bustamante C (2001) The bacteriophage straight phi29 portal motor can package DNA against a large internal force. *Nature* 413:748–752
- Steward A, Toca-Herrera JL, Clarke J (2002) Versatile cloning system for construction of multimeric proteins for use in atomic force microscopy. *Protein Sci* 11:2179–2183
- Tatham A, Shewry PR (2000) Elastomeric proteins: biological roles, structures and mechanisms. *Trends Biochem Sci* 25:567–571
- Thompson RE, Siggia ED (1995) Physical limits on the mechanical measurement of the secondary structure of bio-molecules. *Europhys Lett* 31:335–340
- Tsay JT, Appleman JR, Beard WA, Prendergast NJ, Delcamp TJ, Freisheim JH, Blakley RL (1990) Kinetic investigation of the functional role of phenylalanine-31 of recombinant human dihydrofolate reductase. *Biochemistry* 29:6428–6436
- Tskhovrebova L, Trinick J (2003) Titin: properties and family relationships. *Nat Rev Mol Cell Biol* 4:679–789
- Urry DW, Hugel T, Seitz M, Gaub HE, Sheiba L, Dea J, Xu J, Parker T (2002) Elastin: a representative ideal protein elastomer. *Philos Trans R Soc Lond Ser B* 357:169–184
- Valpuesta JM, Martin-Benito J, Gomez-Puertas P, Carrascosa JL, Willison KR (2002) Structure and function of a protein folding machine: the eukaryotic cytosolic chaperonin CCT. *FEBS Lett* 529:11–16
- Vettiger P, Cross G, Despont M, Drechsler U, Dürig U, Gotsmann B, Häberle W, Lantz M, Rothuizen H, Stutz R, Binnig GK (2002) The “Millipede”-nanotechnology entering data storage. *IEEE T Nanotechnol* 1:39–55
- Viani MB, Schäffer TE, Chand A, Rief M, Gaub E, Hansma PK (1999) Small cantilevers for force spectroscopy of single molecules. *J Appl Phys* 8:2558–2662
- Wallace LA, Matthews CR (2002) Highly divergent dihydrofolate reductases conserve complex folding mechanisms. *J Mol Biol* 315:193–211
- Wang H, Oster G (1998) Energy transduction in the F1 motor of ATP synthase. *Nature* 396:279–282
- Wang K, McClure J, Tu A (1979) Titin: major myofibrillar components of striated muscle. *Proc Natl Acad Sci USA* 76:3698–3702
- Watanabe K, Muhle-Goll C, Kellermayer MS, Labeit S, Granzier H (2002) Different molecular mechanics displayed by titin’s constitutively and differentially expressed tandem Ig segments. *J Struct Biol* 137:248–258
- Weisel JW, Shuman H, Litvinov, RI (2003) Protein-protein unbinding induced by force: single-molecule studies. *Curr Opin Struct Biol* 13:227–235
- Wiborg O, Pedersen MS, Wind A, Berglund LE, Marcker KA, Vuust J (1985) The human ubiquitin multigene family: some genes contain multiple directly repeated ubiquitin coding sequences. *EMBO J* 4:755–759

- Wielert-Badt S, Hinterdorfer P, Gruber H J, Lin JT, Badt D, Wimmer B, Schindler H, Kinne RKH (2002) Single molecule recognition of protein binding epitopes in brush border membranes by force microscopy. *Biophys J* 82:2767–2774
- Wilcox AJ, Choy J, Bustamante C, Matouschek A (2005) Effect of protein structure on mitochondrial import. *Proc Natl Acad Sci* 102:15435–15440
- Williams PM, Fowler SB, Best RB, Toca-Herrera JL, Scott KA, Steward A, Clarke J (2003) Hidden complexity in the mechanical properties of titin. *Nature* 422:446–449
- Wong SS, Joselevich E, Woolley A T, Cheung CL, Lieber CM (1998) Covalently functionalized nanotubes as nanometre-sized probes in chemistry and biology. *Nature* 394:52–55
- Yang G, Cecconi C, Baase WA, Vetter IR, Breyer WA, Haack JA, Matthews BW, Dahlquist FW, Bustamante C (2000) Solid-state synthesis and mechanical unfolding of polymers of T4 lysozyme. *Proc Natl Acad Sci USA* 97:139–144
- Zhang X, Wojcikiewicz E, Moy VT (2002) Force spectroscopy of the leukocyte function-associated antigen-1/intercellular adhesion molecule-1 interaction. *Biophys J* 83:2270–2279
- Zhu C, Long M, Chesla SE, Bongrand P (2002) Measuring receptor/ligand interaction at the single-bond level: experimental and interpretative issues. *Ann Biomed Eng* 30:305–314
- Zhuang X, Rief M (2003) Single-molecule folding. *Curr Opin Struct Biol* 13:88–97
- Zlatanova J, Leuba SH (2003) Chromatin fibers, one at-a-time. *J Mol Biol* 331:1–19
- Zlatanova J, Lindsay SM, Leuba SH (2000) Single molecule force spectroscopy in biology using the atomic force microscope. *Prog Biophys Mol Biol* 74:37–61

Multiphoton-Excitation Fluorescence Microscopy and Membranes

LUIS A. BAGATOLLI

9.1 Introduction

The idea that lipids are simply randomly organized building blocks of membranes that form diffusion barriers between cytoplasm and the outside world has significantly changed since the discovery of the coexistence of stable lipid domains in lipid bilayers, approximately 3 decades ago (Gebhart et al. 1977). Conversely, the consequences of the nonrandom lateral organization of lipid membranes have not been acknowledged, particularly from the biology field, until recently when the raft hypothesis was postulated (Simons and Ikonen 1997; Brown and London 1998; Edidin 2003). At present, however, the connection between membrane lateral structure and membrane function remains obscure. Still, there are simple questions regarding lipids and membranes that still need to be answered. For example, why do cells membranes contain thousands of different molecular lipid species? Why do the molar fractions of these species vary among different membranes? Is there a coherent code still hidden and waiting to be discovered? As Hilgemann (2003) pointed out in his article "Getting ready for the decade of the lipids" "...why not speculate that (phospho)lipids and their metabolites will soon be the subject of an information explosion, similar to that presently occurring for genes and proteins?"

In the last 30 years, there has been extensive research to elucidate the coexistence of lipid domains in membranous systems (mainly liposomes but also cell membranes) using an array of experimental techniques such as fluorescence spectroscopy, differential scanning calorimetry, IR spectroscopy, electron paramagnetic resonance, NMR to mention a few (Lee 1975; Lentz et al. 1976; Mabrey and Sturtevant 1976; van Dijk et al. 1977; Arnold et al. 1981; Blume et al. 1982; Cafrey and Hing 1987; Shimshick and McConnell 1973; Maggio 1985; Maggio et al. 1986; Bagatolli et al. 1997; Vaz et al. 1989, 1990; Bultmann et al. 1991; Almeida et al. 1992; Parasassi et al. 1993; Schram et al. 1996), including theoretical treatments using computer simulations (Ipsen and Mouritsen 1988; Jørgensen and Mouritsen 1995). In general, the experimental techniques just indicated produce mean parameters on the basis of data collected from bulk solution of many liposomes (or cells) and lack information about lipid lateral organization at the level of *single* vesicles, a quality that can be provided by microscopy techniques, in particular fluorescence microscopy. The main advantage in using fluorescence microscopy related techniques in membranes over the traditional experimental approaches is clear: the sensitivity and flexibility of a microscope with the addition of fluorescence spectroscopy allows the collection of spatially resolved information. Ultimately this information bridges membrane mor-

phology with dynamical and structural information obtained at a molecular level using fluorescence spectroscopy (such as lipid mobility and hydration). The additional “visual” information obtained in these experiments has generated new data in the membrane field (in particular for bilayer membrane systems) that have important implications in understanding membrane lateral structure. The aim of this report is to summarize the overall impact of the newly obtained results using fluorescence microscopy in the membrane field from model systems of different composition to native membranes. Also the most relevant aspects of the experimental methodology will be briefly discussed.

9.2 Model Systems

In recent years, several papers have appeared which described the use of giant vesicles as model systems to address biophysical aspects of mainly lipid–lipid, but also lipid–DNA and lipid–protein interactions (Wick et al. 1996; Bucher et al. 1998; Longo et al. 1998; Angelova et al. 1999; Bagatolli and Gratton 1999, 2000a, b; Kowlach et al. 1999; Bagatolli et al. 2000, 2002; Holopainen et al. 2000; Dietrich et al. 2001; Feigenson and Buboltz 2001; Fahsel et al. 2002; Nag et al. 2002; Sanchez et al. 2002; Veatch and Keller 2002, 2003; Koster et al. 2003; Scherfeld et al. 2003; Bacia et al. 2004a, b; Bernardino de la Serna et al. 2004; Girard et al. 2004; Janosch et al. 2004; Kahya et al. 2004; Veatch et al. 2004). One of the reasons why giant vesicles are suitable membrane model systems is their size, on the order of a few tens of micrometers, similar to the size of the plasma membrane of cells. Owing to their size, single vesicles can be directly observed using microscopy-related techniques (such as fluorescence microscopy). Additionally, because experiments are performed at the level of single vesicles, heterogeneity in shape and size and the presence of multilamellar vesicles are ruled out.

One of the significant aspects in using giant vesicles as model systems is the ability to control the molecular composition of the membrane as well as the environmental conditions. For instance, studies of the lateral structure of membranes using giant vesicles as model systems were normally confined to giant vesicles composed of artificial lipids or their mixtures with no more than three to four components (Bagatolli and Gratton 1999, 2000a, b; Kowlach et al. 1999; Feigenson and Buboltz 2001; Dietrich et al. 2001; Veatch and Keller 2002, 2003; Kahya et al. 2004). However, as recently reported in the literature, it is also possible to form giant vesicles from natural lipid extracts (Bagatolli et al. 2000; Dietrich et al. 2001; Nag et al. 2002) and native membranes (Bernardino de la Serna et al. 2004). Additionally, incorporation of membrane proteins into giant unilamellar vesicles (GUVs) composed of lipid mixtures can also be performed (Kahya et al. 2001; Bacia et al. 2004b; Girard et al. 2004; Koster et al. 2003). This last fact allows one to establish an interesting strategy, i.e., to perform comparative studies among artificial lipid mixtures, natural lipid mixtures (both with and without membrane proteins) and finally membranes containing the full composition under controlled environmental conditions (Bernardino de la Serna et al. 2004). The reader can find additional information about

the giant vesicle field in an excellent review by Menger and Keiper (1998) and a book completely devoted to giant vesicles edited by Luisi and Walde (2000).

9.3

Fluorescence Microscopy and Membrane Domains in GUVs

Since 1999, several papers have emerged in the literature applying fluorescence microscopy techniques (epifluorescence, confocal and two-photon-excitation microscopy) to show for the first time images of lateral phase separation (gel/fluid and fluid ordered/fluid disordered) in GUVs composed of different phospholipids, phospholipid binary mixtures, ternary lipid mixtures containing cholesterol, natural lipid extracts and native membranes (Bagatolli and Gratton 1999, 2000a, b; Korlach et al. 1999; Feigenson and Buboltz 2001; Dietrich et al. 2001; Veatch and Keller 2002, 2003; Bernardino de la Serna et al. 2004; Kahya et al. 2004). These papers presented, for the first time, a correlation between micron-sized (visual) domain structure and dynamics under different environmental conditions. Experimental approaches involving fluorescence microscopy and GUVs offer, for instance, a new alternative way to construct lipid phase diagrams (Veatch and Keller 2003; Kahya et al. 2004) for artificial lipid mixtures (i.e., phase diagrams that include visual information about membrane lateral structure).

One of the big challenges in the membrane field is still how to correlate lateral structure information between compositionally simple model systems (normally membranes composed of few lipid species) and compositionally complex mixtures (natural lipid extracts or native membranes) under the same environmental conditions. Because of the complex composition of the natural membranes, this correlation is very difficult to achieve using classical “bulk” techniques. However, the visual information provided by fluorescence microscopy is crucial to be able to perform this correlation. For instance, an interesting example is that reported by Bernardino de la Serna et al. (2004) where similar lateral structure (fluid ordered/fluid disordered phase coexistence) was observed between artificial mixtures composed of 1,2-dioleoyl-*sn*-glycero-3-phosphocholine (DOPC)/1,2-dipalmitoyl-*sn*-glycero-3-phosphocholine (DPPC)/cholesterol and native pulmonary surfactant membranes (full composition, i.e., lipids and proteins). In this report, the lateral structure observed in the native pulmonary surfactant at physiological temperatures is linked with functional aspects of this material. This paper suggests that pulmonary surfactant could be one of the first membranous systems reported where the coexistence of specialized membrane domains may exist as a structural basis for its function (Bernardino de la Serna et al. 2004).

At this point it is clear that the visual data obtained from fluorescence microscopy images of lipid bilayers have generated a new important piece of information regarding lipid domain characteristics (e.g., their shape and size). Although the size and the shape of different lipid domains are important parameters, it is also necessary to obtain information about the *local* lateral structure of these different regions that coexist in the plane of the membrane. For instance, is it possible, having the fluorescence images, to ascertain the local lateral structure of the different lipid domains? Have the laterally ordered (or disordered) lipid domains similar local

lateral properties in compositionally different membranes displaying the same phase coexistence scenario (gel/fluid or fluid ordered/fluid disordered)? Few fluorescence microscopy based experimental approaches were used to determine the local lateral structure of in-plane lipid domains. Using confocal microscopy, Koralach et al. (1999) performed fluorescence imaging of a binary phospholipid mixtures (1,2-dilauroyl-*sn*-glycero-3-phosphocholine, (DLPC)/DPPC) displaying gel/fluid phase coexistence using particular fluorescent probes. The fluorescent probes used in these experiments [$1,1'$ -dijeicosanyl-3,3,3',3't_c tramethylindocarbocyanine perchlorate (DiIC₂₀) and 2-(4,4-difluoro-5,7-dimethyl-4-bora-3a, 4a-diaza-*s*-indacene-3-pentanoyl)-1-hexadecanoyl-*sn*-glycero-3-phosphocholine, (BODIPY-PC)] display different partition properties in the two different lipid phases. Using fluorescence correlation spectroscopy, the authors determined the diffusion constant of each fluorescent probe embedded in the different lipid phases. From the diffusion constant obtained in the different regions of the membrane, the different lipid phases were assigned. This experimental strategy was applied to other lipid mixtures as reported in the literature (Scherfeld et al. 2003; Kahya et al. 2004). One of the main drawbacks of this approach is the requirement to find pairs of fluorescent probes that specifically label the different lipid domains.

9.4 Fluorescent Probes

It is very important to keep in mind that the partition of the fluorescent probes *does not depend on the lipid phase state*. Instead, the partition of fluorescent probes generally depends on the chemical environment of the lipid domains (Bagatolli and Gratton 2001; Bagatolli 2003). For instance, the fluorescent probe Lissamine rhodamine B 1,2-dihexadecanoyl-*sn*-glycero-3-phosphoethanolamine (Rh-PE) shows partition to the fluid phase in 1,2-dimiristoyl-*sn*-glycero-3-phosphocholine (DMPC)/1,2-distearoyl-*sn*-glycero-3-phosphocholine (DSPC) mixtures, while in DLPC/DPPC it shows preferential partition to the gel phase (Bagatolli and Gratton 2000b, 2001). In this last case, both mixtures display gel/fluid phase coexistence.

It is particularly interesting to note what happens in mixtures that display fluid ordered/fluid disordered phase coexistence (e.g., DOPC/sphingomyelin/cholesterol or DOPC/DPPC/cholesterol) where the vast majority of membrane probes (particularly those that are excited in the visible regions) used in epifluorescence/confocal microscopy experiments show preferential partition to the liquid disordered phase. One possible solution to this problem is the incorporation of a small fraction of the ganglioside G_{M1} in the lipid mixtures (Dietrich et al. 2001; Bacia et al. 2004a). With use of fluorescently labeled cholera toxin the G_{M1}-enriched regions were visualized using fluorescence microscopy. The G_{M1}-enriched areas correspond to liquid ordered like phase domains (Dietrich et al. 2001; Bacia et al. 2004a). Alternatively, as recently reported, the fluorescent probe perylene shows preferential partition to fluid ordered regions in DOPC/sphingomyelin/cholesterol mixtures (Baumgart et al. 2003). However, the partition of this fluorescent probe need to be further evaluated in other model systems in order to corroborate if it can be used as a fluorescent marker for fluid ordered phases.

As already pointed out, the partition properties of fluorescent probes are generally independent on the lipid phase state. Therefore, it is particularly difficult to assign the lateral structure of membrane domains in compositionally complex samples (such as natural lipid extracts or native membranes) just using the probe's partition information from simple mixtures. For historical reasons, "bulk" studies of the lateral structure of lipid membranes using fluorescence spectroscopy techniques were done mainly using UV-excited fluorescent probes (e.g., pyrene, diphenylhexatriene, diphenylhexatriene trimethylammonium, perylene, parinaric acid, LAURDAN and PRODAN). However, it is practically difficult to perform fluorescence microscopy experiments (one-photon excitation, i.e., epifluorescence and confocal) with these fluorescent molecules since the extent of photobleaching is high and it is difficult to obtain reliable fluorescence images.

9.5 Two-Photon-Excitation Microscopy

Two-photon-excitation fluorescence microscopy constitutes one of the most promising and fastest developing areas in biological and medical imaging at the optical-resolution level (Diaspro and Sheppard 2002). The first application of two-photon fluorescence microscopy in biology was by Denket al. (1990) in 1990. The benefits of two-photon excitation include improved background discrimination, reduced photobleaching of the fluorophores and minimal photodamage to living cell specimens (Denk et al. 1990; So et al. 1995, 1996; Masters et al. 1999). In their application to membranes, the particular characteristics of multiphoton-excitation fluorescence microscopy allow the use of the UV-excited fluorescent probes discussed earlier in a microscope to fully combine and exploit two important pieces of information: (1) the fluorescence parameters that are sensitive to membrane lateral structure, such as phase-dependent emission shift, fluorescence lifetimes or polarization, as was done in the earlier studies in bulk, i.e., liposome solutions; (2) visual information (morphological and topological data). In two-photon-excitation microscopy, the fluorescent probes that are normally excited to the excited electronic state by UV photons can be excited by simultaneous absorption of two IR photons. As was previously demonstrated, the overall extent of photobleaching in this type of experiment is significantly reduced compared with that in confocal fluorescence microscopy (Bagatolli and Gratton 2001; Bagatolli et al. 2003). Since their introduction, few two-photon-excitation fluorescence microscopy applications have been reported in order to explore the lateral structure of lipid membranes (Bagatolli and Gratton 2001; Quesada et al. 2001; Baumgart et al. 2003). Use of particular fluorescent probes and model membrane systems (ranging from single-component lipid membranes to cell membranes) allows important information regarding the microscopic scenario of lipid-lipid and lipid-protein interactions to be achieved using UV-excited fluorescent probes and two-photon-excitation microscopy. This last fact offer a very consistent picture of relevant events such as lipid phase separation in lipid bilayers and cellular membranes (as discussed later), enzymatic reactions in membranes (Bagatolli et al. 2002; Sánchez et al. 2002) and insertion of peptides in lipid membranes (Fahsel et al. 2002; Janosch et al. 2004). Although other probes and methodologies can be applied

to study membrane lateral heterogeneity, as discussed earlier, in the next sections the discussion will be limited primarily to results obtained with one of these probes, LAURDAN, in many different model system and native membranes using two-photon-excitation fluorescence microscopy.

9.6

LAURDAN Probe: the Tips

As is the case with the vast majority of UV-excited fluorescent probes, confocal or epifluorescence microscopy experiments are difficult to perform with LAURDAN. As already mentioned, the extent of LAURDAN photobleaching under epifluorescence and confocal microscopy is severe, making it almost impossible to collect images on LAURDAN-labeled specimens for more than a few seconds (Bagatolli and Gratton 2001). A way to circumvent this problem is by using two-photon-excitation fluorescence microscopy. Because this fluorescent probe has very interesting spectroscopic and partition properties in membranes, it is possible to perform quantitative spectroscopic studies at each pixel and build an information image related to the sample at hand, be it a living cell, an extended surface polymer or a GUV. Researchers have been interested in this marriage of technologies and examples can be found in the early ratio imaging studies, fluorescence lifetime imaging techniques and fluorescence polarization imaging.

LAURDAN belongs to the family of polarity-sensitive fluorescent probes, first designed and synthesized by Gregorio Weber for the study of the phenomenon of dipolar relaxation of fluorophores in solvents, bound to proteins and associated with lipids (Weber and Farris 1979; Mcgregor and Weber 1986; Parasassi et al. 1986; Lagsagna et al. 1996). When inserted in lipid membranes, LAURDAN displays unique characteristics compared with other fluorescent probes, namely, (1) LAURDAN shows a phase-dependent emission spectral shift, i.e., bluish in the ordered lipid phase and greenish in the disordered lipid phase (this effect is attributed to the reorientation of water molecules present at the lipid interface near LAURDAN's fluorescent moiety), (2) LAURDAN distributes equally into the ordered and disordered lipid phases, (3) the electronic transition moment of LAURDAN is aligned parallel to the hydrophobic lipid chains, allowing use of the photoselection effect to qualitatively discriminate between different lipid phases and (4) LAURDAN is negligibly soluble in water (Bagatolli and Gratton 2001; Bagatolli et al. 2003). The homogeneous LAURDAN distribution in membranes displaying lateral heterogeneity together with the lipid phase-dependent emission spectral shift allows one to obtain lateral packing information *directly* from the fluorescence images. This fact offers a great advantage over the experimental techniques discussed before based on the probe partition to different lipid phases. A way to quantify the extent of water dipolar relaxation, that in turn is related to the phase state of the lipid membrane, is based on a useful relationship between the emission intensities obtained on the blue and the red side of LAURDAN's emission spectrum. This relationship, called generalized polarization (GP) (Parasassi and Gratton 1995; Parasassi et al. 1998), was defined by analogy to the fluorescence polarization function. In the GP function, the relative parallel and perpendicular orientations in the classical polarization function were

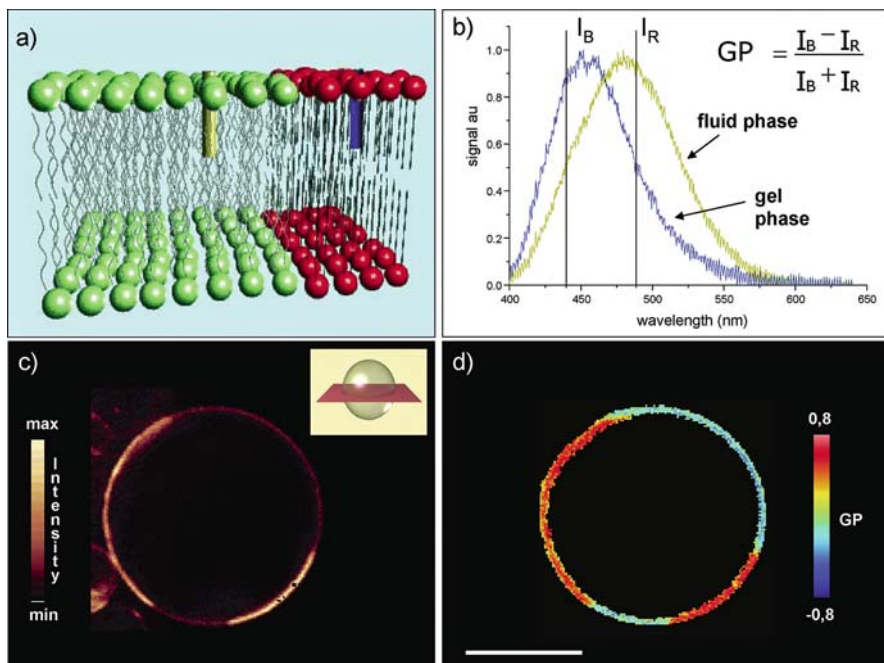


Fig. 9.1. LAURDAN emission properties and the LAURDAN generalized polarization (GP) function. The particular LAURDAN homogeneous distribution in membranes displaying phase coexistence (a) and the phase-sensitive emission shift (b) allows determination of the particular lipid phase coexistence scenario. c LAURDAN fluorescence intensity image of a giant unilamellar vesicle (GUV) composed of a 1,2-dilauroyl-*sn*-glycero-3-phosphocholine (DLPC)/1,2-diarachidoyl-*sn*-glycero-3-phosphocholine (DAPC) 1:1 molar binary mixture obtained using a blue band-pass filter (that selects the emission spectra in the blue region, i.e., the high-intensity area in the image corresponds to LAURDAN emission coming from the lipid gel phase). d LAURDAN GP image of the same phospholipid binary mixture. The high and low GP areas in the image correspond to gel and fluid phase, respectively. Notice the sensitivity of the GP function to the lipid lateral organization. The bar corresponds to 20 μm

substituted by the intensities at the blue and red edges of the emission spectrum (I_B and I_R , respectively) using a given excitation wavelength. It is important to note that no polarizers are required in the experimental setup even though the name of this function contains the word polarization (Parasassi and Gratton 1995; Parasassi et al. 1998). The GP parameter contains information about solvent dipolar relaxation processes which occur during the time that LAURDAN is in the excited state, and is related to water penetration in the lipid interfaces (Parasassi and Gratton 1995; Parasassi et al. 1998). Therefore, GP images can be constructed from the intensity images obtained with blue and green band-pass filters on the microscope, allowing further characterization of the phase state of the coexisting lipid domains (Bagatolli and Gratton 2001; Bagatolli et al. 2003).

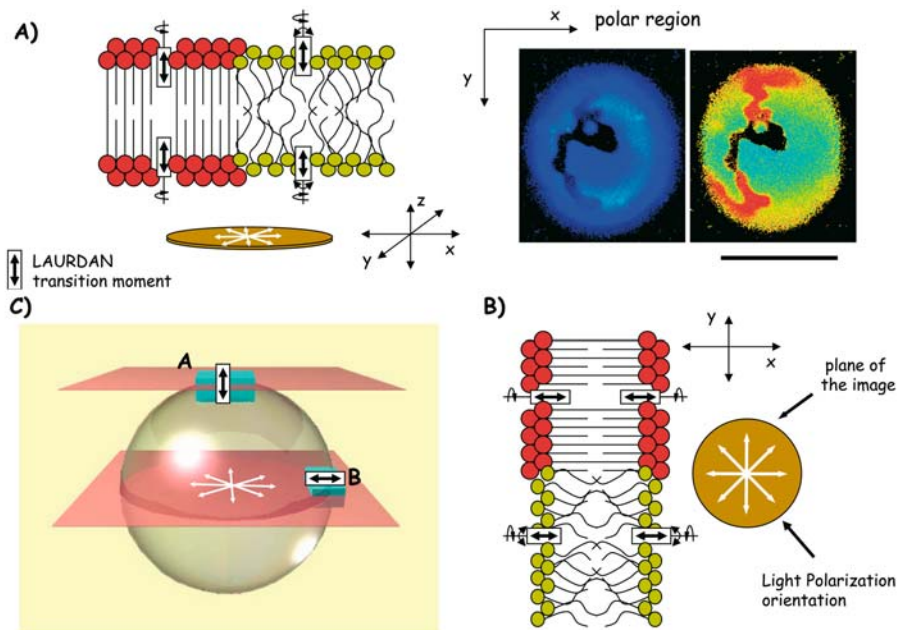


Fig. 9.2. Photoselection effect on LAURDAN-labeled GUVs in the vesicle polar region (**a**). In the GUV's equatorial region the photoselection effect is abolished because of the parallel orientation of the LAURDAN transition moment with respect to the excitation light polarization plane (**b**), see text. These two situations (**a**, **b**) are sketched in three dimensions in **c** by taking into account the excitation light polarization plane (red areas containing the white arrows) that correspond with the scan region plane of the microscope. The lipid mixture in **a** corresponds to GUVs composed of a 1-palmitoyl, 2-oleoyl-*sn*-glycero-3-phosphocholine/ceramide 5:1 molar mixture. The bar corresponds to 20 μm

Figure 9.1 shows the homogeneous partition of LAURDAN between different lipid phases (Fig. 9.1a, c) and the particular emission spectra obtained in the gel and fluid phase regions (Fig. 9.1b). Discrimination of two different fluorescence intensity regions can be observed by using appropriate emission filters Fig. 9.1c. Additionally, computation of the GP function allows further characterization of the different lipid phases (Fig. 9.1d) (Bagatolli and Gratton 2001; Bagatolli et al. 2003). The particular spherical shape of giant vesicles allows application of the photoselection effect to qualitatively distinguish between the different lipid phases. The photoselection effect arises from the fact that only those fluorophores which have excitation transition moments aligned parallel, or nearly so, to the plane of polarization of the excitation light are excited. For example, with use of circularly polarized light as an excitation source, two different pictures can be observed in GUVs displaying gel/fluid phase coexistence at (1) the polar region of the vesicle (Fig. 9.2a) and (2) at the equatorial region of the vesicle (Fig. 9.2b). At the equatorial region of the vesicle the circularly polarized excitation light allows excitation with the same efficiency of all LAURDAN

molecules present in this region of the GUV (Bagatolli and Gratton 2001). In this case the transition moment of the probe is always parallel to the polarization plane. This last fact allows calculation of the GP images without the influence of the photoselection effect, as seen in Fig. 9.1d. On the other hand, at the polar region of the vesicle, only fluorescence coming from the fluid part of the bilayer can be observed (Fig. 9.2a), simply because a component of LAURDAN's transition moment is always parallel to the excitation polarization plane (because of the relatively low lipid order, i.e., the wobbling movement of LAURDAN molecules). At the polar region of the vesicle no fluorescence intensity is observed in the gel phase areas because of the high lipid order (even though LAURDAN molecules are present in this region of the bilayer). The message here is that the photoselection effect allows extraction of qualitative information about lipid phases directly from the intensity images.

The photoselection effect can also be exploited to determine the orientation of the probe transition moment relative to the membrane plane (Bagatolli and Gratton 2000a; Bagatolli et al. 2000). The fluorescence images obtained in the equatorial region of the vesicle using *linearly* polarized light as the excitation source will render particular characteristics depending on the probe orientation in the membrane. For example, in Fig. 9.3 the high-intensity areas in the fluorescence image are populated by fluorescent molecules with their transition moments oriented parallel to the direction of the polarized excitation light (because of the photoselection rule). The immediate conclusion here is that the rhodamine-PE probe has its transition moment oriented 90° in the membrane with respect to that observed for LAURDAN. Additionally, LAURDAN GP images at the equatorial region of the GUV obtained with *linearly* polarized light provide information about coexistence of lipid domains with sizes below the resolution of the microscope (Parasassi et al. 1997; Bagatolli et al. 2003). For a detailed description of the use of LAURDAN GP images to ascertain lipid domains below the resolution of the microscope the reader is encouraged to explore the works of Parasassi et al. 1997 and Bagatolli et al. 2003. To summarize, LAURDAN provides *simultaneous* information about the morphology and phase state of discrete regions in membranes directly from the fluorescence images, an advantage that is not obtained with other fluorescent probes.

9.7

Membrane Lateral Structure in Artificial Lipid Mixtures and Natural Lipid Extracts as Seen by LAURDAN

Since its introduction in 1986, LAURDAN has demonstrated an exquisite sensitivity to detect temperature-induced phase transitions using liposome solutions (Parasassi and Gratton 1995; Parasassi et al. 1998). The first two-photon-excitation fluorescence microscopy study using a membrane model system (multilamellar vesicles) was reported in 1997 by Parasassi et al. (1997). The nonrandom organization of single-component phospholipid bilayers in the fluid phase as well as the coexistence of gel and fluid areas in DOPC/DPPC multilamellar vesicles at room temperature were reported in this pioneering work. Additionally, the lateral structures of red blood cell membranes, renal tubular cell line and renal brush-border and basolateral

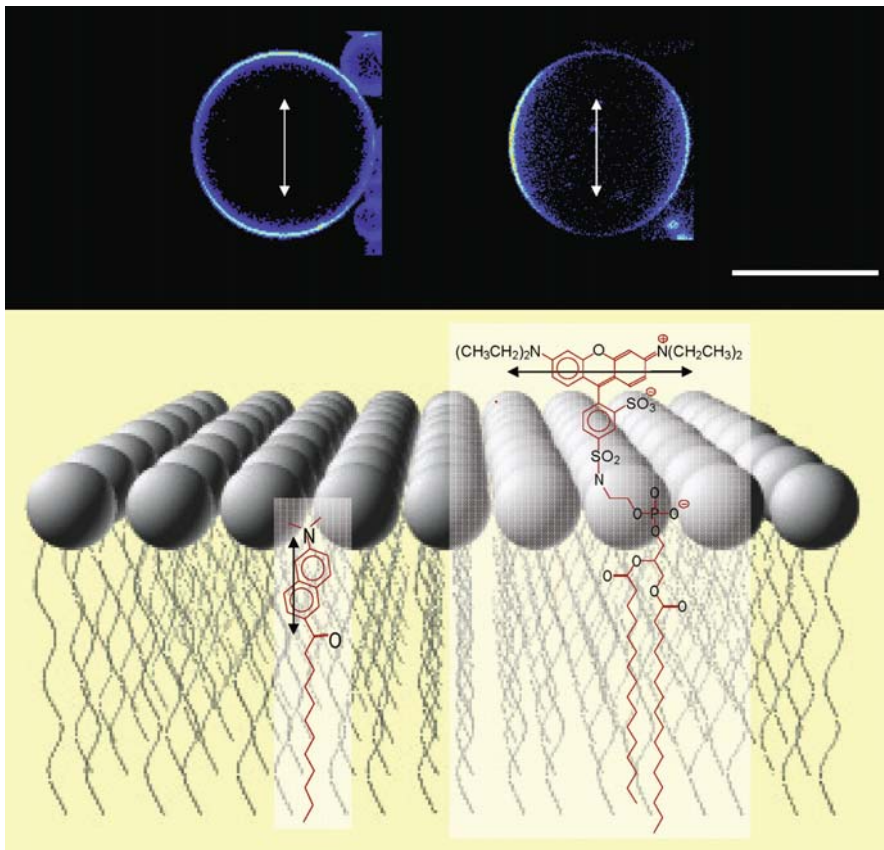


Fig. 9.3. Determination of the orientation of the probe's transition moment with respect to the membrane plane. GUVs are composed of 1,2-dimyristoyl-*sn*-glycero-3-phosphocholine (*DMPC*). GUVs were labeled with LAURDAN (*top left*) and rhodamine Rhodamine PE (*top right*). Sketch of the probe's transition moment orientation (*bottom*) in the membrane. The bar corresponds to 20 μm

membranes using LAURDAN GP were also characterized by using two-photon-excitation microscopy (Parasassi et al. 1997).

The use of GUVs as a model system (instead of giant multilamellar vesicles) improved the characterization of the lateral phase coexistence scenario for several lipid mixtures using the LAURDAN-based fluorescence microscopy approach. From the side of the artificial lipid mixtures, the effect of temperature on the lateral structure of membranes composed of pure phospholipids and their different mixtures at the level of single vesicles was first reported by Bagatolli and Gratton (1999, 2000a, b). In this last case, gel lipid domains of micrometer size with particular shapes depending on the composition of the lipid mixture were observed in GUVs (Bagatolli and Gratton 2000a, b). Additionally, a correlation between domain shape and lipid mis-

cibility was then reported for the different phospholipid binary mixtures displaying gel/fluid phase coexistence (Bagatolli and Gratton 2000b). In this experiment the LAURDAN GP difference between the gel and fluid phases ($\Delta GP = GP_{\text{gel}} - GP_{\text{fluid}}$) for each mixture showed a linear dependence on the phospholipid mixture hydrophobic mismatch (Fig. 9.4). This result was interpreted as evidence of the sensitivity of ΔGP to monitor the compositional differences between the gel and fluid phases among the different mixtures (that in turn are related to the miscibility of the lipid components). Interestingly enough, this relationship correlates with different domain shapes in the different mixtures (mixtures that display hydrophobic mismatch) with similar domain shapes among highly immiscible lipid mixtures (e.g., DLPC/1,2-diarachidoyl-*sn*-glycero-3-phosphocholine (DAPC) and 1,2-dipalmitoyl-*sn*-glycero-3-phosphoethanolamine (DPPE)/DPPC; Fig. 9.4).

The visual information extracted with the experiments just discussed is extremely useful for understanding the lateral structure of membranes composed of complex mixtures. For example, correlations among compositionally complex lipid mixtures displaying gel/fluid (Nag et al. 2002) and fluid ordered/fluid disordered (Dietrich et al. 2001) phase coexistence with binary phospholipid mixtures and cholesterol-containing ternary mixtures respectively were reported using LAURDAN. In particular, the extraction of cholesterol from the brush-border membranes lipid extracts showed, for the first time, a visual change in the phase coexistence scenario (from fluid ordered/fluid disordered phase coexistence to gel/fluid phase coexistence by extraction of cholesterol; Dietrich et al. 2001). Similar experiments were recently reported for pulmonary surfactant membranes and DPPC/DOPC/cholesterol mixtures (Bernardino de la Serna et al. 2004). Examples of this type of information are summarized in Fig. 9.5.

9.7.1 The Importance of Visual Information to Ascertain Lateral Structure in Compositionally Complex Mixtures

The LAURDAN experiments provide a clearer picture of the phase separation phenomenon in compositionally complex systems compared with that obtained from common bulk techniques (Nag et al. 2002). For example, differential scanning calorimetry provides extremely useful thermodynamics parameters to characterize the temperature behavior of lipid mixtures. However, as the number of components in the lipid mixture is increased the data analysis becomes more difficult. A differential scanning calorimetry experiment of a natural complex lipid mixture from bovine lipid extract surfactant is shown in Fig. 9.6. Although important thermodynamics information can be extracted from the differential scanning calorimetry experiment, no detailed information about the physical characteristics of the lipid lateral structure at different temperatures can be obtained using this technique. As observed in Fig. 9.5, from the GU data three different temperature regimes with a particular membrane lateral structure (phase state) are observed and characterized using LAURDAN (Nag et al. 2002). The information obtained from these two tech-

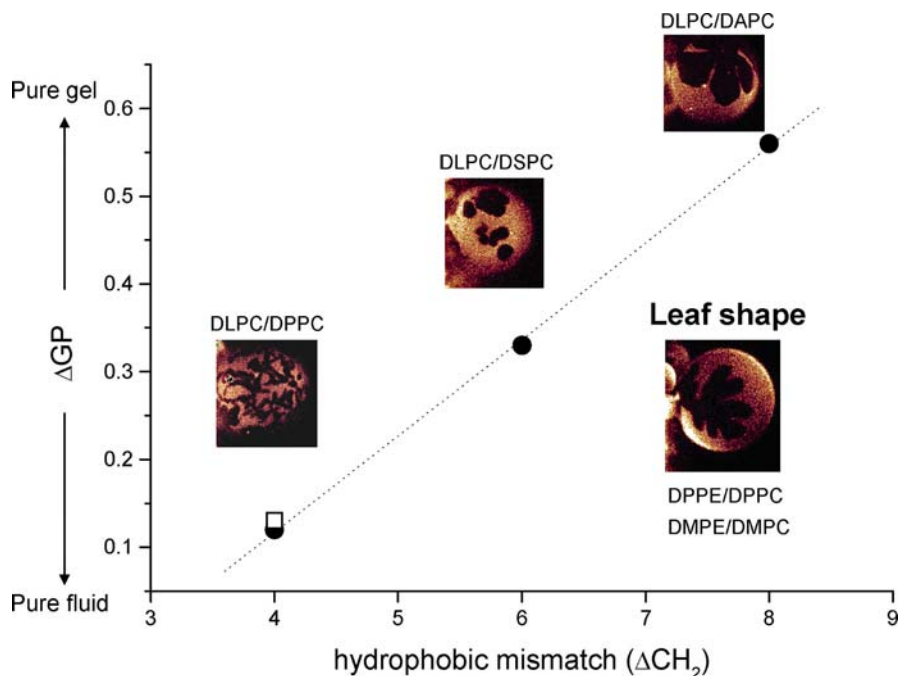


Fig. 9.4. LAURDAN GP differences between the fluid and gel phase domains (ΔGP) versus the hydrophobic mismatch between the components of the binary mixtures. *Circles* DLPC containing GUVs (DLPC/DPPC DPPC, DLPC/DSPC DSPC and DLPC/DAPC 1:1 molar, hydrophobic mismatch equal to 4, 6 and 8, respectively), *squares* DMPC/DSPC mixture (1:1 molar, hydrophobic mismatch equal to 4). ΔGP between the gel and fluid components for 1,2-dimiristoyl-*sn*-glycero-3-phosphoethanolamine (DMPE)/DMPC and DPPE/DPPC (different polar head group/same lipid chain length) was similar to that observed for DLPC/DAPC, showing similar leaf shape

niques is rather complementary and very useful to characterize the lateral structure of particularly complex different lipid mixtures.

9.7.2

LAURDAN in Cell Membranes and Tissues

From the pioneering work of Yu et al. (1996) and Parasassi et al. (1997) LAURDAN was proposed to be very promising in exploring cell membranes. In these studies domains of sizes below, in the same range as and above the microscope resolution limit ($0.3 \mu\text{m}$) were observed in OK cells, red blood cells and brush-border native membranes, respectively (Parasassi et al. 1997). The LAURDAN GP differences observed in compositionally complex mixtures and artificial lipid ternary mixtures containing phospholipids, sphingomyelin and cholesterol were recently

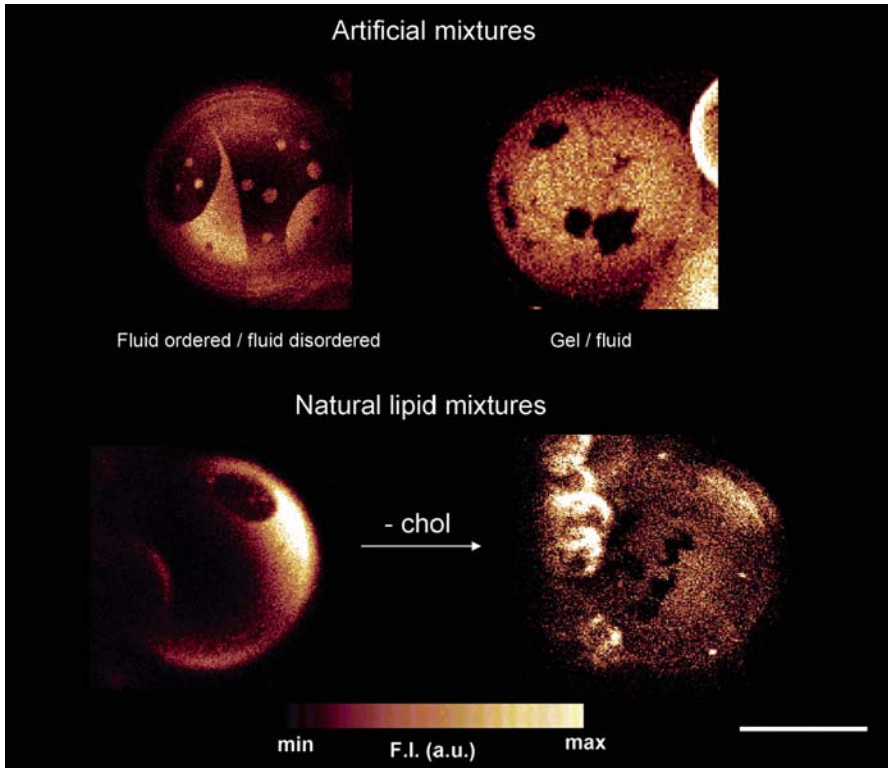


Fig. 9.5. Two-photon-excitation LAURDAN fluorescence images (taken at the polar region of the vesicle) of DOPC/cholesterol/sphingomyelin 1:1:1 molar displaying fluid ordered/fluid disordered phase coexistence (*top left*) and DMPC/DSPC 1:1 molar displaying gel/fluid phase coexistence (*top right*). GUVs composed of brush border membrane lipid extract (*bottom left*) displaying fluid ordered/fluid disordered like phase coexistence. The same membrane after cholesterol extraction (*bottom right*) shows gel/fluid phase coexistence. The *bar* corresponds to 20 μm

used to interpret GP images in cell membranes (Gaus et al. 2003). In this report the LAURDAN GP function was used to directly observed transient micron-sized high GP regions surrounded by low GP areas in living macrophages (Fig. 9.7). This paper demonstrated the presence of lateral phase separation in living cells, supporting strongly the cholesterol effect observed in the model systems (Dietrich et al. 2001). Interestingly enough, this last result is in line with the works reported by Gousset et al. (2002) and Bernardino de la Serna et al. (2004), where micron-sized domains were also observed in platelets upon activation and in native pulmonary surfactant membranes, respectively. Although micron-sized domains are observed in the membranes mentioned before (macrophages, platelets, pulmonary surfactant), generalization of this phenomenon must be done cautiously. Following the literature in this respect, the presence of micron-sized domains in biological membranes seems

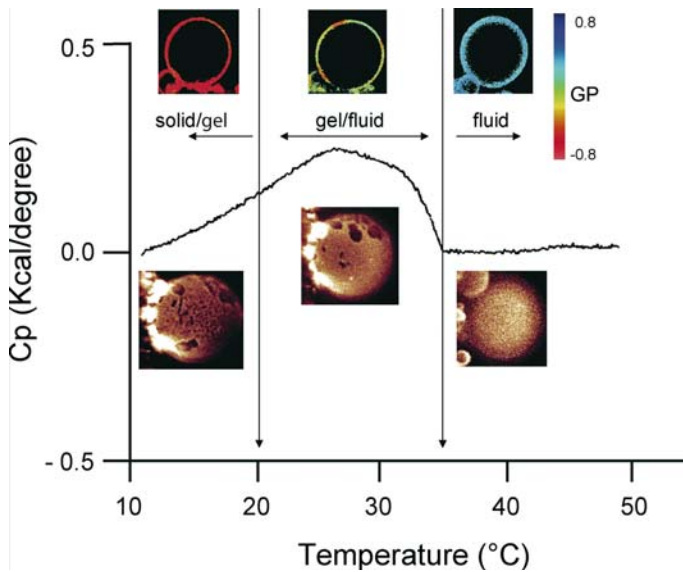


Fig. 9.6. Bovine lipid extract surfactant membranes. The thermogram obtained from differential scanning calorimetry experiments indicates a very broad and complex phase transition temperature region. LAURDAN intensity and LAURDAN GP images reveal additional information showing three different phase transitions (fluid \rightarrow fluid/gel, fluid/gel \rightarrow solid/gel) in the same temperature range used in the differential scanning calorimetry experiments. The diameter of the GUVs is approximately 30 μm

to be the exception instead of the general case. This last fact can be simply related to the membrane composition and the particular functions of the membranes under favorable environmental conditions (such as temperature).

Lastly LAURDAN GP was also applied to tissues. Sun et al. (2004) have demonstrated that both LAURDAN multiphoton polarization and GP can be combined under a two-photon-excitation fluorescence microscope to characterize the structural changes of intercellular lipids in skin tissue. This work demonstrated how the treatment of oleic acid results in a skin surface with a more random packing of lipid molecules, which facilitates water penetration.

9.8 Concluding Remarks

As demonstrated in this chapter, the use of particular UV-fading probes under a two-photon-excitation microscope can be exploited to learn about processes that occur in the complex frame of biological membranes. One of the fundamental issues is to understand the molecular interactions in a simple model system to finally understand the basis of either lipid-lipid or lipid-protein interactions in more com-

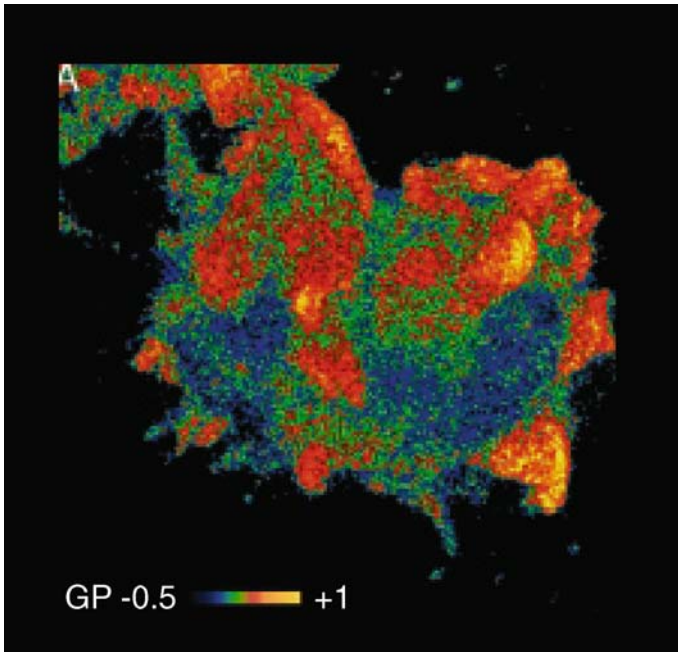


Fig. 9.7. GP image of a living macrophage, 3D-reconstructed pseudocolored GP images of RAW264.7 cell (image plane parallel to cover slip, viewed from above). Notice the different discrete GP regions on the cell membrane indicating membrane lateral heterogeneity. (Adapted from Gaus et al. 2003, copyright 2003 National Academy of Sciences, USA)

plex situations. As discussed, this last type of information renders multiphoton-excitation microscopy a powerful tool to learn about similar phenomena in complex systems such as biological membranes and tissues.

9.9 Summary

In the last few years fluorescence microscopy has become a very useful technique to study the lateral structure of membranes. Several experimental strategies using fluorescence microscopy techniques (epifluorescence, confocal and two-photon excitation) have appeared in the literature employing GUVs as model systems. The advantages and disadvantages of the different experimental approaches, including the use of different fluorescent probes, were discussed in this chapter. In particular, the advantages in using the polarity-sensitive probe LAURDAN as a fluorescent marker under a two-photon-excitation fluorescence microscope to study not only artificial lipid mixtures but also natural lipid extracts and cell membranes was addressed. The visual information obtained with these techniques is extremely useful to correlate

the lateral structure of compositionally simple membranes with that observed in compositionally complex mixtures such as biological membranes.

References

- Almeida PFF, Vaz WLC, Thompson TE (1992) Lateral diffusion and percolation in two-phase, two-component lipid bilayers. Topology of the solid-phase domains in-plane and across the lipid bilayer. *Biochemistry* 31:7198–7210
- Angelova MI, Hristova N, Tsoneva I (1999) DNA-induced endocytosis upon local microinjection to giant unilamellar cationic vesicles. *Eur Biophys J* 28:142–150
- Arnold K, Lösche A, Gawrisch K (1981) ^{31}P -NMR investigations of phase separation in phosphatidylcholine/phosphatidylethanolamine mixtures. *Biochim Biophys Acta* 645:143–148
- Bacia K, Scherfeld D, Kahya N, Schwille P (2004a) Fluorescence correlation spectroscopy relates rafts in model and native membranes. *Biophys J* 87:1034–1043
- Bacia K, Schuette CG, Kahya N, Jahn R, Schwille P (2004b) SNAREs prefer liquid-disordered over “raft” (liquid-ordered) domains when reconstituted into giant unilamellar vesicles. *Biol Chem* 279:37951–37955
- Bagatolli LA (2003) Direct observation of lipid domains in free standing bilayers: from simple to complex lipid mixtures. *Chem Phys Lipids* 122:137–145
- Bagatolli LA, Gratton E (1999) Two-photon fluorescence microscopy observation of shape changes at the phase transition in phospholipid giant unilamellar vesicles. *Biophys J* 77:2090–2101
- Bagatolli LA, Gratton E (2000a) Two photon fluorescence microscopy of coexisting lipid domains in giant unilamellar vesicles of binary phospholipid mixtures. *Biophys J* 78:290–305
- Bagatolli LA, Gratton E (2000b) Two photon fluorescence microscopy of coexisting lipid domains in giant unilamellar vesicles of binary phospholipid mixtures. *Biophys J* 79:434–447
- Bagatolli LA, Gratton E (2001) Direct observation of lipid domains in free standing bilayers using two-photon excitation fluorescence microscopy. *J Fluorescence* 11:141–160
- Bagatolli LA, Maggio B, Aguilar F, Sotomayor CP, Fidelio GD (1997) LAURDAN properties in glycosphingolipid-phospholipid mixtures: a comparative fluorescence and calorimetric study. *Biochim Biophys Acta* 1325:80–90
- Bagatolli LA, Gratton E, Khan TK, Chong PLG (2000) Two-photon fluorescence microscopy studies of bipolar tetraether giant liposomes from thermoacidophilic archaeobacteria *Sulfolobus acidocaldarius*. *Biophys J* 79:416–425
- Bagatolli LA, Binns DD, Jameson DM, Albanesi JP (2002) Activation of dynamin II by POPC in giant unilamellar vesicles: a two-photon fluorescence microscopy study. *J Protein Chem* 21:383–391
- Bagatolli LA, Sanchez SA, Hazlett T, Gratton E (2003) Giant vesicles, LAURDAN, and two-photon fluorescence microscopy: evidence of lipid lateral separation in bilayers. *Methods Enzymol* 360:481–500
- Baumgart T, Hess ST, Webb WW (2003) Imaging coexisting fluid domains in biomembrane models coupling curvature and line tension. *Nature* 425:821–824

- Bernardino de la Serna J, Perez-Gil J, Simonsen AC, Bagatolli LA (2004) Cholesterol rules: direct observation of the coexistence of two fluid phases in native pulmonary surfactant membranes at physiological temperatures. *J Biol Chem* 279:40715–40722
- Blume A, Wittebort RJ, Das Gupta SK, Griffin RG (1982) Phase equilibria, molecular conformation, and dynamics in phosphatidylcholine/phosphatidylethanolamine bilayers. *Biochemistry* 21:6243–6253
- Brown DA, London E (1998) Functions of lipid rafts in biological membranes. *Annu Rev Cell Dev Biol* 14:111–136
- Bucher P, Fischer A, Luisi PL, Oberholzer T, Walde P (1998) Giant vesicles as biochemical compartments: the use of microinjection techniques. *Langmuir* 14:2712–2721
- Bultmann T, Vaz WLC, Melo ECC, Sisk RB, Thompson TE (1991) Fluid-phase connectivity and translational diffusion in a eutectic, two-component, two-phase phosphatidylcholine bilayer. *Biochemistry* 30:5573–5579
- Caffrey M, Hing FS (1987) A temperature gradient method for lipid phase diagram construction using time-resolved x-ray diffraction. *Biophys J* 51:37–46
- Denk W, Strickler JH, Webb WW (1990) Two-photon laser scanning fluorescence microscopy. *Science* 248:73–76
- Diaspro A, Sheppard CJR (2002) Two photon excitation fluorescence microscopy. In: Diaspro A (ed) *Confocal and two photon microscopy, foundations. Applications and advances*. Wiley-Liss, New York, pp 39–73
- Dietrich C, Bagatolli LA, Volovyk ZN, Thompson NL, Levi M, Jacobson K, Gratton E (2001) Lipid rafts reconstituted in model membranes *Biophys J* 80:1417–1428
- Eidin M (2003) The state of lipid rafts: from model membranes to cells. *Annu Rev Biophys Biomol Struct* 32:257
- Fahsel S, Pospiech EM, Zein M, Hazlet TL, Gratton E, Winter R (2002) Modulation of concentration fluctuations in phase-separated lipid membranes by polypeptide insertion. *Biophys J* 83:334–344
- Feigenson GW, Buboltz JT (2001) Ternary phase diagram of dipalmitoyl-PC/dilauroyl-PC/cholesterol: nanoscopic domain formation driven by cholesterol. *Biophys J* 80:2775–2788
- Gaus K, Gratton E, Kable EP, Jones AS, Gelissen I, Kritharides L, Jessup W (2003) Visualizing lipid structure and raft domains in living cells with two-photon microscopy. *Proc Natl Acad Sci USA* 100:15554–15559
- Gebhardt C, Gruler H, Sackmann E (1977) On domain structure and local curvature in lipid bilayers and biological membranes. *Z Naturforsch C* 32:581–596
- Girard P, Pécréaux J, Lenoir G, Falson P, Rigaud JL, Bassereau P (2004) A new method for the reconstitution of membrane proteins into giant unilamellar vesicles. *Biophys J* 87:419–429
- Gousset K, Wolkers WF, Tsvetkova NM, Oliver AE, Field CL, Walker NJ, Crowe JH, Tablin F (2002) Evidence for a physiological role for membrane rafts in human platelets. *J Cell Physiol* 190:117–128
- Hilgemann DW (2003) Getting ready for the decade of the lipids. *Annu Rev Physiol* 65:697–700
- Holopainen JM, Angelova MI, Kinnunen PKJ (2000) Vectorial budding of vesicles by asymmetrical enzymatic formation of ceramide in giant liposomes. *Biophys J* 78:830–838
- Ipsen JH, Mouritsen OG (1988) Modelling the phase equilibria in two-component membranes of phospholipids with different acyl-chain lengths. *Biochim Biophys Acta* 944:121–134

- Janosch S, Nicolini C, Ludolph B, Peters C, Volkert M, Hazlet TL, Gratton E, Waldmann H, Winter R (2004) Partitioning of dual-lipidated peptides into membrane microdomains: lipid sorting vs peptide aggregation. *J Am Chem Soc* 126:7496–7503
- Jørgensen K, Mouritsen OG (1995) Phase separation dynamics and lateral organization of two-component lipid membranes. *Biophys J* 69:942–954
- Kahya N, Pecheur El, de Boeij WP, Wiersma DA, Hoekstra D (2001) Reconstitution of membrane proteins into giant unilamellar vesicles via peptide-induced fusion. *Biophys J* 81:1464–1474
- Kahya N, Scherfeld D, Bacia K, Schwille P (2004) Lipid domain formation and dynamics in giant unilamellar vesicles explored by fluorescence correlation spectroscopy. *J Struct Biol* 147:77–89
- Korlach J, Schwille P, Webb WW, Feigenson GW (1999) Characterization of lipid bilayer phases by confocal microscopy and fluorescence correlation spectroscopy. *Proc Natl Acad Sci USA* 96:8461–8466
- Koster G, van Duijn M, Hofs B, Dogterom M (2003) Membrane tube formation from giant vesicles by dynamic association of motor proteins. *Proc Natl Acad Sci USA* 100:15583–15588
- Lasagna M, Vargas V, Jameson DM, Brunet JE (1996) Spectral properties of environmentally sensitive probes associated with horseradish peroxidase. *Biochemistry* 35:973
- Lee AG (1975) Fluorescence studies of chlorophyll a incorporated into lipid mixtures, and the interpretation of “phase” diagrams. *Biochim Biophys Acta* 413:11–23
- Lentz BR, Barenholtz Y, Thompson TE (1976) Fluorescence depolarization studies of phase transitions and fluidity in phospholipid bilayers. 2 Two-component phosphatidylcholine liposomes. *Biochemistry* 15:4529–4537
- Longo ML, Waring AJ, Gordon LM, Hammer DA (1998) Area expansion and permeation of phospholipid membrane bilayers by influenza fusion peptides and melittin. *Langmuir* 14:2385–2395
- Luisi PL, Walde P (eds) (2000) *Giant vesicles*, 1st edn. Wiley, Chichester
- Mabrey S, Sturtevant JM (1976) Investigation of phase transitions of lipids and lipid mixtures by sensitivity differential scanning calorimetry. *Proc Natl Acad Sci USA* 73:3862–3866
- Macgregor RB, Weber G (1986) Estimation of the polarity of the protein interior by optical spectroscopy. *Nature* 319:70–73
- Maggio B (1985) Geometric and thermodynamic restrictions for the self-assembly of glycosphingolipid-phospholipid systems. *Biochim Biophys Acta* 815:245–258
- Maggio B, Fidelio GD, Cumar FA, Yu RK (1986) Molecular interactions and thermotropic behavior of glycosphingolipids in model membrane systems. *Chem Phys Lipids* 42:49–63
- Masters BR, So PTC, Gratton E (1999) Multiphoton excitation microscopy and spectroscopy of cells, tissues and human skin *in vivo*. In: Mason WT (ed) *Fluorescent and luminescent probes for biological activity*, 2nd edn. Biological techniques series. Academic, New York, pp 416–432
- Menger FM, Keiper JS (1998) Chemistry and physics of giant vesicles as biomembrane models. *Curr Opin Chem Biol* 2:726–732
- Nag K, Pao JS, Harbottle RR, Possmayer F, Petersen NO, Bagatolli LA (2002) Segregation of saturated chain lipids in pulmonary surfactant films and bilayers. *Biophys J* 82:2041–2051
- Parasassi T, Gratton E (1995) Membrane lipid domains and dynamics as detected by LAURDAN fluorescence. *J Fluorescence* 5:59–69

- Parasassi T, Conti F, Gratton E (1986) Fluorophores in a polar medium: time dependence of emission spectra detected by multifrequency phase and modulation fluorometry. *Cell Mol Biol* 32:103
- Parasassi T, Loiero M, Raimondi M, Ravagnan G, Gratton E (1993) Absence of lipid gel-phase domains in seven mammalian cell lines and in four primary cell types. *Biochim Biophys Acta* 1153:143–154
- Parasassi T, Gratton E, Yu WM, Wilson P, Levi M (1997) Two-photon fluorescence microscopy of LAURDAN generalized polarization domains in model and natural membranes. *Biophys J* 72:2413–2429
- Parasassi T, Krasnowska E, Bagatolli LA, Gratton E (1998) LAURDAN and prodan as polarity-sensitive fluorescent membrane probes. *J Fluorescence* 8:365–373
- Quesada E, Acuña AU, Amat-Guerri F (2001) New transmembrane polyene bolaamphiphiles as fluorescent probes in lipid bilayers. *Angew Chem Int Ed Engl* 40:2095–2097
- Sanchez S, Bagatolli LA, Gratton E, Hazlett T (2002) A two-photon view of an enzyme at work: C. *Atrox* PLA2 interaction with single-lipid and mixed-lipid giant unilamellar vesicles. *Biophys J* 82:2232–2243
- Scherfeld D, Kahya N, Schwille P (2003) Lipid dynamics and domain formation in model membranes composed of ternary mixtures of unsaturated and saturated phosphatidylcholines and cholesterol. *Biophys J* 85:3758–3768
- Schram V, Lin HN, Thompson TE (1996) Topology of gel-phase domains and lipid mixing properties in phase-separated two-component phosphatidylcholine bilayers. *Biophys J* 71:1811–1822
- Shimshick EJ, McConnell HM (1973) Lateral phase separation in phospholipid membranes. *Biochemistry* 12:2351–2360
- Simons K, Ikonen E (1997) Functional rafts in cell membranes. *Nature* 387:569–572
- So PTC, French T, Yu WM, Berland KM, Dong CY, Gratton E (1995) Time-resolved fluorescence microscopy using two photon excitation. *Bioimaging* 3:49–63
- So PTC, French T, Yu WM, Berland KM, Dong CY, Gratton E (1996) Two photon microscopy: time-resolved and intensity imaging. In: Wang XF, Herman B (eds) *Fluorescence imaging spectroscopy and microscopy*. Chemical analysis series, vol 137. Wiley, New York, pp 351–374
- Sun Y, Lo W, Lin SJ, Jee SH, Dong CY (2004) Multiphoton polarization and generalized polarization microscopy reveal oleic-acid-induced structural changes in intercellular lipid layers of the skin. *Opt Lett* 29:2013–2015
- Van Dijck PWM, Kaper AJ, Oonk HAJ, de Gier J (1977) Miscibility properties of binary phosphatidylcholine mixtures. A calorimetric study. *Biochim Biophys Acta* 470:58–69
- Vaz WLC, Melo ECC, Thompson TE (1989) Translational diffusion and fluid domain connectivity in a two-component, two-phase phospholipid bilayer. *Biophys J* 56:869–876
- Vaz WLC, Melo ECC, Thompson TE (1990) Fluid phase connectivity in an isomorphous, two-component, two-phase phosphatidylcholine bilayer. *Biophys J* 58:273–275
- Veatch SL, Keller SL (2002) Organization in lipid membranes containing cholesterol. *Phys Rev Lett* 89:268101
- Veatch SL, Keller SL (2003) Separation of liquid phases in giant vesicles of ternary mixtures of phospholipids and cholesterol. *Biophys J* 85:3074–3083
- Veatch SL, Polozov IV, Gawrisch K, Keller SL (2004) Liquid domains in vesicles investigated by NMR and fluorescence microscopy. *Biophys J* 86:2910–2922

- Weber G, Farris FJ (1979) Synthesis and spectral properties of a hydrophobic fluorescent probe: 6-propionyl-2-(dimethylamino)naphthalene. *Biochemistry* 18:3075–3078
- Wick R, Angelova MI, Walde P, Luisi PL (1996) Microinjection into giant vesicles and light microscopy investigation of enzyme-mediated vesicle transformations. *Chem Biol* 3:105–111
- Yu W, So PT, French T, Gratton E (1996) Fluorescence generalized polarization of cell membranes: a two-photon scanning microscopy approach. *Biophys J* 70:626–636

Three-Dimensional Electron Microscopy. The Coming of Age of a Versatile Structural Biology Technique

CARMEN SAN MARTÍN, MIKEL K. VALLE

10.1 Introduction

The last few years have witnessed great advances in structural biology, as shown by the evolution in the number of entries in the protein structure database (<http://www.pdb.org>): from approximately 1,000 entries between 1972 and 1990 to 35,000 today. Most of these protein structures have been solved by X-ray crystallography, some others by NMR. X-ray crystallography exploits the diffracting power of large (approximately millimeter-sized) 3D protein crystals irradiated by an X-ray beam; NMR takes advantage of the response of protein atomic nuclei to changing magnetic fields in solution. These two techniques have the undeniable advantage of giving atomic resolution data on the specimen under study. However, a large majority of biological macromolecules work in complexes too large and too flexible to be easily approached by X-ray crystallography or NMR. In contrast, 3D electron microscopy (EM) can deal with a wide range of specimen sizes (from single proteins to whole organelles; Fig. 10.1) and does not require sample crystallization. Although at present

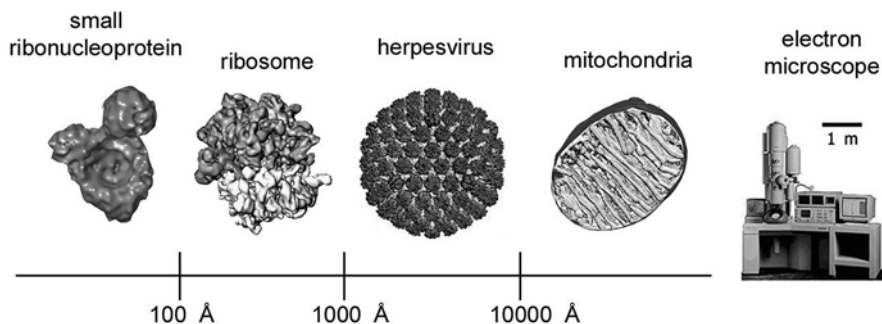


Fig. 10.1. Different scales: some examples of biological objects recently analyzed by 3D electron microscopy (EM) illustrate the versatility of the technique, able to deal with a large variety of geometries and sizes. From left to right, rendering of 3D maps for: spliceosomal U1 small nuclear ribonucleoprotein (Stark et al., 2001) (reproduced with permission from Macmillan Publishers Ltd: Nature, copyright 2001); *Escherichia coli* ribosome (Valle et al., 2003b) (reproduced with permission from Elsevier, copyright 2003); herpesvirus capsid (Zhou et al., 2000) (reproduced with permission from AAAS, copyright 2000) and mitochondria (Frey and Mannella, 2000) (reproduced with permission from Elsevier, copyright 2000).

3D-EM is limited to worse resolution than crystallography or NMR, its versatility has turned it into an excellent technique to analyze the structure of large biological machines. Even the moderate attainable resolution in routine 3D-EM maps permits us to localize structural and functional domains, and to combine this information with previously solved atomic resolution data.

10.2 Basis of 3D-EM

An electron microscope produces very high magnification images of the specimen; however, the images are 2D, whereas the specimen is a 3D object. How is it possible for us to obtain information on the 3D structure of an object using EM images? Moreover, how accurate will this information be? The answers to these questions form the three fundamental pillars of 3D-EM:

1. *EM images contain atomic resolution information. Electron microscopes use accelerated electrons as a “light” source to produce images. In biology, the acceleration is typically achieved by voltages in the 100 kV range. This results in a wavelength (and hence a resolving power) of 0.04 Å for the accelerated electrons. Owing to the aberrations inherent to the magnetic lenses that focus the electron beam, in practice the attainable resolution is never better than 1 Å. Still, this value indicates that the images formed by the electrons do indeed contain information at atomic resolution.*
2. *EM images contain information on the complete object. The depth of field of a transmission electron microscope is on the order of several thousands of angstroms. This is, also, the range of thickness of typical biological samples. This means that the whole specimen is in focus simultaneously, and so the images represent a superposition of structural features corresponding to different heights in the sample. That is, the images formed in the electron microscope are projections of the specimen.*
3. *It is possible to reconstruct the 3D structure of an object starting from its 2D projections. This statement, known as the Radon theorem, was theoretically proven in 1917 (Radon 1917), and is also the principle behind medical imaging techniques like computerized axial tomography and positron emission tomography (Fig. 10.2).*

However, although these three points guarantee that we can obtain atomic-resolution 3D reconstructions from EM images, their practical application in biology is not immediate. This is mainly due to the nature of the biological specimens and their interactions with the electron beam. Biological samples are composed of light atoms, whose interaction with electrons is weak, and so they generate very little contrast in the images. In addition, the environment in the microscope is highly destructive: first, operation of an electron microscope requires high-vacuum conditions, incompatible with the high water content of biological matter; and, second, the ionizing power of the electron beam easily destroys the structure of the irradiated compounds.

Over the years, electron microscopists have tried to solve these problems by designing a number of sample preparation techniques intended to preserve the structure of the object as much as possible under the adverse imaging conditions. Among

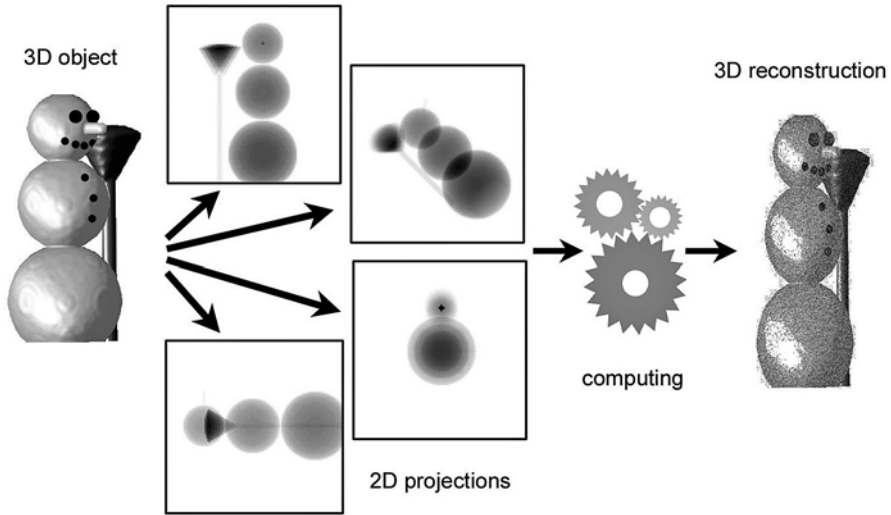


Fig. 10.2. The basis of 3D reconstruction: projections (2D images) of a 3D object can be computationally combined to recover (with more or less accuracy) its 3D structure

them, the most successful from the point of view of 3D-EM was negative staining. This consists in embedding the object in a thin layer of a heavy-metal salt. The heavy metal is excluded from the region occupied by the biological macromolecules, and thus generates contrast by delineating the shape of the exclusion region. This technique is still very useful in 2D, medium-resolution studies. However, it subjects the sample to conditions very different from those of its functional state. Furthermore, it produces images that do not directly reflect the specimen structure, but that of the cast that the staining agent creates around it. For these reasons, it is not considered a suitable technique for obtaining 3D reconstructions at high resolution.

In the late 1980s, the development of cryo-EM (Dubochet et al. 1988) solved some of these limitations. In cryo-EM the biological macromolecules are vitrified in solution by rapid immersion in an appropriate cryogen (usually liquid ethane), and they are kept in this frozen and hydrated state throughout storage, transfer to the microscope and imaging. In this way, the samples conserve their water content, and the images obtained are a direct function of the sample density, not of the staining agent cast. This entails a clear benefit: not only can we observe the surface of the object, but also its internal features. Imaging biological molecules at low temperatures has an additional advantage. The inelastic scattering of electrons by the sample transfers to the specimen amounts of energy high enough to break covalent bonds and produce free radicals. This interaction can modify to a large extent the original structure of the imaged object. The free charged groups are mobile and can spread the damage following a chain reaction. The mobility of these elements is reduced at cryo-EM temperatures, helping to further preserve the structure of the specimen. Still, the electron dose must be kept at minimum levels.

10.3 Recent Evolution of the 3D-EM Field

Although 3D reconstruction methods for EM were developed almost 40 years ago (DeRosier and Klug 1968), obtaining high-resolution information with this technique seemed impossible until 2 decades later. The reason was related to the use of negative staining, which largely altered the relation between the sample in its native state and its images. Another limitation came from the low spatial coherence provided by the available electron sources, thermoionic electron guns. The low coherence produced a fast decay in the signal amplitude, so much so that the images barely contained any information beyond 20-Å resolution. The qualitative jump to high resolution came with the popularization of field-emission guns as electron sources. These guns produce highly coherent electron beams, and consequently the images contain useful information all the way to high frequencies.

One problem that cryo-EM does not solve is the lack of contrast of biological samples. This becomes even worse, because images of frozen hydrated samples are taken with very low electron doses to minimize radiation damage, and the scattering powers of ice and biological material are very similar. The result is that cryo-EM images are very noisy, with a very low signal-to-noise ratio. To overcome this difficulty, a great number of images from chemically identical specimens are averaged. Averaging requires all the images to be first taken into register, since the different individual specimens will in general lie in random positions on the specimen support. In addition, it is necessary to use robust classification techniques to ensure that the average comes from a homogeneous data set. Sophisticated image restoration procedures have to be applied to compensate for the electron microscope aberrations. Several thousands, or tens of thousands, of individual images need to be processed in this way before obtaining a satisfactory 3D reconstruction, making 3D-EM a very computationally demanding technique. This is why the swift evolution of computers and computational methods has been another key factor for the takeoff of EM as a high-resolution technique in biology.

10.4 The Cutting Edge

All the developments described earlier have combined in the last few years to produce the most fruitful era of 3D-EM so far (Baumeister and Steven 2000; Henderson 2004). The first successes came from samples that form 2D crystalline sheets. 2D crystals contain a large number of copies of the same object, already put into register thanks to the specimen intrinsic order. In this way, it was possible to trace the peptide chain of the light-harvesting complex (Kuhlbrandt et al. 1994) and of bacteriorhodopsin (Henderson et al. 1990; Kimura et al. 1997), in cryo-EM maps with 3.0–3.5-Å resolution. After that, about 20 more proteins were reconstructed to resolutions in the 4–8-Å interval, including aquaporin (Murata et al. 2000) and other membrane channels, specimens particularly refractory to X-ray diffraction techniques. Soluble proteins that form crystalline sheets under certain conditions

can also be approached by this methodology, as in the successful determination of the tubulin structure (Nogales et al. 1998).

Reconstruction of 2D crystals suffers from a strong resolution anisotropy. Owing to the physical impossibility of tilting the sample holder in the microscope to angles larger than approximately 60°, there is a lack of data in some orientations. This results in a significant decrease of resolution in the direction perpendicular to the sample plane. Samples that form tubular crystals or highly ordered helical assemblies are particularly advantageous in this respect, since in those cases the images provide a more complete set of projections for the object. Thus, atomic models have recently been obtained using 4-Å-resolution 3D-EM maps for the acetylcholine receptor (Miyazawa et al. 2003) and the bacterial flagellum (Yonekura et al. 2003).

Although they give excellent results regarding resolution, crystalline samples suffer from an important drawback: they require crystallization of the sample. Fortunately, with 3D-EM we can also approach single particle specimens, i.e., isolated biological objects that display low or no symmetry. However, this kind of object usually adopts random orientations on the specimen support, and the difficulty in determining the relative orientations of the thousands of particles one needs to average reduces the practical resolution. Nevertheless, even at 7-Å resolution it was possible to trace the polypeptide chain of the hepatitis B virus core protein (Böttcher et al. 1997; Conway et al. 1997). Icosahedral viruses represent especially favorable cases, because their high symmetry (60-fold) decreases the number of particles required to calculate the reconstruction. In this manner, resolutions better than 10 Å have also been attained for several other viruses, including herpes, one of the largest known viruses (Zhou et al. 2000), the enveloped dengue virus (Zhang et al. 2003), and cytoplasmic polyhedrosis virus (Zhou et al. 2003). The single particle approach applied to low-symmetry samples can reach subnanometer resolutions too, and the results have been pushed to limits difficult to imagine few years ago. That is the case for the astonishing 6-Å-resolution map obtained for the GroEL chaperone in native conditions (Ludtke et al. 2004). The 10-Å objective is also getting closer for particles without any symmetry, as proved by works on ribosome structure (Valle et al. 2003a, b) and the small ribonucleoprotein U1 (U1 snRNP), a particle fundamental for the correct gene reading in eukaryotic organisms, which is in the lower limit of size approachable by 3D-EM (Stark et al. 2001).

In everyday work, the potential of cryo-EM is limited by instabilities of the sample, radiation damage, aberrations of the magnetic lenses, and inaccuracies during the digital processing of the images. In most of the studies, these and other factors constrain the 3D maps to moderate resolutions. In this situation, the combination of 3D-EM reconstructions of whole complexes (at medium resolution) with the structures of each individual component (at atomic resolution) appears as a critical tool. Available atomic coordinates of protein domains or subdomains can be fitted to EM maps of large biological macromolecules, as if solving a jigsaw puzzle (Fig. 10.3). Indeed, this hybrid methodology has already produced models with a precision of around 4 Å for several viruses (Grimes et al. 1997; San Martín et al. 2001; Mukhopadhyay et al. 2003), including the one that causes dengue fever (Kuhn et al. 2002), or the bacteriophage T4 baseplate, composed by 15 different proteins (Kostyuchenko et al. 2003), as well as for the cytoskeleton components tubulin and actin (Steinmetz et al. 1998; Nogales et al. 1999). Subsequent resolution of the same specimens by X-ray crystallography has experimentally confirmed the accuracy of

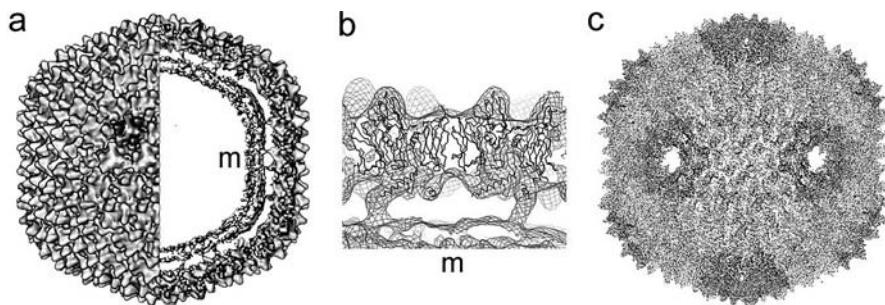


Fig. 10.3. The hybrid approach: a 3D-EM map for bacteriophage PRD1 (San Martín et al. 2001) at moderate resolution (a) supplies enough constraints to fit the atomic coordinates of P3 (b), the major coat protein, and to construct a quasi-atomic model of the viral capsid (c). *m* indicates the presence of an inner membrane

these models (Grimes et al. 1998; Abrescia et al. 2004). The power of this hybrid approach has been demonstrated in difficult scenarios, such as the flexible docking of ribosomal RNA and dozens of ribosomal proteins within cryo-EM maps for the ribosome under large conformational changes (Gao et al. 2003).

10.5 Dynamic Structures and the Fourth Dimension in Cryo-EM

Many biological macromolecules can be envisaged as molecular motors, i.e., they employ biochemical energy to move other molecules in a specific manner and direction. The link between reactions usually comes from conformational changes that macromolecules exploit to perform their tasks. In those cases, it is obvious that understanding the functioning requires following up those movements through the different steps of the process. What necessarily emerges is the addition of a fourth dimension, time, to 3D studies aimed at comprehending the course of action of large molecular assemblies. Cryo-EM is well suited for this purpose, since the structures are determined in near-native conditions and experiments can be designed to trap particular transient states. Here, the 3D exploration by cryo-EM at medium resolution provides a dynamic picture, and the aforementioned combination of the data with available atomic coordinates results in “quasi-atomic” structural models that reveal the biological mechanisms in detail. Since cryo-EM is an averaging technique, it is necessary to merge information coming from individual molecules that display the same conformation. The idea is to capture the dynamic process through “snapshots,” where each 3D map defines one stage of the reaction.

We can find different approaches in the literature that deal with the movement of macromolecules. The straightforward option is to stop *in vitro* reactions while freezing the samples in liquid ethane, the so-called time-resolved cryo-EM. Bacteriorhodopsin 2D crystals were analyzed as early as 1 ms after illumination to define light-driven changes in protein conformation (Subramaniam and Henderson

1999). The feasibility of this technique, however, is currently limited by the nature of the agent that triggers the conformational change. Physicochemical factors, such as light, temperature and pH are easy to apply to the sample in a simultaneous and uniform way. In other cases, where additional molecules participate in the reaction, uneven diffusion and stochastic flows impede to trap a synchronized population of macromolecules.

Sometimes, the system can be slowed down, and the bottleneck created allows the desired synchronic state to be achieved. Following this strategy, the dynamics of herpes simplex virus maturation was studied after blocking the activity of a protease, since protease-deficient procapsids mature slowly *in vitro* (Heymann et al. 2003) (Fig. 10.4). Most dynamic structural studies by cryo-EM, however, employ biochemical interference that stalls the macromolecules in a desired step. The ribosome is the paradigm of this type of characterization, since several natural antibiotics are known to inhibit translation. The incorporation of specific transfer RNAs (tRNAs) to the ribosome was analyzed thanks to the blockade of the process by kirromycin (Valle et al. 2003a). The described rearrangements of ribosome and tRNA uncovered a cascade of interrelated events that clarified the structural basis for fidelity during translation.

Even in cases where the reactions cannot be stopped at the desired step, cryo-EM offers another possibility. Some investigators, rather than synchronize the macromolecules during their activity, have tried to analyze a full dynamic reaction within the same unique sample. In this approach, a cycling reaction takes place in a tube where the molecules alternate between two or more conformations, resulting into a mixed data set. The challenge consists in separating different homogeneous groups of individual particles while processing the whole population. Here, classification tools are indispensable in order to differentiate the variability of the images due to distinct orientations of the molecules from a genuine structural diversity among them. Few attempts have been made following this risky scheme, and in most of them, the success depends on the availability of templates the particles can be compared with. Promising results, however, have been presented with the ATP cycling of chaperonins (Schoehn et al. 2000) or the ratchetlike relative rotation between ribosomal subunits during translocation of tRNAs (Gao et al. 2004).

10.6

Cryo-electron Tomography

So far, for obtaining a 3D reconstruction of a particular biological complex, we have isolated it from the rest of the cell components and have averaged images of thousands of copies. In cryo-EM the final 3D maps reflect the “near-native” state of the molecules *in vitro*. Obviously, the real native conditions should include the biological context and the uniqueness of each object. Up to now, this task has seemed unreachable. But a new exciting era is opening, in which several groups are already setting up the basis for observing the molecular machinery without extracting it from the whole cell, or for obtaining reconstructions of unique objects (without averaging) at an acceptable resolution. Cells and cellular organelles are pleomorphic structures, i.e., they display different shapes despite a common overall design.

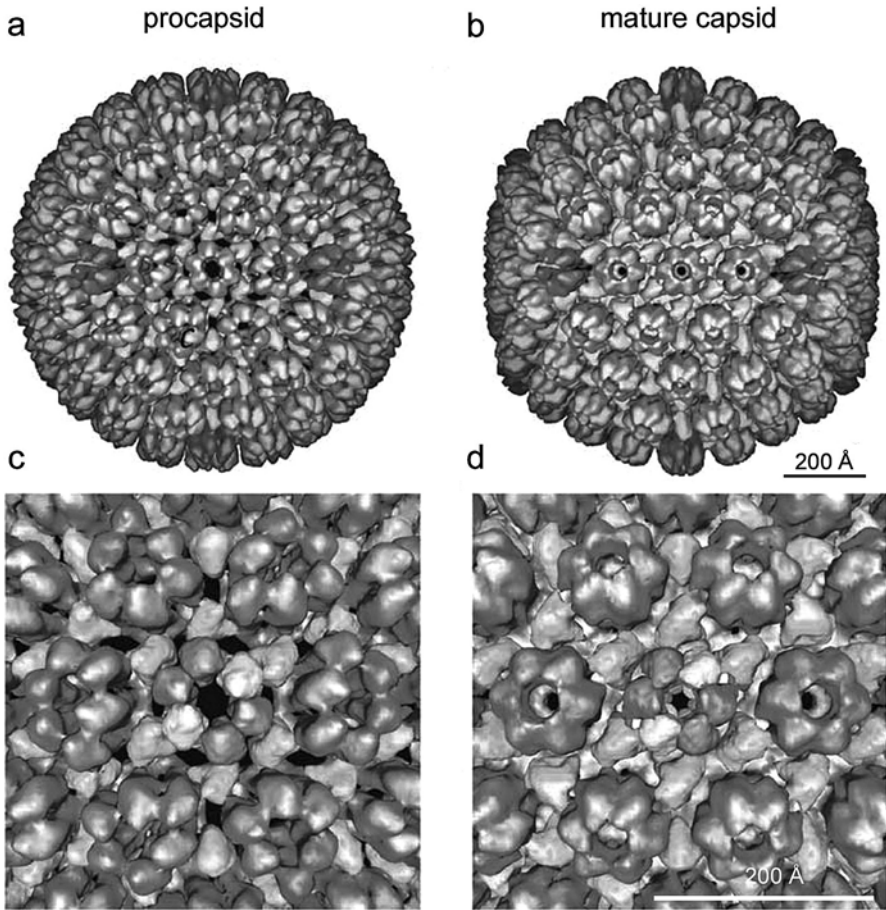


Fig. 10.4. Dynamic studies in 3D-EM: maturation of herpes simplex virus capsid was studied by 3D-EM (Heymann et al. 2003) (reproduced with the permission from Nature Publishing Group, copyright 2003). Several distinct classes of images were detected in a protease-defective viral population. Initial and final states of the described maturation are shown: spherical procapsid (**a, c**) and polyhedral mature capsid (**b, d**)

Clearly, the molecular traffic inside biological vesicles is a dynamic process, and the distribution of macromolecules inside cells changes through time and physiological states. With this in mind, the 3D study of such different objects cannot be faced by averaging techniques; it requires the analysis of unique entities. Cryo-electron tomography (cryo-ET) is a very competent technique to record and analyze 3D images of single biological samples including entire cells. Images of the frozen-hydrated specimen are recorded in tilt series, where the images are projections of the same object viewed from different angles, and can be combined into a 3D density map.

Starting from the principles of cryo-ET, we can anticipate that the wider the angular range covered in the tilt-series and the smaller the angular step between images will yield the best result. By collecting more images, however, the sample damage due to the accumulative electron dose increases too. Limiting the dose is one of the reasons that restricts cryo-ET to resolutions around 4–6 nm. Nevertheless, even in this resolution range valuable information can be extracted, such as for the internal organization of mitochondria (Frey and Mannella 2000). The recently developed automated cryo-ET in helium-cooled microscopes allows us to record data in a highly efficient way and with minimum electron dose. The results are opening a new field that fills the resolution gap between light microscopy and structures resolved at atomic level. The impressive 3D images of eukaryotic *Dictyostelium discoideum* cells (Medalia et al. 2002) depicted an imbricate organization of the cytoskeleton at the level of individual filaments. Additionally, the map showed macromolecular complexes, raising a question about the possibility of identifying individual macromolecules inside crowded cells. In an attempt to prove that this is indeed feasible, thermosomes and proteasomes were encapsulated in liposomes, constructing a simplified model of a cell that was studied by cryo-ET (Frangakis et al. 2002). Interestingly, using a template-matching procedure, the authors separated the two types of molecules with good fidelity (Fig. 10.5).

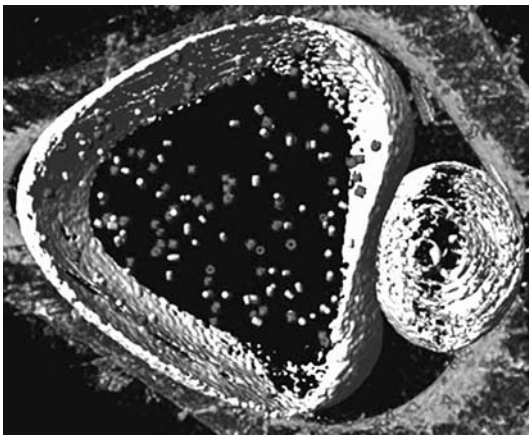


Fig. 10.5. Cryo-electron tomography (*cryo-ET*) of cells: phantom cells containing thermosomes and 20S proteasomes were imaged by cryo-ET (Frangakis et al. 2002) (reproduced with permission from National Academy of Sciences, U.S.A., copyright 2002). This work proved that it was possible to discriminate between the two macromolecules by using a template-matching procedure. In the rendering thermosomes (*dark gray*) and proteasomes (*light gray*) can be seen inside the vesicle in serf

10.7 Concluding Remarks

Analysis of the structure and dynamics of big macromolecular complexes is one of the great challenges of the postgenomic era. It is expected that in the next few years a great number of protein domain structures will be made available through X-ray crystallography and NMR studies, as all branches of biology strive to process the avalanche of data coming from genome sequencing. The understanding of cellular functions requires studying these domains in their context, taking part in the big macromolecular complexes and within the framework of living cells. The contributions from cryo-EM and cryo-ET can bring us to a new scenario, where the researcher could “see” individual molecular motors working within an astonishing complicated factory, the cell. We are convinced that all the possibilities that 3D-EM offers will illustrate these and many other breakthroughs in a not very distant future.

Acknowledgements. We thank José María Carazo for helpful comments on the manuscript. Grants from Comunidad de Madrid (GR/SAL/0645/2004) to M.V. and from Fundación BBVA (2004X578) to C.S.M. are acknowledged too.

References

- Abrescia NG, Cockburn JJ, Grimes JM, Sutton GC, Diprose JM, Butcher SJ, Fuller SD, San Martín, C, Burnett RM, Stuart DI, Bamford DH, Bamford JK. (2004) Insights into assembly from structural analysis of bacteriophage PRD1. *Nature* 432:68–74
- Baumeister W, Steven AC (2000) Macromolecular electron microscopy in the era of structural genomics. *Trends Biochem Sci* 25:624–631
- Böttcher B, Wynne SA, Crowther RA (1997) Determination of the fold of the core protein of hepatitis B virus by electron cryomicroscopy. *Nature* 386:88–91
- Conway JF, Cheng N, Zlotnick A, Wingfield PT, Stahl SJ, Steven AC (1997) Visualization of a 4-helix bundle in the hepatitis B virus capsid by cryo-electron microscopy. *Nature* 386:91–94
- DeRosier DJ, Klug A (1968) Reconstruction of three-dimensional structures from electron micrographs. *Nature* 217:130–134
- Dubochet J, Adrian M, Chang JJ, Homo JC, Lepault J, McDowell AW, Schultz P (1988) Cryo-electron microscopy of vitrified specimens. *Q Rev Biophys* 21:129–228
- Frangakis AS, Bohm J, Forster F, Nickell S, Nicastro D, Typke D, Hegerl R, Baumeister W (2002) Identification of macromolecular complexes in cryoelectron tomograms of phantom cells. *Proc Natl Acad Sci USA* 99:14153–14158
- Frey TG, Mannella CA (2000) The internal structure of mitochondria. *Trends Biochem Sci* 25:319–324
- Gao H, Sengupta J, Valle M, Korostelev A, Eswar N, Stagg SM, van Roey P, Agrawal RK, Harvey SC, Sali A, Chapman MS, Frank J (2003) Study of the structural dynamics of the E coli 70S ribosome using real-space refinement. *Cell* 113:789–801
- Gao H, Valle M, Ehrenberg M, Frank J (2004) Dynamics of EF-G interaction with the ribosome explored by classification of a heterogeneous cryo-EM dataset. *J Struct Biol* 147:283–290

- Grimes JM, Jakana J, Ghosh M, Basak AK, Roy P, Chiu W, Stuart DI, Prasad BV (1997) An atomic model of the outer layer of the bluetongue virus core derived from X-ray crystallography and electron cryomicroscopy. *Structure* 5:885–893
- Grimes JM, Burroughs JN, Gouet, P, Diprose JM, Malby R, Ziéntara S, Mertens PPC, Stuart DI. (1998) The atomic structure of the bluetongue virus core. *Nature* 395:470–478
- Henderson R (2004) Realizing the potential of electron cryo-microscopy. *Q Rev Biophys* 37:3–13
- Henderson R, Baldwin JM, Ceska TA, Zemlin F, Beckmann E, Downing KH (1990) Model for the structure of bacteriorhodopsin based on high-resolution electron cryo-microscopy. *J Mol Biol* 213:899–929
- Heymann JB, Cheng, N, Newcomb WW, Trus BL, Brown JC, Steven AC (2003) Dynamics of herpes simplex virus capsid maturation visualized by time-lapse cryo-electron microscopy. *Nat Struct Biol* 10:334–341
- Kimura Y, Vassilyev DG, Miyazawa A, Kidera A, Matsushima M, Mitsuoka K, Murata K, Hirai T, Fujiyoshi Y (1997) Surface of bacteriorhodopsin revealed by high-resolution electron crystallography. *Nature* 389:206–211
- Kostyuchenko VA, Leiman PG, Chipman PR, Kanamaru, S, van Raaij MJ, Arisaka, F, Mesyanzhinov VV, Rossmann MG (2003) Three-dimensional structure of bacteriophage T4 baseplate. *Nat Struct Biol* 10:688–693
- Kühlbrandt W, Wang DN, Fujiyoshi Y (1994) Atomic model of plant light-harvesting complex by electron crystallography. *Nature* 367:614–621
- Kuhn RJ, Zhang, W, Rossmann MG, Pletnev SV, Corver, J, Lenches, E, Jones CT, Mukhopadhyay, S, Chipman PR, Strauss EG, Baker TS, Strauss JH (2002) Structure of dengue virus: implications for flavivirus organization, maturation, and fusion. *Cell* 108:717–725
- Ludtke SJ, Chen DH, Song JL, Chuang DT, Chiu W (2004) Seeing GroEL at 6 Å resolution by single particle electron cryomicroscopy. *Structure* 12:1129–1136
- Medalia O, Weber I, Frangakis AS, Nicastro D, Gerisch G, Baumeister W (2002) Macromolecular architecture in eukaryotic cells visualized by cryoelectron tomography. *Science* 298:1209–1213
- Miyazawa A, Fujiyoshi Y, Unwin N (2003) Structure and gating mechanism of the acetylcholine receptor pore. *Nature* 424:949–955
- Mukhopadhyay S, Kim BS, Chipman PR, Rossmann MG, Kuhn RJ (2003) Structure of west Nile virus. *Science* 302:248
- Murata K, Mitsuoka K, Hirai T, Walz T, Agre P, Heymann JB, Engel A, Fujiyoshi Y (2000) Structural determinants of water permeation through aquaporin-1. *Nature* 407:599–605
- Nogales E, Wolf SG, Downing KH (1998) Structure of the alpha beta tubulin dimer by electron crystallography. *Nature* 391:199–203
- Nogales E, Whittaker M, Milligan RA, Downing KH (1999) High-resolution model of the microtubule. *Cell* 96:79–88
- Radon J (1917) Über die Bestimmung von Funktionen durch ihre Integralwerte längs gewisser Mannigfaltigkeiten (on the determination of functions from their integrals along certain manifolds). *Ber Verh Saechs Akad Wiss Leipzig Math-Phys Kl* 69:262–277
- San Martín C, Burnett RM, de Haas F, Heinkel R, Rutten T, Fuller SD, Butcher SJ, Bamford DH (2001) Combined EM/X-ray imaging yields a quasi-atomic model of the adenovirus-related bacteriophage PRD1 and shows key capsid and membrane interactions. *Structure* 9:917–930

- Schoehn G, Quate-Randall E, Jimenez JL, Joachimiak A, Saibil HR (2000) Three conformations of an archaeal chaperonin, TF55 from *Sulfolobus shibatae*. *J Mol Biol* 296:813–819
- Stark H, Dube P, Luhrmann R, Kastner B (2001) Arrangement of RNA and proteins in the spliceosomal U1 small nuclear ribonucleoprotein particle. *Nature* 409:539–542
- Steinmetz MO, Hoenger A, Tittmann P, Fuchs KH, Gross H, Aebi U (1998) An atomic model of crystalline actin tubes: combining electron microscopy with X-ray crystallography. *J Mol Biol* 278:703–711
- Subramaniam S, Henderson R (1999) Electron crystallography of bacteriorhodopsin with millisecond time resolution. *J Struct Biol* 128:19–25
- Valle M, Zavialov A, Li W, Stagg SM, Sengupta J, Nielsen RC, Nissen P, Harvey SC, Ehrenberg M, Frank J (2003a) Incorporation of aminoacyl-tRNA into the ribosome as seen by cryo-electron microscopy. *Nat Struct Biol* 10:899–906
- Valle M, Zavialov A, Sengupta J, Rawat U, Ehrenberg M, Frank J (2003b) Locking and unlocking of ribosomal motions. *Cell* 114:123–134
- Yonekura K, Maki-Yonekura S, Namba K (2003) Complete atomic model of the bacterial flagellar filament by electron cryomicroscopy. *Nature* 424:643–650
- Zhang W, Chipman PR, Corver J, Johnson PR, Zhang Y, Mukhopadhyay S, Baker TS, Strauss JH, Rossmann MG, Kuhn RJ (2003) Visualization of membrane protein domains by cryo-electron microscopy of dengue virus. *Nat Struct Biol* 10:907–912
- Zhou ZH, Dougherty M, Jakana J, He J, Rixon FJ, Chiu W (2000) Seeing the herpesvirus capsid at 8.5 Å. *Science* 288:877–880
- Zhou ZH, Zhang H, Jakana J, Lu XY, Zhang JQ (2003) Cytoplasmic polyhedrosis virus structure at 8 Å by electron cryomicroscopy: structural basis of capsid stability and mRNA processing regulation. *Structure* 11:651–663

Subject Index

- (ND) FT-NMR 147
 13C/15N double-labeled proteins 151
 2D-IR 80, 82, 83, 84
 2D spectroscopy 74
 3D-EM 268, 269, 270, 271, 276
- A**
 accordion spectroscopy 148
 AFM 163, 171, 172, 174, 178, 202, 203, 207,
 222, 223, 224, 225, 230, 231, 232, 233, 235
 aggregation 40, 76, 80, 84
 air/water interface 49, 63, 59
 asynchronous 74, 75, 76, 78, 80, 85, 86
 autopeaks 85, 86
- B**
 band 74, 76
 – components 74
 – contour 74
 – cross-correlation 74, 75
 bionanomachines 163, 164, 166, 233
 Brewster angle 52, 55
- C**
 chemical shift 147, 149, 151, 156
 – chemical shift evolution 150
 correlation 74, 75, 76, 83, 84
 correlation peaks 129, 134, 140, 141
 correlation spectra 97, 98, 102, 113, 114
 COSY 104, 106, 131, 134
 cross-peaks 78, 79
 cryo-EM 269, 270, 271, 272, 273, 276
- D**
 denaturation 40, 44
 denaturing agents 81
 dispensing 2, 5, 6, 7, 9, 11, 14
- F**
 folding 34, 36, 37, 39
- G**
 GFT-NMR 147, 152, 156, 157, 158, 159
 gradient 130, 134, 135
 – bipolar 130, 131, 132
 – field 130, 132, 133, 134, 135
 GUV 248, 249, 252, 254, 255, 256, 257, 261
- H**
 heat-capacity 44
 heat capacity 28, 29, 30, 31, 33, 35, 42
 – absolute 42, 43
 heteronuclear magnetization 151
 higher-dimensional heteronuclear FT
 NMR 156
 Homopolypeptides 801
- I**
 in-phase splitting 149
 IR 82, 84
 IRRAS 49, 55, 60, 61
 – data acquisition 52
 instrumentation 68
 – measurements 50, 55, 57
 – simulation 66
 – spectra 49, 56, 62, 63, 67
 labeling 89, 90, 92, 99, 101, 108, 114, 119
 isotope 89, 99, 100
- L**
 Langmuir films 49, 67
 LAURDAN 251, 252, 253, 254, 255, 256, 257,
 258, 259, 260, 261
 lipid domain 247, 249, 250, 253, 255, 256
 lipid phases 250, 252, 254, 255
- M**
 membrane 247
 – domains 251
 – function 247
 – model 248, 251, 255
 – structure 247, 248, 249, 251
 microbatch 13
 microcrystals 18, 19
 microwells 12
 monolayers 49, 52, 55, 59, 60, 61, 63, 65, 66

multidimensional spectroscopy 158
multiplet 95
multiplet components 155

N

nanocrystallization 1, 2, 8, 14, 15, 21
nanodroplet 2
negative staining 269, 270
NMR 89
– heteronuclear 89, 92, 98, 111, 112
NOESY 100, 111, 112, 115, 116
nonequilibrium 185, 202, 223
nozzle 9, 16, 17
nucleation 2, 3, 4, 5, 19, 21

P

peak identification 151
perturbatio 74
perturbation 75, 80, 82
phase 132
phases 131, 133
phenylketonuria 39
Polyproline 80
protein 1, 27, 73, 74, 80, 81
– conformation 273
– crystallization 1, 2, 3, 6, 9, 11, 20
– crystallography 8
– denaturation 31, 39
– energetics 27
– folding 27, 30
– membrane 89
– stability 27
– structures 73
– unfolding 29
pulmonary surfactant 57, 59

R

radiation damage 270, 271
RD (and GFT) NMR 150
RD NMR 148, 151, 152
refolding 183, 211, 217, 218, 232, 235
resonance assignment 98, 100, 101, 105,
106, 108, 113, 116, 119, 159
robotic 14, 17, 18, 20
robotics 16

S

screening 4, 15, 16, 17, 21
secondary structure 80
shape 78
signal-to-noise ratio 52, 67, 68,
130, 134, 135, 141
simulation 80
simulations 76
single-molecule 163, 170, 171, 174, 175, 181,
185, 202, 207, 214, 216, 217, 233
spectral resolution 89, 105, 109
spectrum 75, 76, 80
– dynamic 75
structures 80
synchronous 74, 75, 76, 79, 83, 85, 86

T

temperature 74, 82
thermal denaturation 27, 33
time relaxation 94
transverse relaxation 90, 92, 93, 97, 98
TROSY 90, 92, 93, 95, 97, 98, 101, 102, 104,
105, 106, 108, 109, 111, 112, 114, 115, 118,
119
two-dimensional 129, 130, 133, 135, 136,
139, 140, 142, 144
two-photon 249, 251, 252, 255, 256, 260, 261
two-state 31, 33, 34, 35, 36

U

unfolding 34, 35, 36, 37, 84
urea 81, 82, 83, 84